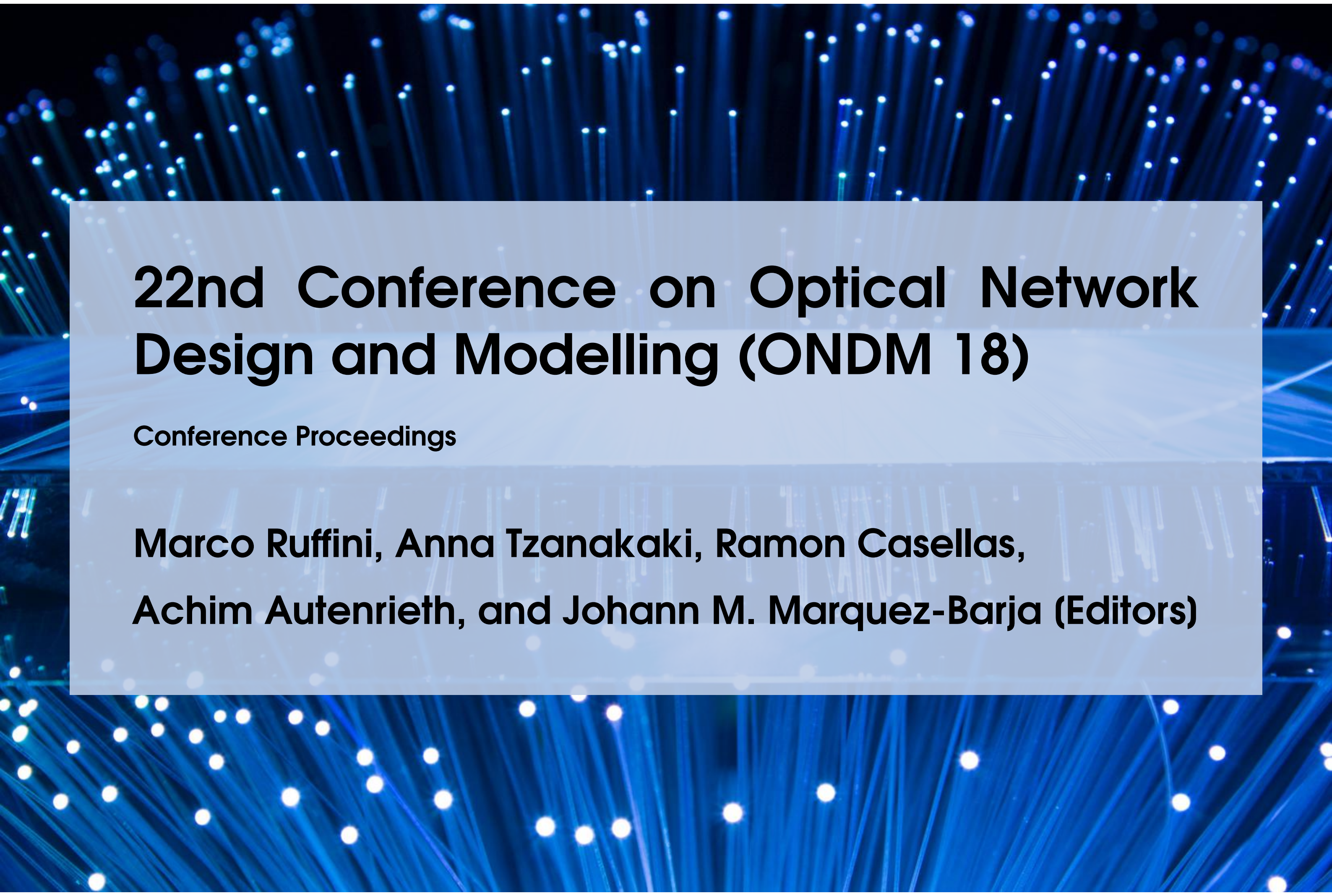


22nd Conference on Optical Network Design and Modelling (ONDM 18)

Conference Proceedings

Marco Ruffini, Anna Tzanakaki, Ramon Casellas,
Achim Autenrieth, and Johann M. Marquez-Barja (Editors)

Technical Co-Sponsors

IEEE ComSoc
IEEE Communications Society

ifip tc6

PATRONS

Fondúireacht Eolaíochta Éireann
Dá bhfuil romhainn
Science Foundation Ireland
For what's next

Fāilte Ireland

TECHNICAL CO-SPONSORS

IEEE ComSoc™
IEEE Communications Society

ifip tc6

ORGANIZERS

CONNECT
Centre for Future Networks

Trinity College Dublin
The University of Dublin

ISBN 978-3-903176-07-2 ©2018 IFIP
May 14-18, 2018
Dublin, Ireland

Contents

I Preface

II Program

III Papers

Preface

ONDM 2018
22nd Conference on Optical
Network Design and Modelling
May 14-17, 2018
Dublin, Ireland

1. Message from the General Chair

Dear colleagues,

It is an honour and a pleasure to welcome everyone to the ONDM 2018 Conference. This is the 22nd edition of the IEEE Optical Network Design and Modelling conference, which is held at the University of Dublin, Trinity College, Ireland, from the 14th to the 17th of May 2018. Trinity College Dublin is one of the oldest Universities in the world, dating 1592, and the technical sessions are held at the Erwin Schrödinger Theatre, named in honour of the pioneer of quantum mechanics who spent much of his career in Dublin.

The conference is technically co-sponsored by the IEEE communications society and the IFIP TC-6, ensuring top quality of the accepted publications. The call for papers was updated this year to include novel topics such as: Slicing, virtualisation and multi-tenancy techniques for optical networks; Machine learning techniques for optical networks; Novel optical node designs including disaggregation and open optical line systems.

The contributed and invited papers were grouped into 9 technical sessions, spread across the 4 days of the event. This year's programme includes two excellent keynote speakers: Andrew Lord (BT) and Gavin Young (Vodafone). This year for the first time we also invited two tutorial speakers: Massimo Tornatore (Polytechnic of Milan) and Josep Prat (Universitat Politècnica de Catalunya), giving an overview on machine learning and WDM access/metro networks, respectively.

The technical program includes two workshops. One is on "Optical technologies in the 5G era", which will see well known speakers, collaborating in H2020 European projects, giving a talk on their recent progress. A second workshop, on "SDN/NFV for optical networks" will discuss topics such as network telemetry, artificial intelligence and network virtualisation. This workshop will be followed by an industry panel, where the speakers will discuss progress, pros and cons of optical network disaggregation.

In addition, this year's event will host a special open public session on Net Neutrality. This free, public seminar will introduce the issues at the heart of the net neutrality debate. We will also discuss how the arrival of 5G technology will potentially impact policy in this area. A selected panel of experts was invited to contribute their opinion on the subject.

Finally, I would like to thank Science Foundation Ireland (SFI) and Failte Ireland for their financial support; the members of the CONNECT research group for their support in preparing the conference; and all the members of the conference committees and the reviewers for their dedicated and passionate work.

Prof. Marco Ruffini
General Chair - ONDM 2018

2. Organizing Committee

- **General Chair**
  Marco Ruffini - CONNECT Centre, The University of Dublin, Trinity College, Ireland
- **TPC Chairs**
  Anna Tzanakaki – University of Athens, Greece, and University of Bristol, U.K.
  Ramon Casellas – CTTC, Barcelona, Spain
  Achim Autenrieth – ADVA Optical Networking, Germany
- **EDAS and Publications Chair**
  Johann M. Marquez-Barja – University of Antwerpen - imec, Belgium
- **Publicity Chairs**
  Ricard Vilalta – Centre Tecnològic Telecomunicacions Catalunya (CTTC), Barcelona, Spain
  Gangxiang Shen - ONTRC, Soochow University, PR China
- **Financial Chair**
  Catherine Keogh – CONNECT Centre, The University of Dublin, Trinity College, Ireland
- **Web Chair**
  Nima Afraz – CONNECT Centre, The University of Dublin, Trinity College, Ireland

3. Steering Committee

Piero Castoldi – SSSA, Italy

Philippe Gravey – Telecom Bretagne, France

Pablo Pavón Mariño – UPCT, Spain

Lena Wosinska – KTH, Sweden

Tibor Cinkler – BME, Hungary

Marco Ruffini - Trinity College Dublin, Ireland

4. TPC Members

Slavisa Aleksic – Hochschule fuer Telekommunikation
Bigomokero Bagula – University of the Western Cape
Johan Bauwelinck – Ghent University – imec
Andrea Bianco – Politecnico di Torino
Luiz Bonani – Universidade Federal do ABC
Aparicio Carranza – New York City College of Technology
Gerardo Castañón – Tecnológico de Monterrey
Isabella Cerutti – Nokia
Jiajia Chen – KTH Royal Institute of Technology
Kostas Christodoulopoulos – National Technical University of Athens
Didier Colle – IMEC – Ghent University
David Coudert – INRIA, I3S, CNRS, Université de Nice Sophia
Filippo Cugini – CNIT
Sandip Das – Trinity College Dublin, Ireland
John Doucette – University of Alberta
Georgios Ellinas – University of Cyprus
Amr Elrasad – Trinity College Dublin
Marija Furdek – KTH Royal Institute of Technology
Maurice Gagnaire – Telecom Paristech
Miquel Garrich – CPqD
Alessio Giorgetti – Scuola Superiore Sant'Anna, Italy
Philippe Gravey – Télécom Bretagne
Ashwin Gumaste – Indian Institute of Technology, Bombay
Hiroaki Harai – National Institute of Information and Communications Technology
Hiroshi Hasegawa – Nagoya University
Yusuke Hirota – National Institute of Information and Communications Technology
Weisheng Hu – Shanghai Jiao Tong University
Brigitte Jaumard – Concordia University

Wojciech Kabacinski – Poznan University of Technology
Ezhan Karasan – Bilkent University
Daniel Kilper – University of Arizona
Ken-ichi Kitayama – The Graduate School for the Creation of New Photonics Industries
Nattapong Kitsuwan – The University of Electro-Communications
Yao Li – University of Arizona
Andrew Lord – British Telecom
Guido Maier – Politecnico di Milano
Ricardo Martinez – Centre Tecnològic de Telecomunicacions de Catalunya (CTTC/CERCA)
Barbara Martini – CNIT
Carmen Mas Machuca – Technical University of Munich
Xavier Masip-Bruin – Universitat Politècnica de Catalunya
Francesco Matera – Fondazione Ugo Bordoni
Branko Mikac – University of Zagreb
Miklos Molnar – LIRMM / University of Montpellier
Paolo Monti – KTH Royal Institute of Technology
Francesco Musumeci – Politecnico di Milano, Italy
Avishek Nag – University College Dublin
Antonio Napoli – Coriant R&D GmbH
Reza Nejabati – University of Bristol
Wenda Ni – Pleora Technologies Inc
Eiji Oki – Kyoto University
Jelena Pesic – Nokia Bell labs
Christina Politi – University of Peloponnese
Carla Raffaelli – University of Bologna
Moises Ribeiro – Federal Universty of Espirito Santo
Cristina Rottondi – Dalle Molle Institute for Artificial Intelligence (IDSIA), Switzerland
George Rouskas – North Carolina State University
Sarah Ruepp – Technical University of Denmark
Mark Ruiz – Universitat Politecnica de Catalunya (UPC)
Dominic Schupke – Airbus
Motoyoshi Sekiya – Fujitsu Laboratories Limited
Gangxiang Shen – Soochow University
Domenico Siracusa – Fondazione Bruno Kessler
Nina Skorin-Kapov – Centro Universitario de Defensa, CUD San Javier
Frank Slyne – Connect Research Centre – Trinity College Dublin
Salvatore Spadaro – Universitat Politecnica de Catalunya (UPC)
Ravi Subrahmanyan – Invisage Technologies
Suresh Subramaniam – The George Washington University
Giuseppe Talli – Tyndall National Institute, University College Cork
Dimitris Varoutas – University of Athens
Emmanouel Varvarigos – University of Patras & Computer Technology Institute
Anna Maria Vegni – Universita Roma 3
Luis Velasco – Universitat Politècnica de Catalunya (UPC)
Krzysztof Wajda – AGH University of Science and Technology
Elaine Wong – The University of Melbourne
Lena Wosinska – KTH Royal Institute of Technology
Sugang Xu – National Institute of Information and Communications Technology
Shuangyi Yan – University of Bristol, United Kingdom

Xin Yin – Ghent University – IMEC, Belgium
Fen Zhou – University of Avignon
Zuqing Zhu – University of Science and Technology of China
Moshe Zukerman – City University of Hong Kong

Program

ONDM 2018 Conference Programme
14-17th May, Dublin, Ireland

General Schedule:

Day 1	**Monday 14th**
8:15	Registration open
9:00 - 9:15	Conference opening and address from Dean of Research
9:15 - 10:15	Tutorial talk by Massimo Tornatore: An Introduction to Machine Learning in Optical Transport networks
10:15-10:35	Coffee Break
10:35- 12:35	TS1: Machine Learning techniques in optical networks. Session chair: Massimo Tornatore
12:35-14:00	Lunch break
14:00 - 15:40	TS2: SDN and network disaggregation. Session chair: Andrew Lord
15:40-16:00	Coffee Break
16:00 - 18:10	TS3: Resilience and security. Session chair: Elaine Wong
18:30-19:30	Social event: Visit to Book of Kells
19:30 - 21:30	Social event: Conference reception in TCD atrium
Day 2	**Tuesday 15th**
9:00 - 10:00	Plenary talk by Gaving Young: Future Proofing the Unified Fibre Access Network: Build, Fill, Perform
10:00-10:30	Coffee Break
10:30- 12:20	TS4: Metro networks. Session chair: Dan Kilper
12:20-13:40	Lunch break

13:40 - 16:00	TS5: Fixed/mobile convergence. Session chair: George Rouskas
16:00-16:20	Coffee Break
16:20 - 18:10	TS6: Elastic and programmable optical networks. Session chair: Manos Varvarigos
18:10 - 19:00	Refreshments
19:00 - 20:30	Open Public session: What is Net Neutrality, and why should we care?
Day 3	**Wednesday 16th**
9:00 - 10:00	Plenary talk by Andrew Lord: The evolution of optical networks in a 5G world
10:00-10:30	Coffee Break
10:30- 12:10	TS7: NFV. Session chair: Mark De Leenheer
12:10-13:40	Lunch break
13:40 - 15:40	Workshop: SDN/NFV for optical networks, Session chair: Ricardo Martinez
15:40-16:10	Coffee Break
16:10- 17:40	Industry panel on optical networks disaggregation
18:45 - 22:00	Social event: Gala diner
Day 4	**Thursday 17th**
9:00 - 10:00	Tutorial talk by Josep Prat: Simple tuneable transmitters for ultra-dense WDM access and metro networks
10:00-10:30	Coffee Break
10:30- 12:10	TS8: Optical access. Session chair: Marco Ruffini
12:10-13:40	Lunch break
13:40 - 15:00	TS9: Data centres. Session chair: Anna Tzanakaki

15:00-15:30	Coffee Break
15:30 - 17:30	Workshop: Optical technologies in the 5G Era. Session chair: Anna Tzanakaki
17:30-18:00	Conference wrap up and closing remarks

Conference Sessions:

Session	Time	Name	Title
TS1: Machine learning techniques in optical networks	10:35 - 11:05	Lihua Ruan; Elaine Wong [Invited]	Machine Intelligence in Allocating Bandwidth to Achieve Low-Latency Performance
Session chair: Massimo Tornatore	11:05 - 11:25	Wei Lu; Hongqiang Fang; Zuqing Zhu	AI-Assisted Resource Advertising and Pricing to Realize Distributed Tenant-Driven Virtual Network Slicing in Inter-DC Optical Networks
	11:25 - 11:45	Antonia Pelekanou; Markos Anastasopoulos; Anna Tzanakaki; Dimitra Simeonidou	Provisioning of 5G Services Employing Machine Learning Techniques
	11:45 - 12:15	Shuangyi Yan; Reza Nejabati; Dimitra Simeonidou [Invited]	Data-driven Network Analytics and Network Optimisation in SDN-based Programmable Optical Networks
	12:15 - 12:35	Tania Panayiotou; Konstantinos Manousakis; Sotirios Chatzis; Georgios Ellinas	On Learning Spectrum Allocation Models for Time-Varying Traffic in Flexible Optical Networks
TS2: SDN and network disaggregation	14:00 - 14:30	Marc De Leenheer ; Yuta Higuchi; Guru Parulkar [Invited]	An Open Controller for the Disaggregated Optical Network
Session chair: Andrew Lord	14:30 - 14:50	Ricard Vilalta; Ramon Casellas; Ricardo Martinez; Raul Muñoz; Young Lee; Haomian Zheng; Yi Lin; Victor Lopez; Luis M. Contreras	Fully Automated Peer Service Orchestration of Cloud and Network Resources Using ACTN and CSO

	14:50 - 15:20	Dan Kilper [Invited]	Disaggregation as a Vehicle for Hyper-scalability in Optical Networks
	15:20 - 15:40	Alan Diaz Montiel; Jiakai Yu; Weiyang Mo; Yao Li; Daniel Kilper; Marco Ruffini	Performance Analysis of QoT Estimator in SDN-Controlled ROADM Networks
TS3: Resilience and security	16:00 - 16:30	Behnam Shariati; Alba Vela; Marc Ruiz; Luis Velasco [Invited]	Monitoring and Data Analytics: Analyzing the Optical Spectrum for Soft-Failure Detection and Identification
Session chair: Elaine Wong	16:30 - 16:50	Róża Goścień; Carlos Natalino; Lena Wosinska; Marija Furdek	Impact of High-Power Jamming Attacks on SDM Networks
	16:50 - 17:10	Bahare Masood Khorsandi; Federico Tonini; Carla Raffaelli	Design Methodologies and Algorithms for Survivable C-RAN
	17:10 - 17:30	Jing Zhu; Carlos Natalino; Marija Furdek; Lena Wosinska; Zuqing Zhu	Control Plane Robustness in Software-Defined Optical Networks Under Targeted Fiber Cuts
	17:30 - 17:50	Aniruddha Singh Kushwaha; Deepak Kakadia; Ashwin A Gumaste; Arun Somani	Designing Multi-Layer Provider Networks for Circular Disc Failures
	17:50 - 18:10	Yosef Aladadi; Ahmed Abas; Abdulmalik Alwarafy; Mohammed Alresheedi	Multi-User Frequency-Time Coded Quantum Key Distribution Network Using a Plug-and-Play System
TS4: Metro networks	10:30 - 11:00	Steinar Bjornstad [Invited]	Optical Ethernet: Can OTN Be Replaced?
Session chair: Dan Kilper	11:00 - 11:20	Masahiro Nakagawa; Kana Masumoto; Hidetoshi Onda; Kazuyuki Matsumura	Photonic Sub-Lambda Transport: An Energy-Efficient and Reliable Solution for Metro Networks
	11:20 - 11:40	Annie Gravey; Djamel Amar; Philippe Gravey; Michel Morvan; Bogdan Uscumlic; Dominique Chiaroni	Modelling Packet Insertion on a WSADM Ring
	11:40 - 12:00	Diogo Sequeira; Luís Gonçalo Cancela; João Rebola	Impact of Physical Layer Impairments on Multi-Degree CDC ROADM-based Optical Networks

	12:00 - 12:20	Dibbendu Roy; Sourav Dutta; Brando Kumam; Goutam Das	A Cost-effective and Energy-efficient All-Optical Access Metro-Ring Integrated Network Architecture
TS5: Fixed/Mobile convergence	13:40 - 14:10	Dimitrios Apostolopoulos; Giannis Giannoulis; Nikos Argyris; Nikolaos Iliadis; Konstantina Kanta; Hercules Avramopoulos [Invited]	Analog Radio-over-Fiber Solutions in Support of 5G
Session chair: George Rouskas	14:10 - 14:30	Bogdan Uscumlic; Dominique Chiaroni; Brice Leclerc; Thierry Zami; Annie Gravey; Philippe Gravey; Michel Morvan; Dominique Barth; Djamel Amar	Scalable Deterministic Scheduling for WDM Slot Switching Xhaul with Zero-Jitter
	14:30 - 14:50	Longsheng Li; Meihua Bi; Wei Wang; Yan Fu; Xin Miao; Weisheng Hu	SINR-Oriented Flexible Quantization Bits for Optical-Wireless Deep Converged eCPRI
	14:50 - 15:20	Ricardo Martinez; Ricard Vilalta; Manuel Requena-Esteso; Ramon Casellas; Raul Muñoz; Josep Mangues-Bafalluy [Invited]	Experimental SDN Control Solutions for Automatic Operations and Management of 5G Services in a Fixed Mobile Converged Packet-Optical Network
	15:20 - 15:40	Lu Zhang; Jiajia Chen; Lena Wosinska; Patryk Urban; Shilin Xiao; Weisheng Hu	Fixed and Mobile Convergence with Stacked Modulation
	15:40 - 16:00	Jiakai Yu; Yao Li; Mariya Bhopalwala; Sandip Das; Marco Ruffini; Daniel Kilper	Midhaul Transmission Using Edge Data Centers with Split PHY Processing and Wavelength Reassignment for 5G Wireless Networks
TS6: Elastic and Flexible optical networks	16:20 - 16:50	Michela Svaluto Moreolo; Josep M. Fabrega; Laia Nadal; Francisco Javier Vílchez [Invited]	Exploring the Potential of VCSEL Technology for Agile and High Capacity Optical Metro Networks
Session chair: Manos Varvarigos	16:50 - 17:10	Mirosław Klinkowski; Grzegorz Zalewski; Krzysztof Walkowiak	Optimization of Spectrally and Spatially Flexible Optical Networks with Spatial Mode Conversion
	17:10 - 17:30	Andrea Tomassilli; Brigitte Jaumard; Frederic Giroire	Path Protection in Optical Flexible Networks with Distance-adaptive Modulation Formats

	17:30 - 17:50	Rodrigo S Tessinari; Didier Colle; Anilton Salles Garcia	Cognitive Zone-Based Spectrum Assignment Algorithm for Elastic Optical Networks
	17:50 - 18:10	Jose-Juan Pedreno-Manresa; José Luis Izquierdo Zaragoza; Filippo Cugini; Pablo Pavon-Marino	On the Benefits of Elastic Spectrum Management in Multi-Hour Filterless Metro Networks
TS7: NFV	10:30 - 11:00	Shireesh Bhat; George N. Rouskas [Invited]	Open Marketplace and Service Orchestration for Virtual Optical Networks
Session chair: Mark De Leenheer	11:00 - 11:20	Tamal Das; Aniruddha Singh Kushwaha; Ashwin A Gumaste; Mohan Gurusamy	Leveraging Optics for Network Function Virtualization in Hybrid Data Centers
	11:20 - 11:50	Hiroaki Harai; Hideaki Furukawa; Yusuke Hirota [Invited]	Hardware-supported Softwarized and Elastic Optical Networks
	11:50 - 12:10	Leila Askari; Ali Hmaity; Francesco Musumeci; Massimo Tornatore	Virtual-Network-Function Placement for Dynamic Service Chaining in Metro-Area Networks
TS8: Optical access	10:30 - 10:50	Nicola Brandonisio; Daniel Carey; Stefano Porto; Giuseppe Talli; Paul Townsend	Burst-Mode FEC Performance for PON Upstream Channels with EDFA Optical Transients
Session chair: Marco Ruffini	10:50 - 11:10	Nejm Eddine Frigui; Tayeb Lemlouma; Stéphane Gosselin; Benoit Radier; Renaud Le Meur; Jean-Marie Bonnin	Optimization of the Upstream Bandwidth Allocation in Passive Optical Networks Using Internet Users' Behavior Forecast
	11:10 - 11:30	Tomoko Kamimura; Sumiko Miyata	Analysis of Mean Packet Delay in DR-MPCP Limited Service Using Queueing Theory
	11:30 - 11:50	Ahmed Helmy; Nitesh Krishna; Amiya Nayak	On the Feasibility of Service Composition in a Long-Reach PON Backhaul
	11:50 - 12:10	Abdulmalik Alwarafy; Mohammed Alresheedi; Ahmed Abas; Abdulhameed Alsanie	Performance Evaluation of Space Time Coding Techniques for Indoor Visible Light Communication Systems

TS9: Data Centre networks	14:00 - 14:30	Konstantinos Kontodimas; Kostas Christodoulopoulos; Emmanouel Varvarigos [Invited]	Resource Allocation in Slotted Optical Data Center Networks
Session chair: Anna Tzanakaki	14:30 - 14:50	Max Curran; Kai Zheng; Himanshu Gupta; Jon Longtin	Handling Rack Vibrations in FSO-based DataCenter Architectures
	14:50 - 15:20	Rui Lin; Joris Van Kerrebrouck; Xiaodan Pang; Michiel Verplaetse; Oskars Ozolins; Aleksejs Udalcovs; Lu Zhang; Lin Gan; Ming Tang; Songnian Fu; Richard Schatz; Urban Westergren; Sergei Popov; Deming Liu; Weijun Tong; Timothy De Keulenaer; Guy Torfs; Johan Bauwelinck; Xin Yin; Jiajia Chen [Invited]	Spatial Division Multiplexing for Optical Data Center Networks

Social Events:

Monday evening (May 14th):

- Visit to Book of Kells: A visit to Trinity's famous Old Library and a chance to see the Book of Kells that remains one of the world's most famous manuscripts with its ornately decorated pages written back around 800 AD.
- Conference Reception: the conference reception is to be held in the Trinity College's Atrium.

Wednesday evening (May 16th):

- Gala dinner: The conference gala dinner will be held at the Belvedere hotel. There will be a three-course dinner accompanied by a beautiful Irish dance show a great opportunity to experience traditional Irish music, dance and food. This spectacular show will give you a flavour of the different styles of Irish dance.

III

Papers

5. Regular papers

Path Protection in Optical Flexible Networks with Distance-adaptive Modulation Formats

A. Tomassilli*, B. Jaumard†, and F. Giroire*
*Université Côte d'Azur, CNRS, Inria Sophia Antipolis, France
†Concordia University, Montreal (Qc) Canada

***Abstract*—Thanks to a *flexible frequency grid*, Elastic Optical Networks (EONs) will support a more efficient usage of the spectrum resources. On the other hand, this efficiency may lead to even more disruptive effects of a failure on the number of involved connections with respect to traditional networks.**

In this paper, we study the problem of providing path protection to the lightpaths against a single fiber failure event in the optical layer. Our optimization task is to minimize the spectrum requirements for the protection in the network. We develop a scalable exact mathematical model using column generation for both shared and dedicated path protection schemes. The model takes into account practical constraints such as the *modulation format*, *regenerators*, and *shared risk link groups*. We demonstrate the effectiveness of our model through extensive simulation on two real-world topologies of different sizes. Finally, we compare the two protection schemes under different scenario assumptions, studying the impact of factors such as number of regenerators and demands on their performances.

I. Introduction

In an Elastic Optical Network (EON), data is distributed over a number of low data rate subcarriers without having to strictly follow the ITU-T fixed wavelength grid. In this way, with a data traffic more and more uncertain and heterogeneous, the spectrum resources can be used more efficiently and with a higher degree of flexibility [1].

With respect to a classical WDM network, EONs impose additional constraints on the structure of the optical path. Indeed, EONs require that contiguous frequency slots are allocated to each connection, which is also the main difference between the Routing and Spectrum Assignment (RSA) and Routing and Wavelength Assignment (RWA) problems. Thus, the already proposed RWA methods are not suitable for EONs. The RSA problem requires to find both an end-to-end optical path and a contiguous subset of frequency slots for each connection request.

Furthermore, EONs open up the possibility of exploiting multiple modulation formats for the different subcarriers. In such a way, the utilization efficiency could be further enhanced [2]. The problem of also determining a modulation format in addition to a routing path and a contiguous segment of spectrum is often referred to as the Routing, Modulation, and Spectrum Allocation (RMSA) problem. The problem is known to be NP-Hard even in the absence of modulation formats [3] and is challenging, even on small instances.

With the increasing efficiency in terms of resource usage, a link may accommodate a larger number of connections in EONs. Hence, the effects of a failure, such as a fiber cut, could be even more disruptive than in traditional networks. Network failures have been widely investigated (see e.g., [4], [5]). In the results of [4], each link experienced, on average, 16 failures per year. If not well managed, a failure may correspond to loss of service to users and loss of revenue. It is thus necessary to provide protection against failures in order to guarantee continuity of service and no violation of SLA requirements. We focus our attention on the *single link failure* scenario, since they are the predominant form of failures in optical networks [6].

Fault management techniques can be grouped into two categories: *restoration* and *protection*. In restoration, the network spare resources are used to reroute the connections affected by the failure. In protection, spare capacity is reserved in advance during connection setup. Restoration schemes use network spare resources more efficiently, but on the other hand, protection schemes have a faster restoration time and guarantee the recovery [7]. We thus study the latter schema.

In *dedicated protection*, there is no spectrum resources sharing between backup lightpaths. Each frequency slot is used for at most one lightpath. In *shared protection*, backup spectrum resources can be shared among different lightpaths if they fail independently. If, on one hand, in shared protection, spectrum resources are used more efficiently [8], on the other hand, in dedicated protection the recovery time is smaller. We thus study both protection schemes in this paper.

Another classification of the protection techniques can be made according to the recovery mechanisms. It could consist in a local repair (i.e., *link protection*) or in an end-to-end repair (i.e., *path protection*). Link protection schemes reroute the traffic only around the failed link. Path protection schemes reroute the traffic through a backup path if a failure occurs on its working path. With path protection, network resources are used more efficiently [6].

We consider the problem of *providing for each connection, a link-disjoint backup lightpath, under both dedicated and shared path protection schemes*. Our model also includes practical parameters such as the modulation format selection and the positions of regenerators. The modulation format of a lightpath adds a constraint on the maximum transmission distance, which may be extended by one or more regenerators if present in the route. One of the key concerns of the network operators is the efficient utilization of the deployed network capacity [1]. Our optimization goal is thus the minimization of the spectrum requirements for the protection.

In this paper, we propose two models for both dedicated and shared path protection against a single link failure. Our resolution strategy is based on a decomposition model using the column generation technique. We show that this technique is effective in dealing with the RMSA problem.

Our contributions can be summarized as follows:

- To the best of our knowledge, we are the first to propose

978-3-903176-07-2 © 2018 IFIP

a *scalable exact method* to solve the problem of providing path protection against a single link failure in elastic optical networks. The method is based on a decomposition model using column generation.
- The model takes into account practical constraints, such as multiple modulation formats, regenerators, and shared risk link groups.
- We compare the shared and dedicated path protection models and evaluate the tradeoff between the resolution time and the effectiveness, in terms of bandwidth utilization.
- We additionally study the impact of the number of regenerators in the network on the bandwidth requirements and on the latencies of both primary and backup lightpaths.

The rest of this paper is organized as follows. In Section II, we review related works in more detail. In Section III, we formally state the problem addressed in this paper. In section IV, we describe our column-generation-based model and show the subproblem to be solved in Section V. In Section VI, we validate our model by various numerical results on two real world topologies of different sizes. Finally, we draw our conclusions in Section VII.

II. Related Work

The problem of providing protection against failures in WDM networks has been widely investigated in the literature, see e.g., [6], [7], [8]. Nevertheless, not enough effort has been made in the context of EONs with multiple modulation formats and flexible spectrum allocation.

Dedicated path protection. The problem of off-line routing and spectrum allocation in flexible grid optical networks with dedicated path protection was studied in [9] and [10]. The optimization goal considered is to minimize the width of spectrum required in the network. In [9], the authors provide both an ILP formulation and a heuristic algorithm to solve the problem. In [10], an evolutionary algorithm metaheuristic is proposed with the aim to support the search for optimal solutions.

Shared path protection. Shared protection for EONs was considered in [11], [12], and [13]. A genetic algorithm metaheuristic with the goal to provide near optimal solutions to the problem of finding a primary and a backup path for each demand is proposed in [11]. The closest works to ours are [12] and [13]. The authors consider exact methods and propose ILP formulations for both dedicated and shared path protection, but with different optimization objectives. In [12], the authors minimize both the required spare capacity and the maximum number of frequency slots used in the network. In [13], the objective is to minimize the width of spectrum required in the network. They propose an ILP formulation in which each demand has a set of candidate pairs of link disjoint routing paths. The ILP model is able to deal with small networks (up to 9 nodes and 26 links). For larger networks, they propose heuristic algorithms based on both jointly and separated assignment of lightpaths to the demands.

Model Scalability. Previous works highlight the fact that finding an optimal or a near-optimal solution to the problem of jointly computing both a primary and a backup path for each demand is a challenging task, even for networks of small sizes and for a small number of demands. For instance, in [13] the authors show the benefits in terms of computing time and accuracy of computing the set of backup paths after the primary path allocation. In order to be able to deal with larger datasets, we adopt a two phase approach. First, we find a working path for each demand, then a backup path under both dedicated and shared protection schemes. We use the column generation technique as a solution approach, as results from [14] evidence the effectiveness of the column generation techniques in obtaining solutions for large instances of the RSA problem (but they do not consider protection against failures). Exact models proposed in the literature are only able to deal with small networks. On the contrary, our model is more scalable and we are able to solve instances with 24 nodes, 43 links, and 120 traffic requests. Moreover, we take into account regenerators and choices of modulation formats, which are not considered in the exact models of the literature.

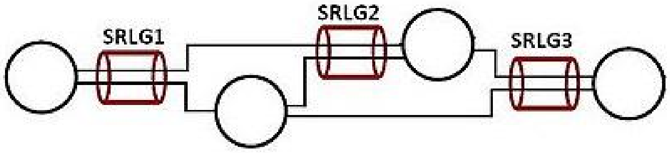


Fig. 1: An example of SRLG Constraints

III. Statement of the RMSA Protection Problem

The RMSA problem assumes an undirected graph $G = (V, L)$ with optical node set V and link set L. We denote by $\omega(v)$ the set of links adjacent to v, for $v \in V$. The bandwidth is slotted into a set S of frequency slots. The traffic is defined by a set K of requests where each request $k \in K$ has a source (s_k), a destination (d_k), and a spectrum demand D_k, expressed in terms of a number of frequency slots. The traffic is assumed to be symmetrical.

The provisioning of the primary lightpaths is given, and we are interested in finding both a dedicated and a shared path protection with minimum spectrum requirements, satisfying the spectrum contiguity and continuity constraints, as well as the following constraints:

• *Shared Risk Link Group (SRLG) constraints*, see Figure 1. Each SRLG constraint is defined by a set of links sharing a common resource, which affects all links in the set if the common resource fails. In the context of optical networks, it refers to a bundle of fiber links going through the same duct and that cannot be used simultaneously for primary and backup provisioning of the same demand. Let $\mathcal{F}$ be the set of all SRLG sets: $\mathcal{F} = \{F$: if ℓ and ℓ' both belong to F, then ℓ cannot be used in a path protecting ℓ' and vice versa $\}$.

• *Modulation constraints* The modulation format can be selected according to the traffic demand and the distance. We consider four modulation formats: BPSK (1 bit per symbol), QPSK (2 bits per symbol), 8QAM (3 bits per symbol), and 16QAM (4 bits per symbol) [15]. For instance, if, for a demand k, we have a request of 250 Gb/s (i.e., $D_k = 20$ assuming the bandwidth of a subcarrier slot as 12.5 GHz), then with BPSK $D_k^{BSPK} = 20$ and with 16QAM, $D_k^{16QAM} = 5$. We consider the following maximum transmission distances: BPSK (9,600 Km), QPSK (4,800 Km), 8QAM (2,400 Km), and 16QAM (1,200 Km). These values are based on the experimental results reported in [16]. Moreover, we assume that a subset of the nodes have regeneration capabilities. Indeed, decisions about the required equipment (i.e., transponders, regenerators, and switches) and its deployment are taken during the planning phase [17].

IV. Path Protection Models

We propose two column generation models relying on lightpath configurations for both dedicated and shared path protection schemes. In the rest of the paper, the two models will be referred to, respectively, as CG_DP and CG_SP.

A lightpath configuration, denoted by π, refers to a backup lightpath, i.e., a backup path, a spectrum slice with s as a starting frequency slot and a modulation format. Denote by Π the set of all possible backup lightpath configurations. Π is decomposed as follows:

$$\Pi = \bigcup_{k \in K} \Pi_k = \bigcup_{k \in K} \bigcup_{s \in S} \Pi_{ks},$$

where Π_k is the set of potentials lightpaths for provisioning request k, and Π_{ks} is the set of potential lightpaths for provisioning request k with a slot slice of width D_k^m according to the selected modulation format m such that s is a starting slot. Note that Π_k contains only *feasible backup lightpaths* for a demand k. We say that a backup lightpath is feasible for k if it does not contain any link in the same shared risk link group of some link of the primary lightpath for k. Each lightpath configuration, or lightpath for short, is denoted by π and is characterized by:

$b^{\pi}_{\ell s}$: indicates if slot s is used on link ℓ in the backup lightpath associated with π.

We assume that working lightpaths are known and described throughput the following parameter:

a^k_{ℓ}: indicates if the primary lightpath of request k goes through link ℓ.

The model uses the following decision variables:

$z_\pi = 1$ if lightpath $\pi \in \Pi$ is selected as a backup path, 0 otherwise.

$x_{\ell s} = 1$ if slot s is used on link ℓ in a backup path, 0 otherwise.

We denote with $\mathcal{LS}^{\text{B}}$ the pairs $(\ell, s) \mid \ell \in L, s \in S$ that can be used for protection, i.e., that are not used by the primary lightpaths.

The objective minimizes the spectrum requirements for the protection, and is written as follows:

$$\min \sum_{(\ell,s) \in \mathcal{LS}^{\text{B}}} x_{\ell s} \qquad (1)$$

Constraints are as follows:

$$\sum_{\pi \in \Pi_k} z_\pi \geq 1 \qquad k \in K \qquad (2)$$

$$z_\pi \in \{0,1\} \qquad \pi \in \Pi \qquad (3)$$

$$x_{\ell s} \in \{0,1\} \qquad \ell \in L, s \in S \qquad (4)$$

Model CG_DP

$$\sum_{k \in K} \sum_{\pi \in \Pi_k} b^{\pi}_{\ell s} z_\pi \leq x_{\ell s} \qquad \ell \in L, s \in S, (\ell, s) \in \mathcal{LS}^{\text{B}} \qquad (5)$$

Model CG_SP

$$\sum_{k \in K} a^k_{\ell'} \sum_{\pi \in \Pi_k} b^{\pi}_{\ell s} z_\pi \leq x_{\ell s} \qquad \ell, \ell' \in L, s \in S$$
$$\{\ell, \ell'\} \not\subseteq F : F \in \mathcal{F}, \ell \neq \ell', (\ell, s) \in \mathcal{LS}^{\text{B}} \qquad (6)$$

Constraint (2) ensures that each request is protected. Constraints (5) and (6) make sure that each slot is never used more than once on each backup fiber link. The difference between the two models relies on these constraints. In the dedicated protection case, two working paths cannot have backup paths going through the same link ℓ and slot s. On the other hand, in the shared protection case, two working paths that are not sharing any link ℓ' can use protection paths going through the same link ℓ and slot s.

V. Solution Design

Given the huge number of variables/columns in the proposed model, we resort to the *Column Generation* method to solve its Linear Programming (LP) relaxation, see, e.g., Chvatal [18], if not familiar with this technique. This technique consists of decomposing the original problem into a restricted master problem - RMP - (i.e., model (1) - (6) with a very restricted number of variables) and one or several pricing problems - PPs. RMP and PPs are solved alternately. Solving RMP consists in selecting the best lightpaths, while solving one PP allows the generation of an improving potential lightpath, i.e., a lightpath such that, if added to the current RMP, improves the optimal value of its LP relaxation. The process continues until the optimality condition is satisfied, that is, the so-called reduced cost that defines the objective function of the pricing problems is non negative for all of them. An ε-optimal solution for the RSA problem is derived by solving exactly the ILP model associated with the last RMP.

Let K_σ denote the set of requests that have the potential to be protected by a lightpath starting at slot σ: $K_\sigma = \{k \in K : \sigma + D_k - 1 \leq |S|\}$. Let D_k^σ be the number of slots needed for request k in K_σ: $D_k^\sigma = D_k$ for $k \in K_\sigma \; : \; \sigma + D_k - 1 = |S|$ and $D_k^\sigma = D_k + 1$ for $k \in K_\sigma \; : \; \sigma + D_k - 1 < |S|$.

Each pricing problem is indexed by a demand k and a starting slot σ, and produces a single potential lightpath for protecting demand k, starting at slot σ.

Definitions of the decision variables are as follows:

$y_\ell = 1$ if link ℓ is used, 0 otherwise

$x_{\ell s}$ indicates if slot s is used on link ℓ or not.

We first describe the model for shared protection. Let $u_k^{(2)}$ and $u^{(6)}_{\ell\ell' s}$ be the values of the dual variables associated with constraints (2) and (6), respectively. The pricing problem can be written as follows:

$$\min \quad 0 \; - \; u_k^{(2)} \; - \sum_{(s,\ell) \in S \times L} \; \sum_{\substack{\ell' \in L: \\ \ell \neq \ell'}} u^{(6)}_{\ell\ell' s} \; a^k_{\ell'} \; x_{\ell s} \qquad (7)$$

subject to:

$$\sum_{\ell \in \omega(s_k)} y_\ell = \sum_{\ell \in \omega(d_k)} y_\ell = 1 \qquad (8)$$

$$\sum_{\ell \in \omega(v)} y_\ell \leq 2 \qquad v \in V \setminus \{s_k, d_k\} \qquad (9)$$

$$\sum_{\ell' \in \omega(v) \setminus \{\ell\}} y_{\ell'} \geq y_\ell \qquad v \in V \setminus \{s_k, d_k\}, \ell \in \omega(v) \qquad (10)$$

$$\sum_{s=\sigma}^{\sigma + D_k^\sigma - 1} x_{\ell s} = D_k^\sigma \, y_\ell \qquad \ell \in L \qquad (11)$$

$$y_\ell, x_{\ell s} \in \{0,1\} \qquad \ell \in L, s \in S. \qquad (12)$$

Constraints (8), (9) and (10) define the routing of the current request. Constraint (11) reserves a contiguous spectrum channel for the current request.

We observe that for each link ℓ:
$x_{\ell s} = y_\ell$ for $s \in \{\sigma, \ldots, \sigma + D_k^\sigma - 1\}$
$x_{\ell s} = 0$ for $s \notin \{\sigma, \ldots, \sigma + D_k^\sigma - 1\}$.
Therefore, the reduced cost can be rewritten:

$$\min \quad 0 - u_k^{(2)} - \sum_{\ell \in L} \left(\sum_{\substack{\ell' \in L: \\ \ell \neq \ell'}} \sum_{s=\sigma}^{\sigma + D_k^\sigma - 1} u_{\ell\ell' s}^{(6)} \right) y_\ell.$$

The first term is a constant for each request, and the second term corresponds to a summation over the links of the network. Therefore, we can solve the pricing problem using the following objective function:

$$\min \quad -\sum_{\ell \in L} \left(\sum_{\substack{\ell' \in L: \\ \ell \neq \ell'}} \sum_{s=\sigma}^{\sigma + D_k^\sigma - 1} u_{\ell\ell' s}^{(6)} \right) y_\ell.$$

where $u_{\ell\ell' s}^{(6)}$ are non-positive dual values. We conclude that, for each request k, the lightpath generator corresponds to a weighted shortest-path problem with link weight: $-\sum\limits_{\ell' \in L: \ell \neq \ell'} \sum\limits_{s=\sigma}^{\sigma + D_k^\sigma - 1} u_{\ell\ell' s}^{(6)}$. As a result, the pricing problem when modulation and regenerators are not taken into account can be solved with a polynomial time algorithm, e.g., Dijkstra's algorithm.

In the dedicated protection case, the only difference lies in the objective function of the pricing problem, defined as:

$$\min \quad 0 - u_k^{(2)} - \sum_{(s,\ell) \in S \times L} u_{\ell s}^{(5)} \, x_{\ell s}$$

where $u_k^{(2)}$ and $u_{\ell s}^{(5)}$ are the values of the dual variables associated with constraints (2) and (5), respectively. As with the shared protection case, the problem can be reduced to finding a shortest path in a weighted graph.

Additional Modulation and Regenerators Constraints. However, if modulation is taken into account, we need to consider the maximum transmission distance constraint according to the considered modulation format. Also, a regenerator may extend the maximum reachable distance with respect to the chosen modulation format.

Each pricing problem is now indexed by a demand k, a starting slot σ, and a modulation format m, and produces a single potential lightpath for protecting demand k, starting at slot σ, if such a lightpath exists. In fact, some demands may not be satisfied, since the reachable distance is not long enough to reach the destination from the source, even in the presence of regenerators.

Regenerators add an additional layer of complexity to the problem. Indeed, without regenerators, for a demand (s, t), we could only consider to solve the subproblem for the modulation formats whose transmission reach is greater or equal to the length of the shortest path between s and t. With the presence of regenerators, this consideration does not apply, since the transmission reach may be increased.

When considering modulation constraints and nodes with regenerator capabilities, the pricing problem becomes a *Minimum-Weight Path Problem with a constraint on the path length.* The Minimum-Weight Constrained Path Problem is proven to be NP-Hard [19]. The problem has been widely studied and efficient algorithms have been proposed (see [20] for a survey on the subject).

Our solving strategy is described as follows. Pricing problems are solved using a modified version of the Label-setting algorithm for the Shortest Path Problem with Resource Constraints [20] based on the dynamic programming approach.

Given a weighted graph $G = (V, E)$, a demand (s, t), the maximum transmission distance according to the selected modulation format, and a set of nodes with regenerator capabilities $V_r \in V$, the algorithm starts from the trivial path $P = (s)$. It is then extended in all the feasible directions considering both the length of the links and the remaining transmission distance from the source s, which may have been increased because of the presence of one or more nodes in the set V_r in the considered path. For each path extension $P' \supset P$, a dominance algorithm is used in order to maintain only a Pareto-optimal set of paths or paths which can be extended to a Pareto-optimal one. When there are no more labels to be processed, the algorithm stops. A solution of minimum cost is selected from the set of all computed paths.

VI. Numerical Results

In this section, we evaluate the accuracy and performance of the proposed models through simulation on two networks of different sizes and according to different types of metrics. The results indicate that our models perform well, with an accuracy better than 1% for CG_DP and 20% for CG_SP in the considered networks. We also compare the performance of the dedicated and shared protection schemes, and show the tradeoff between the time needed to find a solution to the problem in the two cases and the savings in terms of bandwidth overhead.

Data Sets. We conduct experiments on two network topologies: `nobel-US` (14 nodes, 21 links) from SNDlib [21], and `USnet` (24 nodes, 43 links) from [22]. For `nobel-US`, the length of each link is calculated using the GPS coordinates of the nodes, according to the Cosine-Haversine formula. We assume that there is one pair of bidirectional fibers on each link, and the available spectrum width of each fiber is set to be 2000 GHz. We set the bandwidth of a subcarrier slot to 12.5 GHz. We considered four modulation formats: BPSK (binary phase-shift keying), QPSK (quadrature phase-shift keying), 8QAM (8-quadrature amplitude modulation), and 16QAM (16-quadrature amplitude modulation). Similarly, as in [23], we assume transmission distances of 9,600 km for BPSK (M = 1), 4,800 km for QPSK (M = 2), 2,400 km for 8QAM (M = 3), and 1,200 km for 16QAM (M = 4), where M denotes the number of bits per symbol. The number of considered nodes with regenerator capabilities is 5 for `nobel-US` and 10 for `USnet`. Locations are chosen according to the *betweenness centrality*, an index of the importance of an element in the network. It measures the extent to which a node lies on paths between other nodes. Primary paths are computed with the objective of minimizing the total number of used frequency slots in the network. All experiments are run on an Intel Xeon E5520 with 24GB of RAM.

Network	# traffic requests	# slots primary lightpaths	# generated columns		z_{LP}		$\tilde{z}_{\text{ILP}}$	
			CG_DP	CG_SP	CG_DP	CG_SP	CG_DP	CG_SP
nobel-US	20	164	8,735	12,875	292	171.05	292	201
	40	273	15,190	21,744	546	237.1	546	290
	60	457	19,128	28,316	816	328.82	816	430
USnet	40	344	26,828	40,931	574	339.6	574	431
	80	856	39,514	67,936	1,278	557.37	1,278	713
	120	1138	46,938	80,495	1,790	835.55	1,790	1,021

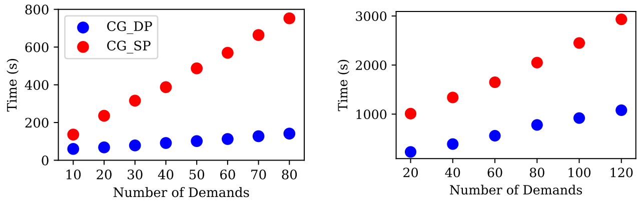
TABLE I: Numerical results for CG_DP and CG_SP.

Performance of CG Models. Table I summarizes the results of the two decomposition models for dedicated and shared protection on the two considered networks. We considered different numbers of demands. The load of each demand is randomly selected according to a uniform distribution within $50 - 200$ Gb/s.

A first difference can be observed in the number of generated columns, revealing the different level of complexity of the two models. This has an impact on the completion time, as can be observed in Figure 2. The large number of generated columns is also a consequence of our solving strategy. In fact, in order to accelerate the time needed to solve the RMP and to find an ILP solution to the last RMP, at each iteration, we remove nonbasic columns from the master problem according to their marginal cost. Thus, the number of iterations increases but, on the other hand, the time needed to find a solution decreases.

Another difference between the two models is the quality of the solution. CG_DP may require twice the number of frequency slots than CG_SP. This is a natural consequence of the different protection strategies. Moreover, the two models exhibit a different level of accuracy as expressed by the ratio of $(\tilde{z}_{\text{ILP}} - z_{\text{LP}})/z_{\text{LP}}$. In the case of CG_DP, it never exceeds 1%, while, for CG_SP, it may go up to 20%. The main reason for the difference in accuracy of the two models is the following. In CG_DP, to reduce the spectrum usage, the goal is to try to use short paths. This leads to fractional solutions with a small number of paths (and often a single one) for each demand. On the contrary, in CG_SP, the goal is to share backup paths as much as possible in order to reduce the value of the objective function. This leads to a large number of fractional paths per demand (sharing frequency slots with backup paths of several other demands) in the optimal fractional solution. The last RMP thus contains a large number of path variables with a nonzero value (often < 0.1) for each demand. Only one of them will be set to 1 per demand, when solving the last RMP as an ILP, leading to a larger gap.

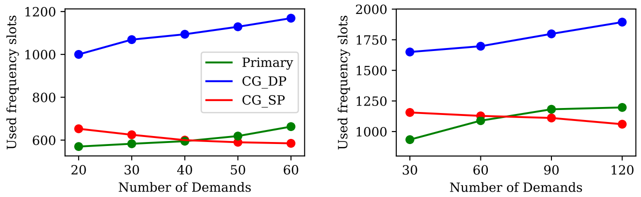
Shared vs. Dedicated Path Protection. We now compare the performances of the two protection schemes. In Figure 3, we study the impact of the number of demands on the resources required by the two protection schemes. We keep the total traffic intensity constant and vary the number of demands. The traffic is set to be 10 Tbps on `nobel-US` and 15 Tbps on `USnet`. As the results indicate, the two protection schemes exhibit a very different behavior. As the number of demands increases, the performance of the shared protection scheme, defined in terms of used frequency slots improves. On the other hand, both the primary lightpaths and the backup lightpaths computed according to the dedicated protection scheme, tend to require more resources as the number of demands becomes larger. This is not surprising, since an increasing number of demands improves the frequency slots' sharing opportunities of the lightpaths. In fact, in the shared protection scheme two link-disjoint primary lightpaths may share frequency slots in their backup paths. The benefits of shared over dedicated path protection is about 20% and 40% in the two networks according to the number of demands. Indeed, the benefits tend to increase with the number of considered demands. These results are similar to the ones reported by [12] and [13].



(a) nobel-US

(b) USnet

Fig. 2: Average completion time as a function of the number of demands

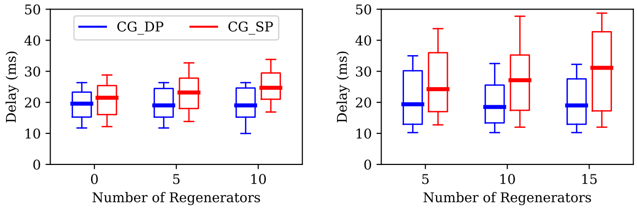


(a) nobel-US

(b) USnet

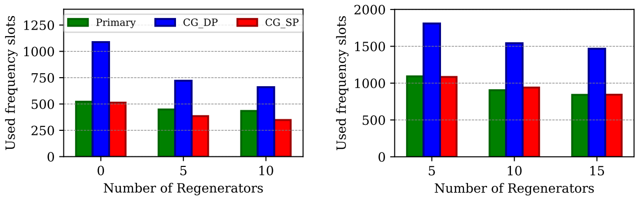
Fig. 3: Average number of frequency slots used as a function of the number of demands

Regenerators and Modulation Formats. Since, in optical networks, regenerators are costly, we are interested in evaluating the impact of the number of regenerators on the lightpaths. In Figures 4 and 5, we study the impact of the number of regenerators on the paths' latencies and on the spectrum requirements for the protection. We consider 50 demands for `nobel-US` and 100 demands on `USnet`. As the number of nodes with regeneration capabilities increases, from 0 to 10 for `nobel-US` and from 5 to 15 for `USnet` (Fig. 5), the spectrum requirements of the primary lightpaths and of the backup lightpaths decrease in both protection schemes. The reason is that a higher number of regenerators allows the lightpaths to use better modulation formats (in terms of bits per symbol) and consequently to use fewer resources. However, when considering lightpaths' latencies, the two protection schemes behave surprisingly in a strikingly different way. While, in the dedicated protection case, backup lightpaths' latencies tend to decrease, in the shared protection case, we observe the reverse phenomena. The explanation is the following. In dedicated protection, backup paths cannot be shared and, thus, the only means to reduce the number of used frequency slots is to use shorter paths. This is what happens when increasing the number of regenerators. Indeed, both primary and backup lightpaths need fewer resources, as they may now use more efficient modulation formats. This leads to increased spare capacity, allowing backup paths to use shorter routes. In shared protection, the situation is different. Indeed, there are *two* ways to reduce the spectrum usage: shorter paths as for DP, but also increased sharing of backup paths. The second way happens to be predominant in our experiments: regenerators allow better



(a) nobel-US (b) USnet

Fig. 4: Path delay distributions under the two protection schemes vs. the number of regenerators.



(a) nobel-US (b) USnet

Fig. 5: Average number of frequency slots used as a function of the number of regenerators

modulation formats and longer routes, leading to better sharing opportunities. As a consequence, the spectrum requirements are reduced, but this comes at the cost of increased lightpath lengths. However, the maximum delay of the backup paths in the shared protection case never exceeds 50 ms, the value often chosen as the maximum allowed delay for a route in networks [24]. As the results indicate, particular attention should be paid to lightpaths' latencies when considering shared path protection, in order not to violate the SLA requirements. Indeed, with the spectrum resources as optimization task, the possibility to share resources may lead to longer paths at the expense of the delay. Note that we could also easily add a constraint in the pricing problem in order to consider only lightpaths under a certain delay requirement.

VII. Conclusion

In this paper, we investigated the problem of providing path protection against a single link failure in elastic optical networks. We presented two decomposition models for both dedicated and shared path protection schemes taking into consideration modulation, regenerators, and shared risk link group constraints. Through extensive simulation, we showed the effectiveness of our models in finding a solution in a reasonable amount of time. Moreover, we studied different metrics in order to compare the accuracy of those models, showing the tradeoff in terms of required bandwidth and latency with the time resources needed by the two protection schemes. Our future works include the further improvement of the model precision and scalability, in order to be able to deal with larger and more complex instances of the problem.

References

[1] M. Jinno, H. Takara, B. Kozicki, Y. Tsukishima, Y. Sone, and S. Matsuoka, “Spectrum-efficient and scalable elastic optical path network: architecture, benefits, and enabling technologies,” *IEEE Communications Magazine*, vol. 47, no. 11, 2009.

[2] M. Jinno, B. Kozicki, H. Takara, A. Watanabe, Y. Sone, T. Tanaka, and A. Hirano, “Distance-adaptive spectrum resource allocation in spectrum-sliced elastic optical path network [topics in optical communications],” *IEEE Communications Magazine*, vol. 48, no. 8, 2010.

[3] M. Klinkowski and K. Walkowiak, “Routing and spectrum assignment in spectrum sliced elastic optical path network,” *IEEE Communications Letters*, vol. 15, no. 8, pp. 884–886, 2011.

[4] D. Turner, K. Levchenko, A. C. Snoeren, and S. Savage, “California fault lines: understanding the causes and impact of network failures,” in *ACM SIGCOMM Computer Communication Review*, 2010.

[5] G. Iannaccone, C.-n. Chuah, R. Mortier, S. Bhattacharyya, and C. Diot, “Analysis of link failures in an ip backbone,” in *Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurment*. ACM, 2002.

[6] S. Ramamurthy, L. Sahasrabuddhe, and B. Mukherjee, “Survivable wdm mesh networks,” *Journal of Lightwave Technology*, vol. 21, no. 4, 2003.

[7] L. Sahasrabuddhe, S. Ramamurthy, and B. Mukherjee, “Fault management in ip-over-wdm networks: Wdm protection versus ip restoration,” *IEEE journal on selected areas in communications*, vol. 20, no. 1, 2002.

[8] D. Zhou and S. Subramaniam, “Survivability in optical networks,” *IEEE network*, vol. 14, no. 6, pp. 16–23, 2000.

[9] M. Klinkowski and K. Walkowiak, “Offline rsa algorithms for elastic optical networks with dedicated path protection consideration,” in *Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), 2012 4th International Congress on*. IEEE, 2012.

[10] M. Klinkowski, “An evolutionary algorithm approach for dedicated path protection problem in elastic optical networks,” *Cybernetics and Systems*, vol. 44, no. 6-7, pp. 589–605, 2013.

[11] ——, “A genetic algorithm for solving rsa problem in elastic optical networks with dedicated path protection,” in *International Joint Conference CISIS12-ICEUTE´ 12-SOCO´ 12 Special Sessions*. Springer, 2013, pp. 167–176.

[12] G. Shen, Y. Wei, and S. K. Bose, “Optimal design for shared backup path protected elastic optical networks under single-link failure,” *Journal of Optical Communications and Networking*, vol. 6, no. 7, 2014.

[13] K. Walkowiak and M. Klinkowski, “Shared backup path protection in elastic optical networks: Modeling and optimization,” in *Design of Reliable Communication Networks (DRCN), 2013 9th International Conference on the*. IEEE, 2013, pp. 187–194.

[14] M. Ruiz, M. Pióro, M. Żotkiewicz, M. Klinkowski, and L. Velasco, “Column generation algorithm for rsa problems in flexgrid optical networks,” *Photonic network communications*, vol. 26, no. 2-3, 2013.

[15] K. Christodoulopoulos, I. Tomkos, and E. Varvarigos, “Elastic bandwidth allocation in flexible ofdm-based optical networks,” *Journal of Lightwave Technology*, vol. 29, no. 9, pp. 1354–1366, 2011.

[16] A. Bocoi, M. Schuster, F. Rambach, M. Kiese, C.-A. Bunge, and B. Spinnler, “Reach-dependent capacity in optical networks enabled by ofdm,” in *Proc. Optical Fiber Communication (OFC), 2009*. IEEE.

[17] A. Kretsis, K. Christodoulopoulos, P. Kokkinos, and E. Varvarigos, “Planning and operating flexible optical networks: Algorithmic issues and tools,” *IEEE Communications Magazine*, vol. 52, no. 1, 2014.

[18] V. Chvatal, *Linear Programming*. Freeman, 1983.

[19] M. R. Garey and D. S. Johnson, *Computers and intractability*. wh freeman New York, 2002, vol. 29.

[20] S. Irnich and G. Desaulniers, “Shortest path problems with resource constraints,” *Column generation*, pp. 33–65, 2005.

[21] S. Orlowski, R. Wessäly, M. Pióro, and A. Tomaszewski, “Sndlib 1.0survivable network design library,” *Networks*, vol. 55, no. 3, 2010.

[22] B. Mukherjee, *Optical WDM networks*. Springer Science & Business Media, 2006.

[23] Z. Zhu, W. Lu, L. Zhang, and N. Ansari, “Dynamic service provisioning in elastic optical networks with hybrid single-/multi-path routing,” *Journal of Lightwave Technology*, vol. 31, no. 1, pp. 15–22, 2013.

[24] F. Giroire, A. Nucci, N. Taft, and C. Diot, “Increasing the robustness of ip backbones in the absence of optical level protection,” in *INFOCOM 2003*. IEEE, 2003, pp. 1–11.

Fixed and Mobile Convergence with Stacked Modulation

Lu Zhang, *Student Member*, *IEEE*, Jiajia Chen, *Senior Member*, *IEEE*, Lena Wosinska, *Senior Member*, *IEEE*, Patryk J. Urban, *Senior Member*, *IEEE*, Shilin Xiao, Weisheng Hu

***Abstract*—We propose a stacked modulation mechanism to realize the transmission convergence of fixed broadband and wireless access networks. As an integrated solution, the direct current biased analog signal for wireless data is multiplied by the digital optical signal carrying fixed broadband service, where the low-speed analog signal becomes an envelope of the high-speed digital signal. At the receiver, the analog signal is firstly recovered with the image edge detection algorithm, after which the digital signal is recovered with the least-square algorithm. This scheme can realize fixed and mobile convergence with a single wavelength, which reduces the access network cost in terms of number of transceivers. Besides, the link capacity and spectrum efficiency are also improved. Moreover, this method is compatible with the existing access networks since the transceivers can be adopted to the scenario by turning off the unused modules. Simulation results show that in case of intensity modulation direct detection system sensitivity, penalty of only 1~2 dB can be obtained for the 10Gbps/λ passive optical network when broadband access services are favored. Meanwhile, the wireless service needs more power than the broadband service to achieve the QoT requirements, and it has a higher penalty (~ 5 dB) compared with broadband access services.**

***Index Terms*—Fixed and mobile convergence, stacked modulation, optical access network, wireless access network.**

I. Introduction

WITH the massive deployment of novel wired and wireless applications, the high bandwidth requirements have put great pressure on communication infrastructure. To solve the capacity bottleneck, optical fiber communication is considered as an attractive technology to enable fixed and mobile convergence (FMC) [1-5], especially point-to-point (PtP) and wavelength division multiplexing passive optical network (WDM-PON), which provide a dedicated channel with ultra-high data rate. FMC allows operators to offer various services through a common infrastructure and hence leads to a great potential for cost saving.

The current research on FMC [4] is mainly addressing the network layer, focusing on the sharing and reconfiguring network infrastructure and functional units to realize the convergence. However, the transmission equipment to support fixed broadband and mobile services is still developed separately. One of the key reasons is that the fixed broadband and mobile systems utilize different transmission technology. For instance, digital signals, such as on-off keying (OOK), are often used for fixed broadband access, while in radio access network analog signals are recognized as promising candidates [5]. Although the fixed and mobile services can be carried out in the common network infrastructure, they need independent channels for their own transmissions. Thus, it is not possible to share the transmission equipment, leading to high cost and low flexibility. To solve these challenges, two major convergence technologies are proposed in physical layer based on digital and hybrid transmission.

The digital transmission convergence solution attempts to digitize the wireless analog signal and then transmit it in the optical fiber, e.g., enhanced common public radio interface (eCPRI) [6] for fronthaul, which can be further multiplexed with the digital fixed broadband data by using time division multiplexing (TDM) or WDM techniques [7-8]. Here some costly devices for high-speed analog-to-digital converter (ADC) and digital-to-analog converter (DAC) are needed. Besides, the multiplexing of digitized wireless and wired signals requires a very high capacity fiber links and worsens the latency performance. On the other hand, the hybrid transmission convergence technology is realized by overlaid modulation techniques [9-11] or simply putting together the baseband optical data and analog wireless data [12,13], which are multiplexed in frequency domain. Overlaid modulation is realized by modulating low speed digital data on top of the high-speed analog data, such as cascading optical lasers [9], modulating control or monitoring signal over analog data signals [10-11], etc. However, specially designed signals for FMC are utilized in [9-13], which is not efficient in the real deployment scenarios where the challenges, such as the modulation crosstalk interference and system inefficiency due

This work is supported by Swedish Foundation for Strategic Research (SSF), Göran Gustafsson Stiftelse, Swedish SRA ICT–TNG project, National Natural Science Foundation of China (#61605047, 61671212, 61550110240, 61271216, 61221001, 61431009, 61775137), and National "863" Hi-tech Project of China.

L. Zhang is with the Optical Network Laboratory, KTH Royal Institute of Technology, Kista 16440, Sweden; and also with the State Key Laboratory of Advanced Optical Communication System and Networks, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: lu8@kth.se; luzhang_sjtu@sjtu.edu.cn).

J. Chen and L. Wosinska are with the Optical Network Laboratory, KTH Royal Institute of Technology, Kista 164 40, Sweden (e-mail: jiajiac@kth.se; wosinska@kth.se;).

P. J. Urban is with the Ericsson Research, Ericsson AB, Stockholm 164 83, Sweden (e-mail: patryk.urban@ericsson.com; Patryk.j.urban@ieee.org).

S. Xiao, W. Hu are with the State Key Laboratory of Advanced Optical Communication System and Networks, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: slxiao@sjtu.edu.cn; wshu@sjtu.edu.cn).

978-3-903176-07-2 © 2018 IFIP

to the additional overhead need to be addressed.

In this paper, we propose to stack analog modulation on top of the digital signals. It leads to a hybrid transmission of both analog and digital signals over a single wavelength channel, which not only enhances the link capacity and spectrum efficiency, but also reduces the network cost compared with two separated transmission systems for fixed broadband and mobile services. Besides, some digital signal processing (DSP) modules can also be switched off to allow compatibility with the legacy infrastructure. The simulation results show that the proposed scheme utilizing an intensity modulation direct detection (IM/DD) system only causes a minor sensitivity penalty (1~2 dB) for the 10Gbps/λ PON after optimization for broadband access services. At the same time, the wireless service needs more power than the broadband service to achieve the QoT requirements, and it has a higher penalty (~ 5 dB) compared with broadband access services.

II. Operational Principles

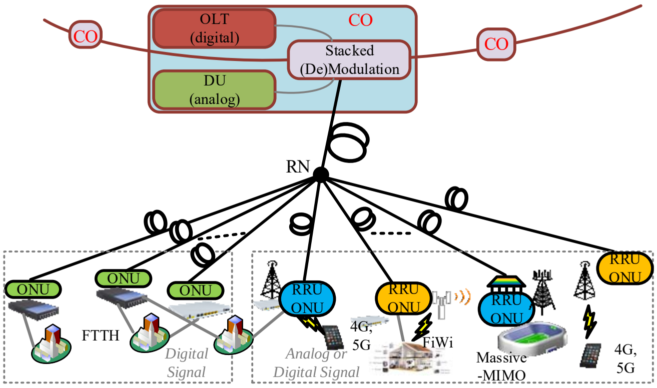


Fig. 1. Schematic architecture supporting FMC by using the proposed stacked modulation scheme.

Fig. 1 shows the schematic architecture supporting FMC by utilizing the proposed stacked modulation scheme. As shown in Fig. 1, the wireless analog signal at distributed units (DU) and optical digital signal at optical line terminals (OLT) are stacked modulated at the integrated central office (CO). The stacked modulated signal is injected into the feeder fiber and broadcasted at the remote node by optical splitter. At each optical network unit (ONU) or wireless access point (e.g., remote radio unit (RRU), base station), the digital or analog signal is received accordingly. For instance, FTTH users (ONUs) would need digital signal transmission, while RRU nodes prefer to deal with analog signals. The optical and radio signals from the user side can also utilize this stacked modulation for the upstream transmission. The system architecture with stacked modulation are described in *Section. II. A*, while the transmitter and receiver design of stacked modulator along with the corresponding mathematical expressions are presented in *Section. II. B and C,* respectively.

A. System Architecture

Fig. 2 shows the architecture of a stacked modulation system for both upstream and downstream transmission. In the downstream direction, the biased wireless access signal (e.g. orthogonal frequency division multiplexing, OFDM) is multiplied by a digital signal coming from the fixed broadband access network (e.g. on-of-keying, OOK), then the multiplied signal is modulated to optical domain with the optical modulator. The signals from different transmitters are multiplexed by WDM. The remote node (RN) delivers the signal to each ONU. Considering NG-PON2, where WDM-PON variant is included, the RN in Fig. 2 can either be power splitter which is compatible with the legacy PON deployment or wavelength-sensitive component (like arrayed waveguide grating AWG) that can split wavelengths to different ONUs. If the splitter is used at the RN, at the receiver a filter is required before photo-detector. The analog signal is first recovered by envelope detection, then, the digital signal is recovered by least-square solution. The principle can also be applied to the upstream, where the stacked signal is transmitted to the central office and recovered by the proposed demodulation schemes.

B. Transmitter Design

Without loss of generality, we consider OFDM as an example for analog signal and OOK (realized by non-return to zero, NRZ format) as an example of digital signal here. Fig. 3 (a) shows the scheme of transmitter based on the proposed stacked modulation for FMC. In such a transmitter, the wireless user data in the frequency domain is modulated by quadrature amplitude modulation (QAM) format. After QAM modulation and subcarriers mapping, data is forwarded to inverse fast Fourier transform (IFFT) implementation. After that, the cyclic prefix (CP) is inserted in front of the multicarrier signal, and then the parallel to series converted (P/S) time-domain signal $\boldsymbol{V_a}$ is multiplied by a extinction factor $\boldsymbol{\alpha}$ to adjust the power ratio between the analog and the digital signal. Then, $\boldsymbol{\alpha V_a}$ is added with a direct current (DC) offset $\boldsymbol{V_{DC}}$ to make the

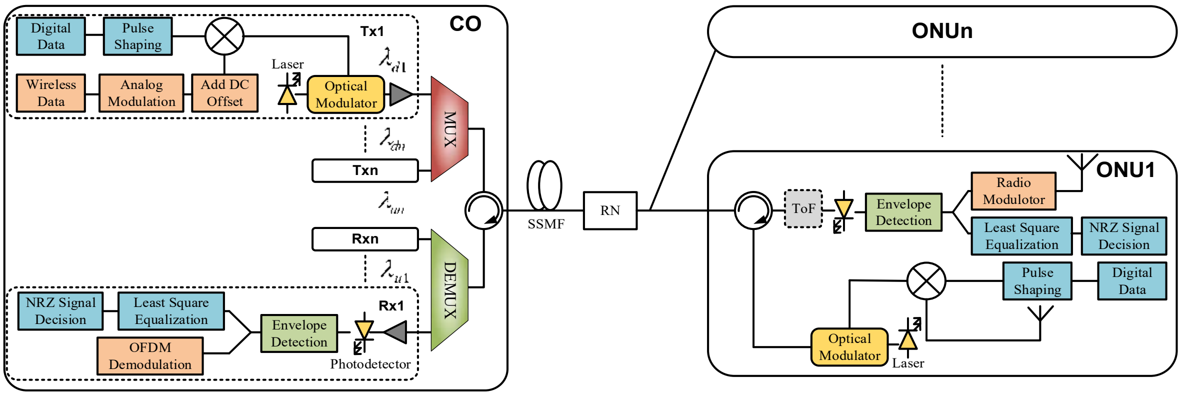


Fig. 2. System architecture of stacked modulation.

minimum value of analog signal larger than the minimum value of digital signal $\boldsymbol{V_d}$. Finally, the analog signals and digital signals are multiplied to generate the stacked modulated signals $\boldsymbol{V_t}$. The mathematical expression is shown in Eq. 1.

$$\boldsymbol{V_t} = (\boldsymbol{V_{DC}} + \alpha \boldsymbol{V_a}) * \boldsymbol{V_d} \tag{1}$$

To be able to also generate pure analog or digital signals, there are two decision modules. First, if the transmitter needs to generate pure analog signals, it directly jumps to the DAC module to generate analog signals. In the second decision module, if the transmitter only needs to generate digital signals, it directly jumps to the DAC module to generate the digital signal. Thus, in case FMC is not needed, this transmitter can also generate either pure digital or analog signals.

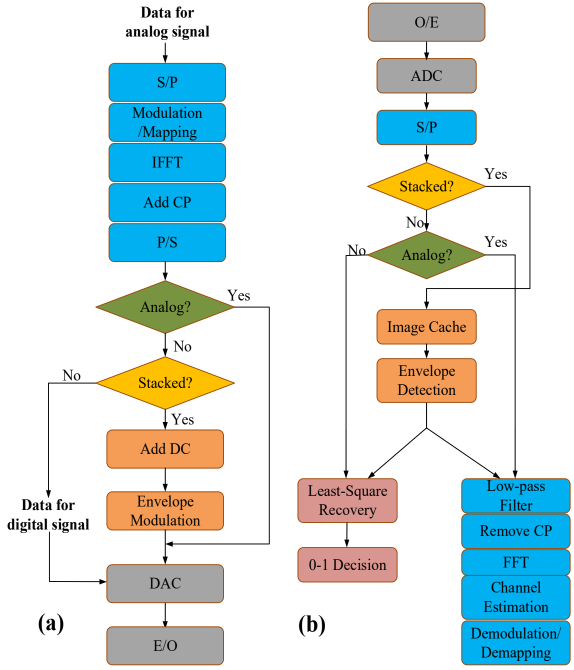


Fig. 3. (a) Transmitter design for the proposed stacked modulation, (b) Receiver design for the proposed stacked modulation.

C. Receiver Design

The stacked modulated signals are transmitted into the fiber, detected by photoelectric detector, and converted to electrical domain. It can be expressed by Eq. 2 where the channel distortions $\boldsymbol{h(t,f)}$ and noise $\boldsymbol{V_w}$ induced by chromatic dispersion and fiber transmission attenuation are considered. In the mathematical model shown in Eq. 2, ⊗ means the convolution for channel response.

$$\boldsymbol{V_r} = \boldsymbol{V_t} \otimes \boldsymbol{h}(\boldsymbol{t}, \boldsymbol{f}) + \boldsymbol{V_w} \tag{2}$$

Fig. 3 (b) shows the receiver design to support the proposed stacked demodulation for FMC. At the receiver side, the analog signals are first abstracted from the received signal $\boldsymbol{V_r}$ by image edge detection algorithm. More specifically, the symbol of the stacked modulated signals is stored as a two-valued image by image cache, and then the envelope of the stacked signals, i.e., the DC-biased analog signals, are recovered by edge detection $\boldsymbol{E}\{.\}$. In this paper, Canny edge detection algorithm [14] is adopted, which followed a list of criteria, such as low error rate, well localized edge points, one response to a single edge, to improve accuracy edge detection. Based on these criteria, the Canny edge detector first soothes the two-valued image of stacked modulated signal to eliminate the noise. Then, it finds the image gradient to recognize regions with high spatial derivatives. The algorithm then tracks along these regions and suppresses any pixel that is not at the maximum. After that, hysteresis is used to track along the remaining pixels that have not been suppressed. Finally, the edge of stacked modulated signal is abstracted and converted to the analog signal after normalization. The algorithm flow is shown in Fig. 4.

```
Algorithm 1:
Input: two-valued image of stacked modulated signal V_r
Output: envelope of stacked modulated signal (analog signal) V_a'
Step 1: Apply Gaussian filter to smooth the image;
Step 2: Recognize regions with high spatial derivatives;
Step 3: Tracks along the regions in Step 2 and suppresses any pixel that is not at
the maximum;
Step 4: Hysteresis:
        Finalize the edge detection E{.} by suppressing the other weak edges .
end
```

Fig. 4. Description of edge detection algorithm.

The recovered analog signals are modulated to the radio frequency, and finally transmitted to the wireless user by antenna. After removing the envelope of the recovered signal, a least-square algorithm is used to recover the digital signals. The mathematic model is shown in Eq. 3.

$$\boldsymbol{V_s} = \begin{cases} \boldsymbol{V_r}(\boldsymbol{V_{DC}} + \alpha \boldsymbol{E}\{\boldsymbol{V_r}\})^{-1}, \text{digital} \\ \boldsymbol{E}\{\boldsymbol{V_r}\} - \boldsymbol{mean}(\boldsymbol{E}\{\boldsymbol{V_r}\}), \text{analog} \end{cases} \tag{3}$$

Meanwhile, demodulation is also made for evaluating the upstream performance of analog signals, e.g., OFDM, the serial-to-parallel (S/P) converted signal first passes through the low pass filter and removes the CP, and then it is sent to the FFT implementation. After subcarriers de-mapping, the pilot based channel estimation (CE) method in our previous work [15] is used. After the CE, the symbols are sent to the QAM demodulation module for user data recovery.

Similar as the transmitter design, there are also two decision modules in the receiver which get the recognition of the modulation types from the network control layer. First, if the signals are stacked modulated, the image edge detection algorithm is utilized to recover the analog signals and then recover the digital signals. If purely analog signal is considered, it directly jumps to the branch for analog signal demodulation. Otherwise, it directly jumps to the branch for digital signal demodulation (e.g. 0-1 decision) module.

III. PERFORMANCE EVALUATION

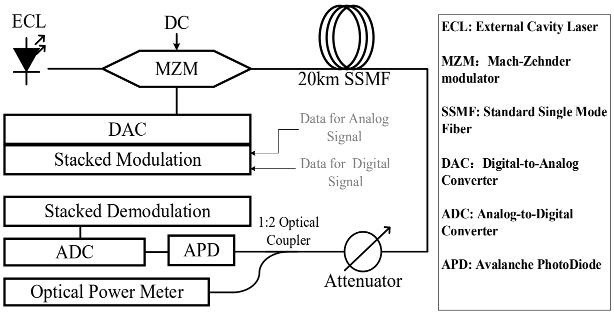

ECL
DC
MZM
20km SSMF
DAC
Stacked Modulation
Data for Analog Signal
Data for Digital Signal
Stacked Demodulation
1:2 Optical Coupler
ADC
APD
Attenuator
Optical Power Meter
ECL: External Cavity Laser
MZM: Mach-Zehnder modulator
SSMF: Standard Single Mode Fiber
DAC: Digital-to-Analog Converter
ADC: Analog-to-Digital Converter
APD: Avalanche PhotoDiode

Fig. 5. Simulation setup

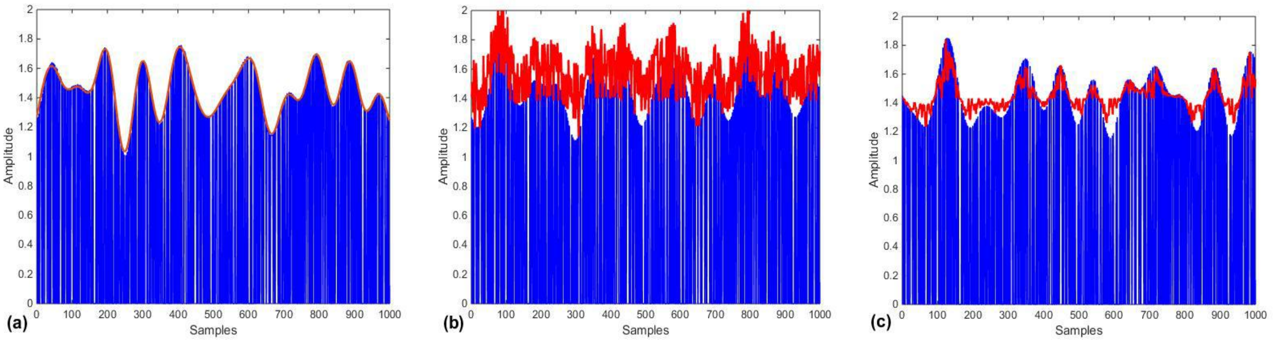


Fig. 6. Envelope detection with (a) image edge detection, (b) bandpass FIR filter and (c) Hilbert transform method. (the original stacked modulated signal (blue curve) and the one detected by different methods (red curve))

To demonstrate the feasibility of the proposed solution, we have implemented our approach in the simulation tools using *MATLAB* 2014b and *Optisystem* 7.0. The OOK signal is randomly generated with 10Gbps data rate. The IFFT point of OFDM signal is 16384 and analog signal bandwidth is 200MHz. The QAM order is 16. The simulation is carried out by optical intensity modulation and direct detection transmission (IM/DD) system. The simulation setup is shown in Fig. 5. The amplified signals are downloaded to the DAC. A MZM is used for modulation with an ECL (15dBm, 1550 nm). Optical signal is transmitted over 20km SSMF. At the receiver, a variable optical attenuator (VOA) is used to change the received optical power for bit-error ratio (BER) measurements. The signal is detected by an APD. Then, the signal is captured by ADC.

First, to confirm the feasibility and effectiveness of the envelope detection in stacked modulation scheme to distinguish the OFDM signals from OOK signals, we test the image edge detection algorithm and compare it with band-pass FIR filter method [4] and Hilbert transform method at ideal channels (i.e. no channel noise). The results are shown in Fig. 6. For comparison, we plot the original stacked modulated signal (blue) and the one detected by different methods (red) in each sub-figure in Fig. 6. It can be easily seen that image edge detection algorithm can perfectly recover the analog OFDM envelope. Moreover, its complexity and overhead are comparable with the other methods. On the other hand, band-pass FIR filter method and Hilbert transform method cannot perform as good as the proposed image edge detection algorithm. Since the multiplication will bring frequency overlap, simple filtering operations cannot filter out the distortions, and the last two schemes cannot mitigate the modulation crosstalk.

Secondly, to evaluate the influence of extinction factor $\boldsymbol{\alpha}$ and DC offset $\boldsymbol{V_{DC}}$ on the stacked modulation format, we have shown the BER performance of OOK and OFDM in terms of α and $\boldsymbol{V_{DC}}$ in Fig. 7. The received optical power levels are -32dBm and -24dBm, respectively. With the increase of extinction factor $\boldsymbol{\alpha}$, the performance of OFDM becomes better since it has more power in the stacked modulation, while the performance of OOK becomes worse. With a smaller value of $\boldsymbol{\alpha}$, the performance of OOK becomes better since its power ratio is larger and the waveform becomes flatter. For DC offset $\boldsymbol{V_{DC}}$, the trend is opposite. When the DC offset $\boldsymbol{V_{DC}}$ increases, there is more power allocated to OOK signals and its performance is enhanced. When $\boldsymbol{V_{DC}}$ decreases, the power of OFDM is higher, and it outperforms the OOK in the stacked modulation. Overall, there is a clear trade-off between digital and analog transmission in stacked modulation format. In the real implementation, the power ratio between digital and analog signals can be flexibly configured to meet requirements on quality of transmission (QoT). If QoT requirement of analog signal is high, then α should be increased and $\boldsymbol{V_{DC}}$ should be decreased. Otherwise, α should be decreased and $\boldsymbol{V_{DC}}$ should be increased.

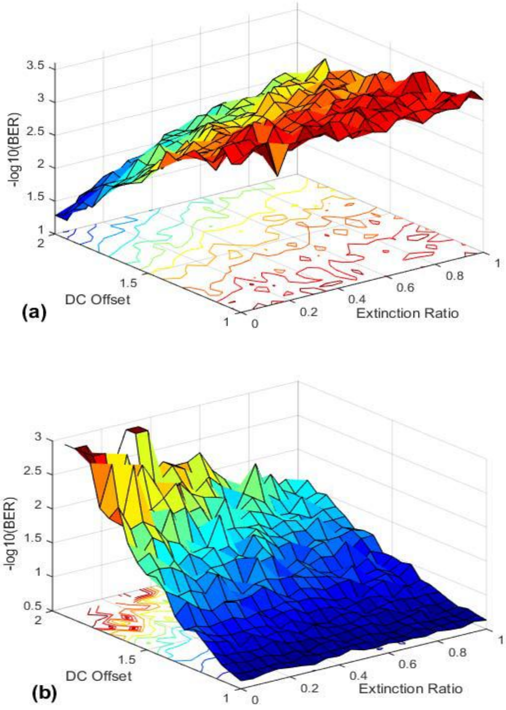


Fig. 7. BER performance of (a) OOK and (b) OFDM in terms of $\boldsymbol{\alpha}$ and $\boldsymbol{V_{DC}}$.

Finally, to demonstrate the impact of stacked modulation on the existing PON and wireless access network (e.g. mobile fronthaul) transmissions, the BER versus received optical power of OFDM and OOK are shown in Fig. 8 and 9, respectively. Here, the extinction factor $\boldsymbol{\alpha}$ and DC offset $\boldsymbol{V_{DC}}$ are chosen to give more power to OOK transmission in PON, which is shown as an example for flexible QoT choice. The common schemes of digital and analog modulations are also realized by turning off modules in stacked modulations. It can be seen from Fig. 8 and Fig. 9 that the power penalty of stacked modulation compared to purely OOK transmission in 10Gbps/λ PON is 1 ~ 2 dB at hard-decision forward error correction (HD-FEC) threshold at BER=3.8e-3 [16], which will not obviously influence the power splitting ratio and access point numbers of the legacy PON infrastructure. Besides, it is also shown that considering physical layer constraints, the eCPRI (digital) transmission [6] can also be supported by stacked modulation.

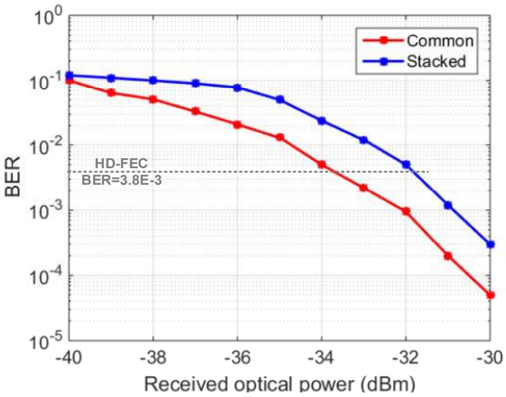


Fig. 8. BER versus optical received power of OOK.

For wireless access network, the power penalty here is 4 ~ 5 dB at HD-FEC threshold. This is because that there is less power allocated to OFDM signals, which also fits the performance analysis of Fig. 7. To improve the QoT of wireless access networks, more power should be allocated to the analog signals. Since the main idea here is suggesting to use PON to load the wireless services while holding the fixed broadband services, QoT requirements of PON is comparatively higher than wireless access network. Otherwise, the parameters can also be adjusted to meet the QoT of wireless access network. Besides, the transmission performance of the envelope of the stacked modulation is also comparable with the purely analog wireless access network, which can also reduce the capacity requirement compared to eCPRI.

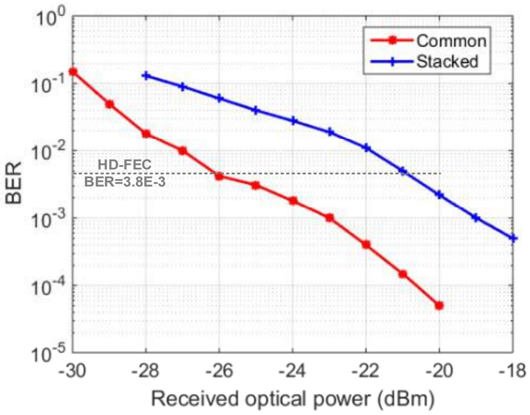


Fig. 9. BER versus optical received power of OFDM.

IV. Conclusions

We propose a stacked modulation mechanism to realize the transmission convergence of fixed broadband and wireless access. The solution is applicable for both upstream and downstream transmission. As an integrated solution, it stacks the analog multicarrier signals on top of the digital signals. This scheme is also compatible with the existing transceivers in wireless and optical access networks. It offers several benefits, such as significantly reduced system cost and enhanced spectrum efficiency, while according to our simulation results the sensitivity penalty for the 10Gbps/λ PON is only 1~2 dB after optimization for broadband services. If the QoT of wireless access is higher, the performance of wireless access can also be improved by increasing the power ratio of wireless signals in stacked modulation.

References

[1] S. Gosselin, A. Pizzinat, X. Grall, D. Breuer, E. Bogenfeld, S. Krauss, J. Alfonso Torrijos Gijon, A. Hamidian, N. Fonseca, and B. Skubic, “Fixed and mobile convergence: which role for optical networks,” IEEE/OSA Journal of Optical Communications & Networking, vol. 7, no.11, pp.1075-1083, 2015.

[2] M. Tornatore, G. K. Chang, and G. Ellinas, “Fiber-Wireless Convergence in Next-Generation Communication Networks,” Springer International Publishing, 2017.

[3] H. S. Nam, and N. I. Park, “Design of Architecture and Signal for Collaboration in Fixed & Mobile Convergence Network,” In Conf. IEEE Collaboration Technologies and Systems (CTS), pp. 574-577, 2016.

[4] J. I. Kani, J. Terada, K. I. Suzuki, and A. Otaka, “Solutions for future mobile fronthaul and access-network convergence,” Journal of Lightwave Technology, vol. 35, no. 3, pp. 527-534, 2017.

[5] ITU-T G-series Recommendations – Supplement 55, “Radio-over-fibre (RoF) technologies and their applications,” 2015.

[6] eCPRI Specification, V1.0, “eCPRI 1.0 specification,” 2017.

[7] D. Iida, S. Kuwano, J. Kani, and J. Terada. "Dynamic TWDM-PON for mobile radio access networks", Optics Express, vol. 21, no. 22, pp. 26209-26218, 2013.

[8] N. Shibata, T. Tashiro, S. Kuwano,N. Yuki,J. Terada, and A. Otaka ,"Mobile front-haul employing Ethernet-based TDM-PON system for small cells", in Conf. OSA/IEEE OFC, pp. 1-3. 2015.

[9] R. Bonk, W. Poehlmann, H. Schmuck, and T. Pfeiffer, “Overlayed-modulation for increased bit rate per carrier wavelength and higher flexibility in access networks”, in Conf. OSA/IEEE OFC, W2A. 60. 2016.

[10] C. Chen, W.D. Zhong, and D. Wu, “Integration of variable-rate OWC with OFDM-PON for hybrid optical access based on adaptive envelope modulation”, Optics Communications, vol. 381, pp. 10-17, 2016.

[11] J.Y. Sung, C.W. Chow, C.H. Yeh, and Y.C. Wang, “Service integrated access network using highly spectral-efficient MASK-MQAM-OFDM coding”, Optics Express, vol. 21, no. 5, pp. 6555-6560, 2013.

[12] G. Shen, R.S. Tucker, and C.J. Chae, “Fixed mobile convergence architectures for broadband access: integration of EPON and WiMAX,” IEEE Commun. Mag. vol. 45, no. 8, pp. 44 –50, 2007.

[13] C. Browning, A. Farhang, A. Saljoghei, N. Marchetti, V. Vujicic, L. E. Doyle, and L. P. Barry, “Converged wired and wireless services in next generation optical access networks,” In Conf. IEEE 19th ICTON, 2017.

[14] J. Canny, “A Computational Approach to Edge Detection”, IEEE Trans. on Pattern Analysis & Machine Intelligence, vol. 8, no. 6, pp. 679-698, 1986.

[15] L. Zhang, S. Xiao, M. Bi, L. Liu, and Z. Zhou, “Channel estimation algorithm for interference suppression in IMDD–OQAM–OFDM transmission systems,” Optics. Commun. vol 364, pp. 129-133, 2016.

[16] ITU -T Recommendation G.975.1, 2004, Appendix I.9.

On the Feasibility of Service Composition in a Long-Reach PON Backhaul

Ahmed Helmy, Nitesh Krishna, and Amiya Nayak

School of Electrical Engineering and Computer Science,
University of Ottawa, Ottawa, Canada
Emails: {ahmed.helmy, nkris012, amiya.nayak}@uottawa.ca

***Abstract*—As network architectures are continuously evolving and being integrated with new technologies to meet the demands and requirements of future applications, many new possibilities and challenges emerge. Additionally, the evolution of user devices has brought other new possibilities, where devices in the same vicinity can offer and exchange services with each other. A vision that has led to developing various service discovery and composition models that aim to satisfy different constraints while ensuring a better quality of service.**

In this paper, we consider service discovery and composition in an optical access network, serving as a backhaul for a wireless front-haul, and examine how it can be greatly affected by the underlying bandwidth allocation scheme. We compare the performances of centralized and decentralized-based service compositions and study their side effects on upstream traffic. Numerical results demonstrate how decentralized allocation can be much better suited for supporting such service models and associated traffic in terms of both service delays and side effects on regular upstream traffic.

***Keywords*—Edge computing, Ethernet passive optical network (EPON), fiber-wireless (FiWi), fog computing, long-reach passive optical networks (LR-PONs), offloading, service composition**

I. Introduction

The evolution and widespread of mobile and IoT devices has led to a new era, where everything is now being reshaped to fit arising trends and better serve tomorrow's requirements. On one hand, network infrastructures have been constantly evolving and going through many reformations to better accommodate the exponential increase in user traffic and to meet application requirements with improved service quality. On the other hand, breakthroughs in the capabilities of edge and mobile devices, in terms of memory, computational power, and storage capacity, had led to broad possibilities of how services can be provided.

As edge devices are becoming smarter and more powerful, with a wider range of functionalities, fog and edge computing paradigms continue to spread, addressing the new demanding application requirements. These new paradigms enable the storage and computational resources of the cloud to be extended all the way to the edge of the network [1]. Not only does this enable resource-constrained devices to offload more tasks with strict latency or location-aware requirements, but it also allows much of their generated traffic to be processed at the network edge instead of being carried to the cloud [2].

The advances made in end devices themselves and their ample widespread has also introduced new possibilities, where devices in the same vicinity can start exchanging services. This new vision has led to developing various service discovery and service composition techniques, where individual services, in a service-oriented environment, can be looked up and properly combined to solve more complex tasks [3]. While the environment can be anything from a mobile network to a deep space research station, the service itself can be any accessible software component, hardware resource, or data segment that can be offered to other devices [4]. This concept brings many potential assets for a wide range of applications and scenarios such as image processing, sharing GPS/internet data, crowd computing, social networking, or aggregating and integrating sensor data to discover meaningful trends such as current weather or traffic conditions [5].

The ubiquity of mobile devices and their evolution into significant service computing platforms has thus drawn much attention in the literature. Some studies focused on the devices' mobility aspect and how it may affect the feasibility of service composition [6]–[8], whereas other studies considered it from an energy consumption perspective [9]. In this paper, we study the feasibility of service composition in optical access backhauls, which are widely believed to play a vital role in tomorrow's infrastructures given their high offered capacities and their cost-effective deployment and operation [10]–[12]. Many architectures have already proposed integrating them with fog computing in order to better serve wireless front-hauls [13]–[15]. Still, the effects of offloading traffic on the network performance as well as the feasibility of service composition itself, in optical access networks, have not yet been addressed. Moreover, many studies only focused on achieving performance gains in the wireless front-end with little regard to the backhaul and its possible bottlenecks [16].

In this paper, we investigate the performance of offloading services and retrieving results, in a *long-reach passive optical network* (LR-PON), when the underlying dynamic bandwidth allocation is either centralized or decentralized. We also study the effect of carrying this new type of traffic on regular traffic. In Section II, we start by formulating our service discovery and composition problem. In Section III, we then demonstrate how

978-3-903176-07-2.00 © 2018 IFIP

service composition can be accomplished in each allocation paradigm. In Section IV, we present numerical results, whereas Section V concludes the study.

II. Problem Formulation and Assumptions

In this section, we first describe the network architecture of the *fiber-wireless* (FiWi) network under consideration before laying out the service composition problem and its relevant assumptions.

A. Network Architecture

LR-PONs were first introduced around a decade ago as a highly cost-effective broadband solution since they can extend the coverages of traditional PONs up to a 100km or more. This allows combining access and metro networks into a single integrated network as well as connecting more users, thereby saving huge operational and capital costs [17]. Fig. 1 illustrates one of the most common LR-PON architectures, in which each access zone is served by dedicated upstream and downstream wavelengths that enable the *optical network units* (ONUs) to communicate with their associated *optical line terminal* (OLT). The ONUs can then be connected to access points or base stations that enable them to serve wireless devices or they can alternatively be connected to wired subscribers [13].

Because the network in the upstream direction forms a multipoint-to-point network, where multiple ONUs transmit toward the OLT through a shared medium, some form of channel arbitration is required to coordinate the ONUs' transmissions and avoid data collisions. For that purpose, *time-division multiple access* (TDMA) has been adopted in most PON standards, where each ONU is periodically allocated a timeslot for transmission. This upstream bandwidth allocation can then either be centralized or decentralized.

B. Edge/Fog Computational and Storage Resources

There have been many architectures in the literature that propose connecting cloudlets or micro-datacenters to optical access networks in various ways [13]–[15]. Alternatively, some studies proposed forming a local cloud from the resources of the optical network itself [18]. Besides being relatively close to each other, the computing and storage resources of the ONUs together exceed that of the OLT by more than eightfold. This led to the idea of making these ample resources accessible to their connected devices.

In this paper, we consider a LR-PON with N ONUs capable of receiving service requests from their connected devices. To formulate our service composition problem, we assume the following with regard to the available computational resources:

- each ONU is capable of running some computational tasks besides its main functions as well as using part of its memory for storage purposes,
- ONUs in the same network are identical (homogenous computing resources),
- an ONU can compose a service from its own offered services or from those currently offered by its connected devices with acceptable battery levels.

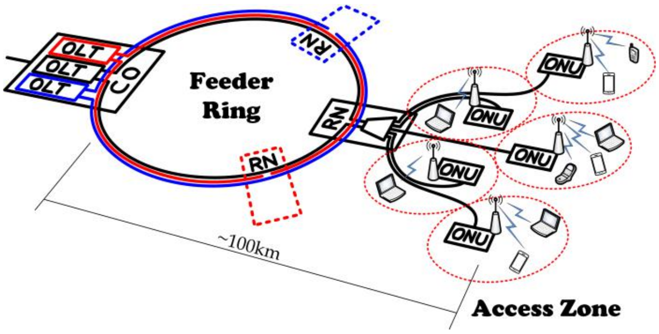


Fig. 1. FiWi architecture with a LR-PON backhaul.

We then assume the following assumptions for a given requested service (offloaded task):

- a task T is requested by a user device connected to an ONU which we call the client ONU or ONU$_c$,
- if the request is accepted, task T is to be carried out using some offloaded data and an associated set of operations that are altogether R bytes in size,
- R is fragmented into Ethernet packets that add up to D bytes in size (including overheads),
- task T may be a composition of multiple services that can be broken into S subtasks or services, which may be composed and run on multiple devices,
- either the OLT (centralized case) or the client's node ONU$_c$ (decentralized case) will be responsible for finding an ONU with enough resources to manage the service composition based on a utility function U,
- the elected composition manager ONU, which we call ONU$_m$, will be responsible for composing the service, collecting the final result (which is E bytes in size), and sending it back to ONU$_c$,

Finally, in our work, we assume that regular upstream traffic is always given higher priority than service traffic. In other words, only unused bandwidth is used for offloading and exchanging edge traffic.

C. Offloading and Service Composition

Before a device can offload a task to the network, the device must first construct a requirement list that specifies the services required and its corresponding QoS requirements for each service. The device then embeds this list into a service request message that it sends to the ONU to which it is connected. Once the ONU receives this service request, four phases are then required to compose the requested service, assuming that the client's ONU does not currently have the resources to do the service composition itself:

1) Composition Manager Selection Phase

Using the cyclic updates collected from other ONUs, either the OLT (in the centralized case) or ONU$_c$ (in the decentralized case) elects an ONU that will manage the service composition. The elected ONU (ONU$_m$) will be the one that currently has the highest available computational resources, offers the requested services, and meets the QoS requirements of these services.

Similar to [4], this ONU election can be based on a utility function *U*, which can be expressed as;

$$U(ONU_i) = \alpha N(ONU_i) + \beta M(ONU_i) + \gamma Q(ONU_i) \qquad (1)$$

where *N* and *M* are the numbers of services currently being offered by the i^{th} ONU itself and by the devices connected to it, respectively, whereas *Q* is a metric that reflects the similarity between the services offered and those requested. α, β, and γ are corresponding weights that may be used to prioritize one utility parameter over the others. After ONU$_m$ is successfully elected, the task has to be passed on to it. ONU$_c$ can then accept the service request from the device, allowing it to offload any task data, before forwarding it to ONU$_m$.

2) *Service Discovery Phase*

In this phase, the selected composition manager performs the service discovery process, in which it investigates all the services offered by its connected devices as well as by itself. The manager then selects the best available services that are to be integrated together to provide the required service. The service discovery phase is thus composed of two main steps: forming a candidate list and ranking each candidate.

The first step requires ONU$_m$ to have a list of the services currently being offered by the devices in its vicinity along with their corresponding QoS metrics. This list may be already available to ONU$_m$ through a locally cached description, which can easily be gathered from its connected devices' periodic updates. Moreover, the services must be represented by their storage or computational capabilities as well as their particular service type (i.e., 100MB memory storage availability or 1GHz processing speed).

The second step, however, must use some information from the service request message to compute the rankings of the available services using a ranking function *R*, such as;

$$R(s_j) = w_c.w_r.sim(s_j, \boldsymbol{S_r}) \qquad (2)$$

where w_c and w_r are weights that reflect the availability of a candidate service and the priority of a required service, respectively, whereas $sim(s_j, \boldsymbol{S_r})$ is a similarity function that matches between a service offered s_j and a set of requested services $\boldsymbol{S_r}$ of similar nature. If multiple instances of the same service exist, the weight w_c can be used to give preference to one over the others based on some metric (e.g., distance: the nearest one or the one having the least number of hops).

3) *Service Integration and Execution Phase*

In this phase, ONU$_m$ coordinates the execution of the selected services in the order specified by the service request message. It also ensures the transfer of intermediate results from one service to another when necessary. Execution can therefore occur in a distributed manner, where partial results received from a service (executed on one device) can be transferred to the following service (executed on another).

4) *Result Collection Phase*

Finally, if one or more of the services produces a result, the final output must be sent back to the device where the service request originated. This means that results first need to be gathered by ONU$_m$ and then sent back to ONU$_c$, to which the device is connected.

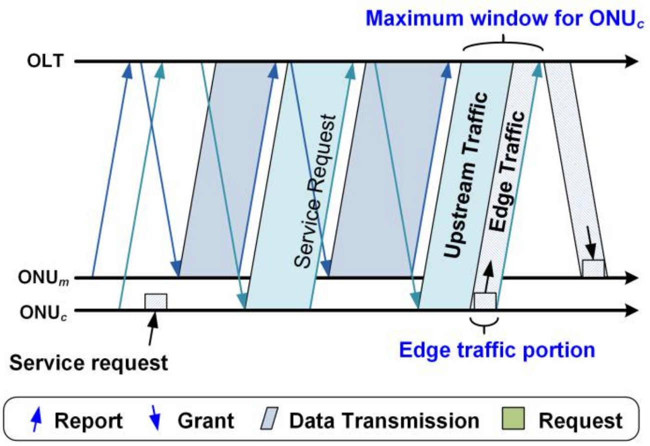


Fig. 2. Centralized-based service composition offloading.

Of these four mentioned phases, the second two phases, in which the service has already been transferred to ONU$_m$, are not dependent on the underlying bandwidth allocation, in contrast to the first and last phases, in which the data is transferred back and forth between ONU$_c$ and ONU$_m$. In the next section, we study how these phases can be carried out in each bandwidth allocation scheme.

III. Service Composition in LR-PONs

The exchanging of service requests and offloaded traffic is different in each allocation scheme. In this section, we examine how it can be carried out with centralized and decentralized bandwidth allocation paradigms.

A. *Centralized DBA – Polling*

Centralized *dynamic bandwidth allocation* (DBA) has been widely considered in the literature [19]–[21], in which, the OLT arbitrates the ONUs' bandwidth allocation. As illustrated in Fig. 2, the OLT basically polls each ONU with a grant message, giving it a window to transmit according to a previously received report message that reflects the ONU's queue status. ONUs, on the other hand, do not need to monitor the network state nor exchange any information, which makes their design relatively simple.

With no direct inter-ONU communications in centralized allocation, ONU$_c$ will have to forward the service request to the OLT, which would then be responsible for selecting the composition manager (ONU$_m$) based on information gathered from its most recently received reports. This, of course, would require ONUs to continuously append the availability of their resources and offered services in all their outgoing reports, something that is not found in a conventional allocation algorithm. Once the OLT elects a composition manager with enough resources to carry out the task, it will accept the service request and start granting ONU$_c$ more upstream bandwidth up to the maximum allowable by its service level agreement. Using this additionally allocated bandwidth, ONU$_c$ will start uploading the relevant task data, as illustrated in Fig. 2. This can last for more than one cycle depending on the ONU's current upstream load and the size of the offloaded data.

After receiving the offloaded data, the OLT forwards this data to ONU$_m$ along with the original service request and its requirement list. ONU$_m$ then carries out the service composition, integration, and execution phases, before sending back the results to ONU$_c$. This again is done by sending them first to the OLT using the excess bandwidth granted by the OLT in its following transmission windows.

B. Decentralized DBA

Because polling forms the basis of centralized DBA, the performance of such allocation greatly depends on the *round-trip times* (RTTs) imposed on the bandwidth negotiation messages exchanged between the ONUs and the faraway OLT. While this does not pose challenges in traditional PONs with 10-20km spans, RTTs become more severe in LR-PONs causing the DBA performance to considerably degrade [19]. Decentralized bandwidth allocation has therefore been proposed as an alternative for LR-PONs, where the ONUs themselves manage the upstream media access instead of having to periodically report their buffer status to the remote OLT and then wait for grants to transmit. However, for the ONUs to successfully manage the upstream media access, they need to communicate together; something that was not needed nor available in the original network design.

One possible way of achieving inter-ONU communications is to place a *fiber Bragg grating* (FBG) near the remote node which selectively reflects back a single wavelength to the ONUs facilitating an *out-of-band* (OOB) multipoint-to-multipoint network. This was shown in [22] to be a viable option for inter-ONU networking and was also used in [23] as the basis for a decentralized media access scheme.

In [23], this inter-ONU communications technique was used to enable the ONUs to take turns transmitting on the upstream wavelength by announcing to each other the durations of their transmissions. This was done by making each ONU send a very short time-stamped frame (a tag) at the beginning of its transmission, announcing how many bytes it intends to transmit without exceeding a certain maximum. This maximum was set by the OLT during an initialization phase according to the ONU's service level agreements. Chances of upstream inter-transmission gaps are then reduced since the time it takes the frame to reach the following ONU on the control channel will be during data transmission on the upstream channel. With no reports to the OLT, the delays in this decentralized scheme are fully independent of the RTTs. Instead, the delays depend on the distances between the ONUs and the reflective device.

In this work, we modify this OOB tagging scheme to allow edge data to be exchanged between ONUs on this additional channel during an offloading or a result retrieval phase. We propose to place a flag in the tag message, which, if toggled by an ONU, will indicate that this ONU needs to transmit edge traffic in the next cycle. As illustrated in Fig. 3, once this flag is toggled, the ONUs switch to another tagging scheme in the next cycle, where all the tags are immediately sent in the beginning of the cycle, thereby giving room for edge traffic to be exchanged. This tagging scheme continues to be used by all ONUs as long as one of them still has a toggled flag in its last tag message. Additionally, the OOB edge window can be shared among multiple ONUs by simply dividing it equally among those ONUs which had their flag toggled in the previous cycle. Alternatively, the ONUs may share the length of their OOB transmissions along with the toggled flag for better OOB utilization and lower edge delays.

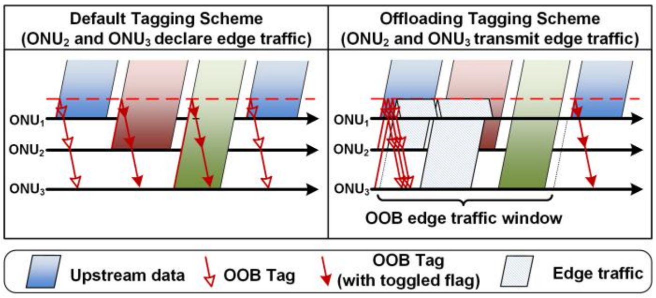

Default Tagging Scheme
(ONU$_2$ and ONU$_3$ declare edge traffic)
Offloading Tagging Scheme
(ONU$_2$ and ONU$_3$ transmit edge traffic)
ONU$_1$
ONU$_2$
ONU$_3$
ONU$_1$
ONU$_2$
ONU$_3$
OOB edge traffic window
Upstream data
OOB Tag
OOB Tag (with toggled flag)
Edge traffic

Fig. 3. Default decentralized tagging scheme and proposed offloading tagging scheme which is provoked when the edge traffic flag is toggled.

Because tags are exchanged in the beginning of the cycle (during the upstream transmission of the first ONU), ONUs cannot transmit more than what had already been announced in their tags. The ONUs may therefore choose to reserve the same transmission windows they used in the previous cycle, by announcing so in their outgoing tags, even though they may not have enough packets yet in their buffers to fully utilize these windows. This however gives each ONU the chance to accommodate some newly arriving packets between the time of sending its tag and the time it starts its upstream transmission.

Inter-ONU communications in the decentralized scheme enable the first and last service composition phases, discussed in Section II, to be carried out in a different manner from its centralized counterpart. Here, ONU$_c$ will be responsible for electing the composition manager from the information received in the last $N-1$ tags. This means that all ONUs need to continuously append their computational status and offered services in any outgoing tag message similar to what is proposed to be done within centralized reports. Using this information, ONU$_c$ directly selects ONU$_m$, without involving the OLT, and broadcasts this selection in its next outgoing tag. This particular tag will not only specify the selected node, but will also have its offloading flag toggled so that the ONU may directly start transferring the service data to ONU$_m$ within the next cycle. Contrary to the centralized scenario, edge data here does not have to go through the OLT. Instead, it is directly broadcasted to all the ONUs on the OOB channel. Once the necessary input data reaches ONU$_m$, the second two phases can then take place similar to the centralized scenario.

After finishing the service integration and execution, ONU$_m$ sends back its output to ONU$_c$ again using the OOB channel in a similar manner as was done in the first phase.

IV. Numerical Results

In our study, we consider a 100km long-reach symmetric Ethernet PON consisting of an OLT and 16 ONUs. The ONUs are placed randomly in the last 5km of a 100km network span, assuming that the FBG is located 95km away from the OLT. ONUs share an upstream wavelength of 1Gbps, whereas from

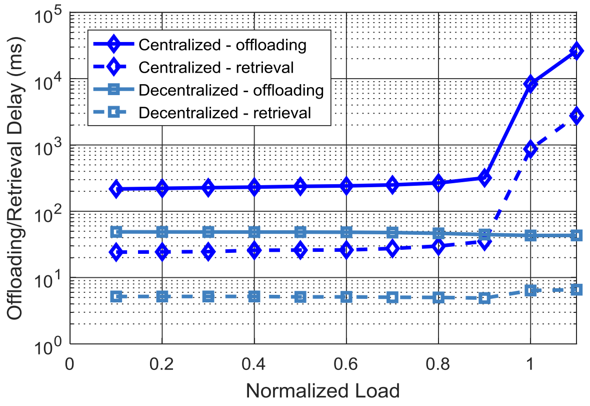


Fig. 4. Offloading and retrieval delays for a 5MB task with a 500KB result.

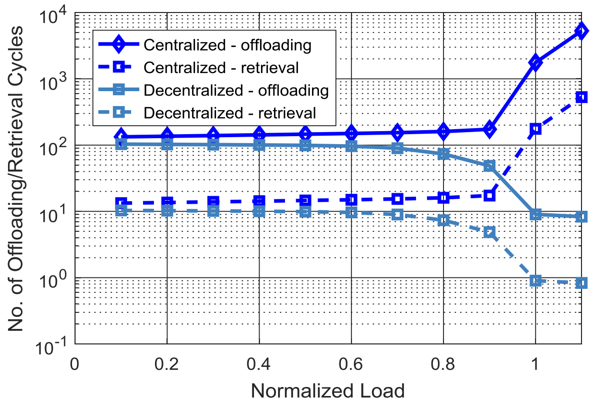


Fig. 5. Number of offloading and retrieval transmission cycles for a 5MB task with a 500KB result.

the access side the end-users have an access rate of 100Mbps. Each ONU has a 10Mbytes buffer, whereas the traffic model used is self-similar Ethernet traffic, constructed from alternating on/off Pareto-distributed streams with a Hurst parameter of 0.8, similar to the traffic model used in [19], [23].

In order to compare the performances of the two schemes, the maximum cycle duration is set to 5ms for both schemes with 5µs inter-transmission guard intervals for both in-band and out-of-band traffic. For the proposed decentralized scheme, we set the OOB transmission rate to 1Gbps, through which ONUs also inform each other of their edge transmission sizes in their outgoing tags with toggled flags. During edge traffic exchange, ONUs reserve the same transmission windows they used in the last normal cycle.

A. General Performance

Fig. 4 illustrates the offloading and retrieval delays for a 5MB task having a 500KB result. It can be seen how, in centralized allocation, the delays increase with the increase of the upstream traffic load, especially at loads greater than 90%. This is because, as the network load increases, unused excess bandwidth in the ONUs transmission windows decreases. With normal upstream traffic having a higher priority, edge traffic would then take longer to transmit and would last for more cycles under heavy loads. This can also be seen in Fig. 5, which shows the number of cycles used to exchange edge traffic for both schemes. On the contrary, edge delays in the decentralized scenario seem to be unaffected as the load increases. In fact, the number of transmission cycles is shown to decrease with increasing the network load. This is because, as the cycle is extended more towards its maximum, a larger OOB edge transmission window is formed.

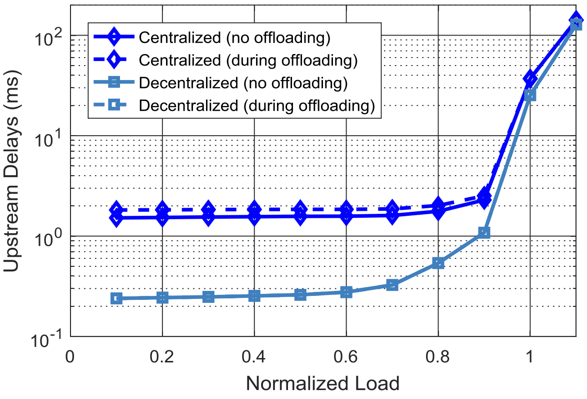


Fig. 6. Effect of service traffic on regular upstream traffic.

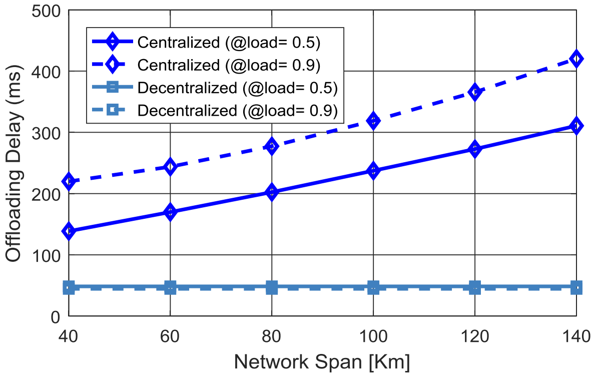


Fig. 7. Effect of network span on offloading delays.

B. Effect on Upstream Traffic Delays

Fig. 6 shows how pre-transmission delays of regular upstream traffic are affected during an offloading phase. Injecting edge traffic on the upstream wavelength is shown to have a significant effect on centralized upstream traffic delays, but has no effect on decentralized delays. This is because injecting edge traffic extends the centralized polling cycle, by the additional excess bandwidth portion used for edge traffic, causing more delays for queued upstream traffic. On the other hand, exchanging edge traffic is implemented out-of-band in the decentralized scheme without causing any cycle extensions.

It is worth mentioning that the effects seen in Fig. 6 are only caused by a single ONU's offloading. The effects will therefore be exaggerated in the centralized scheme when multiple ONUs are concurrently offloading edge traffic to the

OLT. These effects, however, only last while there are ongoing edge traffic transmissions. The overall performance would, therefore, depend on how often the network has to deal with edge traffic as well as the amount of that traffic.

C. *Effect of Extending Network Span on Service Delays*

As was mentioned earlier, centralized allocation is greatly affected by extending the network span. Fig. 7 demonstrates a comparable effect on centralized service delays, where the performance of centralized-based offloading is ultimately degraded as the network span continues to extend. On the other hand, extending the feeder span shows to have no effects on the performance of decentralized-based offloading since the access span is kept constant at 5km.

V. Conclusions

In this paper, we investigated the feasibility of service composition in a long-reach optical access network serving as a backhaul for a wireless front-haul. We studied the delays experienced in offloading service traffic to the composition manager as well as those experienced in retrieving the composed service results. We also examined side effects of service composition traffic on regular upstream traffic.

Because decentralized-based service composition requires no OLT involvement, it has the potential of achieving much lower service delays. Decentralized-based service composition has also shown to have no side effects on regular upstream traffic. These advantages however come at the cost of placing additional transceivers within the ONUs and modifying the architecture to allow inter-ONU communications to take place. Moreover, ONUs themselves have to select the composition manager and may thus be relatively more computationally loaded than in a centralized-based scheme.

On the other hand, centralized schemes may still yet offer some benefits for service composition despite their long delays. For instance, the OLT can easily gain access to ONUs in other access zones to which it may choose to forward service requests instead. Centralized-based service composition may thus offer lower service rejection ratios as well as additional services only available in other access zones. This paper thus opens the door for further studies and calls attention toward a possible hybrid scheme that combines the potential benefits of both centralized and decentralized-based service compositions in these long-reach optical access networks.

References

[1] R. Mahmud and R. Buyya, “Fog computing: a taxonomy, survey and future directions,” *Distrib. Parallel, Clust. Comput.*, pp. 1–28, 2016.

[2] C. C. Byers, “Architectural imperatives for fog computing: use cases, requirements, and architectural techniques for fog-enabled IoT networks,” *IEEE Commun. Mag.*, vol. 55, no. 8, pp. 14–20, 2017.

[3] I. Al Ridhawi, Y. Kotb, and Y. Al Ridhawi, “Workflow-net based service composition using mobile edge nodes,” *IEEE Access*, vol. 5, pp. 23719–23735, 2017.

[4] D. Chakraborty, A. Joshi, T. Finin, and Y. Yesha, “Service composition for mobile environments,” *Mob. Netw. Appl.*, vol. 10, pp. 435–451, 2005.

[5] S. Deng, L. Huang, H. Wu, and Z. Wu, “Constraints-driven service composition in mobile cloud computing,” in Proc. *IEEE Int. Conf. on Web Services (ICWS)*, pp. 228–235, 2016.

[6] D. Kasamatsu, M. Kumar, and P. Hu, “Service compositions in challenged mobile environments under spatiotemporal constraints,” in Proc. *IEEE Int. Conf. on Smart Computing (SMARTCOMP)*, 2017.

[7] S. Deng, L. Huang, J. Taheri, J. Yin, M. C. Zhou, and A. Y. Zomaya, “Mobility-aware service composition in mobile communities,” *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 47, no. 3, pp. 555–568, 2017.

[8] N. C. H. Ngoc, D. Lin, T. Nakaguchi, and T. Ishida, “QoS-aware service composition in mobile environments,” in Proc. *IEEE 7th Int. Conf. on Service-Oriented Computing and Applications (SOCA)*, 2014.

[9] S. Deng, H. Wu, W. Tan, Z. Xiang, and Z. Wu, “Mobile service selection for composition : an energy consumption perspective,” *IEEE Trans. Autom. Sci. Eng.*, vol. 14, no. 3, pp. 1–13, 2017.

[10] K. Bourg, S. Ten, R. Whitman, J. Jensen, and V. Diaz, “The evolution of outside plant architectures driven by network convergence and new PON technologies,” in Proc. *Opt. Fiber Commun. Conf. (OFC)*, 2017.

[11] S. Zhou, X. Liu, F. Effenberger, and J. Chao, “Mobile-PON: a high-efficiency low-latency mobile fronthaul based on functional split and TDM-PON with a unified scheduler,” in Proc. *Opt. Fiber Commun. Conf. (OFC)*, 2017.

[12] J. Li and J. Chen, “Passive optical network based mobile backhaul enabling ultra-low latency for communications among base stations,” *IEEE/OSA J. Opt. Commun. Netw.*, vol. 9, no. 10, pp. 855–863, 2017.

[13] B. P. Rimal, D. Pham Van, and M. Maier, “Mobile-edge computing vs. centralized cloud computing in fiber-wireless access networks,” in Proc. *IEEE INFOCOM*, pp. 991–996, 2016.

[14] W. Zhang, B. Lin, Q. Yin, and T. Zhao, “Infrastructure deployment and optimization of fog network based on microDC and LRPON integration,” *Peer-to-Peer Netw. Appl.*, vol. 10, no. 3, pp. 579–591, 2017.

[15] A. Reaz, V. Ramamurthi, and M. Tornatore, “Cloud-over-WOBAN (CoW): an offloading-enabled access network design,” in Proc. *IEEE Int. Conf. on Commun.*, 2011.

[16] H. Beyranvand, L. Martin, M. Maier, and J. A. Salehi, “FiWi enhanced LTE-A HetNets with unreliable fiber backhaul sharing and WiFi offloading,” in Proc. *IEEE INFOCOM*, pp. 1275–1283, 2015.

[17] S. Zhang, W. Ji, X. Li, K. Huang, and Z. Yan, “Efficient and reliable protection mechanism in long-reach PON,” *IEEE/OSA J. Opt. Commun. Netw.*, vol. 8, no. 1, pp. 23–32, 2016.

[18] Y. Luo, F. Effenberger, and M. Sui, “Cloud computing provisioning over passive optical networks,” in Proc. *1st IEEE Int. Conf. on Commun. in China (ICCC)*, 2012.

[19] A. Helmy, H. Fathallah, and H. Mouftah, “Interleaved polling versus multi-thread polling for bandwidth allocation in Long-Reach PONs,” *IEEE/OSA J. Opt. Commun. Netw.*, vol. 4, no. 3, pp. 210–218, 2012.

[20] H. Song, B. W. Kim, and B. Mukherjee, “Multi-thread polling: a dynamic bandwidth distribution scheme in long-reach PON,” *IEEE J. Sel. Areas Commun.*, vol. 27, no. 2, pp. 134–142, 2009.

[21] A. Mercian, M. P. McGarry, and M. Reisslein, “Offline and online multi-thread polling in long-reach PONs: A critical evaluation,” *J. Light. Technol.*, vol. 31, no. 12, pp. 2018–2028, 2013.

[22] B. D. Manharbhai, A. K. Garg, and V. Janyani, “A flexible remote node architecture for energy efficient direct ONU internetworking in TDM PON,” in Proc. *Int. Conf. Comput. Commun. Electron. (COMPTELIX)*, pp. 453–457, 2017.

[23] A. Helmy and H. Fathallah, “Taking turns with adaptive cycle time a decentralized media access scheme for LR-PON,” *J. Light. Technol.*, vol. 29, no. 21, pp. 3340–3349, 2011.

Handling Rack Vibrations in FSO-based Data Center Architectures

Max Curran, Kai Zheng, Himanshu Gupta, Jon Longtin
Stony Brook University

***Abstract*—To overcome the shortcomings of traditional static (wired) data center (DC) architectures, there have been recent proposals of fully-wireless and reconfigurable architectures, e.g., FireFly, ProjecTor, based on Free-Space Optical (FSO) wireless links. However, there are significant challenges that need to be addressed to make the vision of FSO-based DC architectures a compelling reality. While some of these scientific challenges have been addressed in recent works, one key challenge that has yet to be addressed is how to handle FSO link misalignments due to DC rack vibrations (where the FSO transceivers are placed). The focus of this paper is to comprehensively address this challenge. Particularly, in this work, we present measurement results of a thorough study conducted over a live DC to characterize rack vibrations, and design a novel tracking and pointing (TP) system that is based on received power feedback and has a zero exposed-footprint on the deployment platform (racks). We develop and test a reconfigurable FSO link with our designed TP system and evaluate it over expected rack vibrations; our evaluation results demonstrate the effectiveness of our TP system.**

I. Introduction

Data centers (DCs) are a critical piece of today's networked applications in both private and public sectors (e.g., [5], [6], [11], [12], [14]). A robust *datacenter network fabric* is fundamental to the success of DCs and to ensure that the network does not become a bottleneck for high-performance applications [30]. In this context, DC network design must satisfy several goals: high performance [16], [27], low equipment and management cost [16], [37], robustness to dynamic traffic patterns [38], [43], [28], [41], incremental expandability to add new servers or racks [22], [39], and other practical concerns such as cabling complexity [35], and power and cooling costs [24], [36].

Traditional data center architectures have been based on wired networks; being *static* in nature, these networks have either been (i) *overprovisioned* to account for worst-case traffic patterns, and thus incur high cost (e.g., fat-trees or Clos [19], [27], [16]), or (ii) *oversubscribed* (e.g., simple trees or leaf-spine architectures [17]) which incur low cost but offer poor performance due to congested links. Recent works have tried to overcome the above limitations by augmenting a static (wired) "core" with some flexible links (RF-wireless [28], [43] or optical [41], [18]). These *augmented* architectures show promise, but have offered only incremental improvement in performance due to various limiting factors. Furthermore, all the above architectures incur high cabling cost and complexity [35].

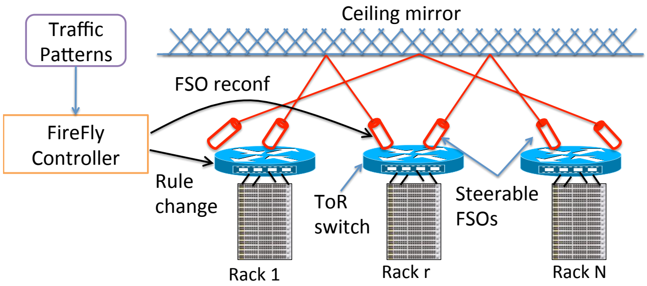


Fig. 1: High-level view of FireFly.

FSO-Based DC Architectures. To overcome the above cost-performance trade-offs and rigidity of past DC architectures, recent work [29], [26] has proposed an *extreme* design point—a *fully flexible*, *all-wireless* inter-rack fabric using *Free-Space Optics (FSO)* communication links. The FSO communication technology is particularly well-suited as it can offer very high data rates (tens of Gbps) over long ranges (>100m) using low transmission power and with small interference footprint [33].

Fig. 1 shows a conceptual overview of *FireFly*, the original FSO-based DC architecture which was proposed by our research group [29]. A number of FSO devices are placed on top of each rack and are connected to the top-of-the-rack switch. Each FSO device assembly is capable of precise and fast steering to connect to FSO devices on other racks. The controller intelligently reconfigures these devices in real-time to adapt to changing network requirements. Since the FSO beams may be obstructed by other devices in the system, FireFly proposes use of a ceiling mirror for beam redirection to ensure clear line-of-sight; in our recent work, we developed alternate line-of-sight techniques that preclude use of a ceiling mirror [21].

There are significant challenges that need to be addressed to make the vision of FSO-based DC architectures a compelling reality. While some of these scientific challenges have been addressed in recent works (e.g., [29], [21], [26]), one key challenge that has yet to be addressed is how to handle FSO link misalignments due to vibrations of DC racks where the FSO transceivers are placed. The focus of this paper is to comprehensively address this challenge.

Handling Rack Vibrations in FSO-Based DCs. Proposed FSO-based DC architectures place FSO transceivers on top of DC racks. Since the racks can vibrate for a variety of reasons (e.g., moving parts in the servers, building and other external

978-3-903176-07-2 © 2018 IFIP

vibrations such as HVAC sysems, humans), the FSO links may fail temporarily due to misalignments as they require very precise alignment for operation. One of the ways to handle this challenge is to use an active tracking and pointing (TP) system which actively corrects for such misalignments based on some feedback. However, in our context wherein tens of FSO devices need to be placed on top of racks with limited space, the key challenge in designing a viable TP system is to ensure that it has minimal/zero physical footprint. In our work, we show that favorable factors in the DC context such as indoor environment and relatively short link ranges make such a TP design feasible.

Paper Contributions and Organization. In the above context, this paper makes the following contributions:

- *Rack Vibration Measurements (§III).* To characterize typical DC rack vibrations, we conduct a measurement study and analysis via motion data collection using accelerometers and IMUs placed on racks in a live DC in our university.
- *Tracking and Pointing System (§IV).* We design a novel tracking and pointing system based on tracking-feedback from received power strength and with zero exposed-footprint, to handle any misalignments due to rack vibrations.
- *Testbed and Evaluation (§V).* We build a link testbed with the proposed TP system and evaluate it for expected misalignments based on our rack-vibration measurement study and analysis.

II. Background and Related Work

In this section, we give an overview of an FSO-based DC architecture with more detailed description of SFP-based FSO link design, and discuss related work.

FSO-based DC Networks. As mentioned in the previous section, FSO-based DC architectures are fully-flexible all-wireless and are based on the key insight that flexibility can facilitate near-optimal performance when done right. The FSO-based DCs are comprised of the following key components, viz., the FSO devices, link steering mechanisms, and the network management techniques. FSO devices needed to create FSO communication links in these architectures need to have a small form factor (so a few tens of them can fit on the top of a rack) and provide high data rates at ranges of up to 50-100m. Prior works [29] demonstrated a design of an FSO link prototype based on SFPs that satisfies the desired requirements; we discuss more details of such SFP-based FSO links in the next paragraph.

For network reconfigurability, the FSO devices are equipped with a mechanism to steer the laser beam from one receiver to the next; for viable performance, this steering must incur very low latency, i.e., on the order of a few milliseconds. In [29], two types of steering mechanisms, viz., switchable mirrors (SMs) and Galvo mirrors (GMs), were explored and their feasibility for FireFly demonstrated, while [26] has explored the feasibility of Digital Micromirror Device (DMD) as a steerable mechanism. Network management of these FSO-based DC architectures involves network design at two different timescales: (i) Preconfiguration of the network with an appropriate number of FSO devices with appropriately pre-oriented steering mechanisms; this preconfiguration is done at coarse (e.g., weekly) timescales and determines possible topologies that can be activated in real-time. (ii) Runtime reconfiguration of the pre-configured network which selects a *runtime topology* and engineers real-time traffic, based on the prevailing network state. These networks offer unprecedented benefits, such as reduced infrastructure cost, increased flexibility, and decreased cabling complexity, and have been shown to perform nearly the same as optimal wired networks.

SFP-based FSO Link Design. Prior works [29], [26], [23], [20] have designed and prototyped SFP-based FSO links to demonstrate feasibility of small-form factor FSO devices suited for FSO-based DC architectures. Below, we discuss this in more detail, as our testbed and TP system builds upon this design. An SFP (small form-factor pluggable) transceiver is a small ($1/2'' \times 1/2'' \times 2''$) and compact commodity optical transceiver [4], widely used to interface optical fibers with network switches. An SFP contains a laser source and a photodetector, for transmitting and receiving respectively. SFPs are available with a variety of laser sources, varying in the wavelength (typically, between 800nm to 1550nm) of the emitted beam as well as the supported data rate (anywhere from 10Mbps to recent variants called CFPs with 100Gpbs [1]). SFP+ refers to an enhanced version of the SFP that supports data rates up to 16Gbps. To create an FSO link using SFPs, the beam emanating from the transmitter SFP is channeled into a short optical fiber which feeds into a collimator. The collimated beam is then launched into free space towards the receiving SFP, where it is captured by another collimating lens and focused back into an optical fiber connected to the receiving SFP.

Related Work. To the best of our knowledge, the only work on measurement of rack vibrations is [40]; however, the focus of [40] is on the impact of vibrations on server hard drive failure, by characterizing vibration level. In contrast, we wish to evaluate impact of rack vibrations on FSO link alignments, and thus, our focus is on measuring motion related characteristics of rack vibrations.

Typical TP mechanisms [31], [32] include a fast steering mirror or gimbal controlled by digital servos [34], [15], [42] for pointing (recent project [7] uses a high-cost SLM), along with some tracking detectors to track the target or beam. Common tracking detectors include positioning sensing diodes [15] (e.g., CCDs [34], photodiode arrays [25]), accelerometers, cameras, GPS [42], etc. Our recent work [20] used a TP mechanism based on GMs and photodiodes in the context of outdoor picocell networks. In our context, we are most interested in developing a TP system that has minimal "exposed" footprint to allow placement of a large number (50+) of devices on top of each DC rack. Thus, we propose a novel tracking mechanism based on (i) the RSSI (relative

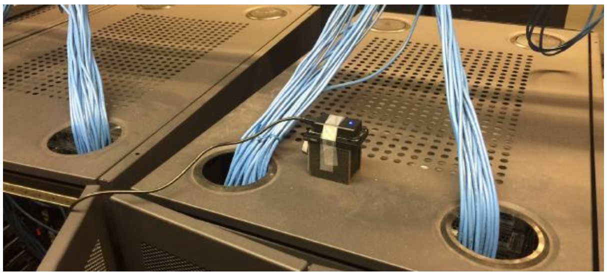


Fig. 2: Inertial measurement unit (IMU) on top of a DC rack.

received signal strength) feedback for tracking (without using any exposed hardware), and (ii) already available steering mechanisms, e.g., GMs, in the FSO links.

III. Vibration Analysis

In this section, we present results of our measurement study conducted over racks in a live DC in our university, to better understand and characterize the vibrations experienced by devices placed on top of the DC racks. We will use the results of this study to evaluate the effectiveness of our proposed TP system.

Measurement Study Setup. To collect rack vibration measurements, we install an accelerometer [2] and an IMU (inertial measurement unit) [8] over various positions on or near racks in a live data center housed in the CEWIT center [3] in our university. The accelerometer measures linear acceleration (and hence, linear displacement via integration) in the three spatial dimensions, while the IMU measures the three angular/rotational accelerations, viz. pitch, yaw, and roll. Thus, together these measurement devices cover all possible motions/vibrations experienced by an object on the rack. The accelerometer and IMU gather measurements at a rate of 512 Hz and 250 Hz respectively, which is sufficient for our purposes.[1] We gather measurements with these devices placed on top and side of three different racks for a continuous 24-hour period.

Data Analysis in Frequency Domain. Once all of the data was collected, the Fourier transform was applied to both the accelerometer and IMU data for data analysis in the frequency domain and to filter out low-frequency noise. In particular, the data in the frequency domain is sent through an appropriately designed Butterworth bandpass filter to remove any low-frequency noise. Then, we integrate the filtered data from both devices to get velocity and displacement measurements.

Integration Validation. In order to measure the accuracy of our integration process, we conduct an experiment in the lab using the accelerometer and a laser displacement sensor (LDS) [10]. We move the accelerometer by hand and measure its actual displacement by the LDS which uses a laser to accurately measure displacement. Then, we compare the displacement measured by LDS with the displacement as computed by integration of the accelerometer measurements; see Figure 3. We observe that the difference between the two displacement results is minimal. This shows that the errors in the displacement result from integrating accelerometer's acceleration data do not accumulate over time, thus validating our integration analysis.

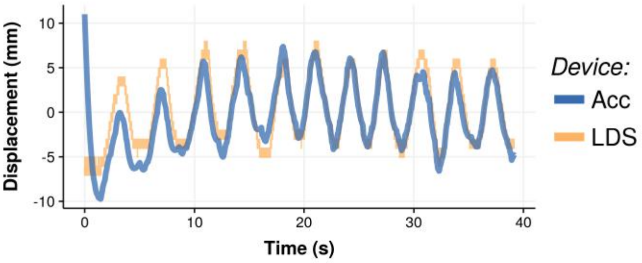


Fig. 3: Difference between the displacement as measured by integration of accelerometer's acceleration and the displacement measurement from the laser displacement sensor.

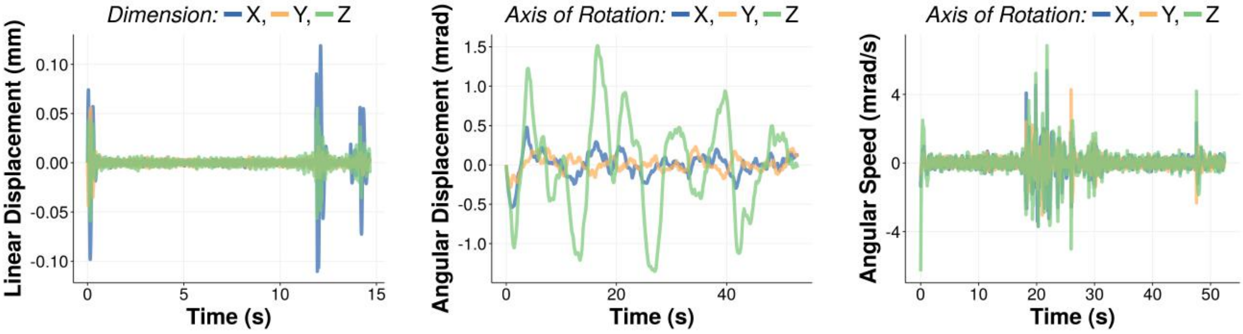
Vibration Results. Below, we summarize our vibration measurement results as related to FSO link's intrinsic movement tolerance.

- Linear displacement (see Figure 4a) was found to be minimal (maximum 0.25mm). This amount of linear displacement can be easily handled by a link's intrinsic movement tolerance (a few millimeters [29]). Moreover, this is subsumed by the angular displacement for links longer than one meter. Thus, we don't analyze linear speed of displacement.
- Angular displacement (see Figure 4b) was observed to be at most 1.5 mrad with an average of 0.9 mrad. This is more than sufficient to disconnect a link with no TP system. For example, on a 10m link, a 1.5 mrad angular deviation of the transmitter would cause a 1.5cm linear displacement at the receiver. This demonstrates a need to have an effective TP system to keep an FSO link continuously operational on DC racks.
- To evaluate the effectiveness of a TP mechanism, we must focus on the angular *speed* of movement, since the TP system incurs a non-zero latency and can therefore only be effective up to some angular speed [20]. Thus, we analyze the angular speed of movement caused by rack vibrations (see Figure 4c); the average rotation speed was found to be 3.30 mrad/s with a maximum being 6.98 mrad/s.

IV. Tracking and Pointing (TP) System

FSO links require precise alignment to function properly, and link misalignment can cause the entire link to stop functioning. Link misalignment is fundamentally caused by movement of the beam at the receiver plane, which is caused by the movement of the FSO transceivers. In a DC environment, rack vibrations can cause the FSO transceivers to deviate from their original positions and thus cause link disconnection. To counter such link misalignments due to rack vibrations, we use an active tracking and pointing (TP) system which uses a *tracking* mechanism to track the beam movement at the

[1] Higher frequency vibrations, if any, will have negligible displacements.



(a) (b) (c)

Fig. 4: Vibration measurements from the accelerometer and IMU placed on top of a DC rack. Here, each plot shows only a small slice of time.

RX and provides feedback to the *pointing* mechanism which corrects any misalignments.

Requirements. The key requirement for our TP systems is that it should be able to keep an FSO link aligned in response to expected rack vibrations, as characterized in the previous section. Any TP system can only be effective up to a certain speed of beam movement in the receiver plane, due to non-zero latency. To relate the beam movement in the receiver plane to motion caused by rack vibrations, we focus on the most dominating component of the vibrations, viz., angular speed and displacement of the TX assembly due to rack vibrations. Thus, we require that the TP system be able to handle vibrations that have an angular speed of at most 7 mrad/s, with an angular displacement of at most 2 mrad. In addition, for a viable TP system in our context, the TP hardware should have a minimal exposed footprint on the top of the rack.

A. Native RSSI-based TP

We propose the following lightweight TP system for use in DC environments:

Tracking Beam Position/Displacement. The goal of the tracking mechanism is to estimate the beam displacement, and use it to provide an appropriate feedback to the pointing mechanism which then corrects the beam alignment. To estimate the beam displacement, we use the RSSI (relative received signal strength) value corresponding to the received power of the beam; this RSSI value is available by "querying" the SFP+ module via the available I^{2}C interface. In our experiments, we used an SFP+ evaluation board [13] to interface with the SFP+ module. Since the evaluation board doesn't need to be exposed (or even placed) on the rack, it has no exposed footprint on the top of the rack. Note that due to the symmetry of the Guassian beam, an RSSI value (or the RX power) corresponding to the current beam "position" is not sufficient to uniquely determine beam's position/displacement. To uniquely determine beam's current position C, we gather RSSI values at *multiple positions around* C, and use these values to uniquely determine C. For example, we can gather RSSI values at C, and at beam positions slightly "north" of C, "south" of C, "east" of C, and "west" of C. (In fact, only 3 of these 5 values are needed to uniquely determine the beam's position/displacement, but more values may be needed to compensate for noise.). To acquire RSSI values at the position C' around C, we "intentionally" move the steering mechanism (GM, in our case) to the desired position C' and read the RSSI value from the SFP+ module. The beam shape and the noise level will help determine the exact positions relative to C that are most effective for our purposes. Once the RSSI values at these nearby positions around C have been recorded, we use either a precomputed-table or a gradient based control algorithm (see below) to determine the correction to be applied to the beam to move it back to the original (aligned) position.

Correction Algorithms. The goal of a correction algorithm is to take the RSSI values from multiple positions around the current beam position C, compute a correction to be applied to the beam, and then apply the correction via the steering mechanism to align the beam back to the original (aligned) position (this is the *pointing mechanism*). This overall process ensures continuous operation of the link at runtime. More formally, a correction algorithm is given a list of $n+1$ tuples, viz. (x_i, y_i, r_i) for $0 \leq i \leq n$, where (x_i, y_i) is the relative position to C, the current beam location, and r_i is the RSSI value at this relative location. We assume that $i = 0$ refers to the current beam location C (i.e., x_0 and y_0 are both 0). Below, we describe two correction algorithms that use these $n+1$ tuples to estimate a beam correction; one of them uses a training phase to precompute a table with (position, RSSI) values, while the other uses the relative differences in RSSI values to compute the beam correction.

Table-Based Correction Algorithm. In this approach, we first have a training phase wherein we query and store (in a table) RSSI values for a wide range of beam locations. The locations queried during the training phase should encompass the entire beam profile, and preferably should be at the finest granularity possible. At runtime, the input $(n+1)$ tuples are then "compared" with the table rows as below to find the row in the table that most closely represents C. In particular, the best match returned by the algorithm is the absolute location

(x_c, y_c) that minimizes the following quantity:

$$\left\| \begin{pmatrix} r_0 \\ \vdots \\ r_n \end{pmatrix} - \begin{pmatrix} T(x_c + x_0, y_c + y_0) \\ \vdots \\ T(x_c + x_n, y_c + y_n) \end{pmatrix} \right\| \quad (1)$$

where $T(x, y)$ is the RSSI value in the table at location (x, y). As (x_c, y_c) is approximately the current location of the beam (with original aligned position being (0,0)), we supply a value of $(-x_c, -y_c)$ to the steering mechanism to move the beam back to the original aligned position. Note that, if the beam is continuously moving, the correction may result in beam moving back to some position *close* to the original aligned location.

Gradient-Based Correction Algorithm. The above table-based correction algorithm's performance depends on the training phase, and thus, may suffer if there is noise or imperfections in the system that changes the RSSI values from what was observed in the training phase. Thus, we also experimented with a gradient-based correction approach, wherein the estimation of correction is (largely) based only on the input tuples. In particular, the algorithm computes the "gradient" within the input tuples, and uses this to move the beam by a constant/input value. Implicitly, the algorithm uses the fact that the beam profile is approximately Gaussian, i.e., like a "hill." Thus, the relative difference between the positions around C can be used to correct the beam from C back to the original position.

More formally, given the $n+1$ tuples (x_i, y_i, r_i), as defined above, we first estimate a gradient $G(i)$ for each $1 \leq i \leq n$ as:

$$G(i) = \frac{r_0 - r_i}{\sqrt{x_i^2 + y_i^2}} \quad (2)$$

We then use these $G(i)$ values to compute the overall correction as follows. First, to account for noise, we only use values of $G(i)$ that are large enough; in particular, for each $G(i)$, we calculate its contribution to the correction $C(i)$ as:

$$C(i) = \begin{cases} V, & \text{if } G(i) \geq K \\ 0, & \text{if } |G(i)| < K \\ -V, & \text{if } G(i) \leq -K \end{cases} \quad (3)$$

Above, V and K are appropriately picked constants. Now, the overall correction (x_c, y_c) is calculated as follows:

$$\begin{pmatrix} x_c \\ y_c \end{pmatrix} = \sum_{i=1}^{n} \frac{C(i)}{\sqrt{x_i^2 + y_i^2}} \begin{pmatrix} -x_i \\ -y_i \end{pmatrix} \quad (4)$$

Note that unlike the previous Table-based algorithm, the goal of the Gradient-based algorithm is not to correct the beam all the way back to the original aligned position (as it doesn't have sufficient information), but to simply move the beam back in the direction of the aligned position.

V. Results

We now describe and evaluate the TP system using a SFP-based FSO link prototype.

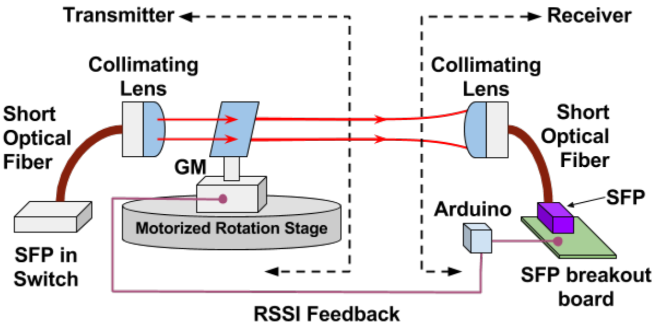


Fig. 5: Schematic of the FSO link tested with the TP system.

FSO Link with TP System Testbed. We setup our FSO link prototype (shown in Figure 5) in a controlled environment with nearly no natural movement, so that we can artificially recreate the movements caused by rack vibrations. We use 10G 1550nm SFP+ with 10 GBASE-ZR interface. We create a 10m unidirectional FSO link, with the other direction connected via a long fiber cable for simplicity. The transmitter assembly is equipped with a GM that is used as the pointing mechanism of the TP system. The entire transmitter assembly is placed on a motorized rotational stage, which allows us to simulate rotational rack vibrations. To access the pins of the SFP+ directly (and query the RSSI values), we use a Timbercon SFP+ evaluation board [13] at the receiver. We connect to the evaluation board a custom Arduino microcontroller, which uses the I^{2}C protocol to fetch the RSSI from the receiver SFP+.

Experimental Results. For our experiments, we used two nearby positions around any current beam position C as input to the correction algorithms; in particular, the nearby positions used were north and east at an "angular distance" of 0.2 mrad from C. To demonstrate the link operation during continuous terminal movement due to rack vibrations, we compute the link's TCP throughput with TP active and TX assembly rotating at varying angular speeds of 0-7 mrad/sec, using a motorized rotational stage, with an amplitude of a few mrads. In particular, we measure the average link throughput (data rate) every second, over a ten minute period using the iPerf3 tool [9].

First, we observed (plots not shown for brevity) that the Table-based algorithm outperformed the Gradient-based algorithm, and thus, we focus on the Table-based algorithm below. Figure 6 shows the CDF of the FSO link's throughput with the TP system active and TX assembly rotating at varying angular speeds (0 to 7 mrad/s). The fixed link (i.e., for 0 mrad/sec speed) achieved an overall average throughput of 9.41 Gbps. We were also able to achieve nearly identical throughput CDF for angular speeds of up to 2.5 mrad/s. Throughput CDF begins to degrade very slightly for angular speed between 4-7 mrad/s, with the average throughput still being near-optimal at about 9.37 Gbps for angular speeds up to 6 mrad/s and about 8.5 Gbps at 7 mrad/s angular speed. To offer further perspective on these results, we analyze our vibration measurement data further and observe that even though the

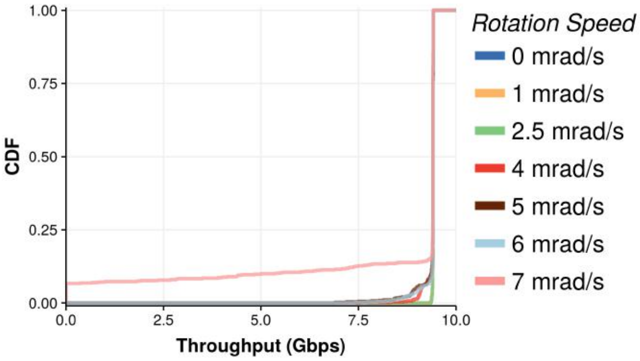


Fig. 6: CDF of the link throughput for various angular speeds of the TX assembly.

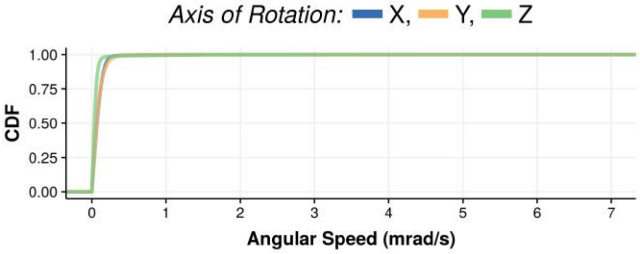


Fig. 7: CDF of the angular speeds observed in our vibration measurement study.

maximum angular speed observed was 6.98 mrad/s, the angular speed went above 6 mrad/s very rarely: only 0.00816% of the time. See Figure 7, which plots the CDF of the angular speeds observed. Thus, in summary, our TP system achieves near-optimal throughput at about 99.992% of the time, with a 10% throughput degradation at remaining times.

VI. Conclusion

In this work, we addressed one of the key challenges that arise in the context of recently proposed FSO-based data center architectures. In particular, we conducted a thorough measurement study of DC rack vibrations, and proposed a novel RSSI-based TP system with zero exposed-footprint that can handle expected rack vibrations. To the best of our knowledge, ours is the best work on handling DC rack vibrations for reliable operation of FSO links placed on DC racks.

References

[1] 100G CFP Optical Transceivers. https://tinyurl.com/yczvt6x7.
[2] Accelerometer. http://www.gcdataconcepts.com/x2-1.html.
[3] CEWIT. http://www.cewit.org/.
[4] Cisco 10GBASE SFP+ Modules. https://tinyurl.com/y894eh9a.
[5] Cisco Global Cloud Index: Forecast and Methodology, 2012 to 2017. http://tinyurl.com/7gnfeeb.
[6] Data center survey. https://tinyurl.com/ycr77yvz.
[7] Hyperion Project, UK. http://projecthyperion.co.uk.
[8] Inertial Measurement Unit. https://tinyurl.com/ybw58762.
[9] iPerf. https://iperf.fr/.
[10] Laser displacement sensor. https://tinyurl.com/yac9fzkj.
[11] Magic quadrant for data center network infrastructure. http://tinyurl.com/mpo3jzt.
[12] NSA Utah Data Center. http://nsa.gov1.info/utah-data-center/.
[13] Timbercon - SFP+ Test Host Board. https://tinyurl.com/y84tvmx9.
[14] US government gives IBM cloud green light to serve agencies. http://tinyurl.com/mx2v2mt.
[15] M. K. Al-Akkoumi. A tracking system for mobile FSO. In *SPIE Proceedings Vol. 6877*, 2009.
[16] M. Al-Fares, A. Loukissas, and A. Vahdat. A scalable, commodity data center network architecture. In *SIGCOMM*. ACM, 2008.
[17] M. Alizadeh and T. Edsall. On the data path performance of leaf-spine datacenter fabrics. In *High-Performance Interconnects (HOTI)*. IEEE, 2013.
[18] K. Chen et al. Osa: An optical switching architecture for data center networks with unprecedented flexibility. *IEEE/ACM Transactions on Networking (TON)*, 2014.
[19] C. Clos. A study of non-blocking switching networks. *Bell Labs Technical Journal*, 1953.
[20] M. Curran et al. Fsonet: A wireless backhaul for multi-gigabit picocells using steerable free space optics. In *MobiCom*. ACM, 2017.
[21] M. Curran and H. Gupta. Providing line-of-sight in a free-space-optics based data center architecture. In *Communications (ICC), International Conference on*. IEEE, 2016.
[22] A. R. Curtis et al. Legup: Using heterogeneity to reduce the cost of data center network upgrades. In *Co-NEXT*. ACM, 2010.
[23] P. Deng et al. MEMS-based beam steerable free space optical communication link for reconfigurable wireless data center. In *Proc. of SPIE Vol*, volume 10128, pages 1012805–1, 2017.
[24] N. Farrington. *Optics in data center network architecture*. University of California, San Diego, 2012.
[25] M. S. Ferraro et al. InAlAs/InGaAs avalanche photodiode arrays for free space optical communication. *Appl. Opt.*, 2015.
[26] M. Ghobadi et al. Projector: Agile reconfigurable data center interconnect. In *SIGCOMM*. ACM, 2016.
[27] A. Greenberg et al. Vl2: a scalable and flexible data center network. In *SIGCOMM*. ACM, 2009.
[28] D. Halperin et al. Augmenting data center networks with multi-gigabit wireless links. In *SIGCOMM*. ACM, 2011.
[29] N. Hamedazimi et al. Firefly: A reconfigurable wireless data center fabric using free-space optics. In *SIGCOMM*. ACM, 2014.
[30] J. Hamilton et al. Data center networks are in my way. *Standford Clean Slate CTO Summit*, 2009.
[31] T.-H. Ho. *Pointing, Acquisition, and Tracking Systems for Free-Space Optical Communication Links*. PhD thesis, University of Maryland, College Park, 2007.
[32] S. V. Kartalopoulos. *Free Space Optical Networks for Ultra-Broad Band Services*. John Wiley and Sons, 2001.
[33] D. Kedar and S. Arnon. Urban optical wireless communication networks: the main challenges and possible solutions. *IEEE Communications Magazine*, 2004.
[34] C. Lv et al. Implementation of FTA with high bandwidth and tracking accuracy in FSO. In *International Conference on Consumer Electronics, Communications and Networks (CECNet)*, 2012.
[35] J. Mudigonda et al. Taming the flying cable monster: A topology design and optimization framework for data-center networks. In *USENIX Annual Technical Conference*, 2011.
[36] R. Niranjan Mysore et al. Portland: a scalable fault-tolerant layer 2 data center network fabric. In *SIGCOMM*. ACM, 2009.
[37] L. Popa et al. A cost comparison of datacenter network architectures. In *Co-NEXT*. ACM, 2010.
[38] A. Singla et al. Proteus: a topology malleable data center network. In *ACM SIGCOMM Workshop on Hot Topics in Networks*, 2010.
[39] A. Singla et al. Jellyfish: Networking data centers, randomly. In *NSDI*, 2012.
[40] X. Tan et al. An advanced rack server system design for rotational vibration (rv) performance. In *Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm), Intersociety Conference on*. IEEE, 2016.
[41] G. Wang et al. c-through: Part-time optics in data centers. In *SIGCOMM*. ACM, 2010.
[42] T. Yamashita et al. The new tracking control system for free-space optical communications. In *2011 International Conference on Space Optical Systems and Applications (ICSOS)*, 2011.
[43] X. Zhou et al. Mirror mirror on the ceiling: Flexible wireless links for data centers. *SIGCOMM*, 2012.

Multi–User Frequency–Time Coded Quantum Key Distribution Network Using a Plug-and-Play System

Y.T. Aladadi[1], A.F. Abas[1], Abdulmalik Alwarafy[1], M.T. Alresheedi[1]
[1]*Department of Electrical Engineering, College of Engineering, King Saud University, Riyadh 11421, Saudi Arabia*

yaladadi@ksu.edu.sa
aabas@ksu.edu.sa
437106913@student.ksu.edu.sa
malresheedi@ksu.edu.sa

Abstract— **In this paper, we propose and analyse a multi-user wavelength division multiplexing technique of frequency-time coded quantum key distribution that uses a plug and play scheme. Numerical simulation results show that the influence of the channel noise is reduced. At the same time, the final key rate per user is enhanced to be close to that of point-to-point link. This performance is the result of simultaneous communications between Alice and four Bobs.**

***Keywords*—Quantum Key Distribution, Plug-and-Play System, Wavelength Division Multiplexing, Frequency-Time Coding, Multi-Wavelength Laser Diode.**

I. Introduction

Quantum Key Distribution (QKD) [1] is a good security solution for optical communication systems. It overcomes the imperfections of classical cryptography by providing a way to securely generate arbitrarily long cryptographic keys using the quantum properties of lights. In the reported literature, the implementations of QKD rely on the polarization coding [1, 2], phase coding [3, 4], frequency coding [5], time coding [6] and entanglement [7]. In case of a polarization coding, the information is carried by the state of polarization (SOP) that should be recovered at the receiver. This technique suffers from Polarization Mode Dispersion (PMD) and Polarization Dependent Loss (PDL)[3]. For the phase coding the quantum bit error rate (QBER) is related to the interference visibility, which is influenced by the noise of channel; therefore, feedback control is needed to stabilize the interferometer[8]. Differential phase coding schemes are introduced to compensate the drawbacks of phase coding schemes [9]. Disadvantages of QKD channels with frequency coding are associated, mainly, with strong levels of carrier and photon subcarriers in one optical fiber and its power grid [10].

A plug-and-play system is a round-trip two-way QKD system that can automatically compensate for the birefringence effect; therefore, it can operate stably for a long period of time without requiring any polarization control in a long optical fiber [4].

Frequency time coding scheme is introduced to reduce the influence of the channel noise [11]. Wavelength division multiplexing (WDM) QKD scheme has been introduced to overcome the inefficiency of splitter. Multi-user QKD systems that employs different wavelengths to transmit an optical pulses to multiple users have been introduced [12-15]. It is known in principle of communication that the final key rate per user decreased as 1/N, where N is the total number of subscribers.

In this paper, we propose and analyses a multi-user wavelength division multiplexing of frequency-time coded (FT) QKD that uses a plug-and-play scheme. QKD based frequency and time coding has lower QBER as compared to other techniques [11]. Combining the plug and play system with WDM maintains the key rate per user to values that are close to that in case point-to-point communication [16].

II. Point-to-Point Frequency-time Coded QKD scheme

In frequency time coded QKD (FT-QKD) [11], the key is encoded in the frequency and time between Alice and Bob. The proposed point-to-point FT-QKD scheme is shown in Fig. 1. There are two laser diodes, LD1 and LD2, which operate at different designed wavelengths. Both lasers are employed for frequency coding.

For the third laser diode (LD3), time delay is introduced for realizing time coding. LD1 and LD2 generate narrow pulses (in frequency domain) with central wavelengths λ_1 and λ_2, respectively as shown in Fig. 2; whereas λ_3 is considered as central frequency of LD3. The bandwidth of the pulse generated by LD3 should be at least the double of that in LD1 or LD3 because the detection gate duration is twice of the width of the pulse sent [11].

978-3-903176-07-2 © 2018 IFIP

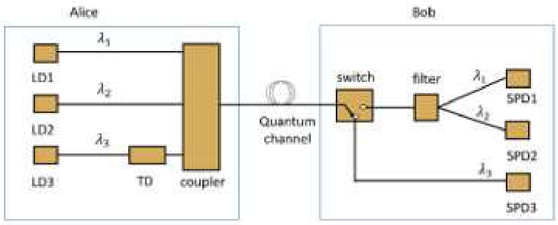


Fig. 1 Schematic of point to point (PTP) FT-QKD system

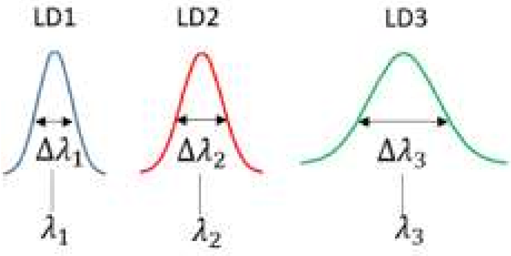


Fig. 2 The frequency domain of laser diodes pulses

To understand how this system works, let us consider that Alice and Bob generate 11 qubits with 11 basis randomly as described in Table I.

Table I: The Bits and Basis Generated By Alice and Bob

Alice's bits:	10000101011
Alice's basis:	00100110000
Bob's bits:	10111010101
Bob's basis:	00110101100

The transmitted photons according to both bits and basis of Alice are shown in Fig. 3. It is clear that when the basis is zero, the frequency coding is selected; whereas time coding is selected when the basis is one.

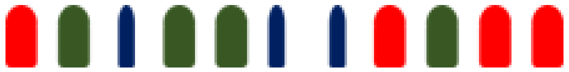


Fig. 3 The transmitted photons

In case of frequency coding, and the bit is zero, the LD1 is fired; whereas LD2 is fired when the transmitted bit is one. In other hand, the time delay is zero when bit is zero and the LD3 is the selected laser according to the basis. When the selected bit is one and LD3 is fired, the time delay (TD) is adjusted to τ.

The transmitted photons are combined at coupler to be transmitted through a quantum channel (QC). The optical switch at Bob works according to Bob's basis. This mean that optical switch operates according to Bob's basis. The received phonons are detected by three single photon detectors that operate at different designated wavelengths. The photons after detection process are shown in fig. 4. It is clear that the received photons by detectors are these with the same basis at both Alice and Bob.

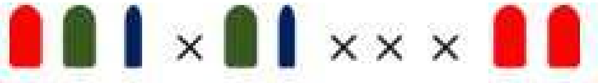


Fig. 4 The received photons

III. System Setup

Fig. 5 shows the proposed system setup. Instead of using single laser diode, multi-wavelength laser diode (MW-LD) is employed. Wavelength selective switch (WSS) is used to select the four pulse signals with differently designated wavelengths generated by MW-LD. As mentioned in Section II, MW-LD1 and MW-LD2 are used in the case of frequency coding; whereas MW-LD3 and TD are employed for time coding. The pulses from three multi-laser are combined using a multiplexer and passed through a circulator (CIR), and subsequently launched into the quantum channel (QC). The variable attenuator (VA) at each Bob is set to a low level and bright laser pulses are emitted by Alice [17]. The transmitted photons pass through two quantum channels, and this makes the distance between Alice and the other four users different. So, time delay and line delay are required to tune the arrival time of the returned pulses in a group to be the same. This helps to reduce the impact of Rayleigh backscattered light [16]. On the other hand, a waiting time will reduce the final key rate.

To understand the principle of the proposed system, let Alice and four Bob generate their bits and basis randomly as shown in Table II.

Table II: The Bits Generated By Alice and the Four Bobs

Alice's bits:	10000101011
Alice's basis:	00100110000
Bob1's bits:	10111010101
Bob1's basis:	00110101100
Bob2's bits:	10111111101
Bob2's basis:	00100101101
Bob3's bits:	10111111101
Bob3's basis:	11001011101
Bob4's bits:	11100000001
Bob4's basis:	10100010100

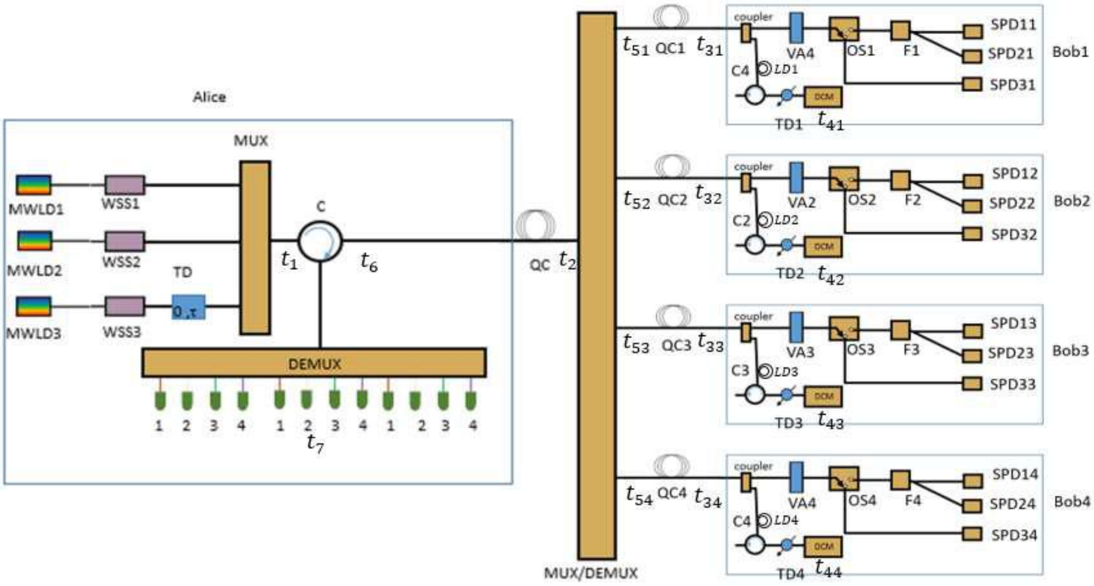


Fig. 1 Schematic of multi-user WDM-FT QKD system

Fig.6 shows the process of pulse transmission at seven positions *t₁*, *t₂*,…, and *t₇*. Time position *t₁* refers to the pulse group after the multiplexer. Then, the pulse group is passed through the QC, before entering the MUX/DEMUX as marked at *t₂*. At t₃, due to the possibility that QC of each user has different length, Bob *i* may receive the transmitted photons before Bob *j* and each one receives the transmitted photons according to his's basis. At *t₄*, the four users complete the reception process, and the driver and control module collects the data and reflects it to Alice with different delay [11]. Therefore, a time delay and line delay are needed to compensate this

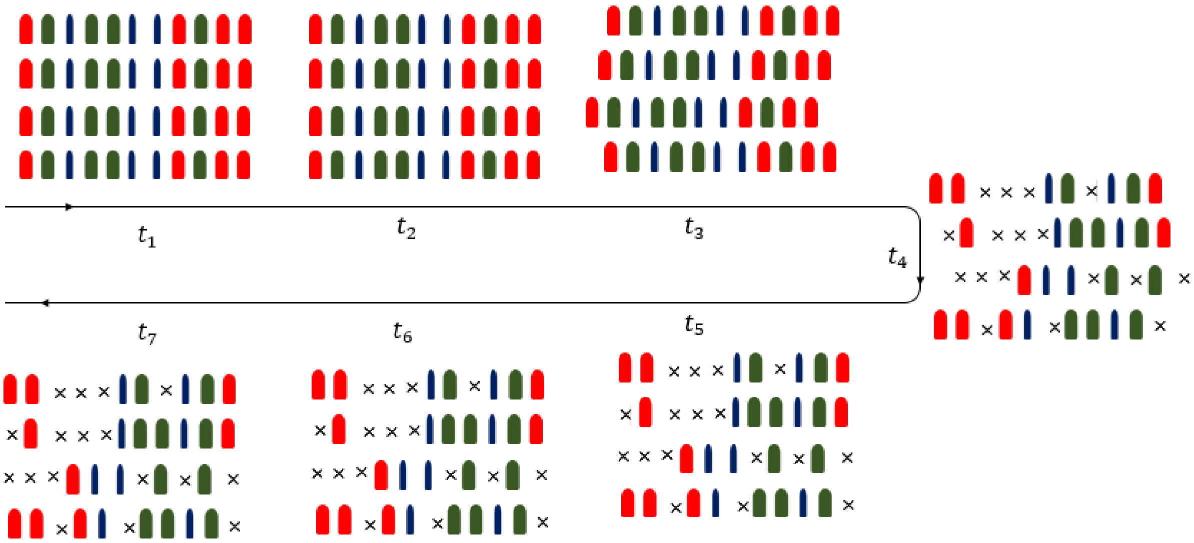


Fig. 2 The pulse transmission process, starting from t_1 to t_7

shortage. It is clear that at t_5, the received photons by all users have the same positions. At t_7, Alice compares her basis with the basis of each user. Then she calculates QBER. If QBER < QBER$_{thr}$, then eavesdropper (Eve) exits, QKD falses and retransmits the photons. Otherwise, if QBER > QBER$_{thr}$, Alice and each user (Bob *i*) obtain the final key that has the same basis after data reconciliation and privacy amplification. According to comparison, the final key rates of Alice and the four Bobs are given in Table III.

Table III: **The Final Key Rates of Alice and the Four Bobs**

7 basis matches.	
Alice's key:	1000111
Bob1's key:	1011001
7 basis matches.	
Alice's key:	1000011
Bob2's key:	1011110
5 basis matches.	
Alice's key:	00101
Bob3's key:	01111
8 basis matches.	
Alice's key:	00000111
Bob4's key:	11000001

IV. Results and Discussion

The sifted key rate and quantum bit error rate (QBER) are the most important parameters used to evaluate a QKD system. The sifted key rate (Raw rate) [17] is given by:

$$R_{raw} = \frac{1}{2} f_r \, \mu t_{AB} t_B \eta_B \tag{1}$$

where μ denotes the mean photon number of each weak coherent pulse, f_r is the pulse repetition rate, t_{AB} is the transmittance of the link from Alice to Bob, t_B is Alice's internal transmittance and η_B is Alice's detector efficiency. R_{raw} is the same for both BB84 protocol (the first implementation method of QKD that uses phase coding or polarization coding) and for FT coding [11, 17, 18]. The difference appears in the QBER, in which QBER of BB84 protocol is given by [16]:

$$QBER_{BB84} = \frac{1-V}{2} + \frac{p_{dark}}{\mu t_{AB} t_B \eta_B} + \sum_{n=0}^{\frac{1}{p_{det}}} p_{after}\left(\tau + n\frac{1}{f_r}\right) \tag{2}$$

where V denotes the visibility of the interference meter, p$_{dark}$ is the probability of a dark count per gate, p$_{det}$ is the probability of a detector click, p$_{after}$ is the probability of an after-pulse over all and τ is the detector's dead time. Both p_{dark} and p_{after} depend on the characteristics of the photon counters. For FT protocol, suppose that the operating wavelength is 1550 nm, and Δt_1 (Δt_2) be 1000 ps, and Δt_3 is 500 ps, and the associated $\Delta\lambda_1$ ($\Delta\lambda_2$) is $8 \times 10^{-3}\ nm$, and $\Delta\lambda_3$ is $16 \times 10^{-3}\ nm$. The detection gate duration is double that of sent pulse duration. Therefore, the effect of time spread and frequency spread from dispersion on detection results can be neglected [11]. So, the first part of Eq. 2 is set to zero when the basis is the same. The QBER of FT protocol is given as:

$$QBER_{FT} = \frac{p_{dark}}{\mu t_{AB} t_B \eta_B} + \sum_{n=0}^{\frac{1}{p_{det}}} p_{after}\left(\tau + n\frac{1}{f_r}\right) \tag{3}$$

In our proposed scheme, a time delay and line delay are taken into account. Therefore, raw kay rate is derived as:

$$\hat{R}_{raw} = \frac{\frac{1}{2} f_r \, \mu t_{AB} t_B \eta_B t_{ex} + \frac{1}{2} f_r \, \mu t_{BA} t_A \eta_A \, \eta_{TD} t_{ex}}{2} \tag{4}$$

and,

$$t_{AB} = 10^{-\frac{\alpha L}{10}} \tag{5}$$

$$t_{BA} = 10^{-\alpha(L+L_{LD})/10} = t_{AB} 10^{-\frac{\alpha L_{LD}}{10}} \tag{6}$$

$$\eta_{TD} = \frac{1}{1 + t_{TD}} \tag{7}$$

where, L is the fiber length between Alice and Bob, LLD is the fiber length of line delay, t_{TD} is time delay at Bob, and t_{ex} is the extra transmittance due to MUX and DEMUX.

Let , $t_B = t_A$, $\eta_B = \eta_A$. So, raw key rate can be described as:

$$\bar{R}_{raw} = \frac{\frac{1}{2} f_r \, \mu t_{AB} t_B \eta_B t_{ex} \left(1 + \ 10^{-\frac{\alpha L_{LD}}{10}} \eta_{TD}\right)}{2} \tag{8}$$

$$\bar{R}_{raw} = \frac{1}{4} f_r \, \mu t_{AB} t_B \eta_B t_{ex} \left(1 + \ 10^{-\frac{\alpha L_{LD}}{10}} \eta_{TD}\right) \tag{9}$$

QBER due to the dark count probability is modified as shown in Eq. 10.

$$\overline{QBER} = \frac{p_{dark}}{2\mu t_{AB} t_B \eta_B t_{ex}} + \frac{1}{2}\sum_{n=0}^{\frac{1}{p_{det}}} p_{after}\left(\tau + n\frac{1}{f_r}\right) + QBER_{BA} \tag{10}$$

$$QBER_{BA} = \frac{p_{dark}}{2\mu t_{BA} t_A \eta_A \eta_{TD} t_{ex}} + \frac{1}{2}\sum_{n=0}^{\frac{1}{p_{det}}} p_{after}\left(\tau + n\frac{1}{f_r}\right) \tag{11}$$

$$\overline{QBER} = \frac{p_{dark}}{2\mu t_{AB} t_B \eta_B t_{ex}}\left(1 + \frac{10^{\frac{\alpha L_{LD}}{10}}}{\eta_{TD}}\right) + \sum_{n=0}^{\frac{1}{p_{det}}} p_{after}\left(\tau + n\frac{1}{f_r}\right) \quad (12)$$

Final key rate of the proposed scheme is given by:

$$\bar{R}_{final} = \frac{1}{4} f_r\, \mu t_{AB} t_B \eta_B t_{ex}(I_{AB} - I_{AE}) + \frac{1}{4} f_r\, \mu t_{BA} t_A \eta_A\, \eta_{TD}(I_{BA} - I_{BE}) t_{ex} \quad (13)$$

$$I_{AB} = 1 - H_2(QBER_{FT}) \quad (14)$$

$$I_{BA} = 1 - H_2(QBER_{BA}) \quad (15)$$

$$I_{AE} = \mu(1 - t_{AB}) + 1 - V \quad (16)$$

$$I_{AE} = \mu(1 - t_{BA}) + 1 - V \quad (17)$$

where I_{AE} denotes the mutual information between Alice and Eve, and $H_2(Q)$ is the binary entropy which is defined as [19]:

$$H_2(Q) = -Qlog_2(Q) - (1 - Q)log_2(1 - Q) \quad (18)$$

The parameters used in the numerical simulation are summarized in Table IV.

Table IV: Simulation Parameters.

Parameter	Value
Pulse repetition rate (f_r)	4MHz
Pulse width	500 ps, 1000 ps
Average number of photons per pulse (μ)	0.1
Transmittance of MUX and DEMUX	0.9
Fiber attenuation coefficient (α)	0.2 dB/km
Detector efficiency at 1,550 nm (η_A)	10%
Probability of dark count (p_{dark})	10–5/gate
Probability of a detector click (p_{det})	0.15%
Detection gate	2 ns
Dead time (τ)	10 μs
Fringe visibility (V)	0.8 , 0.9
Transmittance of Bob's system (t_B)	0.6
After-pulse count probability (p_{after})	4%
Bob' delay line (LDL)	10 km

Fig. 7 shows the QBER of three case, BB84 point-to-point (PTP), FT PTP, and multi user FT system, for different fringe visibility (V). From this figure, it is clear that BB84 protocol is sensitive to V, where decreasing V, gives wore QBER. For example, at L =100 km, QBER is about 0.237 at V=0.9, and 0.287 at V=0.8. The results in Fig. 7 show that independent on the value of V, FT reduces the QBER to 0.187 compared to BB84. The line delay in FT MU system is about 10 km, and this causes a small increase in QBER, 0.24, compared to FT PTP, 0.187. However, FT still better than BB84 for fiber length L< 100 km.

Fig. 8 shows the final key rate against fiber length. Final key rate is very sensitive to V. For instance, at L = 10 km, changing V from 0.9 to 0.8, decreases the final key rate from 4204 b/s, 5601 b/s, and 6182 b/s to 2105 b/s, 4919 b/s and 5425 b/s for BB84 PTP, FT PTP, and FT MU systems, respectively. However, the amount of change in FT system is very small and still work in worse visibility compared with BB84 system. Furthermore, it is clear that the key rate per user in FT MU maintains a high level compared with FT PTP because the communications between Alice and four Bobs can be carried out simultaneously. Meanwhile, when $V =$ 0.9 , and $\bar{R}_{final}$=3000 b/s, the communication distance of FT PTP and FT MU are increased by 9 Km, and 7 km respectively compared to BB84 system. Also, it is clear that the proposed Multi user-QKD network provides a better performance in situations in which all users share a similar quantum channel.

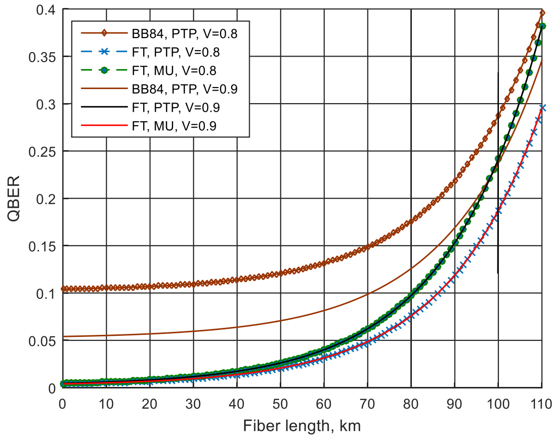


Fig. 7 The QBER versus fiber length between Alice and Bob, for three case, BB84 PTP, frequency time coding PTP, and frequency time coding MU, V=0.8, 0.9, L_{LD}=10 km

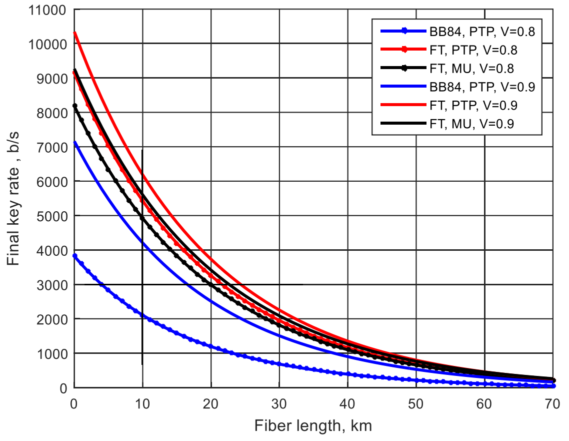


Fig. 8 The final key rate versus fiber length between Alice and Bob, for three case, BB84 PTP, frequency time coding PTP, and frequency time coding MU, V=0.8, 0.9, L_{LD}=10 km

Acknowledgment

The project was supported by King Saud University, Deanship of Scientific Research, College of Engineering Research Center.

V. Conclusion

Frequency and time coding reduce the QBER as compared to BB84 that employs polarization coding or phase coding. Our simulation results show that the FT protocol can work at worse visibility, and offers extra distance. Furthermore, a multi-user WDM-QKD uses the same principle of FT scheme, where QBER is still less than that of BB84 protocol. Meanwhile, the key rate per user maintains a high level compared to point-to-point links. This is because the communications between Alice and four Bobs can be carried out simultaneously.

References

[1] C. H. Bennett and G. Brassard, "Quantum cryptography: Public key distribution and coin tossing," *Theoretical computer science,* vol. 560, pp. 7-11, 2014.

[2] N. Namekata, G. Fujii, S. Inoue, T. Honjo, and H. Takesue, "Differential phase shift quantum key distribution using single-photon detectors based on a sinusoidally gated In Ga As ∕ In P avalanche photodiode," *Applied physics letters,* vol. 91, p. 011112, 2007.

[3] H. Zbinden, N. Gisin, B. Huttner, A. Muller, and W. Tittel, "Practical aspects of quantum cryptographic key distribution," *Journal of Cryptology,* vol. 13, pp. 207-220, 2000.

[4] A. Muller, T. Herzog, B. Huttner, W. Tittel, H. Zbinden, and N. Gisin, "“Plug and play” systems for quantum cryptography," *Applied physics letters,* vol. 70, pp. 793-795, 1997.

[5] J.-M. Merolla, Y. Mazurenko, J.-P. Goedgebuer, and W. T. Rhodes, "Single-photon interference in sidebands of phase-modulated light for quantum cryptography," *Physical review letters,* vol. 82, p. 1656, 1999.

[6] D. Stucki, C. Barreiro, S. Fasel, J.-D. Gautier, O. Gay, N. Gisin, R. Thew, Y. Thoma, P. Trinkler, and F. Vannel, "Continuous high speed coherent one-way quantum key distribution," *Optics express,* vol. 17, pp. 13326-13334, 2009.

[7] F. Xu, B. Qi, Z. Liao, and H.-K. Lo, "Long distance measurement-device-independent quantum key distribution with entangled photon sources," *Applied physics letters,* vol. 103, p. 061101, 2013.

[8] Z. Yuan and A. Shields, "Continuous operation of a one-way quantum key distribution system over installed telecom fibre," *Optics express,* vol. 13, pp. 660-665, 2005.

[9] K. Inoue, E. Waks, and Y. Yamamoto, "Differential phase shift quantum key distribution," *Physical review letters,* vol. 89, p. 037902, 2002.

[10] I. M. Gabdulhakov and O. G. Morozov, "Frequency coded quantum key distribution channel based on photon amplitude-phase modulation," in *Systems of Signal Synchronization, Generating and Processing in Telecommunications (SINKHROINFO), 2017,* 2017, pp. 1-5.

[11] Z. Chang-Hua, P. Chang-Xing, Q. Dong-Xiao, G. Jing-Liang, C. Nan, and Y. Yun-Hui, "A new quantum key distribution scheme based on frequency and time coding," *Chinese Physics Letters,* vol. 27, p. 090301, 2010.

[12] J. Bogdanski, N. Rafiei, and M. Bourennane, "Multiuser quantum key distribution over telecom fiber networks," *Optics Communications,* vol. 282, pp. 258-262, 2009.

[13] P. D. Kumavor, A. C. Beal, E. Donkor, and B. C. Wang, "Experimental multiuser quantum key distribution network using a wavelength-addressed bus architecture," *Journal of lightwave technology,* vol. 24, p. 3103, 2006.

[14] C. Autebert, J. Trapateau, A. Orieux, A. Lemaître, C. Gomez-Carbonell, E. Diamanti, I. Zaquine, and S. Ducci, "Multi-user quantum key distribution with entangled photons from an AlGaAs chip," *Quantum Science and Technology,* vol. 1, p. 01LT02, 2016.

[15] V. Fernandez, K. J. Gordon, R. J. Collins, P. D. Townsend, S. D. Cova, I. Rech, and G. S. Buller, "Quantum key distribution in a multi-user network at gigahertz clock rates," in *Photonic Materials, Devices, and Applications*, 2005, pp. 720-728.

[16] G. Cheng, B. Guo, C. Zhang, J. Guo, and R. Fan, "Wavelength division multiplexing quantum key distribution network using a modified plug-and-play system," *Optical and Quantum Electronics,* vol. 47, pp. 1809-1817, 2015.

[17] D. Stucki, N. Gisin, O. Guinnard, G. Ribordy, and H. Zbinden, "Quantum key distribution over 67 km with a plug&play system," *New Journal of Physics,* vol. 4, p. 41, 2002.

[18] B. Qi, "Quantum key distribution based on frequency-time coding: security and feasibility," *arXiv preprint arXiv:1101.5995,* 2011.

[19] C. Macchiavello, G. M. Palma, and A. Zeilinger, *Quantum Computation and Quantum Information Theory: Reprint Volume with Introductory Notes for ISI TMR Network School, 12-23 July 1999, Villa Gualino, Torino, Italy*: World Scientific, 2000.

Optimization of the Upstream Bandwidth Allocation in Passive Optical Networks Using Internet Users' Behavior Forecast

Nejm Eddine Frigui*, Tayeb Lemlouma†, Stéphane Gosselin*
Benoit Radier*, Renaud Le Meur*, and Jean-Marie Bonnin‡
*Orange Labs, Lannion, France
†University of Rennes 1 IRISA, Lannion, France
‡IMT Atlantique, Cesson Sévigné, France
nejm.frigui@orange.com, tayeb.lemlouma@irisa.fr, stephane.gosselin@orange.com
benoit.radier@orange.com, renaud.lemeur@orange.com, jm.bonnin@imt-atlantique.fr

***Abstract*—The application of classification techniques based on machine learning approaches to analyze the behavior of network users has interested many researchers in the last years. In a recent work, we have proposed an architecture for optimizing the upstream bandwidth allocation in Passive Optical Network (PON) based on the traffic pattern of each user. Clustering analysis was used in association with an assignment index calculation in order to specify for PON users their upstream data transmission tendency. A dynamic adjustment of Service Level Agreement (SLA) parameters is then performed to maximize the overall customers' satisfaction with the network. In this work, we extend the proposed architecture by adding a prediction module as a complementary to the first classification phase. Grey Model GM(1,1) is used in this context to learn more about the traffic trend of users and improve their assignment. An experimental study is conducted to show the impact of the forecaster and how it can overcome the limits of the initial model.**

***Index Terms*—Passive Optical Network (PON), Clustering Analysis, Service Level Agreement (SLA), Grey Model GM(1,1)**

I. Introduction

In the recent years, a change of paradigm in fixed access networks has been experienced. The fast emergence of Passive Optical Networks (PONs) allowed to carry huge amounts of traffic and to offer high bandwidth services to operators' customers. However, the continuous exponential growth of data traffic in the next years as well as the expected widespread integration of Internet of Things, 5G networks, and high-speed services may impact the efficiency of the bandwidth allocation process. Dynamic Bandwidth Allocation (DBA) is currently the mechanism responsible for allocating the up-stream resources in PONs. To optimize the DBA performance, two approaches can be distinguished. The first consists in modifying the way in which the DBA works by acting on the algorithm itself and trying to invent a new mechanism that can overcome the limits of the existing one. The second relies on managing the external parameters of the DBA in a different way without modifying the DBA control algorithm itself at the Optical Line Terminal (OLT) level. The main difficulty of the first approach is the inability to be directly implementable from the operator perspective. Indeed, the DBA as a closed control protocol in the PON network cannot be modified by the network operator who doesn't have the total control of this mechanism due to equipment supplier dependency. In this regard, the second approach looks more suitable in a context of network resources optimization under the control of the network administrator.

In a recent work [1], we have proposed an architecture for optimizing the upstream bandwidth allocation in PON based on the dynamic adjustment of Service Level Agreement (SLA) parameters. The latter represent the input parameters of the DBA algorithm that can be managed by the operator. The idea was to efficiently exploit the bandwidth available in the network by adjusting dynamically the SLA parameters based on the estimation of users' traffic patterns linked to daily life. Clustering analysis was used to identify heavy and light users based on their mean upstream bitrates for a specific time interval (e.g., 5 minutes). Then, an Assignment Index Calculator module was proposed to assign each user to a particular class (heavy, light, or flexible) for all the time series possessed by the network operator. The combination of the clustering analysis and the assignment index calculation allows to have an overall vision of the traffic profile of each user and makes it possible to estimate the possible behavior of a specific user at a specific time. In this case, the reallocation of the SLA parameters can be very useful and advantageous in the context of optimizing PON upstream resources for a specific day period. The evaluation phase that we have conducted in [1] was limited to the analysis of the clustering module in order to select which algorithm gives a better distribution of users. In this work, we continue the evaluation of the model that we have proposed in [1] by analysing the assignment index module and its impact on the user classification phase. Then, we extend the proposed architecture by adding a forecasting module as a part of a second user classification step that we suggest to be an improvement of the first classification method.

The paper is structured as follows. Section II presents some work related to the DBA mechanism optimization, a summary of the initial model that we have proposed in a previous

978-3-903176-07-2 © 2018 IFIP

work, its limits, and the need for a forecasting approach to deal with network users' traffic behavior. Section III presents the enhanced version of the initial model using a forecasting module based on the GM(1,1) model. In section IV, we present simulations used to evaluate the classification techniques as well as the obtained experimental results and their analysis. Finally, we conclude our work in section V.

II. Related Work

The dynamicity of users' traffic patterns lets always network operators thinking about new ideas to make the upstream bandwidth allocation mechanism more efficient. In the research world, many works [2]–[5] have been proposed in this context with the aim of enhancing the DBA overall operation. Despite their contribution at the optimization level, the majority of these works remain theoretical proposals that a network operator cannot directly integrate in its equipment due to the implementation nature of the DBA (a closed control protocol) and the dependency on a specific vendor.

With the trend of using machine learning in the last few years, thoughts are directed towards approaches that have the character of being able to be managed and capable of learning over time. In this context, we have proposed in [1] a new model (Fig.1) for the optimization of PON upstream resources which stems from a very simple idea: analyzing the past customers' behavior based on their historical data to estimate and reallocate their upstream bitrates in the future.

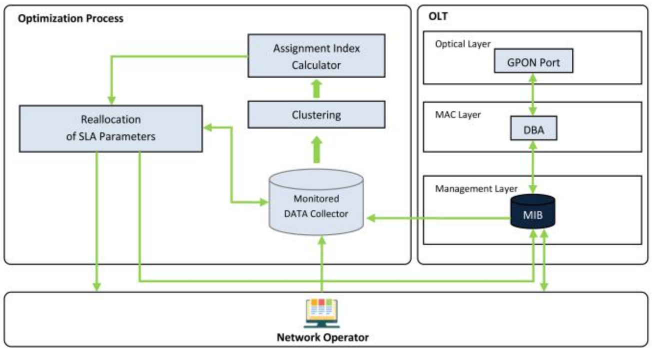


Fig. 1. The proposed model for optimizing PON upstream resources [1]

Indeed, in daily life, traffic patterns of different users may change several times per day. In this case, it is highly possible to have some customers who consume more bandwidth than others for a specific day period. The current DBA allows allocation of PON resources, depending on the instantaneous demand of each Optical Network Unit (ONU). However, each user is limited by predefined Peak Information Rate, which represents one of the SLA parameters which defines the maximum bitrate that a user might benefit from. In this case, when the extra available bandwidth resulting from the presence of light ONUs is greater than the maximum bandwidth authorized to be allocated to heavy users, a part of the extra bandwidth will be lost and not exploited. For this, the idea was to try to exploit this extra bandwidth theoretically untapped by the DBA in order to share it among heavy users without being limited by the constraints of the SLA parameters. As the goal is to propose an implementable approach by the operator in which the DBA algorihm should not be touched, the challenge is then to be able to model the functioning of the DBA while using the historical data provided by the operator and acting only on the DBA externally manageable parameters i.e., SLA parameters.

By analogy with the DBA process, four main components were proposed. The *Monitored Data Collector* gathers the traffic data for each ONU by requesting the Management Information Base (MIB) at regular intervals. This module connects also to the network operator in order to store the historical transmission data. As the DBA refers to the paquet queue status of each ONU to know its needs, two complementary modules responsible for the classification of different users depending on their historical transmission data were proposed. The *clustering* module classifies users into 3 classes: heavy, light, and flexible. Depending on the chosen algorithm, the results may vary. In [1], two well known clustering algorithms namely, K-Means [6] and DBSCAN [7] were evaluated. The results have shown that K-Means using a log_{10}(Bitrate) metric outperforms DBSCAN in terms of a more balanced customer distribution. The clustering module is supposed to be applied per time interval (e.g., 5 minutes). To be able to classify all users based on the entire time series, a second module called *Assignment Index Calculator* was proposed. This module aims to provide the probability for each user to be either heavy, light or flexible. For each day and for each time interval, it analyzes the clustering results. If the user belongs to the heavy class in a given time interval, his probability to be a heavy will increase and likewise for the light class. Then, a calculation process of the average of all probabilities associated to a standard deviation calculation (for validation) is assured to finally assign each user to a specific class. The final users' classification will be used then by the *Reallocation of SLA Parameters* module to define for each heavy user new temporary upstream bitrate allocation in a specific day period.

The purpose of using clustering analysis associated with an assignment index calculation process is to classify PON users depending on their traffic patterns. Although this approach is characterized by its high accuracy in the assignment of a PON user to a certain traffic class, it can lead to an unbalanced user distribution where the majority of ONUs will be assigned to the flexible users class. This may be an impediment to the overall objective which consists in maximizing as much as possible the satisfaction for the majority of customers. In this case, it is preferable to have a significant ratio of heavy and light users in order to maximize the efficiency of the bandwidth usage in the network. To resolve this issue, reference can be made to forecasting-based approaches that deal with the analysis and the prediction of network users' traffic behavior in order to have an idea about the possible future trend of flexible users (whether heavy or light). [8]–[10] represent some works related to the network traffic behavior forecasting in several types of networks. In general, we can distinguish two main techniques for forecasting models: qualitative and quantitative approaches. The first technique relies on the knowledge and the experience of the forecaster who will take the final decision about the expected trend of data. The second one aims to identify the data patterns from the historical dataset in order to predict the future values [11] [12]. Quantitative approaches may be also classified into causal relationship methods and time series ones. While the first category tries to make a relationship between many factors in order to generate the forecasted values, the second is limited to the statistical data that was observed and collected over regular time intervals such as hourly, daily, weekly, monthly, etc [13].

Since the historical data required by the network operator to forceast users' traffic behavior can be easily obtained and processed with the aim to classify the different customers, the focus will be on time series methods and especially on two major forecasting models, respectively Artificial Neural Networks and Grey theory. Artificial neural network (ANN) represents one of the most popular forecasting paradigms [14]. Classified as a machine learning approach, ANN has the ability to learn from complicated data and deduce its pattern and tendency. It can be very appropriate in the context of a knowledge-based learning mechanism that is difficult to specify [15]. By analogy to the human brain, ANN ensure the information process through the interaction of artificial neurons and can interpret the future behavior of a dataset despite the existence of noisy information [11] [15]. As for Grey forecasting theory [16], it was proposed for the first time in 1982. Thanks to its ability to estimate the possible data behavior based only on a few information samples even if they are incomplete, Grey theory becomes one of the most popular prediction approaches used in the research world [17]. The core and the most commonly used model of Grey is known as GM(1,1) [13]. The main task of this model is to identify the mathematical relationship between different points to learn about the behavior of the dataset and to make the right decision about the future trend [11].

In relation with our recent work [1], ANN and Grey model allow both to achieve our main goal concerning the forecasting of customers' traffic behavior. However, we expect that only the Grey Model GM(1,1) remains for the moment the most appropriate for our usecase. This is due to the fact that the dataset we have is limited, which does not represent a problem for Grey systems which can even work with incomplete data. However, neural networks require a very large amount of data to ensure that the forecasted values are statistically accurate, which makes the learning speed slower [11] [18].

III. Enhanced Model Using a Forecasting Module

In this section, we propose the design of the enhanced model inspired by the first model [1] while the main novelty introduced lies in the integration of a forecasting module based on the GM(1,1) model. Fig.2 shows the architecture of the new proposed model.

We expect that the forecasting module will be considered as a second step of the users classification phase as it depends on the results of the clustering analysis and assignment index calculator modules. This new module will be applied particularly to the assignment indexes of flexible users while heavy and light ones will not be involved. Indeed, in our approach, a user can be flexible only if the averages of his assignment indexes to heavy and light users over the supervised days are smaller than 0.5 (the selected threshold). That is, the user does not tend to be neither heavy nor light. Since the index is calculated on the basis of historical data, cases like missing data or the lack of a vision on the traffic trend of users in the future will be often encountered. This can impact the calculation of the assignment index and subsequently the classification of users. In this context, the enhancement of the initial model by using a forecasting approach can be beneficial to have a more reliable and useful approach where a part of flexible users can be assigned to one of the other two classes. Accordingly, the extra bandwidth estimated and the number of beneficiary heavy users will increase automatically. Unlike the initial model where PON users are classified based on the whole set of supervised days, we propose in this work to classify customers per weekdays (from Monday to Friday) and per weekends (Saturday and Sunday). This aims to determine whether the online behavior of PON users is the same for weekdays as it is for weekends.

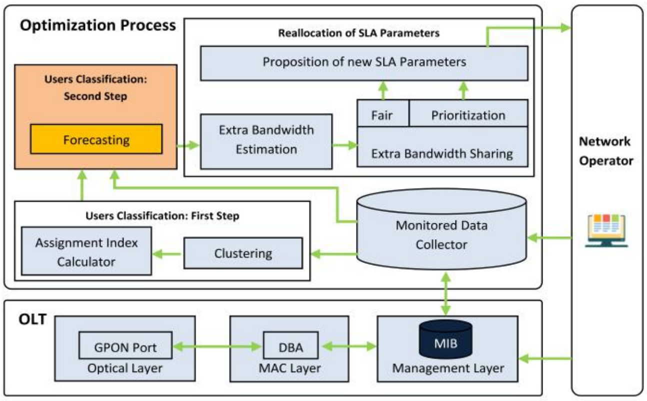

Optimization Process
Reallocation of SLA Parameters
Proposition of new SLA Parameters
Users Classification: Second Step
Forecasting
Extra Bandwidth Estimation
Fair
Prioritization
Extra Bandwidth Sharing
Network Operator
Users Classification: First Step
Assignment Index Calculator
Clustering
Monitored Data Collector
OLT
GPON Port
Optical Layer
DBA
MAC Layer
MIB
Management Layer

Fig. 2. Enhanced model using a forecasting module

The mathematical formulation of the Grey forecasting Model GM(1,1) is illustrated below. We assume the initial data series with $n\,(n \geq 4)$ non-negative values as follows:

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), ..., x^{(0)}(n)) \quad (1)$$

The Accumulated Generating Operation (AGO) is then applied since the initial data series may change randomly while there is a need to know its regular pattern [16]:

$$x^{(1)} = (x^{(1)}(1), x^{(1)}(2), ..., x^{(1)}(n)) \quad (2)$$

Where $x^{(1)}(k) = \sum_{m=1}^{k} x^{(0)}(m), k \in [1, n]$
The original form of the GM(1,1) model is:

$$x^{(0)}(k) + ax^{(1)}(k) = b \quad (3)$$

Where a is the developing coefficient and b is the grey input according to the Grey theory. $x^{(1)}(k)$ can be replaced then by the average of two consecutive neighbours $x^{(1)}(k)$ and $x^{(1)}(k-1)$:

$$x^{(0)}(k) + az^{(1)}(k) = b, k \in [2, n] \tag{4}$$

Where $z^{(1)}(k) = 0,5(x^{(1)}(k) + x^{(1)}(k-1))$
According to the least square method, a and b can be identified as follow:

$$A = \begin{bmatrix} a \\ b \end{bmatrix} = (B^T B)^{-1} B^T Y \tag{5}$$

Where:

$$B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ . & \\ . & \\ . & \\ -z^{(1)}(n) & 1 \end{bmatrix} Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ . \\ . \\ . \\ x^{(0)}(n) \end{bmatrix}$$

By regarding the following differential equation as a shadow of Eq. (4):

$$\frac{dx^{(1)}(k)}{dk} + ax^{(1)}(k) = b \tag{6}$$

The GM(1,1) can be therefore established:

$$\widehat{x}^{(1)}(k+1) = (x^{(1)}(1) - \frac{b}{a})e^{-ak} + \frac{b}{a} \tag{7}$$

As we have applied the AGO in the equation 2, we apply the inverse (IAGO) to obtain the forecasted value of the original data $x^{(0)}$:

$$\widehat{x}^{(0)}(k+1) = (1 - e^{a})(x^{(1)}(1) - \frac{b}{a})e^{-ak} \tag{8}$$

IV. Experimentation and Results

In this section, we proceed to an evaluation of the user classification modules proposed in the initial model [1] and the enhanced one proposed in this work. The objective is to demonstrate that adding a forecasting module can give a more balanced distribution and consequently provide top customers satisfaction. The reference dataset used in this work relies on a real traffic traces collected within the Orange France network. The data collection was ensured by the use of a probe called *OTARIE* and equipped with a DAG (Data Acquisition and Generation) traffic capture card which has an Application Programming Interface (API) that allows reading the packets as they arrive on the network interface. 3447 ONUs belonging to the same OLT were supervised over a period of one month between the 2nd of November and the 3rd of December 2016. Given that the traffic traces do not cover the whole day, the day period theoretically qualified as the busiest which is between 9p.m and 12a.m was selected in order to analyze the behavior of the majority of subscribers.

As the customer traffic pattern is linked to daily life where the online behavior in the weekends may not be the same as the other weekdays, we decide to classify customers per weekday and per weekend. For display reasons, we decide to show the results for a list of Wednesdays as a weekday and a list of Saturdays as day of the weekend. The accomplishment of the experiment relies on the use of Python scripts to evaluate the different algorithms and Matplotlib and Seaborn libraries to plot the different results in the most convenient way. Fig.3 and Fig.4 show the users rate for each class (heavy, light, and flexible) for Wednesdays and Saturdays of the supervised period. These rates are calculated for each day based on the assignment index of each user to the heavy or light classes, calculated once the clustering process based on the K-Means algorithm is finished. This index has been fixed at 0.5 and represents the probability of belonging to a specific class of users. The selected threshold 0.5 is the minimum value that must be chosen to remove any ambiguity concerning the classification step. Indeed, the sum of the assignment indexes to the heavy and light classes is always lower than 1. If we choose a threshold lower than 0.5, we may have cases where both indexes are above the selected threshold and therefore, users will be assigned to both classes at the same time.

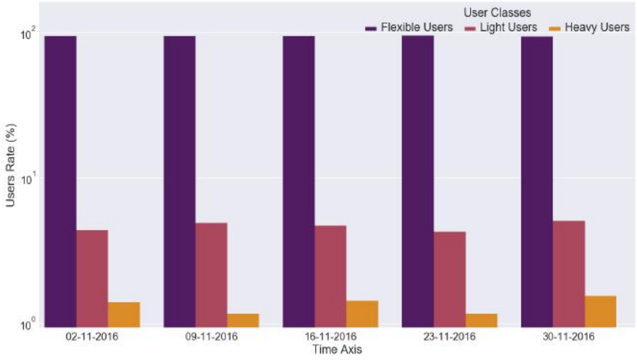


Fig. 3. OLT user distribution for Wednesdays

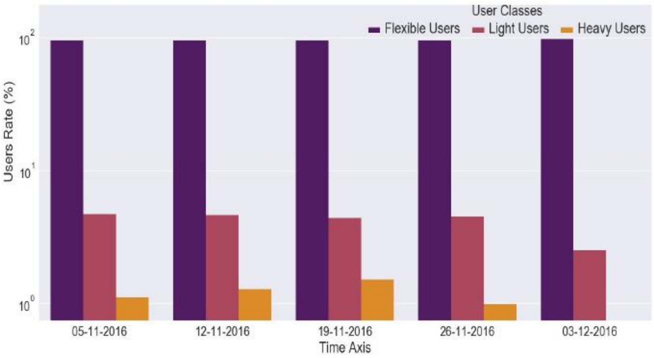


Fig. 4. OLT user distribution for Saturdays

The analysis of the resulting user rates shows an absolute majority of flexible users compared to light and heavy ones as it was expected. This is confirmed for Wednesdays as well as Saturdays. To be able to show the impact of using our forecasting module on the users' rate, we choose to work on a specific OLT PON Port instead of working on the whole OLT. This choice is more appropriate since our optimization approach is supposed to be applied per PON port. As for the whole OLT, Fig.5 and Fig.6 show the users distribution for a PON port that connects 32 subscribers. Fig.7 presents the final distribution of the PON port users based on the average of their assignment indexes over all supervised days.

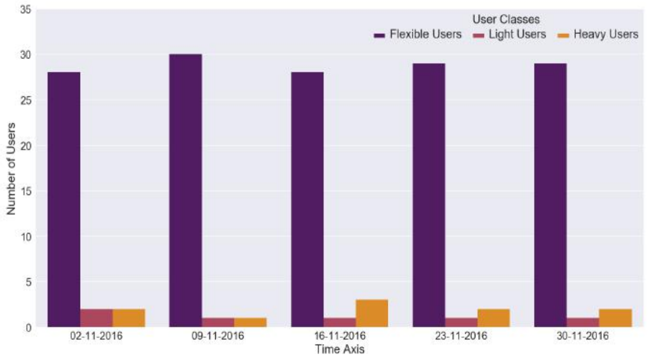


Fig. 5. A PON port user distribution for Wednesdays

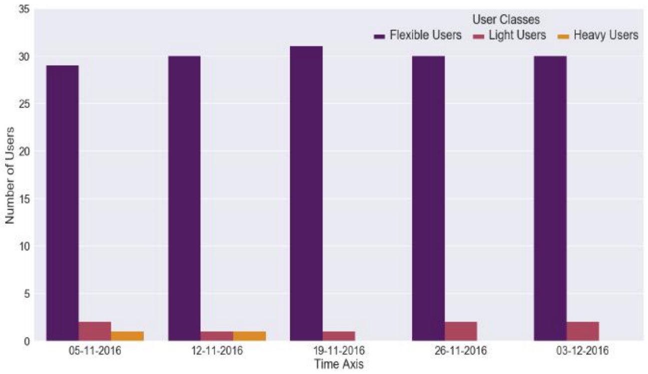


Fig. 6. A PON port user distribution for Saturdays

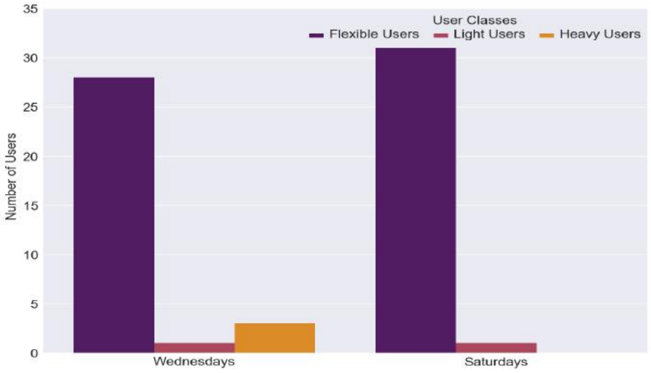


Fig. 7. A PON port user distribution for Wednesdays and Saturdays

As mentioned in section III, the Grey forecasting model takes into account the different user distributions resulted from the combination of clustering analysis and assignment index calculation. While the heavy and light users are already selected with high precision, the flexible ones which represent the majority may tend to one of the other classes if we extend the dataset and generate more indexes. This can influence the final users distribution and consequently the extra bandwidth that will be estimated to be shared among heavy customers. Fig.8 and Table II highlight for a flexible user, the real values of the assignment index to the heavy class for all Wednesdays in the supervised period, and the forecasted values while extending the dataset by 4 weeks. We evaluated our forecasting module based on GM(1,1) using the metrics presented in Table I.

TABLE I
MATHEMATICAL FORMULAS OF FORECASTING METRICS

Forecasting Metric	Mathematical Formula
Residual	$Real\,Value - Forecasted\,Value$
Forecast Error (FE)	$(\lvert Real\,Value - Forecasted\,Value\rvert / Real\,Value) \times 100$
Forecast Accuracy (FA)	$max(0, 100 - FE)$
Mean Forecast Error (MFE)	$\sum_{i=1}^{n} (Real\,Value_i - Forecasted\,Value_i)/n$
Mean Absolute Deviation (MAD)	$\sum_{i=1}^{n} \lvert Real\,Value_i - Forecasted\,Value_i\rvert /n$
Tracking Signal (TS)	$\sum_{i=1}^{n} (Real\,Value_i - Forecasted\,Value_i)/MAD$

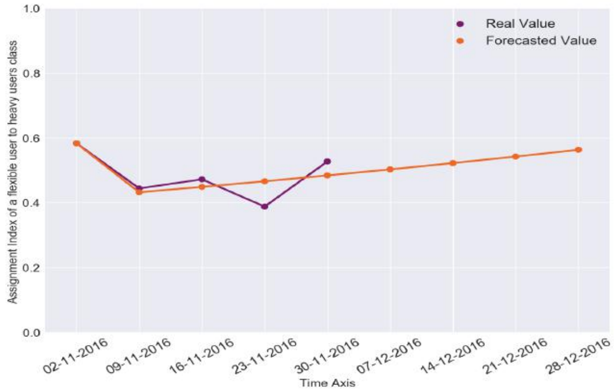


Fig. 8. Forecasted and real assignment indexes of a flexible user to the heavy users class for Wednesdays using GM(1,1)

TABLE II
ASSIGNMENT INDEX (AI) OF A FLEXIBLE USER TO THE HEAVY CLASS: REAL AND FORECASTED VALUES

Day	AI to Heavy Users Real Value	AI to Heavy Users Forecasted Value	Residual	Forecast Error (%)	Forecast Accuracy (%)
02-11-2016	0,583	0,583	0	0	100
09-11-2016	0,444	0,432	0,012	2,702	97,298
16-11-2016	0,472	0,448	0,024	5,084	94,916
23-11-2016	0,388	0,465	-0,077	19,845	80,155
30-11-2016	0,527	0,483	0,044	8,349	91,651
07-12-2016	-	0,502	-	-	-
14-12-2016	-	0,522	-	-	-
21-12-2016	-	0,542	-	-	-
28-12-2016	-	0,563	-	-	-
		Mean Forecast Error (MFE) :	0,0006		
		Mean Absolute Deviation (MAD) :	0,0314		
		Tracking Signal (TS) :	0,095		

The results demonstrate that the average of the assignment indexes to the heavy class for this flexible user increases from 0.4828 (using real values for the supervised days) to 0.505 (using real values for the supervised days and forecasted ones for the next 4 weeks), which qualifies it as a heavy user instead of a flexible one. The different metrics used to evaluate the GM(1,1) show high performances of this model with a good forecast accuracy (the minimum obtained is 80,155 %) and a low Mean Absolute Deviation. The Tracking Signal is generally used to decide if the forecasting model need or not to be reviewed. The low value that we have obtained for this parameter brings confirmation of the good quality of the GM(1,1) model. We performed the same approach for all flexible users belonging to the same PON port. The new users classification is shown in Fig.9. Whereas before using the forecasting model, the rate of flexible users was 87,5% for Wednesdays and 96,87 % for Saturdays, the application of the GM(1,1) shows a more balanced distribution where the flexible users rate decreases to 68,75% for Wednesdays and 75% for Saturdays. The fact that a part of flexible users tends to be heavy more than light for Wednesdays whereas it's the opposite for Saturdays can be explained by making the analogy with the daily life of connected people where the most of them tend to go out on weekends more than weekdays. By looking to the new users' distribution resulted from the application of the Grey model, it's clear that the additional bandwidth that can be estimated will be greater since the number of light users has increased whether for Wednesdays or Saturdays. Additionally, the number of users who will benefit from the bandwidth supplement i.e., heavy users, has also increased, which asserts that the use of the GM(1,1) looks essential and advantageous in a context of satisfying the majority of customers in our approach.

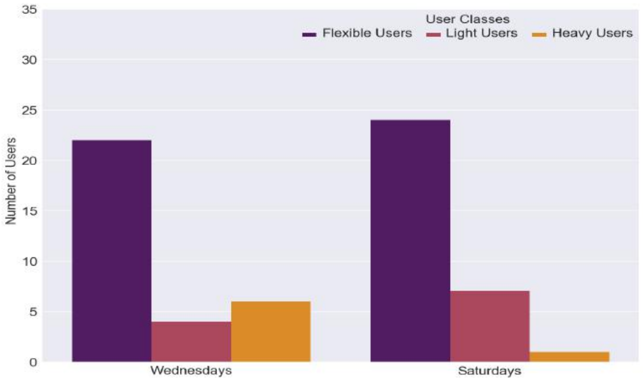


Fig. 9. A PON port user distribution for Wednesdays And Saturdays Using GM(1,1)

V. Conclusions and Future Work

A new approach for enhancing PON users classification based on their traffic patterns has been proposed in this paper. In a previous work, we have proposed a mechanism for reallocating SLA parameters in a PON network based on their online behavior. This mechanism has as objective to optimize the upstream bandwidth allocation process without modifying the DBA itself. The idea was to classify PON users into 3 classes, heavy, light, and flexible and then, try to add an extra bandwidth to heavy users for a specific day period. The classification mechanism was designed based on clustering analysis and an assignment index calculation method. This mechanism is limited by the fact that the majority of users were assigned to the flexible class, which looks like an obstacle in our optimization approach. In this work, we referred to the Grey forecasting theory in order to predict the possible traffic behavior of flexible users in the future with the aim to have a more balanced distribution. Results have shown clearly the advantage of using this predictive approach to improve the final users distribution which impacts directly the extra bandwidth estimation and the number of beneficiary customers. In a future work, we expect to proceed to the whole evaluation of the proposed model taking into account several QoS parameters. We also plan to have a third version of our model based on the self management aspect where our optimization mechanism will be integrated in a real platform and managed by the network itself without any human intervention.

References

[1] N. E. Frigui, T. Lemlouma, S. Gosselin, B. Radier, R. Le Meur, and J.-M. Bonnin, "Dynamic reallocation of SLA parameters in passive optical network based on clustering analysis," in *2018 21st Conference on Innovation in Clouds, Internet and Networks (ICIN) (ICIN 2018)*, Paris, France, Feb. 2018.

[2] D. Nowak, P. Perry, and J. Murphy, "Bandwidth allocation for service level agreement aware Ethernet passive optical networks," in *Global Telecommunications Conference, 2004. GLOBECOM '04. IEEE*, vol. 3, Nov 2004, pp. 1953–1957.

[3] S. i. Choi and J. Park, "SLA-Aware Dynamic Bandwidth Allocation for QoS in EPONs," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 2, no. 9, pp. 773–781, September 2010.

[4] B. Skubic, J. Chen, J. Ahmed, L. Wosinska, and B. Mukherjee, "A comparison of dynamic bandwidth allocation for EPON, GPON, and next-generation TDM PON," *IEEE Communications Magazine*, vol. 47, no. 3, pp. S40–S48, March 2009.

[5] C. M. Assi, Y. Ye, S. Dixit, and M. A. Ali, "Dynamic bandwidth allocation for quality-of-service over Ethernet PONs," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 9, pp. 1467–1477, 2003.

[6] J. MacQueen *et al.*, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, no. 14. Oakland, CA, USA., 1967, pp. 281–297.

[7] M. Ester, H.-P. Kriegel, J. Sander, X. Xu *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Kdd*, vol. 96, no. 34, 1996, pp. 226–231.

[8] P. Cortez, M. Rio, M. Rocha, and P. Sousa, "Multi-scale internet traffic forecasting using neural networks and time series methods," *Expert Systems*, vol. 29, no. 2, pp. 143–155, 2012.

[9] R. Babiarz and J.-S. Bedo, *Internet Traffic Mid-term Forecasting: A Pragmatic Approach Using Statistical Analysis Tools*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 110–122.

[10] V. Alarcon-Aquino and J. A. Barria, "Multiresolution fir neural-network-based learning algorithm applied to network traffic prediction," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 36, no. 2, pp. 208–220, March 2006.

[11] L. Yi, X. Ke, and S. Junde, "Research on forecasting and early-warning methods," in *2013 IEEE 9th International Conference on Mobile Ad-hoc and Sensor Networks*, Dec 2013, pp. 570–576.

[12] M. Uysal and J. L. Crompton, "An overview of approaches used to forecast tourism demand," *Journal of Travel Research*, vol. 23, no. 4, pp. 7–15, 1985.

[13] R. Jouini, T. Lemlouma, K. Maalaoui, and L. A. Saidane, "Employing grey model forecasting gm(1,1) to historical medical sensor data towards system preventive in smart home e-health for elderly person," in *2016 International Wireless Communications and Mobile Computing Conference (IWCMC)*, Sept 2016, pp. 1086–1091.

[14] J. Milojkovi and V. Litovski, "Ann versus grey theory based forecasting methods implemented on short time series," in *10th Symposium on Neural Network Applications in Electrical Engineering*, Sept 2010, pp. 117–122.

[15] G. Zhang, B. E. Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks:: The state of the art," *International journal of forecasting*, vol. 14, no. 1, pp. 35–62, 1998.

[16] J. L. Deng, "Introduction to grey system theory," *J. Grey Syst.*, vol. 1, no. 1, pp. 1–24, Nov. 1989.

[17] X. Wang, L. Qi, C. Chen, J. Tang, and M. Jiang, "Grey system theory based prediction for topic trend on internet," *Engineering Applications of Artificial Intelligence*, vol. 29, pp. 191 – 200, 2014.

[18] J. Cannady, "Artificial neural networks for misuse detection," in *National information systems security conference*, 1998, pp. 368–81.

Designing Multi-Layer Provider Networks for Circular Disc Failures

Aniruddha Kushwaha[#], Deepak Kakadia[*], Ashwin Gumaste[#] and Arun Somani[^]
[#] Dept. of CSE, Indian Institute of Technology, Bombay, Mumbai, India.
[*] Access – Network Performance Group, Google, Mountain View USA.
[^] Dept. of ECpE, Iowa State University, Ames, USA.
Email: aniruddha@cse.iitb.ac.in, dkakadia@google.com, ashwing@ieee.org and arun@iastate.edu

Abstract—**We examine the issue of disaster recovery after zonal outages in core networks, especially IP-over-WDM multi-layer networks. In particular, we consider the network design problem for a regional failure of circular area of radius *R*. Our goal is to design a network that can withstand a randomly located single failure of radius *R*. To this end, we formulate the problem as a constrained optimization problem whose solution for both IP-over-optical networks and pure ROADM-based networks is proposed. Subsequently, we develop an efficient heuristic based on a divide and conquer strategy that gives acceptable results. We also discuss the role of SDN in design and restoration of such networks. Simulation results are showcased over a core network topology thereby realizing the plausibility of such network design.**

I. Introduction

Fault tolerance is an important aspect of wide area networks [1]. Perhaps it is appropriate to say that fault tolerance is the most critical aspect of infrastructure operations on which public services are provisioned [2]. From a telecommunications perspective, node and link failures summarize much of outages. A spectrum of techniques have been developed for protection and restoration of services in the telecommunication domain. Much of these techniques propose 50-millisecond restoration of service post a failure. Complex provider networks are being investigated for node and link failure and a spectrum of protection strategies have been developed towards mitigating node and link failures in mesh networks [3]. Ranging from full capacity per-link 1+1 protection to shared risk link groups (SRLG) [3,4], the protection schemes mostly consider single event outages in networks. These kind of architecture failures are common, though their occurrence is usually not singular. In many of modern provider networks, failures are sizable in numbers and one can map these across different network strata. Such kind of multi-layer protection models are also widely investigated [4,5].

Apart from node and link failure is the issue of mitigating natural and man-made disasters that can take down portions of the network. Unlike network and link failures (which are significantly localized), *disc failures* (implying that an entire region is down) are more difficult to handle [6]. In the case of node and link failures, the failures are often independent of each other implying that two nodes may fail without necessarily having the same set of reasons for the failure. However, in the case of natural or man-made disasters (such as hurricanes, terror-attacks etc.) an entire region is likely to be impacted, which may involve several links and nodes being rendered non-functional. Restoration of a network after an entire region is down is termed as disc protection [7] on account of the approximate circular region (a city or a metropolitan region) that is likely to be impacted during a disaster.

In this paper, we assume a randomly located disc failure of size R. Clearly, we do not know where the failure may occur. Our assumption has the following rationale: in case of both natural and man-made disasters, we want to be able to certify a network to be able to cater to disasters of a certain magnitude. We do want to benchmark a network the worst-case failure such as a natural disaster of Category 5 hurricane or an earthquake of 7 on the Richter scale and man-made disasters of a thermonuclear weapon or a cyberattack on a regional grid [8]. To this end, we want to design a network such that a disc failure of radius R is taken care of by the network design itself – that is to say that post the failure of disc size R, the remaining nodes in the network would continue to be operational.

In section II, we summarize some of the related work pertinent to our disc protection problem. Section III describes the formulation of constrained optimization model that is instructive to the network design problem. Section IV describes an efficient heuristic for network design, which for small-sized networks is fast and gives promising results. Section V describes the impact of SDN on disc protection, while section VI showcases results from a simulations setup and section VII summarizes the paper.

II. Related Work

The area of protection and restoration has been considered by many researchers in the past and there is a rich body of literature available [9,10]. For optical networks, the first body of work revolved around the SONET/SDH concepts of 1+1 and 1:1 protection [9]. Subsequent to these were optical layer protection techniques using wavelength protection. The classical routing and wavelength assignment problem was extended to include protection in [11]. Multi-layer protection was considered in [12,13]. Edge-disjoint wavelength protection and its scope was considered in [14]. A key improvement in protection techniques was the formulation of the Shared Risk Link Group problem in [3]. Improvements of the SRLG problem were considered by many researchers such as [15].

The work that is closest to our work was described in [6]. In that approach, a graph-based approach towards modeling random cuts was developed. The value of that work lies in the computation of geometric probability to random cut lines

978-3-903176-07-2 © 2018 IFIP

leading to reliability computations. The same authors in [16] have extended their work to include max-flow min-cut based approach for disc failures. Our work is different from these efforts in the sense that we consider network design of a known set of nodes and plan on route optimizations without subjecting to traffic variances for a random failure of size R. We consider the network design problem from the practical perspective of ROADM design as well as IP routers and the impact on the number of transponders and line cards. Our solution is practical as it directly considers network equipment. We also consider a variation of the solution by including SDNs.

TABLE 1: PARAMETERS

$\boldsymbol{G(V,E)}$	Network graph of set of $\boldsymbol{V}$ nodes and set of $\boldsymbol{E}$ edges
$\boldsymbol{R}$	Radius of the disaster zone.
$\boldsymbol{Dz^r(v,e)}$	Disaster zone of radius R covering v nodes and e edges in $G(V,E)$
$\boldsymbol{G'(V',E')}$	Network graph after disaster having set of $V' = V \setminus v$ nodes and set of $E' = E \setminus e$ edges.
$\boldsymbol{G_{opt}(V,E)}$	Optimized network graph for protection from a disaster of radius R.
$\boldsymbol{E_{ij}, E'_{ij}}$	Edges between the node i and node j in the graph.$\forall i,j \in \{1,2,3 \dots N\}$
$\boldsymbol{T_{abkm}}$	m^{th} instance of the traffic request between source node a and destination node b using k^{th} path
$\boldsymbol{T_{abkm}(B)}$	Bandwidth required for traffic T_{abkm}
$\boldsymbol{PM^m_{abk}}$	Set of primary links (edges) on k^{th}path between source node V_a and destination node V_b for traffic T_{abkm}.
$\boldsymbol{PM^{m'}_{abk}}$	Set of protection links (edges) on k^{th} path between source node V_a and destination node V_b for traffic T_{abkm} .
$\boldsymbol{\lvert PM^m_{abk}\rvert}$, $\boldsymbol{\lvert PM^{m'}_{abk}\rvert}$	Number of links on the k^{th} path of traffic T_{abkm}
$\boldsymbol{PS_{ab}}$	Set of K paths sorted in increasing order of the path length
$\boldsymbol{l_{ij}}$	Link between node V_i and V_j
$\boldsymbol{Bw^k_{av}}$	Available bandwidth on the k^{th} path
$\boldsymbol{Bw_{ij}}$	Available bandwidth on link l_{ij}.
$\boldsymbol{C_{ij}}$	Total capacity of the edge E_{ij}
$\boldsymbol{R_{ij}}$	It is the number of additional links/ports require for provisioning the protection path over the edge E_{ij}
$\boldsymbol{\delta_{ij}}$	Delay over the link l_{ij}. (link delay + processing delay of node i)

III. OPTIMIZATION MODEL FOR NETWORK DESIGN

Our goal is to build a network that would protect against a randomly located disc failure of some size R occurred due to natural or man-made disasters. Our fundamental assumption is that we do not know which region or disc in the network is likely to fail. However, for sake of classifying a network in terms of ability to be resilient, we would certify a network design to be capable of restoring against a failure of some disc radius R and hence want to come up with a dimensioning model for equipment that can restore services post a disc failure. Hence, we develop a model to compute the additional resources required to restore services in a network of a known topology with a randomly located disc failure of size R. This model would be developed considering practical provider deployments in core and metro networks into consideration.

We assume an optical core that supports WDM technology with Reconfigurable Optical Add Drop Multiplexers (ROADMs) [17], subtending wavelengths into optical fibers. The goal of the optimization model is to reduce the additional resources required (to increase the ROADM pass-through and add/drop ports) that would enable restoration of services post a random disc failure. Since, additional resources directly impact the CapEx (Capital Expenditure) in planning a provider's network our work directly helps a network provider to plan and protect the network based on projected traffic requirements.

TABLE 2:SYSTEM CONSTANTS

$\boldsymbol{N}$	Number of edges $\lvert\boldsymbol{E}\rvert$
$\boldsymbol{l_{bw}}$	Maximum bandwidth of a wavelength
$\boldsymbol{\Delta}$	Maximum permissible delay
$\boldsymbol{W_n}$	Number of wavelengths at each link
$\boldsymbol{t}$	t^{th} Wavelength in set $\{1,2,3,\dots,W_n\}$
$\boldsymbol{t'}$	t^{th} Wavelength in set $\{1,2,3,\dots,2*W_n\}$

TABLE 3: DECISION VARIABLES

$\boldsymbol{Sw_i}$	Size of the electrical switch at node $\boldsymbol{V_i}$
$\boldsymbol{W_{ij}}$	Number of wavelength used at the link ij
$\boldsymbol{l^{ij}_{abkm}, l^{ij'}_{abkm}}$	Link ij for traffic request T_{abkm}
$\boldsymbol{P_{abkm}}$	$\begin{cases}1 \text{ if primary path for traffic } T_{abkm} \text{ exist in } G', \\ 0 \text{ otherwise.}\end{cases}$
$\boldsymbol{P'_{abkm}}$	$\begin{cases}1 \text{ if protection path for traffic } T_{abkm} \text{ exist in } G', \\ 0 \text{ otherwise.}\end{cases}$
$\boldsymbol{\lambda^t_{abkm}, \lambda^{t'}_{abkm}}$	$\begin{cases}1 \text{ if traffic } T_{abkm} \text{ is provisioned over the wavelength t,} \\ 0 \text{ otherwise.}\end{cases}$
$\boldsymbol{\lambda^{tij}_{abkm}, \lambda^{t'ij}_{abkm}}$	$\begin{cases}1 \text{ if } \lambda^{\mathrm{t}}_{\mathrm{abkm}} \text{ or } \lambda^{\mathrm{t}'}_{\mathrm{abkm}} \text{ is provisioned over link ij ,} \\ 0 \text{ otherwise.}\end{cases}$

Optimization model: In our model, we assume a graph $G(V,E)$, of a set of V vertices and E edges extracted from the network topology. We compute the auxiliary graph $G'(V',E')$ by removing the nodes and edges from $G(V,E)$ present in the disaster region of radius R. Since, a disc failure can occur at any geographic location, but to include the disc failure of an arbitrary location in our optimization model, we require the exact node and edge locations which makes the optimization model complex. Therefore, we relax this requirement by assuming the nodes in the network graph G as the center of the disc failures, which actually represents the worst-case of impact post a failure of radius R. The key notations of our constrained optimization model are shown in Tables 1-3. The result of the optimization leads to ROADM/switch size dimensioning required at each node in G, in order to circumvent a circular disc failure. Since the ROADM/switch size depends on the number of wavelengths/links, therefore we model to minimize the number of wavelengths in the network with protection as our objective.

Objective: Our objective is to minimize the average number of wavelengths thereby minimizing the average switch size of the network with protection.

$$\mathbf{min}\frac{1}{N}\sum_{\substack{\forall E_{ij}\\ \forall abkm}}\sum_{\forall t'}\lambda^{t'ij}_{abkm}$$

The above objective function is valid for both ROADMs as well as L2/L3 switches or routers. In the case of switches and routers, the number of ports needs to be minimized. Further, since we are assuming a core network, this means that the number of ports is linearly proportional to the number of wavelengths with the caveat that we do not consider wavelength continuity for a pure IP network or one that allows wavelength conversion through electrical means facilitated by a controller (described later).

Our optimization model is subject to the following constraints:

Path Provisioning constraint: Equation (1) ensures that the primary and protection path does not share any node and link, i.e. node and link disjoint.

$$PM_{abk}^{m} \cap PM_{abk}^{m'} = \emptyset, \forall a, b \in V \quad (1)$$

Capacity constraint: Equation (2) gives the total capacity of the edge E_{ij} by multiplying the number of wavelengths and the bandwidth of the wavelength. Equation (3) and (5) ensures that the total traffic provisioned over an edge does not exceeds the total capacity of the edge. Equation (4) and (6) ensures that a single wavelength is assigned to a traffic request over an edge.

$$C_{ij} = W_n . l_{bw}, \forall E_{ij} \quad (2)$$

$$\sum\nolimits_{\forall t,a,b,k,m} T_{abkm}(B) . \lambda_{abkm}^{tij} . l_{bw} \leq C_{ij}, \forall E_{ij}, \forall t \quad (3)$$

$$\sum\nolimits_{t=1}^{W_n} \lambda_{abkm}^{tij} \leq 1, \forall abkm, \forall E_{ij} \quad (4)$$

$$\sum\nolimits_{\forall t,a,b,k,m} T_{abkm}(B) . \lambda_{abkm}^{t'ij} . l_{bw} \leq C_{ij}, \forall E'_{ij}, \forall t \quad (5)$$

$$\sum\nolimits_{t=1}^{2*W_n} \lambda_{abkm}^{t'ij} \leq 1, \forall abkm, \forall E_{ij} \quad (6)$$

Protection constraint: The constraints in equations (7) and (8) identifies whether the primary or protection path exist for traffic T_{abkm} after the occurrence of disaster.

$$P_{abkm} = \begin{cases} 1, \text{if } \sum_{i,j \in V'} \sum_{t}^{Wn} \lambda_{abkm}^{tij} = |PM_{abk}^{m}| \\ 0, \text{otherwise.} \end{cases}, \forall T_{abkm} \quad (7)$$

$$P'_{abkm} = \begin{cases} 1, \text{if } \sum_{i,j \in V'} \sum_{t}^{Wn} \lambda_{abkm}^{t'ij} = |PM_{abk}^{m'}| \\ 0, \text{otherwise.} \end{cases}, \forall T_{abkm} \quad (8)$$

$$P_{abkm} + P'_{abkm} \geq 1, \quad \forall T_{abkm}, \ \forall a, b \in V' \quad (9)$$

Constraint in equation (9) ensures that for a disaster of radius R occurring anywhere in the network, at least one path is available for the traffic T_{abkm} between node V_a and node V_b.

Wavelength Continuity: Equation (10) and equation (13) ensures that a wavelength is assigned to the primary and protection path. Constraints in equation (11) and equation (14) ensures that only single wavelength is assigned for a traffic request T_{abkm}.

$$\sum\nolimits_{\forall abkm} T_{abkm}(B) . \lambda_{abkm}^{t} \leq l_{bw} . l_{ij} \ \ \forall l_{ij} \in PM_{abk}^{m}, \forall abkm, \forall t \quad (10)$$

$$\Sigma \lambda_{abkm}^{t} \leq 1, \forall abkm, \forall t \quad (11)$$

$$\lambda_{abkm}^{tij} \leq \lambda_{abkm}^{t}, \forall l_{ij} \in PM_{abk}^{m}, \forall abkm, \forall t \quad (12)$$

$$\sum\nolimits_{\forall abkm} T_{abkm}(B) . \lambda_{abkm}^{t'} \leq l_{bw} . l_{ij} \ \ \forall l_{ij} \in PM_{abk}^{m'}, \ \forall abkm, \forall t' \quad (13)$$

$$\Sigma \lambda_{abkm}^{t'} \leq 1, \quad \forall abkm, \forall t' \quad (14)$$

$$\lambda_{abkm}^{t'ij} \leq \lambda_{abkm}^{t'} \ \forall l_{ij} \in PM_{abk}^{m}, \forall abkm, \forall t' \quad (15)$$

Delay Constraint: Constraints in equation (16) and (17) ensures that the total delay over a primary and protection path is within the permissible limit.

$$\sum\nolimits_{\forall \mathrm{l_{ij}} \in PM_{abk}^{m}} \delta_{ij} \leq \Delta, \qquad \forall a, b, k, m \quad (16)$$

$$\sum\nolimits_{\forall \mathrm{l_{ij}} \in PM_{abk}^{m'}} \delta_{ij} \leq \Delta, \qquad \forall a, b, k, m \quad (17)$$

The constrained optimization problem can be mapped to the 2-dimensional bin-packing problem and is hence NP-complete. For large networks or for dynamic traffic requests, an efficient heuristic is needed.

IV.HEURISTIC ALGORITHM

We propose a heuristic algorithm to configure the optimal network and ROADM/switch size such that in case of any disc failure of radius R in the network, the protection path always exists for a traffic originating and subsiding from outside of the disc failure zone. The proposed heuristic takes the network graph $G(V, E)$, ROADM/switch size Sw_i, disc radius R and traffic requests T_{abkm} as input and returns the superimposed network graph covering the protection path for all the affected traffic in disc radius R anywhere in the network.

Algorithm to find the optimal network graph for protection of disc failure of radius R

```
Input: G(V, E), $\Delta, \delta$, Sw, R, $T_{abkm}$ $\forall a, b \in V$
Output: $G_{opt}(V, E)$ for $\forall V$
Compute primary path $PM_{abk}^{m}$ for $\forall T_{abkm}$
Provision $PM_{abk}^{m}$ for $\forall T_{abkm}$
    $Bw_{ij} = Bw_{ij} - T_{abkm}(B)$ where $E_{ij} \in PM_{abk}^{m}$
$\forall i, j = 1,2,3 \ldots N$

For $x$ in range ($V$):
  $Dz_x^r(V_d, E_d)$ is a subgraph of $G(V, E)$ having nodes and edges of V and E in
                                          a circle of radius R centered at node $x$.
  $G'(V', E') = G(V, E)/Dz_x^r(V_d, E_d)$

  For $\forall abkm$ in ($T_{abkm}$)
     If $PM_{abk}^{m} \not\subset G'(V', E')$
          $(G_{opt}(V, E), BW, Sw)$
               $= PROVISION\_BACKUP((G'(V', E'), T_{abkm}, Bw, Sw, \Delta, \delta))$
          $G'(V', E') = G'(V', E') \cup G_{opt}(V, E)$
  $G(V, E) = G(V, E) \cup G'(V', E')$
Return $G(V, E), Sw$
```

To find the optimized network graph, we first calculate the primary path in the graph $G(V, E)$ for all the traffic requests. Traffic requests are provisioned over the corresponding calculated primary paths, if the bandwidth available over the path to accommodate the requested traffic. The algorithm deducts the bandwidth provisioned from the links along the path and only residual bandwidth remains for further provisioning. Since, the location of the disc radius is not known, we calculate the graph $z_x^r(V_d, E_d)$, where V_d are the nodes and E_d are the edges of $G(V, E)$ in the disc failure of radius R centered at node

$x : \forall x \in V$. We create a new auxiliary graph $G'(V', E')$ by removing the common nodes and edges of $z_x^r(V_d, E_d)$ and $G(V, E)$. Now, we compute all the primary paths provisioned in the previous step. If a path does not exist in the new graph $G'(V', E')$ then we invoke *PROVISION_BACKUP* (explained later) by passing the auxiliary graph $G'(V', E')$, the traffic request, the bandwidth and the ROADM/switch size to the *PROVISION_BACKUP* module. The call procedure *PROVISION_BACKUP* calculates and provisions the protection path in $G'(V', E')$ and if needed also adds extra wavelengths by adding an edge in the graph and/or increasing the switch size. The module *PROVISION_BACKUP* returns the graph $G_{opt}(V, E)$ which has the additional wavelengths as the edge of the graph. Thereafter, we superimpose all the $G_{opt}(V, E)$ in $G(V, E)$ to get the final graph.

Algorithm to provision backup path

```
Input: $G'(V',E'), T_{abkm}, Bw, \Delta, \delta$
Output: $G_{opt}(V,E), BW$
Compute all path $PM_{abk}$for traffic $T_{abkm}$ in $G'$, $k = \{1,2,3,\dots,K\}$
$PS_{ab}$ is the set of K paths sorted in increasing order of the path length.
Path provisioned = 0
For t in range K:
    $BW_{av}^t = \min(\{BW_{ij}\})\ E_{ij} \in PS_{ab}(t)$
    Delay=sum($\delta_{ij}$) where $E_{ij} \in PS_{ab}(t)$
    If Delay < $\Delta$
        If $BW_{av}^t > T_{abkm}$
            Provision backup path using $PS_{ab}(t)$
            $Bw_{ij} = Bw_{ij} - T_{abkm}, where\ l_{ij} \in PS_{ab}(t)$
            Path provisioned = 1
            Break;
If Path provisioned = 0
    #We consider that delay is always in permissible limit for $PS_{ab}(0)$
    For all $l_{ij}$ in |$PS_{ab}(0)$|:
        If $Bw_{ij} < T_{ab}$
            $Bw_{ij} = Bw_{ij} + \left\lceil \frac{T_{abkm}}{l_{bw}} \right\rceil$
            $E'_{ij} = E'_{ij} + \left\lceil \frac{T_{abkm}}{l_{bw}} \right\rceil$
            $Sw_i = Sw_i + \left\lceil \frac{T_{abkm}}{l_{bw}} \right\rceil$
            $Sw_j = Sw_j + \left\lceil \frac{T_{abkm}}{l_{bw}} \right\rceil$
    Path provisioned = 1
Return $G'(V',E')$, $Bw$, Sw
```

We also use *PROVISION_BACKUP* to calculate and provision a path. The module computes all the possible paths for a given traffic request in the graph $G'(V', E')$. All the calculated paths are sorted and stored based on their path length. For provisioning the traffic request, all the paths are checked for available bandwidth in the sorted order. In case traffic requests cannot be provisioned on shortest paths, then the next shortest path is evaluated for bandwidth availability. We do so by examining the link of the next shortest path that can suffice for our required bandwidth. Once we find the next shortest feasible path, we provision the traffic by deducting the bandwidth from all the links along the path. In case the available bandwidth over all the paths is not sufficient to provision the traffic request, then extra edges are added in the graph and the ROADM/switch size at the node is increased. This updated graph is provided as the output of the heuristic along with the link bandwidths and ROADM/switch size.

We run multiple iterations of the optimization model and heuristic by considering a different node as center of the disc failure of radius R in each iteration for a particular traffic profile. After running the iterations for all the nodes, we superimpose the additional wavelength requirement of each iteration in the graph G. The result is a network G that is able to protect against a failure of disc size R occurring anywhere in the network.

V. SDN AND DISC PROTECTION

Software defined networking involves the separation of data and control plane with a centralized controller orchestrating the network and planning/provisioning services across a network. The role of SDN in the case of designing networks with disc failures is relegated to traffic routing and equipment optimization. In particular, an SDN controller can plan for optimizing traffic placement across the network post a disc failure. An SDN controller routes traffic based on its atomicity level – i.e. coarse or fine chunks of traffic depending on the controllers' orchestration level are routed along same or different paths thereby optimizing the network. This leads to lower-sized requirements of the equipment in the network – nodes which would be subject to maximum traffic impact post a failure now can potentially be relieved of some of the impact due to better load balancing (post failure). It can hence be said that to design multi-layer network that are robust against disc failures an SDN controller is significantly helpful, if not mandatory. The SDN controller runs the optimization algorithm and comes up with an inventory list for each node. The network is thereafter designed taking into considerations the impact of the SDN controller. In our scheme, when we assume an SDN controller, we drop equations (10-13) on wavelength continuity constraint, thereby facilitating packet-level granularity to be juxtaposed on the network. By doing so (dropping equation 10-13) we are now able to assume that the SDN controller can optimize the traffic routing with respect to disc failures by appropriate sizing (dimensioning) of optical and IP router/switch nodes.

VI. SIMULATION AND RESULTS

In this section, we evaluate the path disrupted by a given disaster of radius in a 30-node core network topology. An optimization model was developed in Gurobi 7.5 with Python support. A separate Python code for the heuristic was developed as a discrete event simulation model. It takes around 20 minutes to run the optimization model in Gurobi running over Linux at HP ProLiant DL380p Gen9, 32GB RAM, 2.9GHz Xeon base server for a given disaster radius and traffic in the network and for a single iteration.

A. Network Model for Evaluation

For the evaluation of our optimization model and the heuristic, we used the network as shown in Fig. 1, which has 30-nodes, 36-links and average nodal degree of 2.4. For a given average hop size, we randomly select source and destination nodes to generate the traffic request in the network. Since the number of wavelengths on a link/edge can give the information about the size of the ROADM/switch, we simulate the model to find the average number of wavelengths used per link. In case the network consists of electrical switches/routers, all the wavelengths can be considered identical and each wavelength adds towards the port-count of the switch. In case of an all-optical network, wavelengths add towards the port-count of the ROADM at each node. For generating the traffic requests in the network, we assume that all the links in the network carry the same number of wavelengths and all the links are bidirectional. We consider load computation in the network to be proportional to the number of edges, number of wavelengths and averaged over the hop count over all source-destination pairs. Further, we assume ROADMs to be of at the most 4-degree.

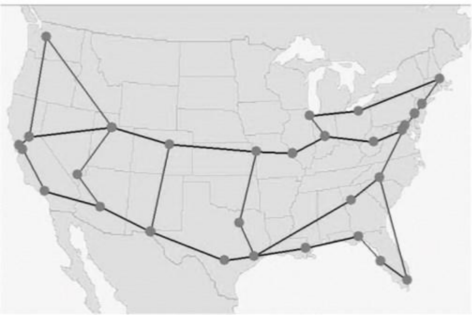


Fig. 1. 30-Node Network topology [18]

In each of our results we consider 20-100 wavelengths with wavelength capacity 10Gbps with services ranging from 10Mbps to 1Gbps selected randomly. Same result is valid for higher wavelength capacity i.e. 100Gbps, as traffic request will be in proportion of wavelength capacity.

It is important to note that the results shown here represent the ROADM/switch size per degree of the node.

B. *Optimization and Heuristic Results*

We analyze the impact of the load on the average ROADM/switch required to protect the network for a disc failure radius of 50 km (typical metropolitan region) for different wavelengths scenarios as shown in Fig. 2. Here, we consider the average hop-length=4 for the traffic generation. From the results, it can be observed that at low loads < 20%, there is no need to deploy additional wavelengths as the wavelengths present already have enough bandwidth to carry the extra traffic in case of disc failure. As we start increasing the load >20%, there is a linear increase in the ROADM/switch size. At high loads > 80%, the required ROADM/switch size starts getting saturated. This behavior is due to the fact that the extra wavelengths added at the load of 80% are not fully utilized and the residual capacity across all the wavelengths is sufficient to accommodate the traffic for load >80%. This result is beneficial for a provider to plan and deploy additional capacity in its network, based on the maximum active load in the network at any given point of time. It is key to note that the heuristic performs well – almost within 15% of the optimal – which is surprising and can only be attributed to the fact that the network size we consider is a small network i.e. 30 nodes. The divergence between the heuristic and optimal would be significant for a large network, say of size 500 nodes, which though is not a typical core network scenario.

We analyzed the impact of the disc radius on the ROADM/switch size requirement as shown in Fig. 3. It is interesting to note that with increase in the disc radius, required ROADM/switch size are either same or it decreases slightly across all the different wavelength scenarios. This behavior can be attributed to the fact that with increase in the disc radius, larger part of the network goes down. As a result, active load in the network gets reduced and already available wavelengths are sufficient to carry the extra added load due to disc failure. This result was found to be valid in both the north-east US as well as Florida peninsula, where nodes are somewhat closer to each other. The importance of this result is that it is useful to benchmark a network against a catastrophe.

We also analyze the impact of the average hop size on the required ROADM/switch size as shown in Fig. 4. It can be observed that for all the wavelength scenarios, there is a linear decrease in the required ROADM/switch size with increase in the average hop length. It can be seen from the result that for an average hop-length of 10, the required ROADM/switch size reduces by ~25% of the size required at average hop length of 2. This is due to the fact that total traffic in the network will be more for the smaller average hop length as compared to the a larger average hop length network. Since, same traffic consumes the bandwidth over multiple hops in larger hop-length network. As a result, a network with higher average hop-length will require less number of additional wavelengths. This result is beneficial for a provider to avoid huge investment incurred in deployment of additional wavelengths when

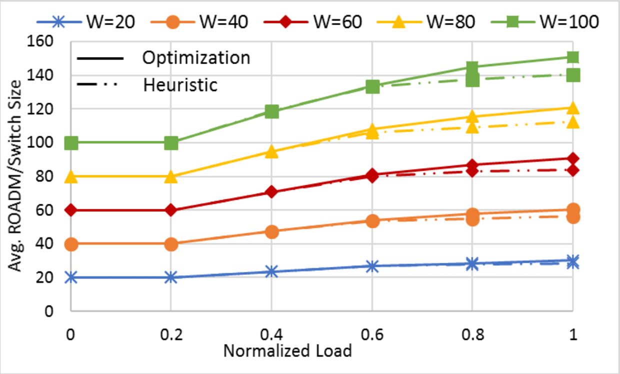


Fig. 2. Effect of load on the ROADM/switch size per nodal degree

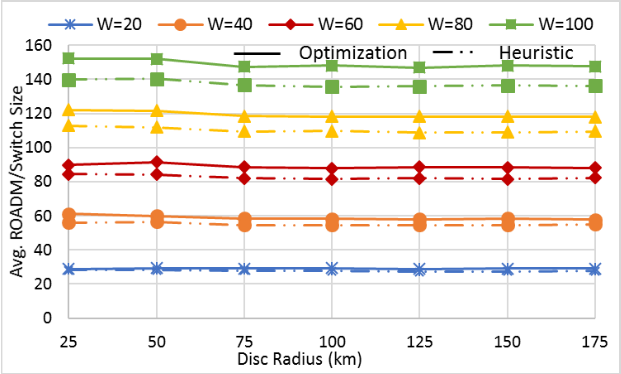


Fig. 3. Impact of disaster radius on the ROADM/Switch size

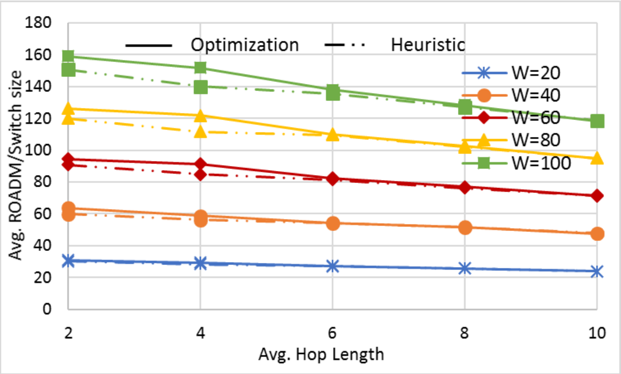


Fig. 4. Impact of the Average hop length on ROADM/Switch size

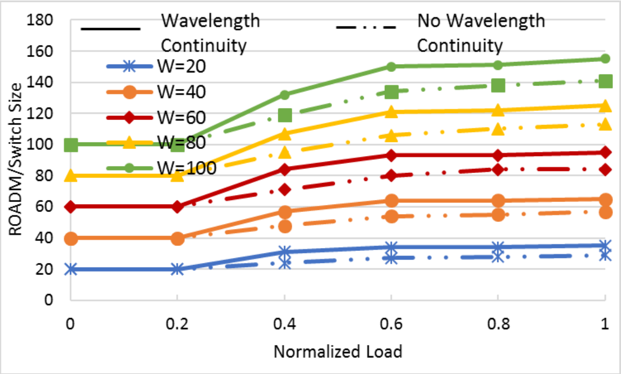


Fig. 5. Effect of wavelength continuity on the ROADM/Switch size

average traffic hop-length is larger in its network.

We also analyzed the impact of wavelength continuity on the ROADM size by adding the constraints in our heuristic model. Shown in Fig. 5 is the effect of the wavelength continuity on the ROADM/switch size. It is observed that addition of the wavelength continuity constraint increases the average ROADM size as compared to the average switch size required without employing the wavelength continuity. This increase in the size is because with no wavelength continuity there is flexibility of choosing a different wavelength at each link of the path. As a result, in case a single wavelength is not available for a traffic request, different wavelengths are selected for each link of the path. With wavelength continuity, a traffic request is provisioned by selecting a single wavelength over all the links in the path. There is possibility that bandwidth required to provision a traffic request is already available on all the links of the path but on different wavelengths, due to unavailability of a single wavelength to provision traffic request a new wavelength is added. As a result, the required ROADM/switch size is increased. This result is important for providers that want to deploy SDN in their networks. The role of an SDN controller would be that of a traffic shaper across the network. This is possible only when the wavelength continuity constraint can be relaxed and traffic at finer granularities can be offered to be provisioned across routes.

VII. Conclusion

In this paper, we have considered the important problem of network design post a failure of a disc of radius *R*. We have formulated this problem as a constrained optimization problem for both optical and IP networks. We have also considered the impact of an SDN controller on this problem. Further, an efficient heuristic is proposed that gives promising results (as compared to the optimal) for core networks. A simulation study validates our finding for different disc sizes, switch size evaluation and hop-length.

Acknowledgment

Aniruddha Kushwaha and Ashwin Gumaste are supported through DST, India. Arun Somani is supported through the Philip and Virginia Sproul Professorship at Iowa State University.

References

[1] D. Zhou and S. Subramaniam, "Survivability in optical networks," *IEEE Network*, vol. 14, no. 6, pp. 16–23, Nov.-Dec. 2000.

[2] J. Borland, "Analyzing the Internet collapse," *MIT Technology Review*, Feb. 2008. [Online]. Available: http://www.technologyreview.com/Infotech/20152/?a=f

[3] J. Doucette, WD Grover, "Capacity design studies of span-restorable mesh transport networks with shared-risk link group (SRLG) effects," SPIE Opticomm 2002, Boston Aug. 2002.

[4] R. Ramaswami and K. Sivarajan, "Optical Networks: A Practical Perspective," Elsevier, 3rd Edn. Morgan Kaufmann, 2010.

[5] L. Velasco, S. Spadaro, J. Comellas and G. Junyent, "ROADM design for OMS-DPRing protection in GMPLS-based optical networks," 6th International Workshop on Design and Reliable Communication Networks-2007, La Rochelle, France, Oct 2007

[6] S. Neumayer and E. Modiano, "Network Reliability With Geographically Correlated Failures," Proc. of Infocom 2010. San Diego, March 2010.

[7] A. Narula-Tam, E. Modiano, and A. Brzezinski, "Physical topology design for survivable routing of logical rings in WDM-based networks," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 8, pp. 1525–1538, Oct. 2004

[8] A Bernstein, D. Bienstock, D. Hay, M. Uzunoglu and G. Zussman, "Power grid vulnerability to geographically correlated failures — Analysis and control implications," In proc. of Infocom 2014, Toronto April 2014.

[9] A. Gumaste, T. Antony, "DWDM network designs and engineering solutions," Cisco Press, McMillan publishers, Dec. 2002.

[10] R. Ramaswami, K. Sivarajan, "Optical Networks: A Practical Perspective," San Mateo, Morgan Kaufmann, 2001

[11] J. Harmatos, P. Laborczi, "Dynamic Routing and Wavelength Assignment in Survivable WDM Networks," in Photonic Network Communications, 2002, Vol 4, No 3-4, pp. 357-376

[12] Q. Zheng and G. Mohan, "Multi-layer protection in IP-over-WDM networks with and with no backup lightpath sharing," in Computer Networks, Vol. 50, No. 3, pp 301-316 April 2006

[13] C. V. Saradhi, M. Gurusamy and L.Zhou, "Differentiated QOS for survivable WDM optical networks," IEEE Communications Magazine, Vol 42, No 5, pp. S8-14, May 2004

[14] C. Taunk, S. Bidkar, C. V. Saradhi, A. Gumaste "Impairment aware RWA based on a K-shuffle edge-disjoint path solution (IA-KS-EDP)," Optical Fiber Communication Conference and Exposition (OFC/NFOEC)-2011, Los Angeles, 6-10 March 2011.

[15] P. Datta and A. K. Somani, "Graph Transformation Approaches for Diverse Routing in Shared Risk Resource Group (SRRG) Failures," in Elsevier Computer networks Journal, Vol. 52, Issue 12, August 2008, pp. 2381-2394.

[16] S. Neumayer, A. Efrot and E. Modiano, "Geographic max-flow and min-cut under a circular disk failure model," Computer Communications 2015. Vol 77 pp 117-127

[17] Online: CDC ROADM Applications and Cost Comparison, OFC 2012. https://www.ofcconference.org/getattachment/188d14da-88ba-4a63-91d6-1cc14b335d8b/CDC-ROADM-Applications-and-Cost-Comparison.aspx

[18] Online: http://www.monarchna.com/topology.html

Leveraging Optics for Network Function Virtualization in Hybrid Data Centers

Tamal Das*, Aniruddha Kushwaha*, Ashwin Gumaste* and Mohan Gurusamy#
* Indian Institute of Technology, Bombay, India
# National University of Singapore, Singapore
{tamaldas, aniruddha}@cse.iitb.ac.in, ashwing@ieee.org, gmohan@nus.edu.sg

***Abstract*—Network Function Virtualization (NFV) has emerged as a hot topic for both industry and academia. NFV offers a radically new way to design and operate networks, by abstracting *physical network functions* (PNFs) to *virtual NFs* (VNFs). This disruptive innovation opens up a wide area of research, as well as introduces new challenges and opportunities – particularly in provisioning VNF forwarding graphs (or *network service chains*), and the resulting VNF placement issue. While forwarding graphs are often provisioned in the packet domain for fine-grained control over the respective traffic, we argue that doing so leads to lower efficiency; instead, provisioning forwarding graphs using optical transport proves to be far more efficient in intra-*datacenter* (DC) scenarios. While optical service chaining for NFV has already been proposed, we emphasize the use of optical bus architectures for the same. We present an architecture conducive for intra-DC NFV orchestration that can easily be extended to inter-DC scenarios. We deploy switchless optical bus architectures in both the frontplane and backplane of the DC. Our design particularly relies on readily available optical components, and scales easily. We validate our model using extensive simulations. Our results suggest that use of optical transport to provision VNF forwarding graphs can result in significant performance enhancement over packet-based electrical switch provisioning, in terms of packet drops and latency.**

I. Introduction

Network Function Virtualization (NFV) along with Software Defined networking (SDN) is considered as the game changers for next generation carrier-networks. While SDN will make the forwarding plane programmable, reduce the cost by including whiteboxes and bring generic agility into the network, NFV will allow the use of virtualization technologies to make complex network functions that existed in hardware to be placed in software. NFV, in some sense, facilitates the commoditization of networks by breaking service chains into network functions that can further be implemented on standard COTS platforms – IT-grade servers. The impact of virtualization is immense – NFV reduces CapEx and OpEx and facilitates better service velocity in terms of provisioning, upgrading, enabling elasticity to service chains. This promise of extreme cost-effectiveness and ability to softwarize the network is what has made NFV a popular research direction, not just in the academic community but also with providers – as evidenced by the ETSI initiative [1]. The NFV framework is undergoing severe consideration across vendors, providers, developers and academia. From a service provider standpoint, the question remains as to which are the best parts of a network to inculcate NFV. Given that at its core, the smallest indivisible part of NFV is the *virtual network function* or VNF – that exists as a standalone software chunk that can be moved around on VMs – the best position for placing a VNF is then the service provider data-center (DC). The thought of placing VNFs in provider data-centers is not new – it was first explored by the CORD project [2], where VNFs are placed in mini-data-centers that are formed by replacing traditional Central Office architecture with a bunch of servers and corresponding switches. While putting VNFs in the CO is a good idea for minimizing equipment churn towards the edge of network, another stress point is at the core of the network, where there is sizable need for network functions as well as storage of data. This is the reason why providers have data-centers in the core of the network, from where they can launch service chains as well as store data. Such a situation warrants that NFV technologies coexist with conventional data-technologies and, moreover, such an arrangement be integrated with the rest of the provider network.

The data-center, hence, becomes a key position in the network where we want to store, process, transport, work-upon data chunks. An ideal backbone data-center must be able to support huge amounts of data and network functions. Scalability is, hence, a key virtue that a DC must possess. Significant amount of research is available on DC design from a pure scalability perspective. We, in this paper, consider DC design from the perspective of both scalability and NFV provisioning. We require a DC to be able to scale to a large number of nodes that support both data storage as well as VNF storage.

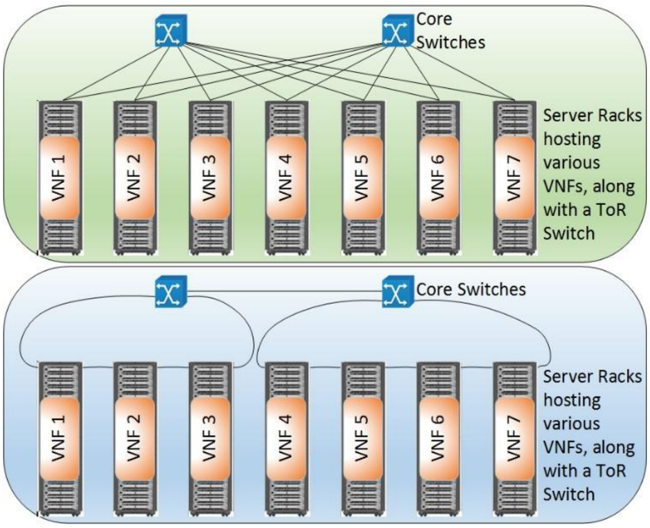

Core Switches
VNF 1
VNF 2
VNF 3
VNF 4
VNF 5
VNF 6
VNF 7
Server Racks hosting various VNFs, along with a ToR Switch
Core Switches
VNF 1
VNF 2
VNF 3
VNF 4
VNF 5
VNF 6
VNF 7
Server Racks hosting various VNFs, along with a ToR Switch

Figure 1: VNF forwarding graphs provisioned using conventional approach (top) and our approach (bottom).

978-3-903176-07-2 © 2018 IFIP

To this end, we proposed in [3] the DOSE architecture that facilitates the creation of a million node DC using optics in both its front plane and backplane. While HELIOS, OSA, WaveCube, FISSION, etc. [4-7] do use optics as an interconnection paradigm between top of the rack (TOR) switches, we go one step further – we use optics in the front plane as well; i.e. to connect servers to each other. Our concept assumes contemporary optics – that is easily available and mature. We do not rely on fast-moving optical devices, instead deploy an interesting architecture that is primarily based on the concept of broadcast and select of data across multiple fiber rings and wavelengths. The resulting DC is then ideally suited to house VNFs in terms of scalability, responsiveness, growth of services and churn in the network.

Our DOSE DC consists of sectors as a fundamental communication unit. A sector could have one-or-more TOR switches that are connected to servers. Sectors are interconnected using fiber rings. We could have as many fiber rings as one would want – subject to only port size of the cascade of sector-fiber switches and OSNR limitations. A sector is allocated a batch of wavelengths on which it perpetually can transmit into any one of the fiber rings. While a sector transmits into just one of the rings – it can receive data from all of the rings. The ring-to-sector interconnection is passive, implying a bus formation that is formed by the use of power-splitter, a combiner (coupler) as the interconnection element between the fiber ring and the sector. In addition to the use of optics in the backplane to connect sectors, we also use optics in the front plane to connect servers. Servers to communicate within a sector may use the TOR electronic switch or may use all-optical wavelength buses for communication. Fig. 1 illustrates the difference between provisioning a network service chain using the conventional spine-and-leaf architecture (top) vs. our proposed bus architecture (bottom). While the service chain needs to visit the core switch between any two VNFs, the same is not true for the bus architecture. The use of front plane optics is a way of saving on wiring as well as reducing latency for communication.

This paper is organized as follows. Section II discusses the related literature, while section III details our DC architecture. We present our simulation framework and results in section IV. Section V contains some concluding remarks.

II. Related Work

In this section, we present some of the related literature in the context of this paper, and highlight our contributions.

Several approaches have been reported in the literature, addressing the VNF placement and deployment problem in the context of NFV. An instrumentation and analytics framework is presented in [8], which shows that use of embedded instrumentation provide opportunities for providers to fine-tune their NFV deployments from both the technical and economic perspectives. A micro-service-based NFV orchestrator TeNOR is presented in [9] that focuses on: (i) automated deployment and configuration of services composed of virtualized functions, and, (ii) management and optimization of networking and IT resources for VNF hosting. Authors in [10] proposed forecast-assisted service chain deployment algorithm that includes the prediction of future VNF requirements. Possibility of minimizing the expensive optical/electronic/optical conversions for NFV chaining in packet/optical datacenters by using on-demand placement of vNFs is identified in [11]. A model for NFV placement is presented in [12] which considers the utilization of links and servers to minimize the maximum utilization over all links and switches. [13] proposed a hybrid architecture (optical/electrical) suited for NFV.

We now summarize some of the leading DC architectures. Several data center architectures have been proposed in recent years. The fat-tree [14] data center architecture proposed a hierarchy of three layers of electrical switches – core, aggregate and edge switches and is the most commonly deployed variant. In the DCell [15] architecture, a server is interconnected with other servers as well as a mini-switch. Servers communicate either through their connection to the mini-switch or through their connections to other servers. In c-Through [16], optical paths between top-of-the-rack (ToRs) switches are shared based on inter-rack traffic demands, while, ToRs are also interconnected with dedicated packet-switched paths. Helios [4] also uses a topology manager to measure traffic and estimate demand of the ToRs, based on which it computes the optimal topology for circuit-switched paths. Architectures like OSA [5], WaveCube [6] use reconfigurable optical devices to create optical circuits at runtime. Reconfiguration delay for these optical devices is a bottleneck. Delay in order of 10ms is too high for latency-intensive or smaller granularity flows.

In FISSION [7], optical backplane consists of number of fiber rings which are divided into sectors. Each sector can receive from all the fiber rings but can transmit only to a single ring. Each sector consists of servers inter-connected using switches in clos architecture. DOSE [3, 17] is an extension of the FISSION, where fiber rings are used to interconnect servers within the sectors, thus leading to both optical backplane as well as optical frontplanes.

Our work is inspired from the DOSE architecture [3], which we apply in this paper in the context of provisioning VNF forwarding graphs. While we largely restrict our discussion in this paper to intra-DC scenarios, it is important to note that our approach can easily be extended to inter-DC scenarios as well.

III. System Architecture

In this section, we detail our datacenter architecture (see Fig. 2) where we deploy the optics in both frontplane and backplane to dynamically provision VNF forwarding graphs of services.

The fiber ring-based DC optical backplane comprises one or more fiber rings. Fiber based backplane support optical buses in a ring configuration. Multiple number of sectors are connected in each fiber ring. Each sector in a fiber is allocated a fix set of wavelengths to transmit in a specific fiber ring. At the receiver of a sector, each ring drops a composite WDM signal constitutes of all the wavelengths of the fiber. This configuration allows a sector to transmit in a single fiber while allowing a sector to receive from all the fibers.

Each sector consists of wavelength selective switches (WSSs) to split the composite WDM optical signal from the backplane fibers to its constituent wavelengths. WSSs are preconfigured to drop only select wavelengths to each sector (being restricted by the EOS port count). These dropped wavelengths are processed by Electro-Optical Switch (EOS) for the service requests. Based on the request of service chain rule and load, EOS forwards the service to one of the k frontplane fiber rings, which are at the k ports of the EOS. Each frontplane fiber ring has m interconnection points, which are the interface points for the ToRs. Each frontplane deploys a unidirectional wavelength bus shared amongst all racks. An interconnection point consists of two couplers (one each for adding/dropping wavelength to/from frontplane) separated by an optical switch as shown in Fig. 2. Being an optical bus-based frontplane, when a ToR/EOS transmits in it, all the downstream ToRs/EOS receive the data. In this scenario all the unintended recipient discards the data based on an electronic match at its receiver. On arriving at a server, the packet is processed and forwarded via the frontplane to the next VNF based on the service's forwarding graph. A rack may host one or more VNFs. Based on VNF forwarding graph, if a service is for the EOS, it is forwarded to the frontplane, and after all of a service's VNFs in that sector are processed, it is thereafter sent to the backplane.

There are different wavelength assignment schemes for the ToR switches in the frontplane fiber ring, based on the number of wavelengths:

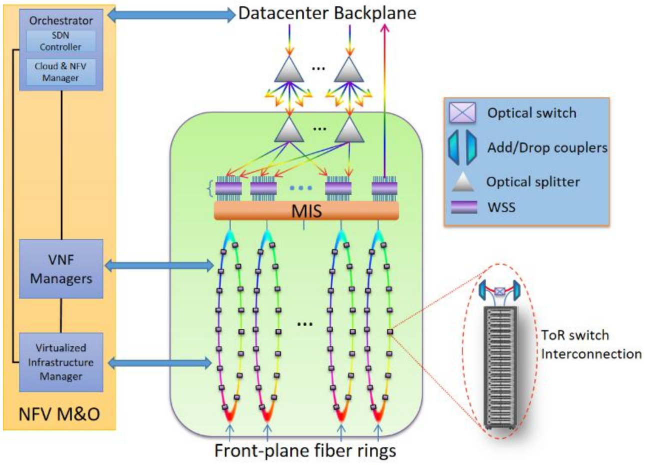


Figure 2: Proposed DC Frontplane architecture.

(a) Single Wavelength: In this wavelength assignment scheme, the frontplane fiber ring has only a single wavelength, where all ToRs are allocated with same wavelength to send and receive traffic. This single wavelength is time-shared between all ToRs and a token-based grant is used to avoid any simultaneous transmission of two or more ToRs/EOS. The arbitration is done using an out-of-band control channel ahead of time. Using a dedicated control channel for token allocation makes the architecture simple and helps in efficiently utilizing the data channel. This scheme also helps in reducing the load on the EOS as number of times a packet visit EOS will be less than the cardinality of its VNF forwarding graph.

(b) Multiple Wavelengths: This wavelength assignment scheme can be further divided into two sub-schemes:

Number of wavelengths = Number of racks: In this sub-scheme, there are a total of m wavelengths in a frontplane, and consequently each ToR is assigned a dedicated wavelength to send and receive traffic. In this case, if a VNF resides on some ToR, then the EOS will use a dedicated wavelength to send the packet to the respective ToR, and after processing the packet, the ToR will forward the packet to EOS for its next VNF processing. Since, each ToR has a dedicated wavelength, this scheme does not require any out-of-band control channel. But, the load on the EOS also increases as the service needs to visit the EOS after each VNF processing.

Number of wavelengths < Number of racks: In this sub-scheme, there are $< m$ wavelengths in a frontplane, which are time-shared to send/receive traffic. Because of time-sharing, each ToR is equipped with a tunable laser. Similar to the single wavelength scheme, an out-of-band control channel is used to arbitrate the pool of wavelengths between the ToRs and the EOS. This scheme also helps in reducing the load on the EOS as the ToR belonging to next service chain rule in the downstream can be directly reached without visiting the EOS.

Our architecture assumes an SDN-based central controller for service provisioning, which interfaces with the VNF manager for the instantiation and management of the VNFs on servers (see Fig. 2). The SDN controller populates the service chain rules and gather statistics to/from the EOS and ToRs. It shares the flow statistics and new service request information with the VNF manager. Based on the load and service request, VNF manager instantiates the VNFs on the server and shares this information with SDN controller for service provisioning and resource management.

IV. Optimization Model for Backplane Wavelength Assignement

In this section, we formulate an optimization model to deduce the backplane wavelength assignment. The goal is to connect most pair of sectors across the backplane using the minimum wavelengths. Our list of parameters and decision variables part are listed in Table 1 and Table 2, respectively.

Table 1: List of Parameters

Parameter	**Meaning**
$\boldsymbol{W_i}$	Wavelength of type i
$\boldsymbol{F_j}$	Fiber j
$\boldsymbol{S_k}$	Sector k
$\boldsymbol{\alpha}$	Wavelength Multiplicity in the backplane
$\boldsymbol{\beta}$	Contention factor at a sector's drop ports
$\boldsymbol{W}$	Number of wavelengths per backplane fiber
$\boldsymbol{n}$	Number of sectors per backplane fiber
$\boldsymbol{\gamma_p}$	Number of add ports at sector S_p

Table 2: List of Decision Variables

Variable	**Meaning**
$\boldsymbol{\lambda_{ij}^{pq}}$	Wavelength of type W_i in fiber F_j from sector S_p to sector S_q.

The objective of our optimization model is to ensure maximum connectivity between every pair of sectors across the backplane, i.e.,

$$\max \sum_{i,j,p,q(\neq p)} \lambda_{ij}^{pq}$$

subject to the following constraints.

Each wavelength can be used at most once across a sector's add ports. For instance, a sector cannot transmit on the same wavelength on different fibers in the backplane. Moreover, a wavelength added by an ingress sector in the backplane may be dropped at a maximum of α sectors. Thus, each of the λ_{ij}^{pq} originating from ingress sector S_p on wavelength W_i can connect to atmost α egress sectors, i.e.,

$$\forall i,p, \sum_{j,q(\neq p)} \lambda_{ij}^{pq} \leq \alpha$$

An ingress sector can only transmit on wavelengths from a single fiber in the backplane. This results from the physical constraint that a sector's ADD WSS can be connected to only a single fiber, and consequently a sector's add wavelengths cannot be added across multiple fibers in the backplane. Thus, for a given ingress sector, there exists a unique fiber in the backplane on which it transmits, i.e.,

$$\forall p, \exists! j: \sum_{i,q(\neq p)} \lambda_{ij}^{pq} > 0$$

Such a uniqueness constraint can be handled by LP solvers using a Special Ordered Set (SOS) of type One.

Each wavelength in the backplane can be used by atmost one sector. This eliminates the case of multiple sectors transmitting on the same wavelength in the same backplane fiber. Thus, there exists a unique sector which transmits on a particular wavelength in a backplane fiber, i.e.,

$$\forall i,j,\exists! p: \sum_{q(\neq p)} \lambda_{ij}^{pq} \neq 0$$

Such a uniqueness constraint can be handled by LP solvers using a Special Ordered Set (SOS) of type One.

The number of distinct wavelengths added from a sector is bounded by the number of add ports at the sector. Let us first define few auxiliary variables to formulate this constraint.

$$\forall i,p: d_1^{ip} = \sum_{j,q(\neq p)} \lambda_{ij}^{pq} \text{ and } d_2^{ip} = \begin{cases} 0, \text{if } d_1^{ip} = 0 \\ 1, \text{otherwise} \end{cases}$$

Here, d_1^{ip} denotes the cardinality of the set of egress sectors receiving from sector S_p on wavelength W_i via the backplane, whereas d_2^{ip} is a binary variable which determines whether wavelength W_i is used by sector S_p to transmit in the backplane. The stated constraint can then be formulated in terms of these auxiliary variables as:

$$\forall p: \sum_{i} d_2^{ip} \leq \gamma_p$$

Each backplane fiber has W wavelengths, each of which can potentially be dropped at α sectors. Thus, a fiber can drop at up to αW port across n sectors, i.e. $\frac{\alpha W}{n}$ ports per sector. In addition, an egress sector is configured to receive at most β wavelengths of the same type from the backplane, i.e. it can receive up to $\frac{\alpha W}{n\beta}$ distinct wavelengths. Thus,

$$\forall q,j, \sum_{i,q(\neq p)} \lambda_{ij}^{pq} \leq \frac{\alpha W}{n\beta}$$

Given the contention factor, at most β backplane fibers can drop the same wavelength at an egress sector, i.e.,

$$\forall q,i, \sum_{j,p(\neq q)} \lambda_{ij}^{pq} \leq \beta$$

Each wavelength in a backplane fiber is dropped at upto α sectors.

$$\forall i,j, \sum_{p,q(\neq p)} \lambda_{ij}^{pq} \leq \alpha$$

Each sector has at least one drop port and at least one add port connected to the backplane.

$$\forall s: \sum_{i,j,p(\neq s)} \lambda_{ij}^{ps} \geq 1 \text{ and } \sum_{i,j,q(\neq s)} \lambda_{ij}^{sq} \geq 1$$

Atmost one wavelength connects every pair of sectors in the backplane.

$$\forall p,q(\neq p): \sum_{i,j} \lambda_{ij}^{pq} \leq 1$$

To compute the backplane wavelength assignment for a million server DOSE datacenter, the above formulation takes ~20 seconds on an Intel Quadcore i7 CPU@3.5GHz with 16GB RAM.

V. Simulation and Results

In this section, we evaluate our proposed DC design using a Python-based discrete event simulation, and discuss the observed results. We simulate a DC with the well-known fat-tree architecture [13] and compare its performance (primarily in terms of metrics such as latency and packet drops) with that of our proposed DC architecture.

Simulation Model: For a given number of servers, we generate the corresponding fat-tree DC. We assume one of the PODs interface with the Datacenter Inteconnection Point (DCIP), and is thus the source/sink of all DC traffic. Although edge, aggregate and core switches in a fat-tree DC are considered the same, for the sake of comparison, we consider all server-edge switch links at 1Gbps, all edge switch-aggregate switch links at 10Gbps and all aggregate switch-core switch links at 100Gbps. As traffic enters the DC via the DCIP, it visits various servers in succession depending on its VNF forwarding graphs, and on completion, exits the DC through the DCIP. We assume each server to host a single VNF. In the rest of this section, we refer to this case as the "FatTree" scenario.

Similarly, we also generate a DC network with our proposed architecture, comprising of sectors, each of which hosts a bunch of frontplanes, which in turn consist of a bunch of racks, while the sectors are interconnected via backplanes. Here too, we assume one sector to interface with the Internet (via DCIP), and is thus the source/sink of all DC traffic. We assume each frontplane to host all VNFs, one per server rack. As traffic enters the DC via DCIP, it visits the least-loaded frontplane in the DC and on completion, exits the DC via the DCIP. An EOS has three port types, namely, (a) backplane ports (to receive traffic from the backplane), (b) add ports (to send traffic to the backplane, and, (c) frontplane ports (each hosting a unique frontplane). In the backplane, each wavelength drops traffic at two sectors (we term this a "*wavelength multiplicity*" of 2). If two sectors are not directly connected (i.e. via a single-hop) via a backplane wavelength, we consider multi-hopping routing to route traffic between them. The server-ToR switch links are assumed at 1Gbps, the frontplane rings are assumed at 10Gbps, and the backplane rings are assumed to be at 100Gbps. In the rest of this section, we refer to this case as the "DOSE" scenario.

In both scenarios, to service a particular VNF requirement of a network service chain, of the many servers hosting the

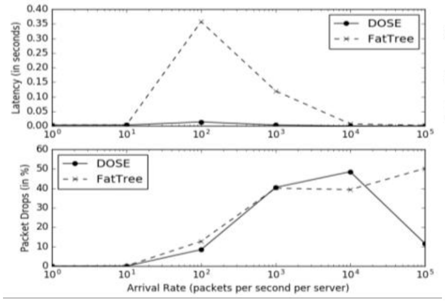


Fig. 3. Effect of Load on DOSE and FatTree DCs

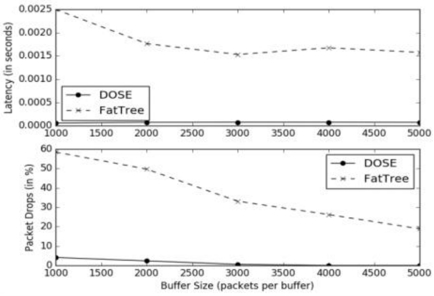


Fig. 4. Effect of buffer size on DOSE and FatTree DCs

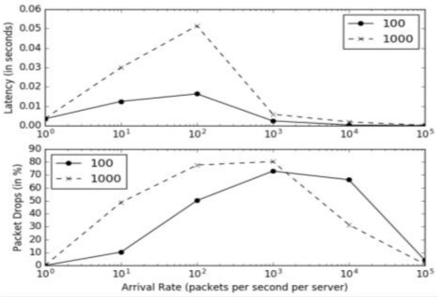


Fig. 5. Effect of network size on DOSE DC

required VNF, we choose the one with the least loaded path. Every switch/server port is assumed to have two buffers – one each for sending and receiving. All ports are assumed to be bidirectional. For a fair comparison between the two DC networks, we ensure the same traffic footprint in both scenarios. Each DC network is generated with 100 servers. Although for a given number of servers a unique topology is possible for the fat-tree architecture, the same is not true for our proposed architecture. In this evaluation, we consider an arbitrary topology for the DOSE DC, while deriving an optimal topology remains our future work. For a given number of VNFs, we generate the same number of services, each with a random service chain of varying lengths. All services are assumed to have the same priority level. Packet sizes are generated from an exponential distribution with a mean of 250 bytes, while packet arrivals are assumed to be Poisson distributed. A number of packet generators are placed on the Internet-facing side of the DCIPs to pump traffic composed of various services into the DC. A lookup delay of 300 nanoseconds and an average processing latency of 200 microseconds per packet is assumed. The port buffer capacities are considered proportional to their port rates, starting at 256kB for 1Gbps (~1000 packets per buffer) and so on. Leveraging the optical bus-based backplane in the DOSE DC, we consider each wavelength to be dropped to 2 sectors. We term this as "*wavelength multiplicity*". A wavelength multiplicity of 1 would imply a point-to-point connection (lightpath). To eliminate statistical errors, all results are averaged over 5 distinct traffic patterns.

Effect of Load: Fig. 3 contrasts the effect of load on the two DC architectures in terms of average end-to-end latency (in seconds, top) and average packet drops (in %, bottom). Both metrics are computed across all services. We vary the packet arrival rate per server from 1 to 100,000 packets per second. Note that the abscissa is plotted in log-scale.

The latency gradually increases from low to medium loads and drops thereafter. The decrease in latency from medium to high loads might seem counter-intuitive at first, but can be reconciled when observed in sync with the corresponding packet drops. At medium to high loads, packet drops significantly increase, and consequently lesser service chains are fully served. As a result, packets are either promptly dropped (resulting in higher packet drops), or promptly served (resulting in lower latencies). The benefit of DOSE over FatTree architecture is most pronounced at medium loads. Thus, in terms of latency, both packet and optical scenarios perform similarly at low and high loads, while benefit of optical backplane and frontplane is most pronounced at medium loads.

The average packet drops (or blocking probability) increases from low to high loads for both DC architectures, though the difference between the two is not much pronounced. In conclusion, while optics help bring down the latency, it does not improve the blocking probability as much.

Effect of Buffer Size: Fig. 4 plots the impact of buffer size on the two DC architectures. We vary the buffer size from a 1000 packets to 5000 packets in steps of 1000, and note the observed effect on the two performance metrics. These plots consider an arrival rate of 100,000 packets/second per server.

Both latency and packet drops decrease with rise in buffer size, the former only slightly while the latter considerably. This can be explained as follows. Larger the buffer, more packets can be stored, resulting in an increase in the observed end-to-end latency. An increase in buffer size essentially means more packets are accommodated, and in turn lesser packets dropped. The impact of buffer size is more pronounced for fat-tree architecture than for DOSE. This is attributed to the fact that the scope for betterment in latency/packet is rather low in case of a DOSE DC.

Effect of Network Size: Fig. 5 plots the effect of network size for a DOSE DC. The optical bus architecture employed in the DOSE DC significantly reduce the simulation run time, as compared to the fat-tree architecture; so much so that simulating a fat-tree network over servers becomes infeasible. Hence, the effect of network size could only be studied for the DOSE DC. We consider a 100 and 1000 node DOSE DC network, and vary the packet arrival rate per server from 1 to 100,000 packets per second, and plot the observed affect. Note that the abscissa is plotted in log-scale.

A larger topology leads to longer paths resulting in higher latencies. At low loads, larger topologies seem to lower the packet drops due to the larger cumulative buffer capacity across the network, although the buffer per server/switch remains the same. However, at medium to high loads, the packet drops tend to increase with larger topologies.

Effect of Service Chain Length: Fig. 6 plots the effect of service chain length on the two DC architectures. We vary the service chain length from 1 to 10, and note the observed effect on the two performance metrics. These plots consider an arrival rate of 1,000 packets/second per server. We generated a mix of 10 services each with a service chain length from 1 to 10.

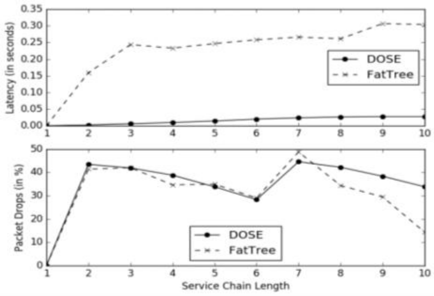


Fig. 6. Effect of service chain length on DOSE and FatTree DCs

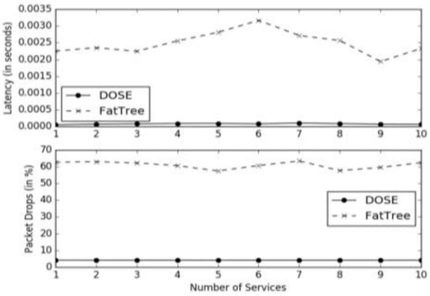


Fig. 7. Effect of number of services on DOSE and FatTree DCs

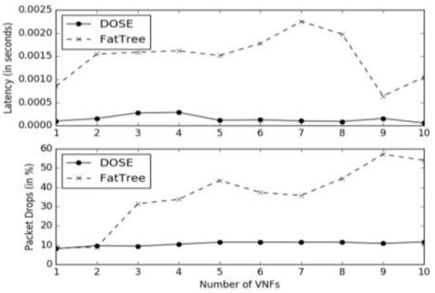


Fig. 8. Effect of number of VNFs on DOSE and FatTree DCs

With increasing service chain length, the latency increases in both the DC architectures, as longer service chains lead to longer service latencies. However, the latency for the DOSE DC is significantly and consistently less than that of a FatTree DC. With increasing service chain length, the packet drops largely increases, with no tangible improvement offered by a DOSE DC over a FatTree DC.

Effect of Number of Services: Fig. 7 plots the effect of varying number of services provisioned using the two DC architectures. We vary the number of services from 1 to 10, and plot the observed effect on the two-performance metrics. These plots consider an arrival rate of 100,000 packets/second per server.

The latency as well as the packet drops remain largely agnostic to the number of services in a DOSE DC, while a fat-tree DC seems to be slightly impacted. Thus, DOSE significantly outperforms fat-tree over a wide range of services.

Effect of Number of VNFs: Fig. 8 plots the effect of varying number of VNFs in the DC considering both architectures. We vary the number of VNFs from 1 to 10, and for a given number of VNFs, we generate as many services with varied service chain lengths. These plots consider an arrival rate of 100,000 packets/second per server.

Growing number of VNFs increases both the latency as well as the packet drops in case of a fat-tree DC, while a DOSE DC is hardly affected by the same. The performance of a DOSE DC is again better than that of a fat-tree DC across a varied range of VNFs.

VI. Conclusion

In this paper, we proposed a novel approach to provision network service chains for intra-DC scenarios. Our architecture heavily relies on optics, and deploys a switchless optical bus design both in the frontplane as well as in the backplane. Compared to the case with packet-based provisioning of network service chains, our architecture offers higher bandwidth due to use of optical fibers, as well as traffic-agnostic, as only the ports and not the links needs to be upgraded from time to time, unlike the packet-based scenario. We validated our model using extensive simulations, and compared our design with packet-based provisioning in terms of relevant metrics such as packet drops, latency as well as effect of design parameters such as buffer size, service chain length, topology size, etc. We observe that optics can play a significant role in improving the provisioning the VNF forwarding graphs for NFV.

References

[1] ETSI NFVISG Whitepaper, “Network Functions Virtualization (NFV) – Network Operator Perspectives on Industry Progress,” SDN and OpenFlow World Congress 2013, Frankfurt, Germany.

[2] CORD Project [Online]. Available: https://opencord.org

[3] Gumaste, A. Kushwaha and T. Das, “DOSE: Double optics single electronics data-center using a switchless optical frontplane and backplane,” IEEE International Conference on Communications (ICC), 2016, Malaysia.

[4] N. Farrington et al., "Helios: a hybrid electrical/optical switch architecture for modular data centers." ACM SIGCOMM Computer Communication Review, Vol. 40, No. 4, pp. 339-350, 2010.

[5] K. Chen et al., “OSA: an optical switching architecture for data center networks with unprecedented flexibility,” IEEE/ACM Transactions on Networking, Vol. 22, No. 2, pp. 498-511, 2014.

[6] K. Chen *et al.*, “WaveCube: A scalable, fault-tolerant, high performance optical datacenter architecture,” in *Proc. IEEE Conf. Comput. Commun.*, Hong Kong, 2015, pp. 1903–1911.

[7] A. Gumaste et al., “On the Unprecedented Scalability of the FISSION (Flexible Interconnection of Scalable Systems Integrated using Optical Networks) Datacenter,” To appear in IEEE Journal of Lightwave Technology, 2016.

[8] P. Veitch, M. J. McGrath and V. Bayon, "An instrumentation and analytics framework for optimal and robust NFV deployment," in IEEE Communications Magazine, vol. 53, no. 2, pp. 126-133, Feb. 2015.

[9] J. F. Riera et al., "TeNOR: Steps towards an orchestration platform for multi-PoP NFV deployment," 2016 IEEE NetSoft Conference and Workshops (NetSoft), pp. 243-250, Seoul, 2016,

[10] Sun, Quanying, et al. "Forecast-Assisted NFV Service Chain Deployment based on Affiliation-Aware vNF Placement."

[11] Xia, Ming, et al. "Network function placement for NFV chaining in packet/optical datacenters." Journal of Lightwave Technology, vol 33, no. 8, pp 1565-1570, 2015.

[12] A. Mohammadkhan et al., “Virtual function placement and traffic steering in flexible and dynamic software defined networks,” ser. LANMAN 2015. IEEE, pp. 1–6, 2015.

[13] Xia, Ming, et al. "Optical service chaining for network function virtualization." IEEE Communications Magazine 53.4 (2015): 152-158.

[14] M. Al-Fares, A. Loukissas and A. Vahdat, “A scalable, commodity data center network architecture,” ACM SIGCOMM Computer Communication Review, Vol. 38, No. 4, 2008.

[15] C. Guo et al., “Dcell: a scalable and fault-tolerant network structure for data centers,” ACM SIGCOMM Computer Communication Review, Vol. 38, No. 4, 2008.

[16] G. Wang et al., “c-Through: Part-time optics in data centers,” ACM SIGCOMM Computer Communication Review, Vol. 40, No. 4, 2010.

[17] A. Kushwaha, T. Das and A. Gumaste, “Does it Make Sense to put Optics in Both the Front and Backplane of a Large Data-Center?,” Optical Fiber Communication Conference (OFC), Los Angeles, California, March 2017.

Impact of High-Power Jamming Attacks on SDM Networks

Róża Goścień*, Carlos Natalino†, Lena Wosinska†, and Marija Furdek†
*Department of Systems and Computer Networks, Faculty of Electronics,
Wroclaw University of Science and Technology, Wroclaw, Poland. E-mail: roza.goscien@pwr.edu.pl
† Optical Networks Laboratory (ONLab), Royal Institute of Technology (KTH), Stockholm, Sweden.
E-mail: {carlosns, wosinska, marifur}@kth.se

Abstract—**Space Division Multiplexing (SDM) is a promising solution to provide ultra-high capacity optical network infrastructure for rapidly increasing traffic demands. Such network infrastructure can be a target of deliberate attacks that aim at disrupting a large number of vital services. This paper assesses the effects of high-power jamming attacks in SDM optical networks utilizing Multi-Core Fibers (MCFs), where the disruptive effect of the inserted jamming signals may spread among multiple cores due to increased Inter-Core CrossTalk (ICo-XT). We first assess the jamming-induced reduction of the signal reach for different bit rates and modulation formats. The obtained reach limitations are then used to derive the maximal traffic disruption at the network level. Results indicate that connections provisioned satisfying the normal operating conditions are highly vulnerable to these attacks, potentially leading to huge data losses at the network level.**

Index Terms—**High-power jamming attacks, optical network security, space division multiplexing.**

I. Introduction

Space Division Multiplexing (SDM) [1], [2] has been identified as a promising solution to the capacity crunch driven by the fast growth of bandwidth-intensive services. SDM enables ultra-high capacity in optical networks by utilizing a number of spatial resources, which can refer to multiple cores inside the same cladding of Multi-Core Fibers (MCFs); multiple modes inside the same core of Few-Mode Fibers (FMFs); or parallel single-mode fibers in the same bundle [3]. In weakly-coupled MCFs, which are in the focus of this work, each core within the fiber is used as a distinct communication channel, assuming sufficiently low interference between neighboring cores [4]. Key parameters determining the maximum transmission reach of optical signals in MCF are Amplified Spontaneous Emission (ASE) noise and Inter-Core CrossTalk (ICo-XT) [5].

As the critical infrastructure enabling a plethora of vital societal services, optical networks can be an enticing target of deliberate attacks aimed at service disruption [6]. High-power jamming attacks, in which an attacking signal is inserted into the network via, e.g., direct access to the fiber plant, monitoring ports, or by bending the fiber, can be harmful to optical networks deploying different technologies. In networks based on Wavelength Division Multiplexing (WDM), this attack affects co-propagating user signals by increasing the Inter-Channel CrossTalk (ICh-XT) among channels traversing the same fiber (core) [6]. In SDM-based networks, the damaging potential of jamming signals can not only affect signals inside the same core, but it can also propagate to signals in adjacent cores via increased ICo-XT. The primary requirement for increasing the network robustness to attacks is to evaluate the harmful effects caused by attacks and to quantify the damage they can cause to the network. While the damage from jamming attacks and the ways of increasing the level of physical-layer security in optical networks have been investigated in the context of Single-Mode Fibers (SMFs) [7]–[9], the harmful effects of jamming attacks in MCF-based SDM networks have not been studied so far.

To provide an assessment of the vulnerability of SDM networks to high-power jamming attacks, we evaluate the disruptive effects of jamming attacks to legitimate co-propagating signals in MCF. We first identify the maximum signal reach limited by ASE noise and ICo-XT under normal operating conditions. We then calculate the reduction of the maximum reach due to increased ICo-XT as a function of the power of the jamming signal, as well as the modulation format and bit rates of the legitimate signals. Using the developed model and ICo-XT-imposed reach limitations, we evaluate the overall traffic losses due to the physical-layer disruptions imposed by jamming attacks in the European backbone network, thus bounding the maximum extent of damage caused in the considered network. Results show that individual connections are highly vulnerable to the high-power jamming attacks, especially the ones with more complex modulation formats or longer reaches. At the network level, the attacks can disrupt a significant number of connections, causing the loss of huge amounts of data.

The remainder of this paper is organized as follows. Related works on physical-layer security aspects in optical networks are reviewed in Sec. II. Sec. III presents an assessment of the reach limitations of optical channels in an MCF imposed by ASE noise and by attacks causing excessive ICo-XT. Sec. IV expands the analysis to a network-wide scenario and evaluates the maximum possible traffic disruption. Finally, Sec. V concludes the work and presents guidelines for further investigation.

II. Related Work

As a promising solution to overcome the upcoming capacity crunch, SDM networks have been the subject of several studies

978-3-903176-07-2 © 2018 IFIP

focusing on a range of aspects from fiber manufacturing to the efficient spectrum management. Due to significant architectural differences, several management strategies need to be revisited, such as resource allocation algorithms, e.g., Routing and Wavelength Assignment (RWA) and Routing and Spectrum Assignment (RSA) algorithms, used in WDM and Elastic Optical Networks (EONs), respectively, need to be revisited to be suitable for SDM networks.

The work in [7] investigates the intra- and inter-channel CrossTalk (XT) effects caused by the injection of high-power jamming signals in WDM all-optical networks and shows their harmful effect to the performance of the optical channels. In [9], the authors propose approaches to decrease the overall damage caused by attacks through tailored, attack-aware routing and/or wavelength assignment. The work in [8] proposes a design strategy that enhances the conventional Dedicated Path Protection (DPP) with attack-awareness. The above-mentioned studies show that physical-layer security can be enhanced while using the same amount of optical resources as conventional, resource-saving approaches. However, these works consider a WDM optical network where the damaging effects of jamming signals stay confined in a single fiber core. In SDM networks, signal interference among adjacent cores cannot be neglected, particularly in the presence of high-power jamming signals.

The ICo-XT seems to be the main SDM drawback and limitation, which can affect the maximum transmission distance depending on the applied modulation format and bit rate. Therefore, the ICo-XT assessment is a crucial issue, and different ICo-XT models have been proposed in the literature. For instance, the authors of [10] and [11] apply very precise models, which allow estimating ICo-XT level for a particular core and transmission distance (from a source node) as a function of fiber physical characteristics, current transmission distance and number of adjacent cores.

The models can be simplified assuming the worst-case ICo-XT scenario (i.e., the core with the highest number of adjacent cores), as well as applied to find transmission reaches of different modulation formats in the presence of ICo-XT. By these means, the authors of [5] assess the modulation transmission reach as a function of the ICo-XT, the modulation format and its XT tolerance for different MCFs. Then, the work in [5] proposes a design strategy that considers SDM networks by considering the transmission reach limitations in MCFs. The work in [12] considers SDM networks and proposes an attack-aware Routing, Spectrum and Core Assignment (RSCA) strategy for design and provisioning. The strategy avoids assigning the same spectrum slot to potentially harmful signals and trusted signals if they traverse adjacent cores. This approach reduces the risks from ICo-XT impairment and related vulnerability of trusted channels. These works focus on the design and connection provision in SDM networks, but do not investigate the potential disruption caused by the ICo-XT in the presence of a malicious high-power jamming signal attack to connections provisioned considering normal operating conditions.

TABLE I
OSNR AND ICO-XT SIGNAL REQUIREMENTS [5].

	BPSK	QPSK	16-QAM	64-QAM
$OSNR_{min}$ [dB]	4.2	7.2	13.9	19.8
$XT_{dB,max}$ [dB]	-14	-17	-23	-29
P_S = 1 mW	L_{span} = 100 km	G = 20 dB	NF = 5.5 dB	

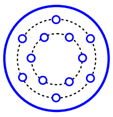


Fig. 1. 12-core double-ring MCF [14].

Different from the previous works in the literature, this work investigates traffic realized considering normal operating conditions and the most spectrally efficient modulation format is affected by the maximum reach limitations imposed by ICo-XT. We first provide an analysis of the maximum transmission reach of signals limited by ASE noise and ICo-XT under normal conditions and in the presence of a high-power jamming signal. Then, we evaluate how the reduction of signal reach disrupts traffic at a network level.

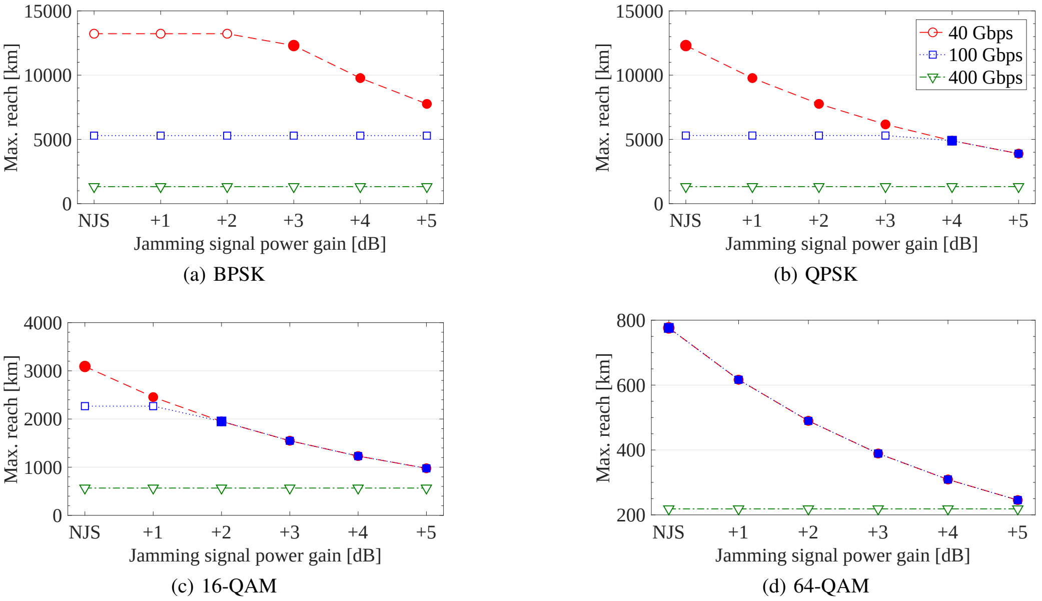
III. THE IMPACT OF JAMMING ATTACKS ON TRANSMISSION REACH IN MCF

In this section, we provide a methodology to calculate the transmission reach limitations of optical signals traversing MCF and quantify the reduction in the reach caused by jamming signal-induced ICo-XT.

The maximum transmission reach of an optically amplified signal is limited by several impairments which guide the selection of the bit rate and modulation format for each network connection. The two dominant limiting factors in MCF networks are ASE noise and ICo-XT [5], considering that the network has Digital Signal Processing (DSP)-enabled receivers, which are capable of compensating chromatic and polarization-mode dispersion, nonlinear channel backpropagation to compensate intra-channel nonlinearities, and balanced channel power.

Optical Signal-to-Noise Ratio (OSNR) requirements, which largely depend on the ASE noise, tighten with the increasing complexity of modulation formats, where more complex and spectrally efficient modulation formats require higher OSNR to achieve acceptable Bit Error Rate (BER) values. The transmission reach limitation due to noise is also inversely proportional to the signal bit rate, i.e., signals with higher bit rates have a shorter reach. The reach limitation due to ASE is calculated using (1), where P_S is the average optical power per channel, L_{span} is the distance between the equally spaced line amplifiers, $OSNR_{min}$ is the required OSNR at the receiver side (summarized in Table I), h is Planck's constant, f is the optical signal frequency, G and NF are the amplifier gain and noise factor, and R_S is the symbol rate [5], [13].

$$L_{max,OSNR} = \frac{P_S \cdot L_{span}}{OSNR_{min} \cdot h \cdot f \cdot G \cdot NF \cdot R_S} \tag{1}$$



(a) BPSK

(b) QPSK

(c) 16-QAM

(d) 64-QAM

Fig. 2. Maximum transmission reach limited by OSNR (white-faced markers) or XT (color-faced markers) for different bit rates. No Jamming Signal (NJS) represents the case where there is no jamming signal present in the fiber.

The reach limitation due to ICo-XT is a function of the modulation format only, where more complex modulation formats are more sensitive to ICo-XT, independent of the bit rate. This limitation is calculated using (2), where $XT_{dB,max}$ refers to the XT limit of the modulation format (described in Table I) and $XT_{dB,1km}$ refers to the fiber unitary ICo-XT (accumulated by transmission over 1 km) [5], [14].

$$L_{max,XT} = 10^{\frac{XT_{dB,max} - XT_{dB,1km}}{10}} \tag{2}$$

This analysis considers the balanced power scenario and the corresponding model from [5] as a baseline and investigates the ICo-XT effects of a harmful jamming signal with different power levels present in the fiber. Effects of the jamming signal on OSNR limits are considered negligible.

Table I describes the set of parameters, assumed as in [5], where user signals are transmitted at 1550 nm over a 12-core double-ring structure MCF with one propagation direction (see Fig. 1), yielding worst aggregate ICo-XT ($XT_{dB,1km}$) of -61.9 dB [5], [14]. A 4 dB penalty margin is also assumed for both OSNR and XT limits [5]. The considered transponder types support bit rates of 40 Gbps, 100 Gbps and 400 Gbps, as well as BPSK, QPSK, 16-QAM and 64-QAM modulation formats. The maximum transmission reach is calculated for the attack-free setup and for the worst-case attack scenario where the harmful jamming signal is inserted in one of the fiber cores in the inner ring, potentially affecting the signals in four adjacent cores via increased ICo-XT. The power gain of the jamming signal is varied from 1 to 5 dB to mimic attacks with different intensities.

Fig. 2 shows the maximum transmission reach for the different bit rates and modulation formats in the 12-core double-ring MCF showed in Fig. 1. In each scenario, the transmission reach of user signals is determined by the most limiting factor between OSNR and ICo-XT, denoted with white-faced and color-faced markers, respectively.

It is interesting to note that 400 Gbps signals are not affected by the considered attacks regardless of the used modulation format or the power of the jamming signal. This is because OSNR severely limits the reach of 400 Gbps signals already in normal operating conditions, and the attack-induced ICo-XT levels are not sufficient to exceed this limitation. However, as the modulation complexity increases, the ICo-XT limitation for 400 Gbps signals tightens and approaches the OSNR limitation. The trends for 400 Gbps signals across Figs. 2a-2d indicate that the power gain of the jamming signal should be above 5 dB to violate the OSNR threshold and impose reach limitations on these signals.

The less restrictive OSNR constraints allow for a longer reach of 40 and 100 Gbps channels, making these channels more likely to be limited by ICo-XT. The reach of 40 Gbps signals using QPSK, 16- or 64-QAM (Figs. 2b, 2c and 2d) is limited by ICo-XT even in the attack-free scenario (note the color-faced markers for the NJS case). For instance, as the power gain of the jamming signal increases, the maximum reach of 40 Gbps 64-QAM signals (Fig. 2d) decreases significantly, dropping by 20% already for 1 dB jamming signal power gain, and by 68% for 5 dB gain. Similar decrease (68%) is also experienced by 40 Gbps signals using QPSK and 16-QAM for jamming signal with 5 dB gain. For 40 Gbps QPSK signals, the drop is of 41% for jamming signal with 5 dB gain.

The reach of 100 Gbps signals in the attack-free scenario is limited by OSNR for all modulation formats but 64-QAM (Fig. 2d), where it is shaped by ICo-XT. Compared to normal operating conditions, jamming signal with 5 dB power gain reduces the reach of 100 Gbps signals by 26% (QPSK, Fig. 2b) to 68% (64-QAM, Fig. 2d). The transmission reach reduction caused by a malicious signal shown in Fig. 2 indicates the level of disruption of individual connections which are established to satisfy the normal operating conditions, and gives an insight into the safety margins that should be considered to take this reduction into account.

IV. Network-Wide Traffic Disruption Caused by Jamming Attacks

After determining the impact of high-power jamming to the maximum transmission reach of individual connections, using the model and the assumptions from Sec. III, we now investigate the worst-case damage from a jamming attack at the network level. First, we describe the scenario and assumptions considered in this work. Then, we assess the disruption caused by the attack scenarios considered.

A. Network Scenario and Assumptions

We perform numerical experiments on the Euro28 network topology with 28 nodes and 82 links with an average length of 625 km, shown in Fig. 3. All physical links are assumed to be realized with 12-core double-ring MCFs (see Fig. 1) supporting elastic spectrum allocation with 12.5 GHz granularity and independent switching policy as in [3]. A 12.5 GHz guard-band is used between neighboring signals. Each demand can be supported by one transponder capable of serving the requested bit rate, i.e., traffic splitting/grooming is not supported. The available transponder bit rates are the same as considered in Sec. III, i.e., 40, 100 and 400 Gbps.

Each traffic matrix consists of randomly generated demands with a total traffic volume of 800 Tbps. The source and destination nodes of connection demands are uniformly distributed among all node pairs and the requested bit rate follows uniform distribution in the range between 10 and 400 Gbps.

To assign routes and spectral resources to each demand, we apply the Spectrum-Spatial Allocation (SSA) algorithm from [15], aimed at minimizing the total network spectrum usage. The algorithm begins by sorting the demands in the descending order of their bit rates. For each demand, up to 30 candidate paths are computed, and associated with a modulation format and the number of required spectrum slices. The modulation format assignment follows the Distance-Adaptive Transmission (DAT) rule from [15] aimed at maximizing the spectral efficiency and minimizing the number of required regenerators. The number of slices required per candidate path is calculated as a function of demand bit rate and the applied modulation format, using the model from [5]. During the SSA, the transmission reach is calculated using the procedure described in Sec. III for the attack-free scenario. Regenerators are deployed at network nodes only for demands which cannot be established otherwise, and do not perform spectrum/modulation conversion. The SSA heuristic then selects the candidate path, the cores and the spectrum for each demand which result in the lowest total spectrum usage.

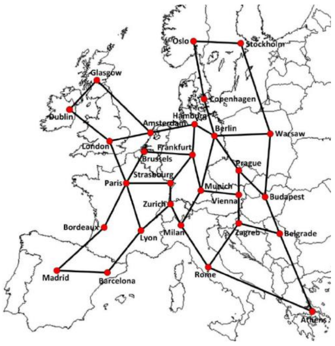


Fig. 3. Euro28 network topology with 12-core MCF.

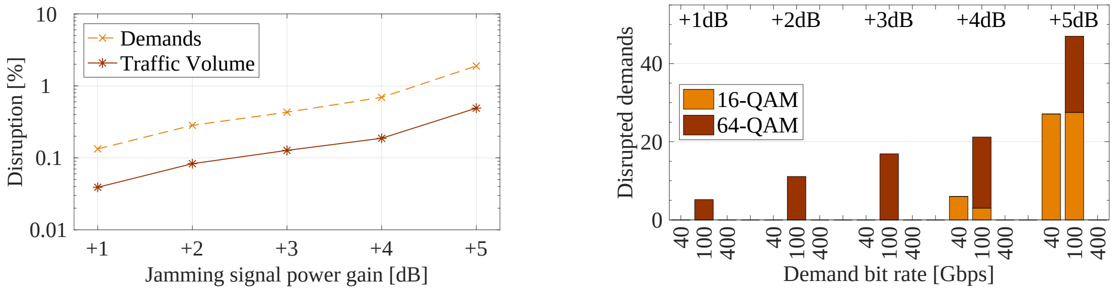
TABLE II
Traffic Matrices Bit Rates and the Modulation Formats Allocated to Satisfy the Demands.

Modulation	Bit Rate (%)			Total
	40	100	400	
BPSK	0.23	0	0	0.23
QPSK	3.55	6.15	69.39	79.1
16-QAM	4.39	8.67	6.93	19.99
64-QAM	0	0.51	0.16	0.68
Total	8.17	15.34	76.49	

Table II presents the distribution of the randomly generated traffic matrices in terms of bit rates. All presented results are averaged over ten different traffic matrices. For the considered traffic matrices, more than 75% of the demands are served by 400 Gbps bit rate channels, which are not affected by the analyzed attack scenarios, as shown in Sec. III. In such settings, less than 25% of the total traffic is vulnerable to an attack-induced reduction of transmission reach according to the results presented in Fig. 2. Table II also shows modulation formats selected by the SSA algorithm. For the considered SSA algorithm, 79.1% of the demands are realized using QPSK, followed by nearly 20% utilizing 16-QAM. BPSK and 64-QAM are applied to less than 1% of the demands.

B. Traffic Disruption Assessment

Considering the network scenario and assumptions presented in Sec. IV-A, we investigate the extent of disruption caused by jamming signals with different power gain values. Similar to Sec. III, the jamming signal is considered to have power gain of 1 to 5 dB relative to the legitimate signals. We consider a worst-case attack scenario where the jamming signal traverses all fiber links in the topology. While in reality the spreading of the jamming signal can be thwarted at intermediate nodes, this assumption allows us to assess an upper bound on the possible network disruption caused by this type of attacks. For each demand, we verify whether the demand is disrupted by the attack or not, considering the reach limitations presented in Sec. III. A demand is considered as disrupted if its path length exceeds the maximum reach constraint imposed by the jamming attack. The results for different attack scenarios are shown in Fig. 4. Fig. Fig. 4a shows the percentage of disrupted demands and traffic volume. Nearly 2% of all demands can be disrupted by a jamming signal with 5 dB power gain, carrying 0.5% of the total network traffic volume. Considering that the total network traffic is 800 Tbps, up to 4 Tbps can be disrupted, causing huge data losses. Moreover, if we consider only the demands



(a) Percentage of demands and traffic volume disrupted by the attack.

(b) Number of demands disrupted by the attack according to their bit rate and modulation format.

Fig. 4. Percentage, bit rate and modulation format of the demands disrupted by the attack.

vulnerable to the attack, i.e., excluding 400 Gbps signals, the percentage of disrupted demands can reach up to 8%.

Fig. 4b presents the number of disrupted demands according to their modulation format and bit rate. Only 16-QAM and 64-QAM demands are affected, which is in line with their vulnerability analysis shown in Figs. 2c and 2d. 100 Gbps signals are the most sensitive to jamming. On average, five 100 Gbps signals are affected already when considering attacks with 1 dB gain, while this number increases to 47 for 5 dB power gain. Jamming signals at 4 and 5 dB gain affect 40 Gbps demands as well, disrupting 6 and 27 demands, respectively.

V. Conclusions

This paper investigates the extent of disruption caused by high-power jamming attacks to legitimate traffic in a SDM network. We quantify the attack-induced reduction of maximum transmission reach for different bit rates and modulation formats, as well as the resulting traffic losses at the network level. The study provides an insight into the safety margins that could be considered to mitigate traffic losses and increase SDM network security. The results show that the correct modulation format is crucial not only for the spectrum efficiency, as shown in the related works, but is also of utmost importance for the resiliency of demands against high-power jamming signal attacks.

Further studies are needed to understand how different optical network technologies affect the vulnerability to physical layer attacks. In particular, the migration from WDM to SDM optical networks may require new approaches to guarantee the security of the optical layer. Moreover, the different extent of disruptions can be observed depending on the considered traffic matrices and network topology, as well as the applied SSA algorithm. Finally, in addition to jamming signal attacks, other kinds of physical layer attacks need to be studied in order to offer high security and minimize the network vulnerability.

Acknowledgment

This article is based upon work from COST Action CA15127 ("Resilient communication services protecting end-user applications from disaster-based failures – RECODIS") supported by COST (European Cooperation in Science and Technology) and the Celtic-Plus project SENDATE-EXTEND funded by VINNOVA. Róża Goścień was supported by the National Science Centre, Poland, under Grant 2015/19/B/ST7/02490 and by statutory funds of Department of Systems and Computer Networks, Wroclaw University of Science and Technology.

References

[1] T. Mizuno, H. Takara, K. Shibahara, A. Sano, and Y. Miyamoto, "Dense space division multiplexed transmission over multicore and multimode fiber for long-haul transport systems," *IEEE/OSA J. Lightwave Techn.*, vol. 34, no. 6, pp. 1484–1493, Feb 2016.

[2] W. Klaus, B. J. . Puttnam, R. S. Luis, J. Sakaguchi, J.-M. D. Mendinueta, Y.Awari, and N. Wada, "Advanced space division multiplexing technologies for optical networks [invited]," *IEEE/OSA J. Optical Commun. Netw.*, vol. 9, no. 4, pp. C1–C11, Apr 2017.

[3] M. Klinkowski, P. Lechowicz, and K. Walkowiak, "Survey of resource allocation schemes and algorithms in spectrally-spatially flexible optical networking," *Opt. Switch. Netw.*, vol. 27, pp. 58–78, Sep 2017.

[4] K. Saitoh, T. Fujisawa, and T. Sato, "Design and analysis of weakly- and strongly-coupled multicore fibers," *Proc. Photonic Netw. and Devices*, pp. NeTu2B.5.1 – 3, Jul 2017.

[5] J. Perelló, J. M. Gené, A. Pagès, J. A. Lazaro, and S. Spadaro, "Flex-grid/SDM backbone network design with inter-core XT-limited transmission reach," *IEEE/OSA J. Opt. Commun. and Netw.*, vol. 8, no. 8, pp. 540–552, Aug 2016.

[6] N. Skorin-Kapov, M. Furdek, S. Zsigmond, and L. Wosinska, "Physical-layer security in evolving optical networks," *IEEE Com. Mag.*, vol. 54, no. 8, pp. 110–117, August 2016.

[7] Y. Peng, Z. Sun, S. Du, and K. Long, "Propagation of all-optical crosstalk attack in transparent optical networks," *Opt. Eng.*, vol. 50, no. 8, pp. 085 002.1–3, August 2011.

[8] M. Furdek, N. Skorin-Kapov, and L. Wosinska, "Attack-aware dedicated path protection in optical networks," *IEEE/OSA J. Lightwave Techn.*, vol. 34, no. 4, pp. 1050–1061, February 2016.

[9] N. Skorin-Kapov, M. Furdek, R. A. Pardo, and P. P. Mariño, "Wavelength assignment for reducing in-band crosstalk attack propagation in optical networks: ILP formulations and heuristic algorithms," *European Journal of Operational Research*, vol. 222, no. 3, pp. 418 – 429, 2012.

[10] A. Muhammad, G. Zervas, and R. Forchheimer, "Resource allocation for space-division multiplexing: Optical white box versus optical black box networking," *Journal of Lightwave Technology*, vol. 33, no. 23, pp. 4928–4941, Dec 2015.

[11] L. Zhang, N. Ansari, and A. Khreishah, "Anycast planning in space division multiplexing elastic optical networks with multi-core fibers," *IEEE Communications Letters*, vol. 20, no. 10, pp. 1983–1986, Oct 2016.

[12] J. Zhu and Z. Zhu, "Physical-layer security in MCF-based SDM-EONs: Would crosstalk-aware service provisioning be good enough?" *IEEE/OSA J. Lightwave Techn.*, vol. 35, no. 22, pp. 4826–4837, Nov 2017.

[13] R. J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," *IEEE/OSA J. Lightwave Techn.*, vol. 28, no. 4, pp. 662–701, Feb 2010.

[14] A. Sano *et al.*, "409-tb/s + 409-tb/s crosstalk suppressed bidirectional mcf transmission over 450 km using propagation-direction interleaving," *Opt. Express*, vol. 21, no. 14, pp. 16 777–16 783, Jul 2013.

[15] R. Goścień, K. Walkowiak, and M. Klinkowski, "Distance-adaptive transmission in cloud-ready elastic optical networks," *IEEE/OSA J. Opt. Commun. and Netw.*, vol. 6, no. 10, pp. 816–828, 2014.

Modelling packet insertion on a WSADM ring

A. Gravey*‡, D. Amar *§, P. Gravey*§, and M. Morvan*§
*IMT Atlantique, Brest, France
email: firstname.lastname@imt-atlantique.fr
‡UMR CNRS 6074 IRISA, France, §UMR CNRS 6285 Lab-STICC, France

B. Uscumlic and D. Chiaroni
Nokia Bell Labs, Paris Saclay, France
firstname.lastname@nokia-bell-labs.com

***Abstract*—The WDM slotted Add/Drop Multiplexer (WSADM) technology relies on time slotted WDM rings, where a slot can carry a single WDM packet. All stations can insert and receive these WDM packets. This differs from previous architectures in which packets were carried over a single wavelength, while multiple packets could be carried in a single slot, thus taking advantage differently of the WDM dimension. The WSADM architecture is expected to reduce costs by exploiting low cost technologies. We propose mathematical models for evaluating the performance offered by WSADM optical packet rings, under two different packet insertion policies. In the slot reservation mode, a station can only use the slots that are periodically reserved for its exclusive usage. In the opportunistic insertion mode, a station can use any slot that is neither reserved, nor already occupied. These modes are bench-marked with a channel reservation mode in which each wavelength is dedicated to a single station.**

Keywords*—Wavelength Division Multiplexing, Optical Packet Switching, Metropolitan Area Network, Network Performance

I. Introduction

The distribution/aggregation network segment, also called Metropolitan Area Network (MAN), is expected to be particularly impacted by the current traffic growth. Different traffic sources, varying from small Digital Subscriber Line Access Multiplexers (DSLAMs) to large data centers, generate highly variable types of traffic, which favors using packet-based transport technologies in MANs. Ethernet rings with specific protection protocols are often considered. The main issues with “opaque” networks are, on the one hand their high energy consumption, and on the other hand the Ethernet packet granularity that is convenient for Metro Access areas, but too fine for Metro Core ones. Optical Packet/Burst Switching (OPS/OBS) technologies have been for many years considered as potential options for combining sub-wavelength granularity and optical transparency but the lack of viable optical buffering technologies has precluded implementing them. However, time-slotted OPS rings such as TWIN [1], POADM [2] and OPST [3] have been shown to provide both an efficient use of transmission resources and carrier-grade performance without optical packet buffering. Nevertheless, these technologies rely on custom optical components (in particular on fast-tunable burst-mode emitters) that are not currently commercially available.

In order to rely on more widely available components, the WDM slotted Add/Drop Multiplexer (WSADM) technology has been recently proposed [4]. The key optical devices required in a WSADM are integrated multi-wavelength laser sources that are fully in line with the trend of optoelectronic industry, and Semiconductor Optical Amplifiers (SOA) gates that are naturally suited to operate on WDM packets because of their wide optical bandwidth. Preliminary CAPEX comparisons have suggested that WSADM technology could compete favorably with existing electronic packet technologies and other OPS/OBS options [4], [5].

To the best of our knowledge, the network performance of WSADM technology has yet to be assessed. The present paper proposes a set of models for assessing WDM packet insertion performance in a WSADM ring. Packet insertion has a structuring impact on the global performance, as all inserted packets travel transparently till their destination, resulting in loss-less transfer and deterministic latency once packets are inserted. The introduction of WSADM raises several questions. For example, the comparison between a purely opportunistic insertion mode and a fully (or partially) deterministic one: how do these modes impact on network performance, in particular on latency and what is the impact of the number of wavelengths on their respective merits? More generally, the stringent requirements on latency, notably in the framework of future 5G deployments, make worth performing a detailed analysis of the packet insertion process in a candidate technology for future metro/aggregation networks.

Section II describes the network architecture considered in this work. Section III presents the various mathematical models developed for WSADM networks. A partial validation by simulation of the models is presented in the next section. The main performance assessments are summarized in section V and conclusions are drawn in section VI.

II. Network Architecture

We consider WDM packets as described in [4]. Multiple Service Data Units (SDU) are aggregated within a single Packet Data Unit (PDU); a typical SDU is e.g. an Ethernet Frame. To be transported over the optical ring, the PDU is split over K wavelengths, and not carried over a single wavelength as in TWIN, POADM and OPST.

The network is controlled by both a fast (i.e. real time) control realized in line, and a slower, although dynamical, control realized thanks to a SDN controller. The fast control is implemented through a control channel carried over a separate wavelength, and synchronized with the data channel: during a time slot, both a control packet and a data packet (which carries, or not, a PDU) are transmitted. The SDN controller provides a “provisioning oriented” type of control: it is in charge of station provisioning, of specifying the control information associated to PDUs before insertion and of specifying the operation (reception, pass-through, erasure) associated with PDUs carried over the ring. A similar “provisioning oriented” control has been described in a different context in [6]. As the present paper focuses on transfer plane performance, it shall not provide a detailed specification of the SDN control.

We assume that each station presents a single D Mbit/s interface (typically, in a metro network, $D = 10$ Gbit/s), which

978-3-903176-07-2 © 2018 IFIP

is equal to the rate of a single wavelength. Let Z be the size, in bytes, of a PDU. Z should be large enough to contain many Ethernet frames, in order to avoid segmentation/reassembly and to limit the proportion of resources wasted due to the fixed overhead necessary for each packet (guard-band, preamble and framing). On the other hand, Z should not be too large. Indeed, in order to limit the latency due to the network, the time taken to fill a PDU by SDUs shall most likely be limited by a timer, unless it is filled before the timer runs out. Were Z too large, either latency would be negatively impacted by an overly long timer value, or PDUs would be systematically sent partially filled, timers having run out before the PDU was full, thus wasting resources. For the sake of generality, define $T = Z/D$ to be the time it takes to fill a PDU at rate D. T is split into K slots, where K is the number of channels over which a packet is split to be transmitted; T/K is thus slot duration.

Stations are organized into a single unidirectional ring, in which each station can both insert and extract PDUs from the optical packet ring. Once a PDU is inserted, it cannot be lost till it is received by the final destination station, as it is passed transparently through the transit stations; therefore, PDUs are not lost within the network; a PDU could however be lost within a station, due to insertion buffer overflow. End-to-end PDU latency is the sum of the sojourn time in the insertion buffer and of the (fixed) propagation delay between source and destination stations (typically in the order of 0.1-1 ms). The performance offered to PDUs is thus mostly characterized by the performance of the PDU insertion process. The performance offered to SDUs also depends on how SDUs are aggregated in PDUs, and on whether timer-based policies are implemented, or not, in order to control latency. This is not considered in the present paper which focuses on the performance offered to PDUs in terms of latency and jitter.

III. Modelling Packet Insertion

Insertion performance is first driven by the PDU arrival process. As we consider a metro network, where each station aggregates the traffic of thousands of customers, it is justified to assume that PDUs arrive according to a Poisson process with parameter Λ. Let $\gamma_j(x)$ be the probability that j PDUs arrive during an interval of duration x:

$$\gamma_j(x) = e^{-\Lambda x}\frac{(\Lambda x)^j}{j!} \tag{1}$$

The number of arrivals during an interval of duration x is thus Poisson with parameter Λx. Insertion performance also depends on slot availability, characterized by the insertion mode applied to PDUs. We shall benchmark two slot insertion modes, "slot reservation" and "opportunistic insertion", with a classical channel reservation mode, in which each wavelength is dedicated to a station.

A. Slot Reservation Mode

In the slot reservation mode, the PDU can be inserted only on a slot that is marked as being available for its class. It is assumed that there is a reserved slot every R slot. Let a "reservation period" start at the beginning of a reserved slot, and end just before the next reserved slot. If at least one PDU is in the system at the beginning of the reservation period, there is an exit at the end of the reserved slot. Otherwise, no PDU is served during the period. We assume that system capacity is finite of size B. In the following, we shall derive the distribution for N_r, number of PDUs in system at the beginning of a reservation period, M_r, number of PDUs seen by an arriving PDU, P^r_{loss}, the probability that an arriving PDU finds B PDUs in the system and W_r, sojourn time of a PDU which enters the system.

We first derive the transitions probabilities for N_r. As system capacity is B, N_r varies between 0 and B, and the transitions are as follows:

$$\begin{aligned}
P^r(0,i) &= \gamma_i\left(\frac{RT}{K}\right) && (B-1) \geq i \geq 0\\
P^r(0,B) &= \sum_{j=B}^{\infty}\gamma_j\left(\frac{RT}{K}\right) &&\\
P^r(n,i) &= \gamma_{i+1-n}\left(\frac{RT}{K}\right) &&\\
&\qquad B \geq n > 0, && (B-1) \geq i \geq 0\\
P^r(n,B) &= \sum_{j=B-n+1}^{\infty}\gamma_j\left(\frac{RT}{K}\right) && B \geq n > 0
\end{aligned}$$

Let $\pi^r = \{\pi^r_i, 0 \leq i \leq (B-1)\}$ be the probability distribution for N_r; π^r is numerically derived by solving $\pi^r P^r = \pi^r$.

In order to derive the distribution for M_r, let us consider the probability that, knowing that a PDU arrives during $[0, RT/K[$, it arrives in the interval $[x, x+dx[$, and that exactly j other PDUs arrived before it, in the same reservation period. As the arrival process is Poisson, within a reservation period of length RT/K, the probability that the PDU arrives during an interval of length dx is Kdx/RT. The date of arrival x and the number of arrivals between the beginning of the period and x are related as follows:

$$\begin{aligned}
P\Big(\text{tagged arrival in } [x, x+dx[, j \text{ arrivals during } [0,x[\Big)&\\
= \frac{Kdx}{RT}\gamma_j(x)&
\end{aligned}$$

The previous joint probability is independent from the state of the system at the beginning of the period. M_r depends both on N_r (number of PDUs in the system at the beginning of a period), and on whether the tagged PDU arrives before the end of the reserved slot, or not. Indeed, if the tagged PDU arrives after the end of the reserved slot, and if $N_r > 0$, one PDU has been served before the arrival of the tagged PDU; on the other hand, if it arrives during $[0, T/K[$ it sees all the PDUs present in the system at time 0. Let ν^r_k be the probability for $\{M_r = k\}$. For k smaller than B,

$$\begin{aligned}
\nu^r_k = \frac{K}{RT}\Big[\pi^r_0 &\int_0^{RT/K}\gamma_k(y)dy\\
+\sum_{n=1}^{k}\pi^r_n &\int_0^{T/K}\gamma_{k-n}(y)dy\\
+\sum_{n=1}^{k+1}\pi^r_n &\int_{T/K}^{RT/K}\gamma_{k-n+1}(y)dy\Big]
\end{aligned}$$

which yields, after integrating (1):

$$\nu^r_k = \frac{K}{\Lambda RT}\Big[\pi^r_0\sum_{k+1}^{\infty}\gamma_j\Big(\frac{RT}{K}\Big) + \sum_{n=1}^{k+1}\pi^r_n\gamma_{k-n+1}\Big(\frac{T}{K}\Big)$$

$$-\pi^r_{k+1}+\sum_{n=1}^{k+1}\pi^r_n\sum_{j=k-n+2}^{\infty}\gamma_j\Big(\frac{RT}{K}\Big)\Big] \qquad (2)$$

The loss probability P^r_{loss} is ν^r_B, and can be derived similarly:

$$\begin{aligned}P^r_{loss}=\frac{K}{\Lambda RT}\Big[\pi^r_0\sum_{i=B+1}^{\infty}(i-B)\gamma_i\big(RT/K\big)\\ +\sum_{n=1}^{B}\pi^r_n\sum_{j=B-n+1}^{\infty}\gamma_j\big(T/K\big)\\ +\sum_{n=1}^{B}\pi^r_n\sum_{i=B-n+2}^{\infty}(i+n-B-1)\gamma_i\big(RT/K\big)\Big] \qquad (3)\end{aligned}$$

The number of PDUs seen by an arriving PDU which is not lost is distributed as $\nu^r_k/(1-\nu^r_B)$ ($k<B$). The mean sojourn time of such a PDU is then derived using Little's formula:

$$E(W_r)=\frac{1}{\Lambda(1-\nu^r_B)}\sum_{k=0}^{B-1}k\nu^r_k \qquad (4)$$

In order to derive the distribution for W_r, let U_r be the time between the arrival of the tagged PDU and the end of the reservation period. Let also $A_r(U_r)$ be the number of PDUs arriving before the tagged PDU in the same reservation period. The distribution for W_r depends on both N_r and U_r. If N_r is null, the tagged PDU is delayed only by $A_r(U_r)$ PDUs. On the other hand, if N_r is positive, it is also delayed by the (N_r-1) PDUs arrived in previous reservation periods (one PDU is served during the reservation period). W_r is thus equal to the sum of the time till the end of the reservation period during which it arrived (U_r and k reservation periods, where k is the number of PDUs which are in system when the PDU arrives, and which are not served during the reservation period during which the tagged PDU arrived, and its own service time T/K. By conditioning on N_r and on the number of other PDUs that arrived before a tagged PDU, in the same period (which are independent), we can directly obtain the distribution for W_r, valid for k smaller than $(B-1)$ and x in $\Big[0,\frac{RT}{K}\Big[$.

$$\begin{aligned}P\Big(W_r\in\Big[\tfrac{(kR+1)T}{K}+x,\tfrac{(kR+1)T}{K}+x+dx\Big[\Big)=\\ \frac{Kdx}{RT(1-\nu^r_B)}\Big(\pi^r_0\gamma_k(RT/K-x)\\ +\sum_{n=1}^{k+1}\pi^r_n\gamma_{k-n+1}(RT/K-x)\Big) \qquad (5)\end{aligned}$$

For $k=(B-1)$ we need to ensure that the tagged PDU is not lost, which could occur for x larger than $(R-1)T/K$ (i.e. the tagged PDU arrives before a PDU present in the system at the beginning of the reservation period is served). The next result is thus only valid for x in $\Big[0,\frac{(R-1)T}{K}\Big[$:

$$\begin{aligned}P\Big(W_r\in\Big[\tfrac{((B-1)R+1)T}{K}+x,\tfrac{((B-1)R+1)T}{K}+x+dx\Big[\Big)\\ =\frac{Kdx}{RT(1-\nu^r_B)}\Big(\pi^r_0\gamma_{B-1}(RT/K-x)\\ +\sum_{n=1}^{B}\pi^r_n\gamma_{B-n}(RT/K-x)\Big) \qquad (6)\end{aligned}$$

Lastly, the sojourn time in a system of capacity B is upper bounded by BRT/K which implies that:

$$P^r(W_r\in[x,x+dx[)=0 \qquad x\notin[T/K,BRT/K] \qquad (7)$$

B. Opportunistic Insertion Mode

Under the opportunistic insertion mode, once the station decides that a PDU should be inserted, it inserts the PDU on the first available slot. A slot is unavailable either because it already carries a PDU, or because it is reserved to be used by another class of PDUs. In order to obtain a tractable model for opportunistic insertion, we assume that slot availability is modelled by a Bernoulli process with parameter q_K. A PDU which arrives and finds an empty system only starts its service at the beginning of the next slot; then, if the slot is available (with probability q_K), the service finishes at the end of this slot ; otherwise, the service lasts at least another slot. More precisely, the service time is equal to $l, l>0$ with probability $q_K(1-q_K)^{l-1}$ (geometric distribution with parameter q_K). In the following, we shall derive the distribution for N_o, number of PDUs in system just at the end of a slot, M_o, number of PDUs seen by an arriving PDU, P^o_{loss}, the probability that an arriving PDU finds B PDUs in the system and W_o, sojourn time of a PDU which enters the system.

As system capacity is B, N_o cannot be larger than B. Transition probabilities are as follows:

$$\begin{aligned}P^o(0,i)&=\gamma_i(T/K) && i\le B-1\\ P^o(0,B)&=\sum_{j=B}^{\infty}\gamma_j(T/K)\\ P^o(n,i)&=(1-q_K)\gamma_{i-n}(T/K)\\ &\quad+q_K\gamma_{i-n+1}(T/K) && 1\le n\le B, i\le B-1\\ P^o(n,B)&=(1-q_K)\sum_{j=B-n}^{\infty}\gamma_j(T/K)\\ &\quad+q_K\sum_{j=B-n+1}^{\infty}\gamma_j(T/K) && 1\le n\le B\\ P^o(n,i)&=0 && \{i>B\}\cup\{n>B\}\end{aligned}$$

Let $\pi^o=\{\pi^o_i,0\le i\le(B-1)\}$ be the probability distribution for N_o; π^o is numerically derived by solving $\pi^oP^o=\pi^o$.

M_o differs from N_o, as PDUs can arrive before x, arrival time of a tagged PDU, in the same slot. However, thanks to the fact that a Poisson process is memory-less, the arrival process of PDUs after the beginning of the slot is independent from N_o. We can thus derive ν^o_k, the probability for $\{M_o=k\}$, by summing on n and integrating on x. For k smaller than B:

$$\begin{aligned}\nu^o_k&=\frac{K}{T}\sum_{n=0}^{k}\pi^o_n\int_0^{T/K}\gamma_{k-n}(x)dx\\ &=\frac{K}{\Lambda T}\sum_{n=0}^{k}\pi^o_n\Big(\sum_{i=k-n+1}^{\infty}\gamma_i(T/K)\Big) \qquad (8)\end{aligned}$$

The loss probability P^o_{loss} is ν^o_B and can be derived similarly:

$$\begin{aligned}\nu^o_B&=\frac{K}{T}\sum_{n=0}^{B}\pi^o_n\int_0^{T/K}\sum_{j=B-n}^{\infty}\gamma_j(x)dx\\ &=\frac{K}{\Lambda T}\sum_{n=0}^{B}\pi^o_n\sum_{i=B-n+1}^{\infty}\gamma_i(T/K)(i-B+n) \qquad (9)\end{aligned}$$

The number of PDUs seen by an arriving PDU which is not lost is distributed as $\nu_k^o/(1-\nu_B^o)$ for $k<B$. Little's formula yields the mean sojourn time of such a PDU:

$$E(W_o) = \frac{1}{\Lambda(1-\nu_B^o)} \sum_{k=0}^{B-1} k\nu_k^o \tag{10}$$

In order to derive the distribution for W_o, let U_o denote the time elapsed between the arrival of the tagged PDU and the beginning of the next slot. W_o is equal to the sum of U_o, of the tagged PDU's service time, and of the time it takes to serve the PDUs which are in system when the PDU arrives, and whose service does not stop at the end of the tagged slot. In particular, this implies that W_o is larger than T/K.

$$P(W_o \in [x, x+dx[) = 0 \qquad x < T/K$$

Note that if $N_o = 0$, no PDU can be served during the slot, even if PDUs arrive before the tagged PDU in the same slot. Otherwise, a PDU is currently being served and its service can finish at the end of the slot with probability q_K. Let S_k be the service for the k^{th} PDU to be served, S be the service for the tagged PDU and $\bar{S}_1$ be the remaining service time for the PDU currently being served at the beginning of the slot, if any. The sojourn time W_o of an arriving PDU which is not lost is thus derived as follows:

$$\begin{aligned}
&P\big(W_o \in [iT/K+x, iT/K+x+dx[\big) = \frac{1}{(1-\nu_B^o)} \\
&\Big(\sum_{j=0}^{min(B-1,i-1)} P\big(N_o = 0, \\
&\qquad j \text{ arrivals during } [0, T/K-x[, \\
&\qquad U_o \in [x, x+dx[, S_1 + S_2 + .. + S_j + S = iT/K\big) \\
&+ \sum_{j=1}^{min(B-1,i)} \sum_{n=1}^{j} P\big(N_o = n, \\
&\qquad (j-n) \text{ arrivals during } [0, T/K-x[, \\
&\qquad U_o \in [x, x+dx[, \bar{S}_1 + S_2 + .. + S_j + S = iT/K\big)\Big)
\end{aligned}$$

$\bar{S}_1$, S_k and S are independent. S_k and S are identically distributed, but $\bar{S}_1$ follows a different distribution. Due to the memory-less property of the geometric distribution, $\bar{S}_1$ is equal to $l, l \geq 0$, with probability $q_K(1-q_K)^l$. Thanks to the memory-less property of the Poisson process and of the geometric distribution, we know that what happens before the beginning of the slot (which determines N_o), what happens during the slot (which determines U_o and potential arrivals during $[0, T/K-U[$), and what happens after the slot (which determines the value for the sum of geometrically distributed services) are independent. We finally obtain:

$$\begin{aligned}
&P\big(W_o \in [iT/K+x, iT/K+x+dx[\big) = \frac{1}{(1-\nu_B^o)} \qquad (11) \\
&\Big(\sum_{j=0}^{min(B-1,i-1)} \pi_0^o \gamma_j(T/K-x) \binom{i-1}{j} q_K^{j+1} (1-q_K)^{i-1-j} \\
&+ \sum_{j=1}^{min(B-1,i)} \sum_{n=1}^{j} \pi_n^o \gamma_{j-n}(T/K-x) \binom{i}{j} q_K^{j+1} (1-q_K)^{i-j}\Big)
\end{aligned}$$

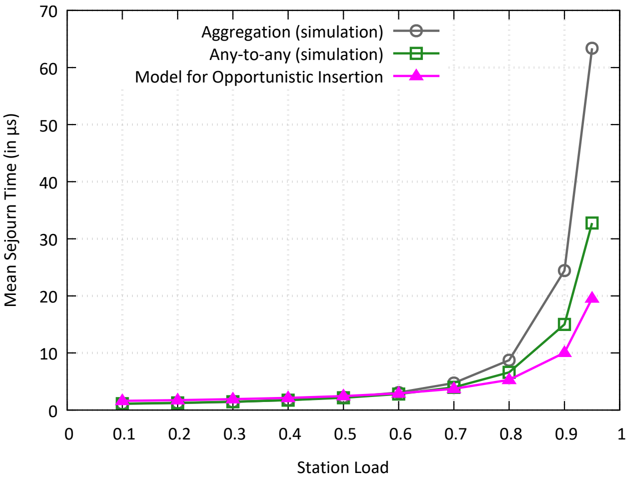


Fig. 1: Mean Sojourn Time: model versus simulations

C. Channel Reservation Mode

A typical benchmark for WSADM systems corresponds to dedicating a data channel to a given station. The behaviour of this system is modelled by an $M/D/1$ queue, with load $\rho = \Lambda T = \lambda$. Both the $M/D/1$ and the $M/D/1/B$ queues are well known models. In particular, the distribution for the number of PDUs seen in the system by an arriving customer M_c (which is also the stationary number of customers N_c in the $M/D/1$ queue thanks to the PASTA property) is given below (see section 5 in [7]).

$$\begin{aligned}
&\pi_0^c = \quad 1-\lambda \qquad\qquad \pi_1^c = \pi_0^c(e^\lambda - 1) \qquad (12) \\
&\pi_n^c = \pi_0^c\Big(e^{n\lambda} + \sum_{j=1}^{n-1}(-1)^{n-j} e^{j\lambda}\Big[\frac{(j\lambda)^{n-j}}{(n-j)!} + \frac{(j\lambda)^{n-j-1}}{(n-j-1)!}\Big]\Big) \\
&\qquad \text{if } n \geq 2
\end{aligned}$$

Moreover (see section 8.2.3 in [7]), the distribution for the sojourn time W_c in the station can also be explicitly derived:

$$\begin{aligned}
P(W_c \leq t) &= (1-\lambda)\sum_{i=1}^{k} e^{-\Lambda(iT-t)} \Lambda^{i-1} \frac{(iT-t)^{i-1}}{(i-1)!} \\
&\qquad t \in [kT, (k+1)T[, \quad k \geq 1 \\
&= \quad 0 \quad t < T \qquad (13)
\end{aligned}$$

The mean sojourn time is given by

$$E(W_c) = T\Big(1 + \frac{\lambda}{2(1-\lambda)}\Big) \tag{14}$$

IV. Validity of Queueing Models

There is no need to check the validity of the slot reservation model, as long as an exact reservation period can be maintained in a real life scenario. Note that this may not always be possible as all reservations have to be organized into a single schedule, which may not always ensure a perfect periodicity for all reservations. Further studies are requested to assess the impact of the schedule design. The validity of the opportunistic insertion model, addressed in Fig. 1, is however more questionable as slot availability depends on the activity of the other stations whereas it is modelled in Section III-B by a Bernoulli process with parameter q_K. A ns3 simulation software has been developed in order to assess the global performance of a WSADM network. A WSADM ring is simulated, with a varying number of stations (link length between two stations $= 4$ km). Each station generates PDUs according to a Bernoulli process. PDUs are stored in a finite buffer of size $B = 99$, making the loss probability negligible. T and K are respectively equal to 10μs and to 10. Each simulation runs during 1 second. Fig. 1 compares the mean sojourn times obtained by the model of Section III-B with the sojourn times measured by simulation in two scenarios. In the "any-to-any" scenario, the sojourn time is measured in one station of a WSADM ring of 20 stations, exchanging traffic in an any-to-any scenario. In the "aggregation" scenario, the sojourn time is measured in a station, which sees the traffic aggregated from 10 other stations. For each value of $\lambda = \Lambda T$, we assume both in the opportunistic insertion model and in the simulations that the tagged station experiences the same mean slot availability. Fig.1 shows that the model is quite close to simulation results as long as $\lambda \leq 0.8$, although it is too optimistic at high load. The model is also closer to the simulation in the "any-to-any" case than in the "aggregation" case.

V. Assessing Packet Insertion Performance

This section provides a performance analysis of a WSADM ring based on the previous models. We focus on traditional MAN scenarios, in which WSADM rings link stations that aggregate the traffic of a large number of customers (at least several tens of thousands of customers for a MAN access ring, up to several hundreds of thousands of customers for a MAN core ring). Although MANs are usually statically dimensioned, data center interconnection may necessitate a more dynamic operation of these networks in the future. This is why the flexible control plane considered for WSADM could be beneficial, compared with a static channel reservation case.

A. Impact of the number of WDM channels

Consider a station generating PDUs according to a Poisson process with parameter Λ. Two cases are analysed: in the first case ($\Lambda T = 0.8$), a full wavelength channel allocation makes sense ("channel reservation"), whereas in the second case ($\Lambda T = 0.4$), it would represent a significant overallocation. Both WSADM insertion modes offer the same amount of resources to the station, i.e. $R = 1/q_K$. We assume that B ensures that the loss probability is negligible. The mean sojourn times in the three models (given respectively in equations (4), (10), (14)) represent the mean time taken to insert a PDU on the MAN for the three considered modes. Actually, if the buffer is infinite, a closed-form formula for the mean sojourn time in slot reservation mode is given by

$$E(W_r) = \frac{T}{K}\left(1 + \frac{KR}{2(K - R\lambda)}\right) \qquad (15)$$

Indeed, the slot reservation model waiting time is quite close to an $M/D/1$ queue with a service time equal to the reservation period RT/K; however, as the service in the WSADM model is provided only at the beginning of the reservation period, statistically, it is necessary to add half a reservation period, and lastly to add the service time T/K.

Fig. 2 and Fig. 3 assume that for each K, both offered traffic and provided resources (i.e. $Kq_K = K/R$) are fixed, which means that $q_K = 1/R$ decrease with K. Both figures represent the mean sojourn times versus K, (i.e. increasing WSADM capacity), bench-marking with the channel reservation mode. The sojourn time for channel reservation mode is obviously independent from K, as a single channel is dedicated to the station, whatever is K; in that case, the sojourn time is constant, equal to $T(1+\frac{\lambda}{2(1-\lambda)})$. Moreover, the mean sojourn time decreases with K in the slot reservation mode while it increases with K in the opportunistic mode. In Fig. 2, the sojourn time for slot reservation mode is smaller than the one for channel reservation mode except for $K = 1$ (in this particular case, both opportunistic and slot reservation cases can use each slot, but as service is slotted, the mean sojourn time exceeds the $M/D/1$ sojourn time by $T/2$). In both figures, the mean sojourn time varies quickly for small values of K, but the variation is smaller for larger values of K. The limit value for the sojourn time in the opportunistic case corresponds to the $M/M/1$ sojourn time, as a geometrically distributed service converges to an exponential service, and as slotting service times has little impact for a small slot size; in the present case, this limit value is $\frac{T}{K/R-\lambda}$. Using equation (15), we see that the limit value for $E(W_s)$ is $\frac{T}{2(K/R-\lambda)}$ which is half the limit value for the opportunistic insertion case. We also see that the slot reservation mode at high load is almost equivalent to the channel reservation mode. At a low station load (as in Fig. 3), channel reservation is an inefficient use

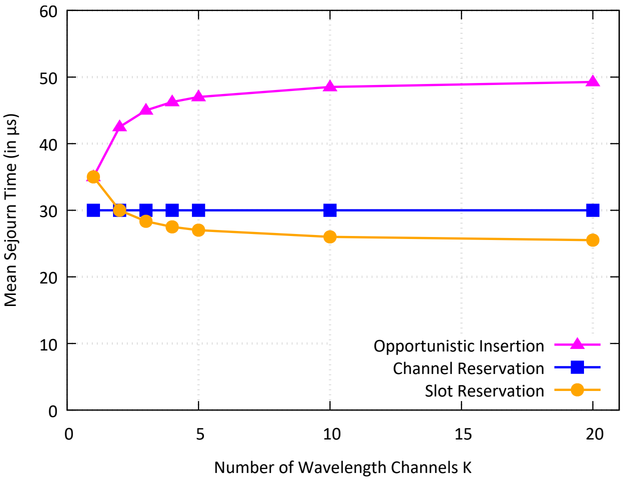


Fig. 2: Mean Sojourn Time versus K; high station load

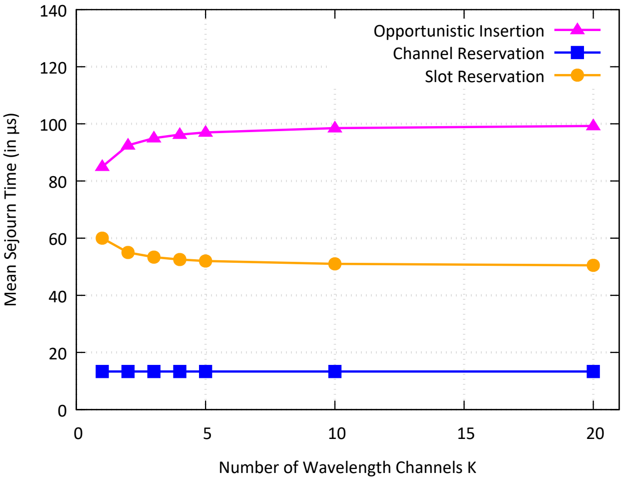


Fig. 3: Mean Sojourn Time versus K; low station load

of optical resources, although it provides of course a better performance than both reservation and opportunistic modes.

B. Supporting transport level performance

Performance objectives for Ethernet Frames are provided by the MEF for different performance tiers [8]. The performance delivered by a WSADM network to SDUs is not fully assessed in the present paper as the aggregation of SDUs within PDUs is not taken into account. However, it is possible to dimension the network, for the various modes, by setting some objectives for PDU transfer that are significantly smaller than the performance objectives set for Ethernet Frames for the Metro Performance Tier (PT1), i.e. spanning up to 250km. In Table I, the first line is extracted from [8], whereas the second line corresponds to the targets we set for the WSADM network. Dimensioning is performed by identifying the amount of resources ensuring a given level of latency and jitter (assuming that B is large enough to neglect PDU loss). In the present case, jitter is defined as a quantile on the insertion delay (propagation delay does not vary). Dimensionning thus consists in identifying how much resources are necessary to ensure that $P(W > 250\mu s)$ is smaller than 10^{-3} versus λ (the 99.9 percentile is selected as in [8]). The case $K = 10$ is considered, as advocated by [4].

One-way Performance Objectives for the Metro Portion			
	Loss	Delay	Jitter
MEF 23.2 [8]	10^{-4}	10ms	3ms
WSADM PDU level	0	2.5ms (propagation) 0.25ms (insertion)	0.25ms insertion

TABLE I: Specifying performance objectives for PDU transfer

Fig. 4 depicts dimensioning for varying λ. The benchmark circuit allocation case corresponds to the horizontal line, as a full channel is allocated for all λ values. Using (13), it is assessed that channel allocation supports the set target up to $\lambda = 0.86$. Opportunistic and slot reservation insertion modes are in most cases more efficient than channel reservation, especially for medium and small λ values; they are also more flexible due to their sub-wavelength granularity. The slot reservation mode is more efficient than the opportunistic insertion mode, especially for small λ values. If the constraint is relaxed (i.e considering a larger target delay for the quantile), this difference would however decrease.

VI. Conclusion

Models for assessing the transfer plane performance in a WSADM network have been derived. They focus on the PDU (or slot) level performance that is governed by the PDU insertion process, as PDUs experience neither loss nor jitter once inserted. The models assume that PDU arrive according to a Poisson process, which is a realistic assumption in a metro network. Slot reservation and opportunistic insertion have been considered, and bench-marked with channel allocation. Both modes have been shown to easily support MEF performance targets, and to present significant resource allocation gains compared to a classical channel allocation.

As we assume a constant channel bit rate, the global ring capacity is proportional to K. This implies that, as insertion latency is less impacted by K, selecting K should be mainly be determined by techno-economic issues. As an example, a ring with 10-channel transponders would have the same capacity as 10 rings with single-channel transponders and would deliver

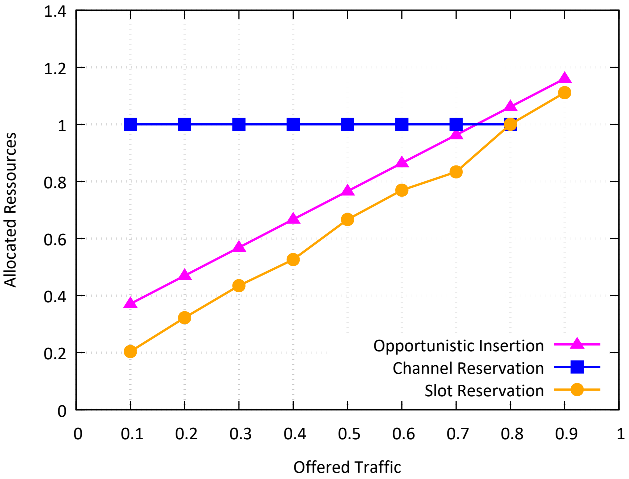


Fig. 4: Resources ensuring $P(W > 250\mu s)$ smaller than 10^{-3} versus λ for $K = 10$

a similar insertion latency (slightly shorter when using a reservation mode and slightly larger in case of opportunistic insertion). However, the cost benefits of using integrated WDM transponders and a single SOA for the 10-wavelength band, as discussed in [5], together with the benefit of managing a single ring, clearly favour WSADM.

Regarding WSADM, the reservation mode slightly outperforms the opportunistic mode in terms of insertion latency and resource usage. However, dimensioning for the opportunistic mode is quite simple: it only implies ensuring that enough resources are available for each station. On the other hand, dimensioning for the slot reservation mode is more complex as it relies on building a global schedule taking into account all flows, each with its own period. However, the two modes are not exclusive as the opportunistic mode only uses slots that are neither already occupied, nor reserved, which makes the WSADM technology quite flexible.

Acknowledgment

The work has been carried out in the framework of the N-GREEN project (ANR-15-CE25-0009-0x), supported by the French Research Agency.

References

[1] I. Widjaja, I. Saniee, R. Giles and D. Mitra *Light core and intelligent edge for a flexible, thin-layered, and cost-effective optical transport network*, Communications Magazine, 2003.

[2] B. Uscumlic, A. Gravey, M. Morvan and P. Gravey, *Impact of peer-to-peer traffic on the efficiency of optical packet rings*, BROADNETS 2008;

[3] J. Dunne, T. Farrell and J. Shields, *Optical Packet Switch And Transport: A new metro platform*," ICTON 2009.

[4] D. Chiaroni and B. Uscumlic, *Potential of WDM packets*, 2017 International Conference on Optical Network Design and Modeling (ONDM 2017), Budapest, May 2017.

[5] A. Triki, A. Gravey, P. Gravey and M. Morvan *Long-Term CAPEX Evolution for Slotted Optical Packet Switching in a Metropolitan Network*, 2017 International Conference on Optical Network Design and Modeling (ONDM 2017), Budapest, May 2017.

[6] L. Sadeghioon, A. Gravey, B. Uscumlic, P. Gravey and M. Morvan, *Full featured and lightweight control for optical packet metro networks*, IEEE/OSA Journal of Optical Communications and Networking, volume 7, number 2, A235-A248 (2015).

[7] D. Gross, J. Shortle, J. Thompson, and C. Harris, *Fundamentals of Queueing Theory*, 2008.

[8] MEF, *Implementation Agreement MEF 23.2 Carrier Ethernet Class of Service - Phase 3*, 2016.

Performance Evaluation of Space Time Coding Techniques for Indoor Visible Light Communication Systems

Abdulmalik Alwarafy*, Mohammed Alresheedi*, Ahmad Fauzi Abas*, and Abdulhameed Alsanie*
* Department of Electrical Engineering, King Saud University, Saudi Arabia
Emails: {437106913@ksu.edu.sa, malresheedi@ksu.edu.sa, aabas@ksu.edu.sa, sanie@ksu.edu.sa}

***Abstract*—In this paper, the performance of visible light communication (VLC) systems, employing Space Time Block Coding (STBC) and Repetition Coding (RC) techniques for an indoor environment is investigated and analyzed. The indoor channel impulse response is taken into account assuming line-of-sight (LOS) and Non-LOS (NLOS) scenarios. The proposed systems employ multiple transmit light emitting diodes (LEDs) with one and two photodetectors (PDs). Various physical arrangements and placements of the LEDs and PD within the indoor scenario are considered. Simulation results show that, for a specific LEDs and PDs arrangement, RC techniques outperform the respective STBC techniques. Furthermore, a 2x2 multiple-input multiple-output (MIMO) VLC system implementing Alamouti STBC is investigated and compared with the RC scheme using a single receiver. It is shown that adding another PD can achieve a signal-to-noise ratio (SNR) improvement of about 5 dB and 2 dB over the Alamouti and RC schemes with a single PD, respectively.**

***Index Terms*—Visible Light Communications, Alamouti Space Time Block Coding, Repetition Coding, Performance Evaluation.**

I. Introduction

Visible light communication (VLC) systems provide means of delivering both high data rate and illumination services over indoor or short-distances outdoor environments, as shown in Fig. 1. The high data rates are supported due to the higher spectral efficiency since VLC systems have a vast amount of unregulated bandwidth and a limited coverage that enables extensive frequency re-use. Additionally, the short carrier wavelength and large square-law photodetector (PD) used in VLC systems enable a spatial diversity that reveals immunity against multipath fading [1]. The maximum transmitted power in VLC systems is governed by safety considerations, and the noise arising from conventional fluorescent lamps and sunlight will limit the maximum achievable optical signal-to-noise ratio (SNR) [1], [2].

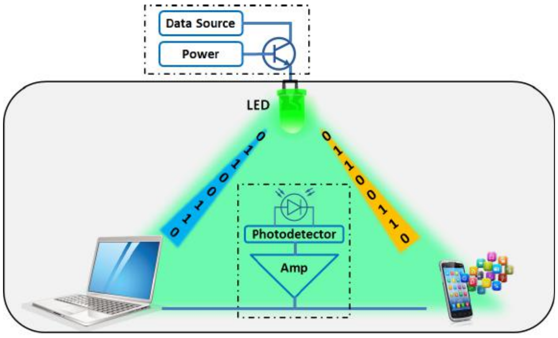


Figure 1: The concept of VLC system.

Indoor VLC systems are characterized by smaller distances and they are free from atmospheric degradations; however, VLC links suffer from interference induced by multipath propagation. Hence, the performance of VLC systems can be significantly enhanced by utilizing multiple transmit light emitting diodes (LEDs) and receive PDs at either/both ends of VLC terminals. It is more convenient, however, to add these LED elements at the transmitter side to provide both data communication and the necessary illumination. Hence, multiple-transmit LEDs VLC systems are becoming more attractive and this has led researchers to explore the Multiple-Input Multiple-Output (MIMO) techniques for VLC systems with or without Space Time Block Coding (STBC) [2]–[9] as well as the Repetition Coding (RC) techniques [4], [5], [10]. STBC and RC techniques have proven to be promising for the VLC systems [1], since they can increase capacity and improve the performance without any increase in transmitting optical power, and with only a simple linear processing at the receiving end [3], [11].

Previous researches in the literature have considered the infrared (IR) optical wireless communication (OWC) systems, in which the works have studied the STBC [1], [12], the RC [12], [13], and the MIMO [1], [13] systems. Moreover, there have been a limited and specific number of studies of the potential of using STBC and RC for VLC systems. The work in [5] demonstrated the feasibility of binary system employing only Alamouti STBC with one camera receiver in an outdoor image-sensor-based VLC system. The authors in [10] considered a Multiple-Input Single-Output (MISO) VLC system utilizing only RC scheme in which they added a pilot bit to ensure a reliable blind estimation of channel coefficients. The work in [4] proposed a 2×2 MIMO system with only Alamouti STBC in a VLC system using image sensor-based direct detection (DD) with a high-speed camera. The authors in [2] introduced spatial modulation (SM) into layered STC that is used in image sensor-based VLC systems. The work in [6] proposed a design of linear space codes for an indoor MIMO VLC with two transmitters and multiple receivers. The authors in [7] considered RC and SM coding schemes, and they tried to optimize the placement and power of the LEDs in a 4×4 MIMO configuration to obtain a uniform SNR for the desired BER and data rate. However, these works that are

978-3-903176-07-2 © 2018 IFIP

related to the STBC VLC are limited to the Alamouti STBC and they do not consider higher order coding schemes such as the 4×4 STBC. Furthermore, the works that are related to the RC do not provide any comparison with the STBC scheme to describe or emphasize which one of these two coding techniques is the best to be used with the VLC systems. Additionally, none of these works has devoted to the impact of the line-of-sight (LOS)/Non-LOS (NLOS) scenarios or the LED/PD arrangements on the VLC system performance.

In this paper, a comprehensive numerical performance analysis is conducted for both the STBC and RC, in order to investigate which one of these coding techniques is the best to be used with VLC systems. The performances of these two coding schemes are quantitatively and qualitatively compared considering LOS and NLOS scenarios with various LEDs/PD arrangements. We first consider the Alamouti STBC, 4×4 STBC, and RC with one PD for the LOS and NLOS scenarios. Then, the performance of the Alamouti STBC is studied for the 2×2 MIMO VLC system in a LOS scenario. Each of these proposed systems is analyzed by obtaining the simulation results in terms of SNR vs bit error rate (BER).

The rest of this paper is organized as follows. Section II provides the proposed system model of the VLC system under the STBC and RC techniques with a special focusing on the Alamouti STBC. Section III gives a comprehensive description of the simulation procedure for the LOS and NLOS scenarios that are considered in this paper, along with the performance analysis and discussion of the results obtained. Finally, section IV summarizes the paper.

II. STBC and RC VLC System Model

In this work, we consider VLC systems using intensity modulation and direct detection (IM/DD) equipped with N_T transmit LEDs per array and one or two PDs per receive array. Fig. 2 shows the block diagram of the VLC system considered in this paper. The proposed system is studied using STBC (either Alamouti or 4×4 STBC) and RC schemes. It should be noted that most of the analysis considers the Alamouti scheme; however, the 4×4 STBC follows the same concept. In IM/DD VLC scheme, the LEDs require positive and real modulated symbols since the LED cannot differentiate the phase of the input signals [1], [3], [4], [12], [14]. Therefore, the binary phase shift keying (BPSK) to on-off keying (OOK) Mapper block is used to generate the OOK sequences x_1 and x_2 from the BPSK sequences s_1 and s_2, respectively, as shown in Fig. 2. These OOK sequences are then applied to either the STBC encoder (Alamouti or 4×4) or the repetition encoder. The transmitted OOK symbols at two consecutive symbol periods 1 and 2 from each element of the two-LED array in the Alamouti STBC (or at the four consecutive symbol periods 1, 2, 3, and 4 from each element of the four-LED array for the 4×4 STBC) are shown at the top right corner of Fig. 2. This Alamouti scheme is called the modified orthogonal Alamouti STBC [4], [12], [14]. The received signals from PD1 and PD2 at the first symbol period after DD are denoted by r_1 and r_2, respectively, and the respective signals at the second symbol period are denoted by r_3 and r_4 as shown in the bottom right corner of Fig. 2. For the RC, on the other hand, the same OOK symbol is transmitted from all the available LEDs at a particular symbol period [12], [13] as shown in the top right corner of Fig. 2 for the $N_T = 2$ RC case. Note that the same logic is applied for $N_T > 2$. In STBC or RC techniques, there is no additional optical power needed, since the power will be equally divided between all the N_T LEDs [1]. The OOK encoded optical signals will then propagate through the diffused VLC optical channel.

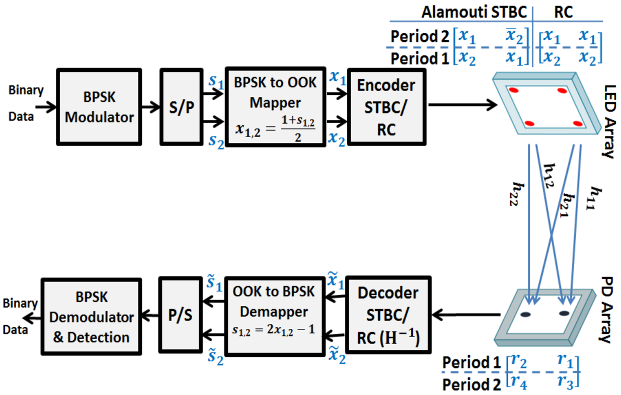


Figure 2: The block diagram of the proposed VLC system that employs STBC and RC techniques.

A. Alamouti STBC System Model

Although we analyze the performance of both 4×4 and Alamouti STBC, we focus on the Alamouti orthogonal STBC with either one or two PDs, as shown in Fig. 2. We assume background noise limited optical receivers in which the shot noise caused by background radiation is dominant relative to the thermal noise [1], [3], [12]. Since the optical channel does not introduce any nonlinearity [3], the overall noise components are modeled as an additive white Gaussian noise (AWGN) [3], [12]. Based on these assumptions, the received electrical signals after DD from the two PDs at the two symbol times, are given by [12], [14]:

$$\begin{aligned} r_1 &= \frac{R}{N_T}\Big(h_{11}x_1 + h_{12}x_2\Big) + n_1, \\ r_2 &= \frac{R}{N_T}\Big(h_{21}x_1 + h_{22}x_2\Big) + n_2, \\ r_3 &= \frac{R}{N_T}\Big(h_{11}\overline{x}_2 + h_{12}x_1\Big) + n_3, \\ r_4 &= \frac{R}{N_T}\Big(h_{21}\overline{x}_2 + h_{22}x_1\Big) + n_4. \end{aligned} \tag{1}$$

where R is the PD responsivity. If only one PD is used, we have only r_1 and r_3 in Eq. (1). To obtain the LOS indoor optical wireless channel DC gains h_{ij} for a single LED, the modified Monte Carlo method is used with the arrangement shown in Fig. 3(a), so we have [1], [3], [7], [13]:

$$h_{LOS} = \begin{cases} P_{TX}\frac{(m+1)A_{PD}}{2\pi D^2}cos(\phi)cos^m(\theta) & 0 \leq \phi \leq \Psi_{\frac{1}{2}} \\ 0 & \phi > \Psi_{\frac{1}{2}} \end{cases} \tag{2}$$

where P_{TX} is the transmit optical power, m is the mode number of the Lambertian source, which is related to the half power semi angle ($\Phi_{\frac{1}{2}}$) of the LED by $m = -ln2/cos(\Phi_{\frac{1}{2}})$, $\Psi_{\frac{1}{2}}$ is the field-of-view (FOV) semiangle of the PD, D is the distance between the LED and the PD, A_{PD} is the effective area of the PD, and θ and ϕ are the irradiance and incident angles, respectively as depicted in Fig. 3(a).

At the decision logic of the Alamouti ST decoding, it is assumed that the receiver has a perfect knowledge of the VLC optical channel DC gains [3], [4], [12], [14]. Therefore, the decision statistics formed from the PDs at the two symbol periods are given by [14]:

$$\begin{aligned}\widetilde{x}_1 &= \sum_{i=1}^{N_T} h_{i1}r_i + \sum_{i=1}^{N_T} h_{i2}r_{i+2} - \sum_{i=1}^{N_T} h_{i1}h_{i2},\\ \widetilde{x}_2 &= \sum_{i=1}^{N_T} h_{i2}r_i + \sum_{i=1}^{N_T} h_{i1}r_{i+2} + \sum_{i=1}^{N_T} h_{i1}^2,\end{aligned} \tag{3}$$

Finally, the maximum likelihood (ML) decision is made separately on each of the transmitted information signals x_1 and x_2 using the metric [4], [14]:

$$m(\tilde{x}_i, x_i) = (\tilde{x}_i - x_i)^2 + (h_{11}^2 + h_{12}^2 - 1)x_i^2, \qquad i = 1,2 \tag{4}$$

and the respective decision rule is to choose $x_i = \hat{x}_i$ if:

$$\begin{aligned}(\tilde{x}_i - \hat{x}_i)^2 + (h_{11}^2 + h_{12}^2 - 1)\hat{x}_i^2 &\le (\tilde{x}_i - x_i)^2\\ +(h_{11}^2 + h_{12}^2 - 1)x_i^2. &\quad x_i \neq \hat{x}_i\end{aligned} \tag{5}$$

B. RC System Model

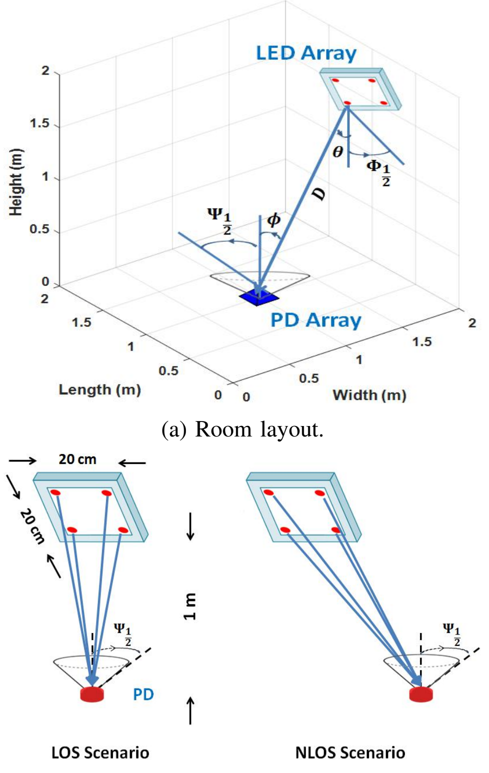
One advantage of using IM in the VLC systems is that transmit diversity can be realized through RC [12], [13], [15]–[17]. In RC, the same OOK signal is simultaneously transmitted from all the available N_T LEDs as shown in the top right corner of Fig. 2 for the $N_T = 2$ RC case. Since the optical channel DC gains h_{ij}'s are real and positive, the intensities coming from the several independent transmit LEDs in RC will add up at the PD side [12], [13]. In RC, the received signal over a single symbol period is given by [12]:

$$r = \frac{R}{N_T}\sum_{i=1}^{N_T} h_i x + n, \tag{6}$$

The major advantage of RC scheme is that it combines the faded signals before noise accumulation, unlike the SIMO scheme which combines the noisy faded signals. Therefore, the performance of RC is better than SIMO scheme [12], [17].

III. Results and Analysis

In this section, the performance of the two coding techniques discussed above will be investigated and analyzed considering various LOS and NLOS scenarios with different LEDs/PD configurations. For the STBC, the modified Alamouti and 4×4 STBC will be considered with one PD at the receiver side. Whereas, the RC is considered with $N_T = 2, 3,$ and 4 LEDs per array and one PD. The BER vs SNR performance results of all these coding schemes are then qualitatively and quantitatively compared. Finally, the impact of using the 2×2 MIMO with Alamouti STBC on system performance will be demonstrated in LOS scenarios.



(a) Room layout.

(b) LOS and NLOS scenarios.

Figure 3: Simulation setup: (a) room configuration (b) LOS and NLOS scenarios.

A. Simulation Setup

Fig. 3(a) shows the communication setup of the VLC system that is considered in the simulation, which is equipped with LED and PD arrays. The room has dimensions of 4 m × 4 m × 3 m. The LED array has a first order Lambertian pattern and is oriented vertically towards the floor. The rest of the simulation parameters are shown in Table I.

B. Performance Evaluation of the LOS Scenarios

In the LOS case, we further consider three scenarios and investigate their impacts on the VLC system performance. These scenarios are: the effects of changing the LEDs spacing within the array, the effects of the separation distance between the LED array and the PD, and the effects of the implementation of 2×2 MIMO Alamouti STBC scheme.

1) Effects of LEDs Spacing Within the Array:

The spacing between LED elements within transmit array

Table I: Simulation Parameters

Parameter	Value
Room dimensions	(4, 4, 3) m
P_{TX}	30 dBm
Responsitivity	1
$\Phi_{\frac{1}{2}}$	70°
PD area	1 cm^2
$\Psi_{\frac{1}{2}}$	90°
Electrical baseband modulation	BPSK

must be deliberately adjusted in such a way, small spacing is required so that the LEDs can be integrated in the same end-device, whereas, large spacing is required to exploit the spatial diversity. To carry out this study, we consider the LOS scenario shown in Fig. 3 (b), in which the single PD is placed at the midpoint of the LOS view of the LED array with a 1 m LED-PD separation distance, while the LED spacing is varied from 20 cm to span the whole area of the room's roof.

Fig. 4 shows the performance of all the coding schemes for the case when the LED spacing is 20 cm. At a fixed BER, the RC with $N_T = 2$ outperforms the modified Alamouti STBC. For example, at BER= 10^{-3}, the SNR for RC is around 18 dB compared to around 22 dB for the Alamouti STBC, which means that the RC requires less SNR of around 4 dB, hence it is more power efficient. The performance of the 4×4 STBC is identical to the $N_T = 2$ RC, and worse than the $N_T = 4$ RC. Furthermore, the performance of RC rapidly increases as the number of transmitting LEDs increases. These results clearly conclude that RC is performing better than the STBC when considering the same N_T; therefore, they are the best to be used with VLC systems. Similar conclusions have been reported in [1], [12], [13] for the infrared OWC. It is worthy to mention that these trends are applicable for all the case studies considered in this paper. The major difference is how much reduction/enhancement in SNR achieved in each case study. The performance results of the LOS scenario shown in Fig. 4 are considered as the reference for the quantitative analysis with other scenarios. To examine the impact of increasing the LEDs spacing on the systems performances, Fig. 5 shows the simulation results when the spacing increases to 1 m, for the same PD position (i.e. in the midpoint of the LEDs LOS view) and the same LED-PD separation distance (i.e. 1 m). Increasing the LEDs spacing will deteriorate the performance for all the coding schemes, since we require additional SNR (or power) to achieve a fixed BER. The reason is that the contribution of the LOS component intensity decreases as the LED elements go away from the PD. For example, compared with the reference scenario in Fig. 4, to maintain the same BER of 10^{-3}, an increase in the SNR of about $18 - 13 = 5$ dB is required for $N_T = 4$ RC and around $28 - 22 = 6$ dB for the modified Alamouti STBC. To capture the general trends of SNR as a function of the LEDs spacing, the position of the PD was fixed to be at the midpoint of the LEDs' LOS view with a LED-PD height of 1 m, while allowing the LEDs' spacing to span over the whole area of the roof (for the 4 LEDs elements or the whole length of the roof for the 2 LED elements). It was shown that the SNR that is required to maintain a particular BER increases approximately linearly as a function of the LEDs spacing for all the coding schemes.

2) Effects of Separation Distance Between LED Array and PD:

In this section, the impact of changing the separation distance between the LED array and PD, on the VLC system performance is investigated. The simulation scenario for this case is similar to the LOS scenario shown in Fig. 3(b) (i.e. the LED spacing is 20 cm and the PD is placed at the midpoint of the LED array LOS view). The only difference is that the separation distance between the LED array and the PD will now be changed. Fig. 6 shows the performances of all the coding schemes when the separation distance is 3 m. Compared with the reference scenario that is shown in Fig. 4, as the separation increases, the SNR that is required to achieve a particular BER increases. For example, compared with Fig. 4, to achieve a BER of 10^{-3}, we need an extra SNR (or power) of around $32 - 13 = 19$ dB for $N_T = 4$ RC and around $41 - 22 = 19$ dB for Alamouti STBC. If, however, the separation distance is decreased to 50 cm, the results shown in Fig. 7 are obtained. A huge enhancement is now achieved in the SNR performance. For instance, a SNR reduction of around $13 - 3 = 10$ dB for $N_T = 4$ RC and around $22 - 13 = 9$ dB for Alamouti STBC is obtained when comparing Fig. 4 with Fig. 7. Therefore, simulation results show that the separation distance between the LED array and PD play a crucial role in the VLC systems. To study the general trends of SNR as a function of the LED-PD separation distance, the

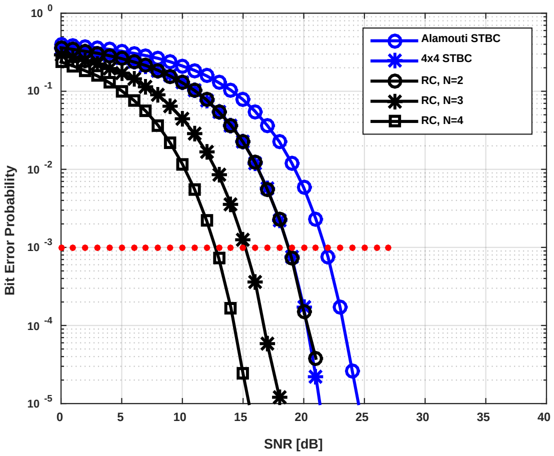


Figure 4: BER vs SNR performance for the LOS scenario shown in Fig. 3(b), with LED spacing of 20 cm, LED array-PD separation of 1 m. This case is considered as the reference for the quantitative analysis.

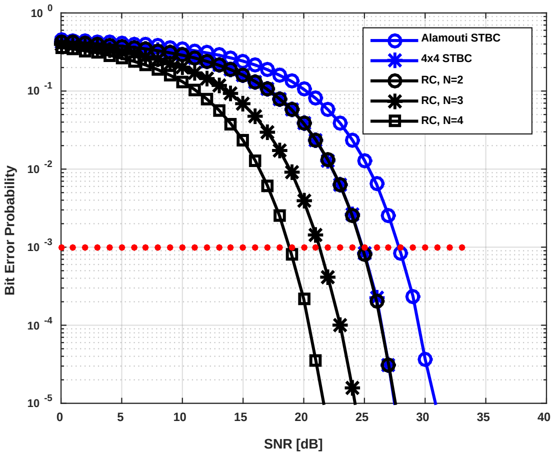


Figure 5: BER vs SNR performance for the LOS scenario shown in Fig. 3(b), with LED spacing of 1 m, LED array-PD separation of 1 m.

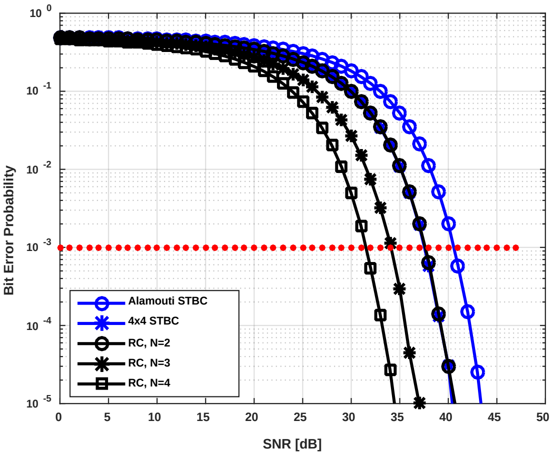


Figure 6: BER vs SNR performance for the LOS scenario shown in Fig. 3(b), with LED spacing of 20 cm, LED array-PD separation of 3 m.

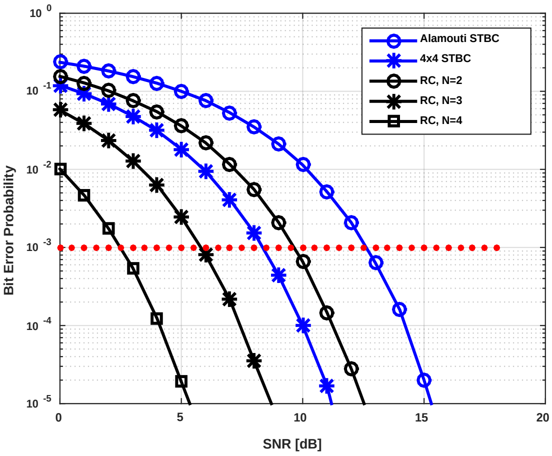


Figure 7: BER vs SNR performance for the LOS scenario shown in Fig. 3(b), with LED spacing of 20 cm, LED array-PD separation of 50 cm.

spacing between the LEDs within array were fixed to 20 cm and the PD was positioned at the midpoint of the LED array LOS view, while allowing the LED array to go apart from the PD to reach the room's roof. It was shown that the SNR that is required to maintain a specific BER increases logarithmically as a function of the LED-PD separation distance. It was also shown that the RC with $N_T = 4$ requires the least amount of additional SNR at any specific separation distance, whereas Alamouti STBC requires the highest additional SNR.

3) 2x2 MIMO Alamouti STBC for VLC:
This section investigates the performance of VLC systems when implementing 2×2 MIMO (i.e. two LEDs per transmit array and two PDs per receive array) with Alamouti STBC. The simulation layout of this case is similar to the LOS scenario shown in Fig. 3(b), except another PD was added and placed 10 cm away from the previous one. Fig. 8 shows the simulation results of three systems: 2×1 Alamouti, 2×2 MIMO Alamouti, and 2×1 RC. The 2×2 MIMO Alamouti STBC has the best performance for any fixed level of BER, followed by the RC then by the 2×1 Alamouti. For example, at a BER of 10^{-3}, the 2×2 MIMO system requires around $22 - 17 = 5$ dB and around $19 - 17 = 2$ dB SNR less than the 2×1 Alamouti and the 2×1 RC, respectively. This enhancement, however, comes with an additional complexity at the receiver side.

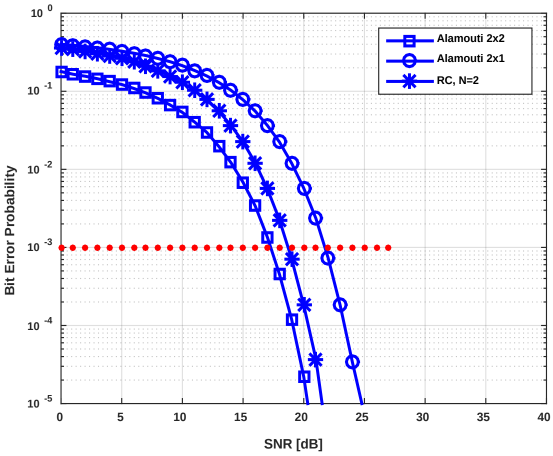


Figure 8: 2×2 MIMO Alamouti STBC BER vs SNR performance for the LOS scenario shown in Fig. 3(b).

C. Performance Evaluation of the NLOS Scenarios

This section investigates the partial NLOS scenario in which only the first reflected paths are considered at the PD. We analyze the effects of the position of the PD with respect to the LOS view of the LED array. Fig. 3(b) shows the scenario in which the single PD is positioned "or misaligned" 90 cm outside the LOS view of the LED array, and Fig. 9 shows the simulation results obtained for all the considered coding schemes. The performance drastically deteriorates due to the decrease in the received optical intensities as the PD is not within the LOS view of the LED array. Comparing these results with the LOS results shown in Fig. 4, to achieve a fixed BER of 10^{-3}, an extra SNR of around $23-13 = 10$ dB for $N_T = 4$ RC and around $32 - 22 = 10$ dB for Alamouti STBC are required. The results show that even if the PD is slightly placed in a NLOS communication links with respect to the LED array, the VLC system encounters severe reduction in the performance. In general, it was observed that the SNR that is required to achieve a fixed BER, increases linearly as a function of the PD distance from the edge of the LED array for all the considered coding schemes.

IV. Conclusion

This paper presents performance analysis of the STBC and RC techniques for VLC systems. It is shown that the performance of RC is better than the STBC in a single PD reception case; however, if MIMO VLC implemented, STBC outperforms the RC at an expense of additional complexity at the receiver side. The effects of LOS and NLOS scenarios as

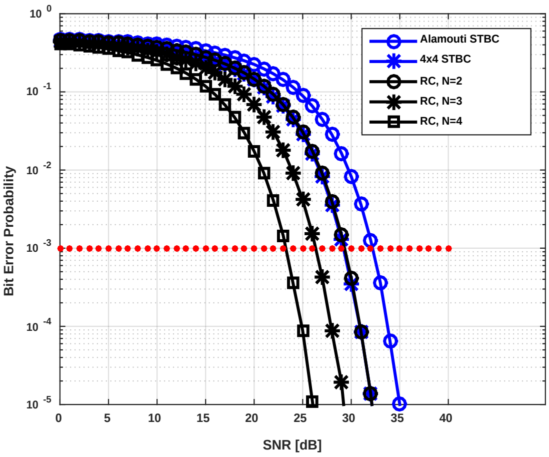


Figure 9: BER vs SNR performance for the NLOS scenario shown in Fig. 3(b), with LED spacing of 20 cm, LED array-PD separation of 1 m and the PD is placed 90 cm apart from the LED array edge.

well as the LEDs/PD physical arrangements on VLC system performance was also investigated. Three parameters were investigated which heavily contribute to the VLC performance which are: the spacing of LEDs within the array, the position of the PD with respect to the LOS view of the LED array, and the LED-PD separation distance. Simulation results show that even if the PD is slightly placed in NLOS communication links with respect to the LED array, the performance of VLC system encounters severe deterioration. Furthermore, proper placement of the PD could enhance the SNR up to 19 dB in LOS scenarios. Our future work is to investigate and analyze the performance of imaging angle diversity for MIMO VLC systems.

Acknowledgment

The project was supported by King Saud University, Deanship of Scientific Research, College of Engineering Research Center.

References

[1] G. Ntogari, T. Kamalakis, and T. Sphicopoulos, "Performance analysis of space time block coding techniques for indoor optical wireless systems," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 9, 2009.

[2] K. Masuda, K. Kamakura, and T. Yamazato, "Spatial modulation in layered space-time coding for image-sensor-based visible light communication," in *IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2016, pp. 1–6.

[3] Z. Ghassemlooy, W. Popoola, and S. Rajbhandari, *Optical wireless communications: system and channel modelling with Matlab®*. CRC press, 2012.

[4] Y. Amano, K. Kamakura, and T. Yamazato, "Alamouti-type coding for visible light communication based on direct detection using image sensor," in *IEEE Global Communications Conference (GLOBECOM)*, 2013, pp. 2430–2435.

[5] K. Ebihara, K. Kamakura, and T. Yamazato, "Spatially-modulated space-time coding in visible light communications using 2× 2 LED array," in *IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)*, 2014, pp. 320–323.

[6] X.-c. Gao, J.-k. Zhang, and J. Jin, "Linear space codes for indoor MIMO visible light communications with ML detection," in *IEEE 10th International Conference on Communications and Networking in China (ChinaCom)*, 2015, pp. 142–147.

[7] M. K. Jha, A. Addanki, Y. Lakshmi, and N. Kumar, "Channel coding performance of optical MIMO indoor visible light communication," in *IEEE International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2015, pp. 97–102.

[8] M. Biagi and A. M. Vegni, "Enabling high data rate vlc via MIMO-LEDs PPM," in *Proc. of* 4th *IEEE Globecom* 2013 *Workshop on Optical Wireless Communications (OWC)*, Atlanta GA, USA, Dec. 2013, pp. 1058–1063.

[9] M. Biagi, A. M. Vegni, S. Pergoloni, P. M. Butala, and T. D. Little, "Trace-orthogonal PPM-space time block coding under rate constraints for visible light communication," *Journal of lightwave technology*, vol. 33, no. 2, pp. 481–494, 2015.

[10] Y.-J. Zhu, Z.-G. Sun, J.-K. Zhang, Y.-Y. Zhang, and J. Zhang, "Training receivers for repetition-coded MISO outdoor visible light communications," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 1, pp. 529–540, 2017.

[11] H. Bölcskei, *Space-time wireless systems: from array processing to MIMO communications*. Cambridge University Press, 2006.

[12] M. Safari and M. Uysal, "Do we really need OSTBCs for free-space optical communication with direct detection?" *IEEE Transactions on Wireless Communications*, vol. 7, no. 11, pp. 4445–4448, Nov. 2008.

[13] T. Fath and H. Haas, "Performance comparison of MIMO techniques for optical wireless communications in indoor environments," *IEEE Transactions on Communications*, vol. 61, no. 2, pp. 733–742, 2013.

[14] M. K. Simon and V. A. Vilnrotter, "Alamouti-type space-time coding for free-space optical communication with direct detection," *IEEE Transactions on Wireless Communications*, vol. 4, no. 1, pp. 35–39, 2005.

[15] E. J. Lee and V. W. Chan, "Part 1: Optical communication over the clear turbulent atmospheric channel using diversity," *IEEE journal on selected areas in communications*, vol. 22, no. 9, pp. 1896–1906, 2004.

[16] S. G. Wilson, M. Brandt-Pearce, Q. Cao, and M. Baedke, "Optical repetition MIMO transmission with multipulse PPM," *IEEE journal on Selected Areas in Communications*, vol. 23, no. 9, pp. 1901–1910, 2005.

[17] S. M. Navidpour, M. Uysal, and M. Kavehrad, "BER performance of free-space optical transmission with spatial diversity," *IEEE transactions on Wireless Communications*, vol. 6, no. 8, 2007.

Impact of Physical Layer Impairments on Multi-Degree CDC ROADM-based Optical Networks

Diogo G. Sequeira[1], Luís G. Cancela[1,2], and João L. Rebola [1,2]
[1] Optical Communications and Photonics Group, Instituto de Telecomunicações, Lisbon, Portugal
[2] Department of Information Science and Technology, Instituto Universitário de Lisboa (ISCTE-IUL), Portugal
Emails: dgsao@iscte-iul.pt; luis.cancela@iscte-iul.pt; joao.rebola@iscte-iul.pt

***Abstract*—Nowadays, optical network nodes are usually based on reconfigurable optical add/drop multiplexers (ROADMs). Due to exponential growth of internet data traffic, ROADMs have evolved to become more flexible, with multi-degree and their add/drop structures are now more complex with enhanced features, such as colorless, directionless and contentionless (CDC). In this work, the impact of in-band crosstalk, optical filtering and amplified spontaneous emission noise on the performance of an optical network based on multi-degree CDC ROADMs is studied considering 100-Gb/s polarisation division multiplexing quadrature phase-shift keying signals for the fixed grid. We show that, an optical signal can pass through a cascade of 19 CDC ROADMs, based on a route and select architecture with 16-degree, until an optical signal-to-noise ratio (OSNR) penalty of 1 dB due to in-band crosstalk is reached. We also show that the ASE noise addition, due to the increase of the number of CDC ROADMs, is more harmful in terms of OSNR penalty than in-band crosstalk.**

Keywords: ASE noise, CDC ROADMs, coherent detection, in-band crosstalk, optical filtering, PDM-QPSK.

I. INTRODUCTION

The exponential growth of internet data traffic due to the increase of the number of devices, cloud and video-on-demand services, has been putting fibre optic network technologies in a continuous development to support all the data generated. Technologies, such as dense wavelength-division multiplexing, optical coherent detection, polarisation division multiplexing (PDM) and advanced digital signal processing (DSP) are now fundamental to achieve the huge transport capacities required by the overall telecommunications infrastructure [1].

In addition to these technologies, the reconfigurable optical add/drop multiplexers (ROADMs) nodes evolution is also very important to support this exponential growth. In the past, the network nodes were static and their configuration was manual. Nowadays, these nodes became more reconfigurable with colorless, directionless and contentionless (CDC) features [2], that improves the routing and switching functionalities in the optical nodes, making them more dynamic and reliable.

On the other hand, the optical network physical layer impairments (PLIs) require a comprehensive study since the optical signal along its path, passes through optical fibre links as well as optical components inside the ROADMs, such as optical switches, (de)multiplexers and splitters/couplers. The losses, noises and interferences generated in these links accumulate along the light-path degrading the optical signal transmission. In particular, the imperfect isolation of switches and filters inside the ROADMs leads to signal leakages that originate interfering signals known as crosstalk signals. One of the crosstalk types that becomes enhanced in an optical network and degrades the optical network performance is the in-band crosstalk [3]. This type of crosstalk occurs when the interfering signals have the same nominal wavelength as the primary signal but are originated from different sources, so that this impairment cannot be removed by filtering. In an optical network based on ROADMs, the in-band crosstalk will accumulate over the ROADM cascade and can limit the number of nodes that the signal passes in the network [4]. In the literature, some studies were performed to address the impact of the in-band crosstalk on the optical network performance, however with a simple ROADM model [5] or not considering the ROADM add/drop structures with the CDC features [6].

In this work, the impact of in-band crosstalk generated inside multi-degree CDC ROADMs on the network performance is studied through Monte-Carlo simulation. Polarisation division multiplexing quadrature phase-shift keying (PDM-QPSK) signals at 100-Gb/s for the fixed grid are considered. This study is performed by properly modelling the in-band crosstalk generation inside the ROADMs. Different ROADM architectures, namely broadcast and select (B&S) and route and select (R&S) architectures [6], as well as, different add/drop structures, based on multicast switches (MCSs) and wavelength selective switches (WSSs) [7], are considered.

This paper is organized as follows. Section II describes the model for studying the in-band crosstalk inside a ROADM node, and the number of in-band crosstalk terms generated inside a ROADM is quantified, for both B&S and R&S architectures. Details on the ROADM transponder, as well as, on the ROADM add/drop structures are also provided in this section. In section III, the PLIs such as the optical filtering, amplified spontaneous emission (ASE) noise and in-band crosstalk in an optical network based on multi-degree CDC ROADMs are studied and their impact on the network performance is assessed. Finally, in section IV, the conclusions of this work are presented.

II. MODELLING THE IN-BAND CROSSTALK INSIDE A ROADM

The main focus of this section is on the in-band crosstalk generation inside a ROADM node. In subsections II.A and II.B, we will describe, respectively, the ROADM transponder features and the ROADM add/drop structures. Subsection II.C deals with the in-band crosstalk generation inside a ROADM node.

978-3-903176-07-2 © 2018 IFIP

A. ROADM Transponder

In this sub-section, we present the main blocks of the coherent receiver of a ROADM transponder, used for detecting the optical signal that is dropped in a ROADM. Fig. 1 depicts the block diagram of the coherent receiver for a single polarisation of the signal. The coherent receiver with dual polarisation consists of two polarisation beam splitters connected with two structures identical to the one depicted in Fig. 1. In this work, we assume that the optical receiver is ideal, so the receiver performance can be assessed considering only the structure of Fig. 1 for a single polarisation of the signal [8].

The structure of the optical coherent receiver is formed by a 2×4 90º hybrid, which has $E_r(t)$ and $E_{LO}(t)$, respectively, the complex envelope of received signal and local oscillator (LO) electrical fields as inputs. The received electrical signal corresponds to the signal under test, the primary signal, dropped by a ROADM. The 2×4 90º hybrid is followed by two balanced photodetectors. The hybrid, which is modelled as in [8], is composed by four 3 dB couplers and a 90º phase shift in the lower branch, which allows the receiver to decode the in-phase and quadrature signal components of the received currents, respectively, $I_i(t)$ and $I_q(t)$ in Fig. 1. An electrical filter is placed after the balanced photodetector, to reduce the inter--symbolic interference and the noise power, consequently, improving the signal-to-noise ratio [9]. In this work, we use a 5*th* order Bessel filter as the receiver electrical filter, which is a typical filter used in several studies [10]. The −3 dB bandwidth of this filter is set equal to the symbol rate. After electrical filtering, the signal is sampled by an analog-to-digital converter before going to a DSP (not shown in Fig. 1). Finally, a decision on the transmitted symbol is taken at the decision circuit.

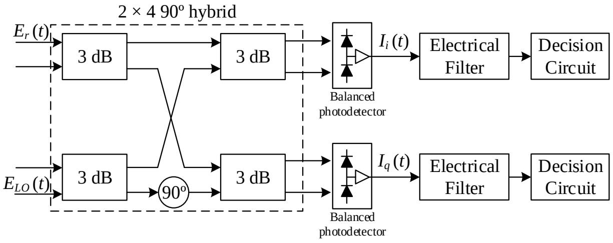


Fig. 1. Coherent receiver block diagram for a single polarisation QPSK signal.

B. ROADM Add/Drop Structures

In this subsection, we present the internal structure of a ROADM add/drop structure based on both MCSs and WSSs. Fig. 2 shows a generic internal structure of (a) MCSs and (b) WSSs that can be used in the drop section of a CDC ROADM [7]. As we can observe from this figure, the MCSs are based on 1×*M* splitters and *N*×1 optical switches. As such, they are not wavelength selective as the WSS structures. On the other hand, the WSS structures have higher costs. However, in terms of in-band crosstalk generation, since inside a *N*×*M* WSS, the interfering signals pass through the isolation of two WSSs, the interferers are second order interferers, instead of the first order interferers that appear on the *N*×*M* MCSs outputs.

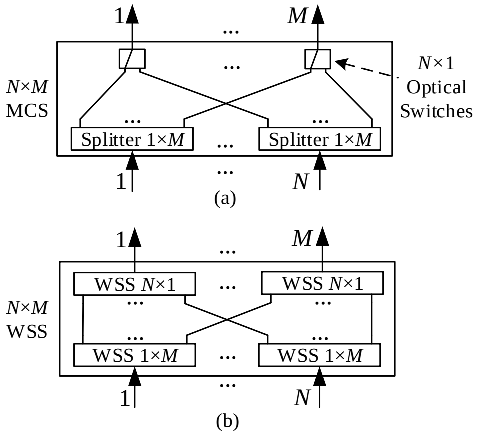


Fig. 2. ROADM drop section structure based on (a) MCS and (b) WSS.

C. In-Band Crosstalk Generation inside a ROADM

For studying the number of crosstalk terms generated inside a ROADM with degree *R*, we consider, a four-node star network with a full-mesh logical topology as depicted in Fig. 3. As a worst-case scenario, we assume that the central ROADM, node 2, communicates with other nodes using the same wavelength, λ1. This means that, the wavelength λ1 reaching node 2, is dropped and new optical signals with the same wavelength λ1 are added and directed to the ROADM outputs.

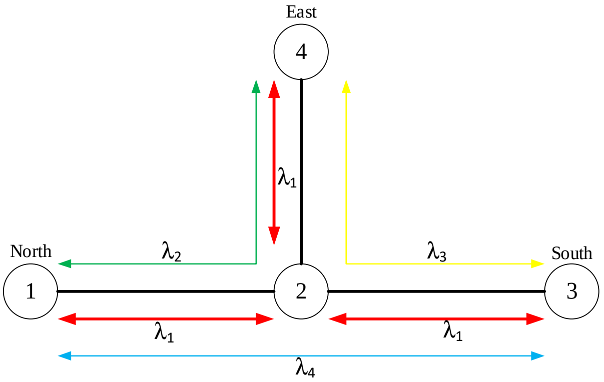


Fig. 3. Four-node star network with a full-mesh logical topology.

Fig. 4 represents the structure of the ROADM designated by node 2 in Fig. 3, a 3-degree CDC ROADM based on a R&S architecture, i.e., with WSSs both at its inputs and outputs, and with WSSs-based add/drop structures. The crosstalk generation inside the ROADM is also represented. From this figure, we can observe that all in-band crosstalk terms originated with wavelength λ1 are second order terms (identified with number 2). In this case, in each drop port, where wavelength λ1 is dropped, we find two in-band crosstalk terms, coming from the other two ROADM inputs. At the ROADM outputs, the output wavelength λ1 in each direction is impaired by four in-band crosstalk terms, two of them arising from the ROADM inputs and the other two are generated from the presence of wavelengths λ1 at the add section.

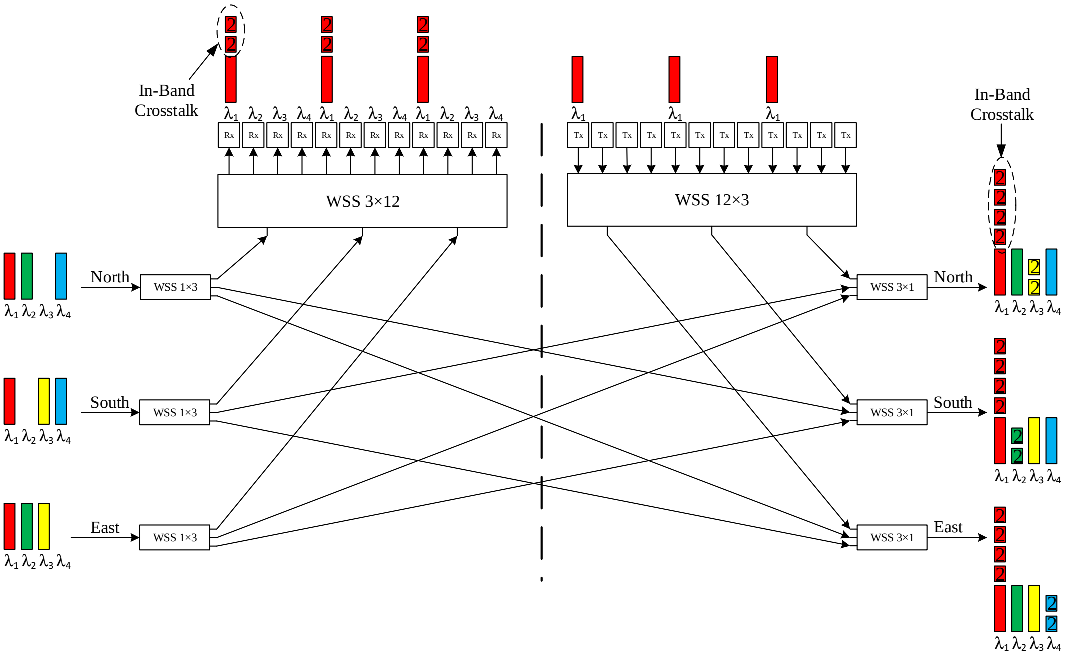


Fig. 4. Node 2 structure – a 3-degree CDC ROADM based on a R&S architecture with WSSs-based add/drop structures.

The conclusions taken from Fig. 4, for a 3-degree CDC ROADM, can be generalized for a *R*-degree ROADM. In Tables 1 and 2, the number of in-band crosstalk terms generated inside a *R*-degree C, CD and CDC ROADM with MCSs and WSSs-based add/drop structures, for both B&S (Table 1) and R&S (Table 2) architectures is presented. From Tables 1 and 2, we can conclude that, for a CDC ROADM, the WSS-based add/drop structures are the best choice in terms of minimising the in-band crosstalk generation. For both studied architectures, the interfering signals generated with these add/drop structures are mainly of second order. In summary, to minimises the crosstalk generation inside multi-degree CDC ROADMs, the R&S architecture with WSSs-based add/drop structures seems to provide the best solution.

Table 1. Number of in-band crosstalk terms generated inside a *R*-degree ROADM based on the B&S architecture.

	Drop ports		Outputs	
	1st order	2nd order	1st order	2nd order
C	-	-	$R-1$	-
CD	$R-1$	-	$2(R-1)$	-
CDC (MCSs)	$R-1$	-	$2(R-1)$	-
CDC (WSSs)	-	$R-1$	$R-1$	$R-1$

Table 2. Number of in-band crosstalk terms generated inside a *R*-degree ROADM based on the R&S architecture.

	Drop ports		Outputs	
	1st order	2nd order	1st order	2nd order
C	-	-	-	$R-1$
CD	$R-1$	-	$R-1$	$R-1$
CDC (MCSs)	$R-1$	-	$R-1$	$R-1$
CDC (WSSs)	-	$R-1$	-	$2(R-1)$

III. Physical Layer Impairments Impact

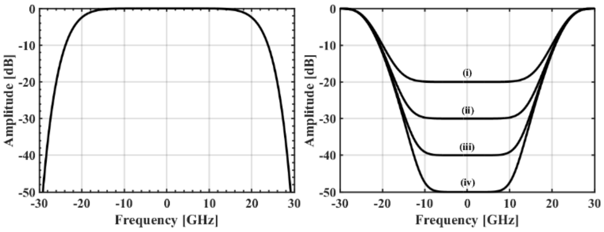
In this section, the impact of in-band crosstalk, optical filtering and ASE noise in a cascade of multi-degree CDC ROADMs based on the R&S architecture, the architecture that minimises the generation of in-band crosstalk, with MCSs and WSSs-based add/drop structures is studied. The main goal of this study is to investigate the maximum number of ROADMs that an optical signal can pass until the degradation of these PLIs causes an optical signal-to-noise ratio (OSNR) penalty higher than 1 dB. The OSNR penalty is the difference between the imposed OSNR in each optical amplifier, to reach a target bit error rate (BER) of 10^{-3}, with and without the PLIs. The signal referred in this work as the primary signal corresponds to the signal that is taken as a reference to study the PLIs. We consider a non-return-to-zero (NRZ) QPSK signal with 50-Gb/s in a single polarisation (which corresponds to 100-Gb/s in dual polarisation) as the primary signal.

We start, in subsection III.A, by characterising the optical filters used to model the ROADM components. This will permit to obtain the crosstalk level at the end of an optical network. In subsection III.B, the optical filtering impact on a cascade of CDC ROADMs considering only one amplification stage is studied. In subsection III.C, the impact of ASE noise is studied with optical amplifiers at every ROADM inputs and outputs.

A. Optical Filters used to Model the ROADM Components

We consider two types of optical filters to model the ROADM components, the passband $H_p(f)$ and the stopband $H_b(f)$ filters. The signals that pass through the ROADM components (e.g. WSSs) are filtered by the passband filter, while the signals that the ROADM component blocks are filtered by the stopband filter. The optical passband filter is modelled by a 4th order Super-Gaussian optical filter [11] with

−3 dB bandwidth (B_0) equal to 41 GHz, usually used for the 50 GHz channel spacing [5]. The optical stopband filter is modelled by the inversion of the optical passband filter and by setting the blocking amplitude, in dB, with B_0 equal to approximately 48 GHz. Fig. 5 shows the transfer functions of these filters, Fig. 5 (a) for the passband filter and Fig. 5 (b) for the stopband filter with different blocking amplitudes (i) −20 dB, (ii) −30 dB, (iii) −40 dB and (iv) −50 dB.



(a) (b)

Fig. 5. Transfer function of the (a) optical Super-Gaussian 4th order passband filter $H_p(f)$ and (b) optical stopband filters $H_b(f)$ with different blocking amplitudes (i) −20 dB (ii) −30 dB (iii) −40 dB and (iv) −50 dB.

With these passband and stopband filters, we can model the effect of the ROADM node on both the express and add/drop signals and also on the crosstalk signals. For the add/drop signals, considering the MCS structure, the signals pass through one passband filter, while with the WSS structure, the signals pass through two passband filters. The express signals are filtered by two passband filters, at the ROADM input and output WSSs. Regarding the crosstalk signals, second order terms pass through two stopband filters inside the ROADM, while first order terms pass through only one stopband filter.

Now, we can evaluate the crosstalk level at the end of an optical network composed by 32 cascaded CDC ROADMs with MCSs and WSSs-based add/drop structures, for several ROADM degrees and for the blocking amplitude of −20 dB, as shown in Table 3. Studies with lower blocking amplitudes (−50, −40 and −30 dB) exhibit crosstalk levels at the end of a cascade of 32 CDC ROADMs, that will lead to a negligible performance degradation. For a blocking amplitude of −20 dB, the total crosstalk level, as defined in [3], at the end of a network with 32 16-degree CDC ROADMs is −5.2 dB and −13.3 dB, respectively, with MCSs and WSSs add/drop structures.

Table 3. Total crosstalk level at the end of an optical network composed by 32 cascaded CDC ROADMs for a blocking amplitude of −20 dB.

	Total crosstalk level [dB]	
ROADM degree (R)	MCSs	WSSs
2-degree	−18.7	−35.4
4-degree	−13.3	−21.6
8-degree	−9.4	−16.3
16-degree	−5.2	−13.3

B. Impact of Optical Filtering and In-Band Crosstalk on a ROADM Cascade with Only One Amplification Stage

In this subsection, we study the impact of optical filtering and in-band crosstalk on a ROADM cascade with only one amplification stage at the end of the optical network, as shown in Fig. 6. The optical (de)multiplexers represented in Fig. 6 are for 50 GHz channel spacing and are modelled by the optical passband filter $H_p(f)$ represented in Fig. 5 (a).

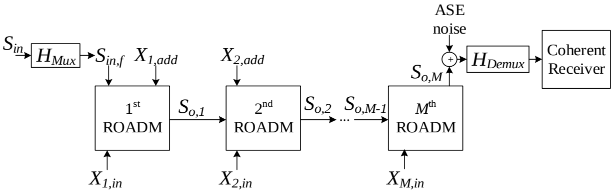


Fig. 6. Schematic model of an optical network composed by M ROADMs with only one amplification stage.

The signals represented by $X_{M,in}$ and $X_{M,add}$ are, respectively, the in-band crosstalk signals from the ROADM inputs and add section. The signal $S_{in,f}$ is the primary signal after passing through the optical multiplexer. The primary signal that appears at Mth ROADM output is called $S_{o,M}$, and the ASE noise is added to this signal. In this work, the ASE noise is considered to be an additive white Gaussian noise.

We start by studying the impact of the optical filtering due to the amplitude distortion introduced by the optical passband filters cascade without the in-band crosstalk impairment. Fig. 7 depicts the BER as a function of the required OSNR for a 50-Gb/s NRZ QPSK signal that passes through 2, 4, 8, 16 and 32 ROADM nodes. The OSNR difference between the case where the primary signal passes 2 nodes, our reference case, and the other cases, represents the OSNR penalty due to the optical filtering, which is estimated for a BER of 10^{-3}. This penalty is represented in Fig. 7 by δ_F for the case of 32 ROADM nodes (green curve) and is approximately 1.2 dB.

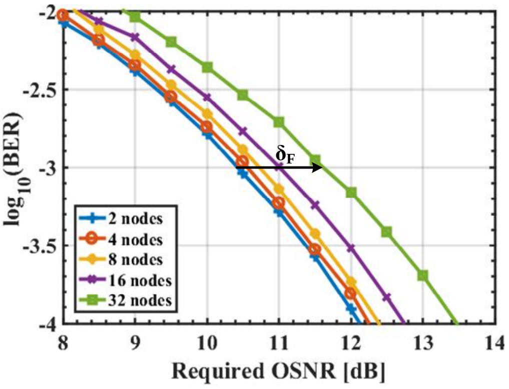
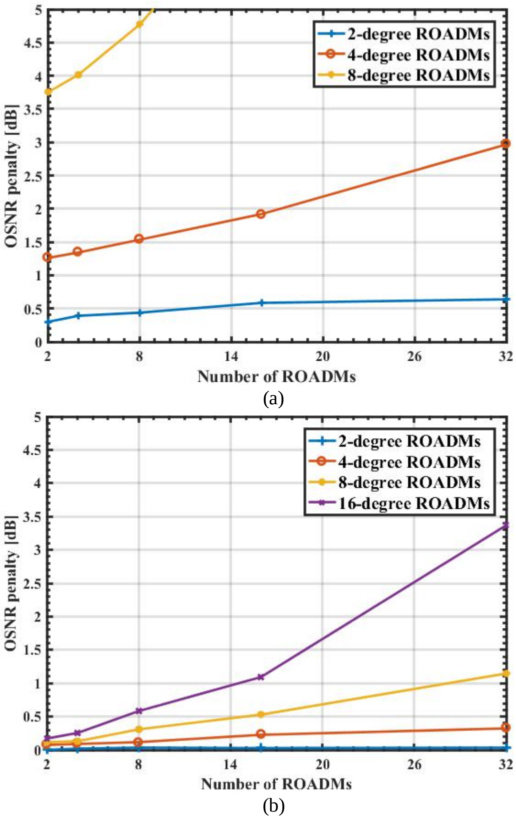


Fig. 7. BER as a function of the required OSNR for a 50-Gb/s NRZ QPSK signal that passes through a cascade of ROADM nodes.

After having evaluated the optical filtering penalty without the in-band crosstalk, we add the in-band crosstalk signals to our simulation model to estimate the OSNR penalty due to in--band crosstalk [12], considering a ROADM-based network. Fig. 8 depicts the OSNR penalty as a function of the number of ROADM nodes, with the ROADM degree as a parameter, considering a blocking amplitude of −20 dB and (a) MCSs and (b) WSSs-based add/drop structures. For 16-degree ROADMs and MCSs-based add/drop structures, the OSNR penalty is higher than 5 dB at the end of 2 cascaded ROADMs, and it is not represented in Fig. 8 (a). Note that, for a 4-degree ROADM with MCSs-based add/drop structures, the OSNR penalty due to in-band crosstalk after 2 nodes is about 1.3 dB. This OSNR penalty is higher than at the end of 32 ROADMs with 2, 4 and 8-degree with add/drop structures based on WSSs. When we have WSSs-based add/drop structures, the number of ROADMs associated with a 1 dB OSNR penalty is 15 for 16-degree ROADMs and 28 for 8-degree ROADMs. The 2 and 4-degree ROADMs provide penalties below 0.5 dB at the end of 32 nodes. For case with 4-degree (red curve in Fig. 8 (b)), the OSNR penalty is only 0.4 dB. On the other hand, in the same case but with add/drop structures based on MCSs, the OSNR penalty due to the in-band crosstalk is 3 dB.



(a)

(b)

Fig. 8. OSNR penalty as a function of the number of ROADMs nodes, with the ROADM degree as a parameter, for a blocking amplitude of −20 dB and with the add/drop structures based on (a) MCSs and (b) WSSs.

C. Impact of ASE Noise and In-Band Crosstalk on a ROADM Cascade with Amplification Stages at every ROADMs Inputs and Outputs

In this subsection, we analyse the impact of the ASE noise and in-band crosstalk on a network composed by a cascade of CDC ROADMs in a more realistic scenario. In this scenario, there are optical amplification stages at the inputs of all the ROADMs to compensate the path losses, and at the ROADMs outputs to compensate the losses inside the nodes [13], as depicted Fig. 9. In Fig. 9, the path of the primary signal, since it is added until it is dropped, is represented by the red line. Notice that, in the network represented in Fig. 9, an increase on the number of ROADMs, besides leading to a higher number of interferers, also leads to a substantial increase of the ASE noise power. To study this effect, we plot in Fig. 10 the total signal power evolution as a function of the number of 16-degree CDC ROADM nodes that the signal passes for three cases: 1) considering only the signal power evolution, 2) considering the signal power plus the ASE noise power and 3) considering the signal power plus ASE noise and in-band crosstalk powers. These results are plotted considering WSSs-based add/drop structures and a blocking amplitude of −20 dB. This figure shows that, the ASE noise power is by far superior to the in-band crosstalk power, as we can check by comparing the curves with crosses and diamonds.

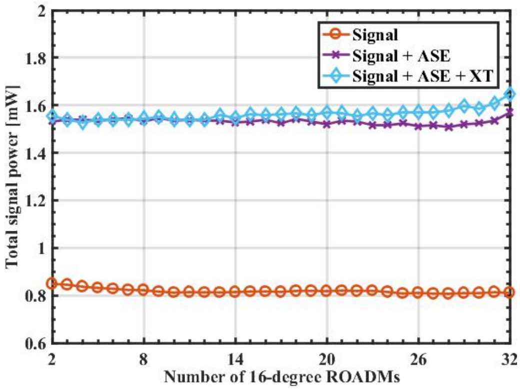


Fig. 10. Total signal power as a function of the number of 16-degree CDC ROADMs with WSSs-based add/drop structures.

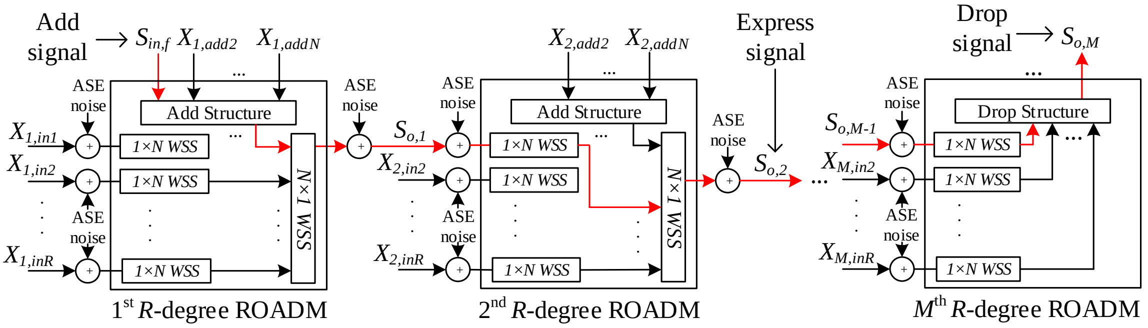
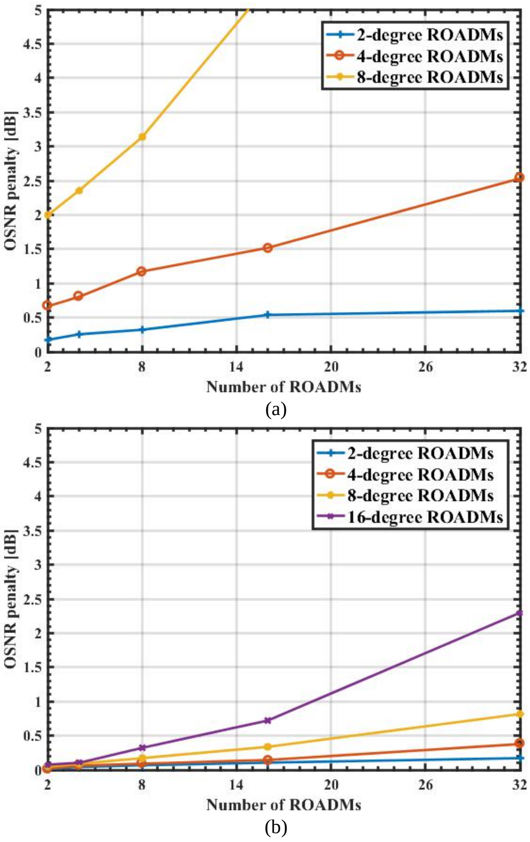


Fig 9. Schematic model of a cascade of *M* multi-degree CDC ROADM based on the R&S architecture with in-band crosstalk signals and ASE noise addition.

Fig. 11 depicts the OSNR penalty as a function of the number of ROADMs nodes for stopband filters with blocking amplitude of −20 dB and add/drop structures based on (a) MCSs and (b) WSSs in a network depicted in Fig. 9. By comparing the results depicted in Figs. 8 and 11, we can observe that the OSNR penalty due to the in-band crosstalk obtained in Fig.11 is lower than the penalty obtained in Fig. 8. For example, in subsection III.B, the number of CDC ROADMs with WSSs-based add/drop structures, that can be reached associated with a 1 dB OSNR penalty is 15, for 16-degree ROADMs, and 28, for 8-degree ROADMs. In this subsection, Fig. 11 (b) shows that, for 16-degree ROADMs, the maximum number of nodes reached is 19 nodes. For 8-degree ROADMs, a 1 dB OSNR penalty is not reached at the end of 32 nodes.



(a)

(b)

Fig. 11. OSNR penalty as a function of the number of ROADMs for a blocking amplitude of −20 dB and add/drop structures based on (a) MCSs and (b) WSSs with amplification at every ROADMs inputs and outputs.

IV. Conclusions

In this work, the performance of an optical network based on multi-degree CDC ROADMs impaired by in-band crosstalk, ASE noise and optical filtering has been investigated considering a 100-Gb/s QPSK signal for the fixed grid. The ROADM model considers different architectures and add/drop structures.

It is shown that the R&S architecture is the most robust architecture in terms of the in-band crosstalk generated inside multi-degree CDC ROADMs. With ROADM add/drop structures based on WSSs, for the R&S architecture, we only found second order in-band crosstalk terms either at ROADM outputs and drop ports. For the B&S architecture first order in--band crosstalk terms appear at the ROADM output.

Our results have shown that, for a BER of 10^{-3}, the OSNR penalty due to the optical filtering, without the in-band crosstalk effect, is approximately 1.2 dB when the optical signal passes through 32 CDC ROADMs. In a more realistic scenario, with amplification at all ROADMs inputs and outputs, the system degradation is mainly due to the ASE noise accumulation, making the in-band crosstalk impact lower than in networks with one amplification stage. In this realistic scenario, for CDC ROADMs with add/drop structures based on WSSs, the number of cascaded ROADMs nodes that leads to a 1 dB OSNR penalty degradation is 19, for 16-degree ROADMs. For 8-degree ROADMs, the OSNR penalty does not reach 1 dB at the end of a network with 32 nodes.

Acknowledgments

This work was supported by Fundação para a Ciência e Tecnologia (FCT) of Portugal within the project UID/EEA/50008/2013.

References

[1] K. Roberts *et al.*, “High capacity transport-100G and beyond,” *J. Lightw. Technol.*, vol. 33, no. 3, pp. 563-578, Feb. 1, 2015.

[2] S. Gringeri, *et. al.*, "Flexible architectures for optical transport nodes and networks," *IEEE Commun. Mag.*, vol. 48, no. 7, pp. 40-50, Jul. 2010.

[3] L. Cancela, *et. al.*, "Analytical tools for evaluating the impact of in-band crosstalk in DP-QPSK signals," *NOC 2016*, Jun. 2016.

[4] S. Tibuleac and M. Filer, "Transmission Impairments in DWDM Networks With Reconfigurable Optical Add-Drop Multiplexers," *J. Lightw. Technol.*, vol. 28, no. 4, pp. 557-598, Feb. 15, 2010.

[5] M. Filer and S. Tibuleac, "Generalized weighted crosstalk for DWDM systems with cascaded wavelength-selective switches," *Opt. Exp.*, vol. 20, no. 16, pp. 17620-17631, Jul. 2012.

[6] M. Filer and S. Tibuleac, "N-degree ROADM architecture comparison: Broadcast-and-select versus route-and-select in 120 Gb/s DP-QPSK transmission systems,” *OFC 2014*, Mar. 2014.

[7] H. Yang, *et. al.* "Low-cost CDC ROADM architecture based on stacked wavelength selective switches," *J. Opt. Commun. Netw.*, vol. 9, no. 5, pp. 375-384, May 2017.

[8] M. Seimetz and C. Weinert, "Options, Feasibility, and Availability of 2 × 4 90º Hybrids for Coherent Optical Systems," in *J. Lightw. Technol.*, vol. 24, no. 3, pp. 1317-1322, Mar. 2006.

[9] M. Seimetz, *High-Order Modulations for Optical Fiber Transmission*, T. Rhodes, Ed. Atlanta: Springer, 2009.

[10] S. Yao, *et. al.* "Performance comparison for NRZ, RZ, and CSRZ modulation formats in RS-DBS Nyquist WDM system," *J. Opt. Commun. Netw.*, vol. 6, no. 4, pp. 355-361, Apr. 2014.

[11] C. Pulikkaseril, *et. al.* “Spectral modeling of channel band shapes in wavelength selective switches,” *Opt. Exp.*, vol. 19, pp. 8458–8470, Apr. 2011.

[12] B. Pinheiro, *et. al.* "Impact of in-band crosstalk signals with different duty-cycles in M-QAM optical coherent receivers,” *NOC 2015*, pp. 1-6, Jul. 2015

[13] T. Zami, "Current and future flexible wavelength routing cross-connects," *Bell Labs Tech. J.*, vol. 18, pp. 23-38, Dec. 2013.

Scalable Deterministic Scheduling for WDM Slot Switching Xhaul with Zero-Jitter

Bogdan Uscumlic[1], Dominique Chiaroni[1], Brice Leclerc[1], Thierry Zami[2], Annie Gravey[3], Philippe Gravey[3], Michel Morvan[3], Dominique Barth[4] and Djamel Amar[3]

[1]Nokia Bell Labs, Paris Saclay, France, *firstname.lastname*@nokia-bell-labs.com
[2]Nokia, Paris Saclay, France, thierry.zami@asn.com
[3]IMT Atlantique Bretagne-Pays de la Loire, Brest, France, *firstname.lastname*@imt-atlantique.fr
[4]DAVID, Université de Versailles Saint-Quentin-en-Yvelines, Versailles, France, dominique.barth@uvsq.fr

***Abstract*— A low-cost WDM slot switching "N-GREEN" network is studied for the Xhaul application. We assess the impact of inter-slot intervals on the jitter in N-GREEN and propose a deterministic scheduler ensuring a zero-jitter performance as needed by CPRI traffic. The scheduling is then implemented in the form of an Integer Linear Program and as a scalable heuristic, and these tools are used for the evaluation of the scheduler performances. The results show important savings and improvements in cost, energy consumption, latency and jitter using N-GREEN w.r.t. state-of-the-art Ethernet Xhaul.**

Keywords*— *5G; Xhaul; zero-jitter; deterministic scheduling; scalability; WDM slot switching.

I. Introduction

The Ethernet-based fronthaul scheduling problem has recently become a topic of many research groups. Although Ethernet is a mature technology, Ethernet-based fronthaul exploiting statistical multiplexing has difficulties to support synchronization constraints, low jitter (<65ns) and latency (<100 µs) for the CPRI (Common Public Radio Interface) traffic [1] in this network segment.

Recently, a new WDM slot switching technology called WDM Slotted Add/Drop Multiplexer (WSADM), investigated in the ANR N-GREEN project, has been proposed for the 5G fronthaul/Xhaul networks [2], [3]. The WSADM technology of the N-GREEN project exploits the WDM transparency for the transit traffic to lower the cost of optical components and adopt off-the-shelf devices. The WDM slot technology (in which the data is carried simultaneously over 10 wavelengths [2]) has the potential to provide a low-cost [3] and performant fronthaul/Xhaul. However, the problem of the deterministic scheduling of isochronous (CPRI) traffic over a time slotted ring (such as adopted in N-GREEN), where each time slot starts after a fixed size inter-slot interval ("guard time", T_G), has not yet been solved. Furthermore, the scalable scheduling method has not yet been proposed. Indeed, the scalability of the scheduling mechanism is needed to reduce the network reconfiguration time, since the scheduler will be implemented at the SDN (software defined networking) controller, enabling the logically centralized network control.

Our main contributions are as follows. For a first time, the impact of guard time on the jitter performance is considered in the scope of deterministic scheduling. Then, a deterministic scheduler with zero-jitter is proposed for N-GREEN. For this scheduler, a solution is proposed in the form of Integer Linear Program (ILP), enabling to achieve the scheduling at optimal network cost. Next, a heuristic algorithm based on the greedy approach is designed as an alternative scalable solution for the same scheduling. Finally, by using the previously developed tools, we evaluate the cost, jitter and latency advantages of the N-GREEN technology when compared to a state-of-the-art Ethernet Xhaul.

The remainder of the paper is organized as follows. In Section II we present the N-GREEN network and node architecture and define the properties of the WDM transponders (WDM-TRX) used for the network operation. Section III defines the scheduling solution that we propose for the WDM slot switching Xhaul network. In Section IV, we detail the mathematical model (based on ILP) implementing the previous scheduling algorithm. Section V introduces the greedy algorithm for the scalable scheduling, while Section VI provides detailed numerical results, evaluating the cost and the power consumption of the N-GREEN network. Finally, concluding remarks are provided in Section VII.

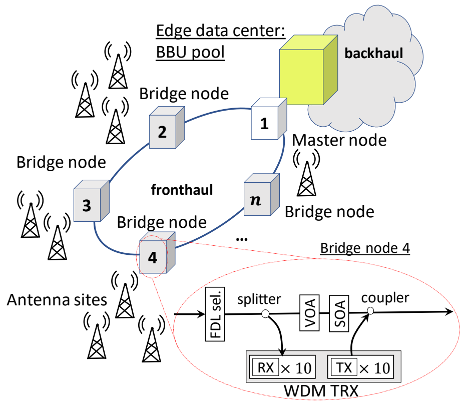


Fig. 1. N-GREEN network architecture (example of the application in the fronthaul)

978-3-903176-07-2 © 2018 IFIP

II. N-GREEN NETWORK ARCHITECTURE

An example of the N-GREEN ring in the fronthaul and N-GREEN's node architecture is illustrated in Fig. 1. Network is composed of n nodes, one of which is the master node, having a role of hub and interconnecting the ring with the edge cloud. The other nodes are called the "bridge nodes".

The N-GREEN node is composed of: a) splitters and couplers for the traffic reception/extraction via a set of WDM-TRX transponders operating at 100 Gbit/s (a single WDM-TRX operates physically with 10 single-wavelength receivers, RX, and 10 single-wavelength transmitters, TX; each RX and TX operate at 10 Gbit/s); b) variable optical attenuator (VOA) and semiconductor optical amplifier (SOA) for WDM slot blocking/bypassing and c) Fiber Delay Line (FDL) selection box, for adjusting the propagation time of the transit traffic to enable the processing of the control channel [2].

N-GREEN ring of a dozen of nodes operates on a time slot basis (with time slot duration of T_S=1 µs). The time slots carry the "WDM slots" which encapsulate and transport the client data simultaneously over 10 wavelengths. Note that the privileged solution for N-GREEN consists in using the WDM-TRX transponders (case called: "CONF. 1" as in Fig. 1), but to validate the techno-economic interest of N-GREEN, we also consider the case "CONF. 2" where the TRX transponders at 10 Gbit/s are used (1 TRX is composed of 1 single-wavelength RX and of 1 single-wavelength TX). The inserted WDM slots are separated by a guard time of T_G=50 ns. Regarding the control plane, the network uses a dedicated wavelength channel for the node and network synchronization and for the transport of the control information, that is processed at each network node. The master node is connected to the SDN controller.

For statistical scheduling, WDM slot switching has been already demonstrated as highly cost efficient when compared to an electronic Ethernet technology or to a single channel approach [2]. Indeed, WDM slots allow to reduce the costs of the active components, either immediately for the optical gates that can serve for the entire ring bandwidth, either by benefiting from the savings resulting from the integrations of WDM-TRX transponders. Furthermore, N-GREEN network offers the same functionalities as Ethernet (through a processing of all the bus), while being in line with the recent Datacom technology evolution (towards WDM-TRX).

III. ZERO-JITTER AND DETERMINISTIC LATENCY SCHEDULING FOR N-GREEN XHAUL

The scheduling of CPRI (or isochronous) traffic in N-GREEN consists in: a) aggregating different flows over the same transmission resource (wavelength or waveband) so that the resource use is maximized; b) properly positioning the CPRI packets within subsequent time slots to avoid any jitter and latency due to the packet insertion process and c) including the perturbation caused by the guard time in the jitter guarantees.

To address the scheduling problem, we first note that the positions of the scheduled subsequent CPRI packets in the N-GREEN time slots will vary from slot to slot, since the period of the CPRI flow is not necessarily the same as the time slot duration T_S. This means that the transmission resource allocated to the transport of CPRI flows needs to be "continuous". The continuity can be assured either by allocating a physical channel to this transport (as seen so far) or by allocating a virtual channel to it (in CONF. 1 only). For instance, to ensure a 10 Gbit/s transport, we could either allocate a wavelength of 10 Gbit/s to it, or use a periodic "temporal slice" of sufficient capacity, e.g. we could use a 100 Gbit/s WDM TRX during 1/10 of the time. For a physical channel, the jitter introduced by the guard time T_G (and limited to the value of T_G), in a general case, is experienced when inserting the CPRI traffic into the optical medium. For a virtual channel, the guard time does not impact the scheduling (as it is not perceived during the emission by the source), and scheduling with zero-jitter is then achieved.

The scheduling that we propose is based on the optimal solution in form of ILP and on the greedy algorithm, defined over a scheduling cycle of size S, and minimizing the network cost (equal to the cost of transponders and channels, transponders being the most expensive components). To precisely account for the CPRI packet position within a time slot, we introduce a notion of "subslot", i.e. we suppose that each time slot is composed of exactly L equally sized "subslots" (L is a given parameter). Let us suppose that S=$m \cdot L$, where m is the number of time slots after which the scheduling is repeated cyclically by all network nodes. Next, for CPRI flow d, exchanged between nodes i and j, we define the parameters: $\alpha_{(i,j)}^{d}$- the number of subsequent subslots corresponding to a transmission duration of a single CPRI packet and $\beta_{(i,j)}^{d}$ – the number of subsequent subslots between the consecutive CPRI packets (both calculated from L), and $F_{(i,j)}^{d}$ – the number of times the CPRI packets have to be scheduled in a single scheduler cycle S (calculated from S). To get S, we define it as the *least common multiple* of numbers $\alpha_{(i_1,j_1)}^{d_1}+\beta_{(i_1,j_1)}^{d_1}$, $\alpha_{(i_2,j_2)}^{d_2}+\beta_{(i_2,j_2)}^{d_2}$, …, L. Finally, the value of S allows us to find m.

Note that the previous scheduling can be used also for CoE (CPRI over Ethernet) traffic. In the following, the scheduling is realized by two mathematical models: 1) as an optimal ILP program, and 2) as a scalable greedy algorithm.

IV. ILP MODEL FOR THE COST OPTIMIZED SCHEDULING

This section introduces the ILP model for calculating the slot allocation according to the previous scheduling mechanism. The resulting scheduling is cost optimized, and has the minimum cost of channel interfaces and channels, needed to support all the traffic flows in the network. We define channel interface either as a transponder (for physical channels) or as a duration of a single temporal slice of the transponder (for virtual channel). For instance, in CONF. 1 for virtual channels, the cost of channel interface is the same as the cost of a single virtual channel, and is lower than the cost of WDM-TRX. In CONF. 1 for waveband channels and in CONF. 2, the cost of channel interface is the same as WDM-TRX and TRX cost, respectively. The input parameter to the ILP model is also the traffic matrix, defining the number of slots needed for each traffic flow. The output of the ILP model is the scheduling algorithm that shall be applied by each network node, and the network configuration, expressed in the number of the required channel interfaces and channels at each node and in the network. The full list of the input parameters for the ILP model is provided in Tab. I. For instance, the scheduling cycle size and the cost parameters are defined in this table.

Next, in Tab. II we can see all the output variables that are used for the integer linear program. Some variables are not included in Tab. II, since they are auxiliary.

Tab. III provides the list of the linear constraints that build the ILP model. Eq. (1) is the objective function. This function minimizes the overall cost of the N-GREEN ring, i.e. the cost of the channel interfaces and channels required in the ring. Equation (2) is the traffic-load constraint, ensuring that right amount of ring capacity is allocated to each isochronous traffic flow. Constraint (3) ensures the periodic slot allocation for the isochronous traffic demands. The constraint (4) implements the logical "IF-THEN-ELSE" condition (connecting the variables $p_q^{(i,j),d,s}$ and $B_q^{(i,j),d,s}$), so it uses the auxiliary variables and constants. Each slot can be allocated only once, per each link and channel, which is ensured by constraint (5). Constraint (6) ensures that sufficient number of transmitters and receivers are allocated at each network node.

TABLE I. INPUT PARAMETERS

Input Parameters	Definition
$G(V,E)$	A directed graph representing the unidirectional N-GREEN ring, where V is the set of nodes, E is the set of (unidirectional) links; ($\lvert V\rvert = n$ in Fig. 1, where $\lvert.\rvert$ is the cardinality notation);
Q	Set of channels (wavelengths/wavebands or virtual channels);
S	Size of the scheduling cycle (in number of subslots);
$\pi_{(i,j)}$	Set of the links in the ring belonging to the routing path between the nodes i and j;
$D_{(i,j)}$	Set of CPRI (isochronous) traffic demands d between nodes i and j (D_{max} – the maximum number of demands between any source-destination pair (i,j));
$\alpha_{(i,j)}^d$	The number of subsequent subslots corresponding to a transmission duration of a single CPRI packet for the demand d between source-destination pair (i,j);
$\beta_{(i,j)}^d$	The number of subsequent subslots between the consecutive CPRI packets for the demand d between source-destination pair (i,j);
$F_{(i,j)}^d$	The number of times the CPRI packets (for the demand d between source-destination pair (i,j)) are to be scheduled in a single scheduler cycle S ;
C_t	Cost of channel interfaces;
C_q	Cost of a wavelength (CONF. 2), of a waveband or of a virtual channel (CONF. 1);
M_1 , M_2	Large constants.

TABLE II. OUTPUT VARIABLES

Output Variables	Definition
$B_q^{(i,j),d,s}$, $p_q^{(i,j),d,s}$	Binaries, equal to 1 if the demand d between (i,j) is routed over channel q starting from/by using (respectively) the subslot s, and equal to 0 otherwise;
t_q^i, r_q^i, u_q^i	Binaries, equal to 1 if transmitter, receiver, channel interface (respectively) at channel q is used at node i for the transport of traffic, and eq. to 0 otherwise;
y_q	Binary, equal to 1 if channel q is used in the ring for the transport of traffic, and equal to 0 otherwise.

Finally, the constraint (7) ensures the allocation of the correct number of channel interfaces at each node and channels in the network.

TABLE III. ILP FORMULATION

No.	Constraint Definition
(1)	$Min\left(\sum_{i\in V}\sum_{q=1}^{Q} C_t u_q^i + \sum_{q=1}^{Q} C_q y_q \ \right)$
(2)	$\sum_{q=1}^{Q}\sum_{s=1}^{S} B_q^{(i,j),d,s} = F_{(i,j)}^d \, , \forall (i,j) \in V^2, \forall d \in D_{(i,j)}$
(3)	$B_q^{(i,j),d,s_1}{=}B_q^{(i,j),d,s_2}, \forall q \in Q, \forall (i,j) \in V^2, \forall d \in D_{(i,j)}, (\forall s_1, s_2)\begin{pmatrix} 0 \le s_1, s_2 \le S \quad \wedge \\ s_2 \equiv (s_1 + k \cdot (\alpha_{(i,j)}^d + \beta_{(i,j)}^d)) mod\ S \end{pmatrix} (\forall k, F_{(i,j)}^d > k \ge 0\ )$
(4)	$0 \le p_q^{(i,j),d,s_2} - B_q^{(i,j),d,s_1} + M_1 \cdot z_q^{(i,j),d,s_1}, B_q^{(i,j),d,s_1} \le 0 + M_2 \cdot (1 - z_q^{(i,j),d,s_1}), \forall q \in Q, \forall (i,j) \in V^2, \forall d \in D_{(i,j)}, (\forall s_1, s_2)\begin{pmatrix} 0 \le s_1, s_2 \le S \quad \wedge \\ s_2 \equiv (s_1 + k) mod\ S \end{pmatrix} (\forall k, \alpha_{(i,j)}^d > k \ge 0\ )$
(5)	$\sum_{d\in D_{(i,j)}} \sum_{(i,j):l\in\pi_{(i,j)}} p_q^{(i,j),d,s} \le 1\, , \ \forall q \in Q, \forall s \in S \ \ , \forall l \in E$
(6)	$\sum_{j\in V:l\in\pi_{(i,j)}} \sum_{\substack{d:\ i=source(d),\\ d\in D_{(i,j)}}} \sum_{s=1}^{S} p_q^{(i,j),d,s} \le S \ \ \cdot t_q^{i,l},$ $\forall i \in V, \forall q \in Q, \forall l \in E;\ \sum_{l\in E} t_q^{i,l} \le \lvert E\rvert t_q^i,$ $\sum_{j\in V} \sum_{\substack{d:\ i=dest(d),\\ d\in D_{(i,j)}}} \sum_{s=1}^{S} p_q^{(j,i),d,s} \le S \ \ \cdot r_q^i,$ $\forall i \in V, \forall q \in Q$
(7)	$t_q^i + r_q^i \le 2 \cdot u_q^i, \forall i \in V, \forall q \in Q;$ $\sum_{i\in V} t_q^{i,l} \le 1, \forall q \in Q, \forall l \in E; \sum_{i\in V} u_q^i \le \lvert V\rvert \cdot y_q\ ,\ \forall q \in Q$

V. HEURISTIC SOLUTION FOR THE SCALABLE SCHEDULING IN N-GREEN XHAUL

The cost-optimized scheduling described in the previous section consumes important computing time (>> 1 minute) when the ILP program is implemented, but no formal proof about its computational complexity is available yet. For the real network implementation of the scheduling, it would be excellent to have a faster algorithm, that is adapted to a fast execution at the SDN controller. With goal of addressing this problem, in the current section we propose a greedy algorithm, that addresses the same scheduling problem that is already introduced, and that has the same input parameters and output variables as the ILP (with exception of the constants M_1 , M_2).

The name of the proposed algorithm is "Greedy Source Cost Minimization" (GSCM), and its pseudocode is provided in Tab. IV. The algorithm name comes from its intention to ensure the minimization of the number of channel interfaces (transponders) and consequently the cost of each source in the network. The "group of flows" in the algorithm is defined as a set of flows generated by the same source. In steps 2-6 of the algorithm, the group of flows are taken in the decreasing order and routed according to the first-fit channel assignment method. Next, in steps 7-13, the algorithm applies the same procedure, but this time for each network node separately, which has shown to have cost saving advantages for more centralized traffic matrices (like "hub and spoke" scheme, as introduced in the following section). Finally, in the steps 14-16, the algorithm GSCM chooses its best solution.

TABLE IV. PSEUDOCODE OF THE ALGORITHM: GREEDY SOURCE COST MINIMIZATION (GSCM)

```
1: **Input:** the same as for the ILP model (with exception of the constants $M_1$, $M_2$ which are not needed)
2: Sort the groups of flows in order of decreasing traffic rate
3: for each group of flows taken in this order do
4:     First-fit channel assignment of the group of flows, by always assigning first the resources (channel interfaces and channels) to the highest flow in the group of flows
5: end for
6: Compute the overall network cost $C_{TOT}$
7: for each network node A
8:     Sort the groups of flows exchanged between nodes A and its destinations in order of decreasing traffic rate
9:     for each group of flows taken in this order do
10:        First-fit channel assignment of the group of flows, by always assigning first the resources to the highest flow in the group of flows
11:    end for
12: end for
13: Compute the new overall network cost $C_{TOT}'$
14: if ($C_{TOT}' < C_{TOT}$ ) then
15: Accept the new design
16: end
```

VI. NUMERICAL RESULTS

In this section, we report the simulation results obtained by using both the ILP and GSCM tools, to estimate the cost and energy consumption of the N-GREEN network.

Regarding the traffic matrix, CPRI or isochronous traffic flows, for the packet size of 1250 bytes and 10 Gbit/s channel capacity, are chosen from the sets $A_1 = \{2.5, 5, 10\}$ Gbit/s [1] and $A_2 = \{1.67, 3.33, 6.67\}$ Gbit/s, that have different basic periods. Under these assumptions, for different simulations, we consider one of the following four traffic scenarios:

a) Scenario 1: The traffic profile is "hub-and-spoke" (more precisely we suppose that each bridge node communicates with the master node, and vice versa, and that the amount of traffic exchanged in both directions in this communication is the same). Next, $m = 4, L = 1, S = 4$. Each bridge node sends the traffic from set A_1 (to the master node), resulting in a uniform and symmetric traffic. b) Scenario 2: The traffic profile is hub-and-spoke, with $m = 4, L = 3, S = 12$. Each bridge nodes sends three randomly chosen flows from the set $A_1 \cup A_2$. c) Scenario 3: The source, destination and flow type are all randomly distributed. Each bridge nodes sends three randomly chosen flows from the set $A_1 \cup A_2$. We suppose that $m = 4, L = 3, S = 12$. d) Scenario 4: The traffic profile is hub-and-spoke, with $m = 4, L = 3, S = 12$. Each bridge nodes sends three randomly chosen flows from the set $A_1 \cup A_2$, for different ring size n.

A. *The results obtained by using the ILP model*

We implement the ILP model in IBM CPLEX software (for $n = 6, C_t = 1$ [a.u.] for 10 Gbit/s channel interfaces (CONF. 1 or 2), $C_q << C_t$, i.e. $C_q = 0.1$ [a.u.]) and report the optimal simulation results in Figs. 2-5, for Scenarios 1 and 2. In all the simulations, the cost C_q corresponds either to the cost of a wavelength or of a virtual channel. Indeed, for such channels, the value of cost C_q is the same, since the equivalent capacity of a wavelength or of a virtual channel is the same and equal to 10 Gbit/s. Finally, in all scenarios in this work, the simulations are run for the gradually increasing traffic, by adding to the traffic matrix in each step the traffic contribution (the sent and received traffic) for the next bridge node in the network, until all the bridge nodes have been accounted for.

Fig. 2 shows the number of channel interfaces and channels in the network, for Scenarios 1 and 2. The number of channel interfaces is higher than number of channels for both scenarios, while Scenario 2 is more expensive. For Scenario 2, the traffic is random, resulting in a non-linear increase of the number of channel interfaces and channels. The channel occupancy (of physical channels, i.e. wavelengths or virtual channels) in N-GREEN is illustrated in Fig. 3. The channel occupancy reaches high values (≈90%) meaning that the N-GREEN network with the proposed scheduler is highly efficient in the resource use.

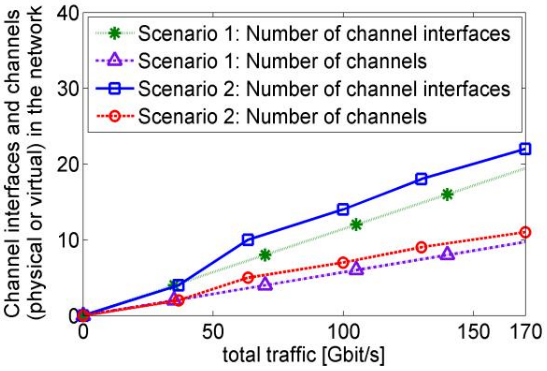


Fig. 2. Number of channel interfaces and channels (for Scenarios 1 and 2)

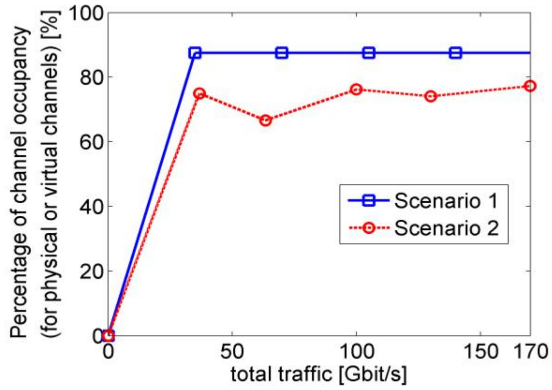


Fig. 3. Channel occupancy [%] in N-GREEN (for Scenarios 1 and 2)

The number of transponders in Scenario 2 for an equivalent Ethernet ring (with TRXs at 10 Gbit/s and first-fit wavelength allocation) and N-GREEN are compared in Fig. 4. For N-GREEN we consider two cases: 1) CONF. 2, i.e. nodes equipped with several TRX at 10 Gbit/s (with physical channels, i.e. with single wavelength per TRX) and 2) CONF.

1, i.e. nodes equipped with several WDM-TRX at 100 Gbit/s (supporting 10 virtual channels at 10 Gbit/s). Fig. 4 shows the savings of up to 9 times savings when the WDM TRX are used.

Next, the transponder cost is compared in Fig. 5, for the following assumptions:

1) Since burst mode TRX operation at 10 Gbit/s results in a small additional logic [4], the cost of TRX in Ethernet and N-GREEN (CONF. 2, physical channels at 10 Gbit/s) is approximately the same and equal to $C_t = 1$ [a.u.];

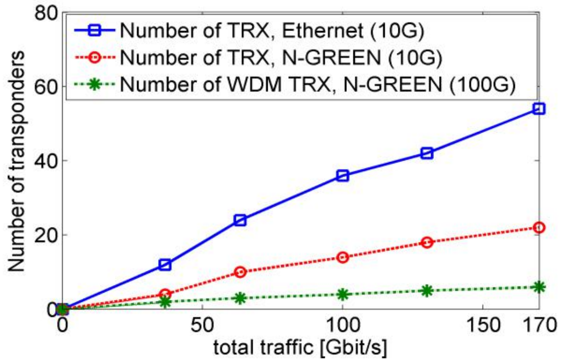


Fig. 4. Number of transponders for Scenario 2. Results show potential savings of N-GREEN vs Ethernet (up to 9 times) in number of transponders.

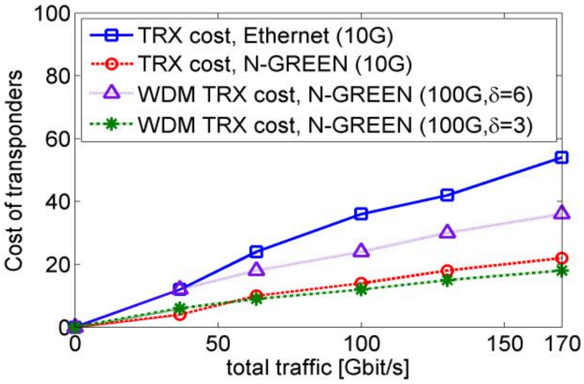


Fig. 5. Transponder cost for Scenario 2. Results show potential savings of N-GREEN vs Ethernet in the cost of transponders (savings up to 33% when WDM-TRX are used).

TABLE V. COMPARISON BETWEEN N-GREEN AND ETHERNET FRONTHAUL

	N-GREEN	Ethernet
Jitter at insertion	0 or $\leq T_G$ (for physical channels)	~1μs [1]
Jitter in transit	0	Depends on traffic load and network size
Latency at insertion & transit	Fixed	

2) The cost of WDM-TRX (CONF. 1, virtual channels) is supposed to be $\delta \cdot C_t$ (where parameter $\delta \in$[1,10]).

Fig. 5 shows important savings of N-GREEN in transponder cost w.r.t. Ethernet for physical channels. For virtual channels, the savings up to 33% are achieved for $\delta \leq 6$, which seems achievable since the cost of a single 10 Gbit/s TRX is ≈250 $ [5] and the cost of WDM-TRX is expected to be ≤6·250 =1500$ thanks to the laser integration design of these devices.

Finally, Tab. V compares the sources of jitter and latency in N-GREEN and Ethernet. While variable in Ethernet [1], latency is fixed in N-GREEN in transit and at the insertion, and the jitter is different than zero (and limited to T_G) only for the physical channel scheduling in N-GREEN. When channel is virtual the jitter is equal to zero.

B. *The validation of the GSCM algorithm*

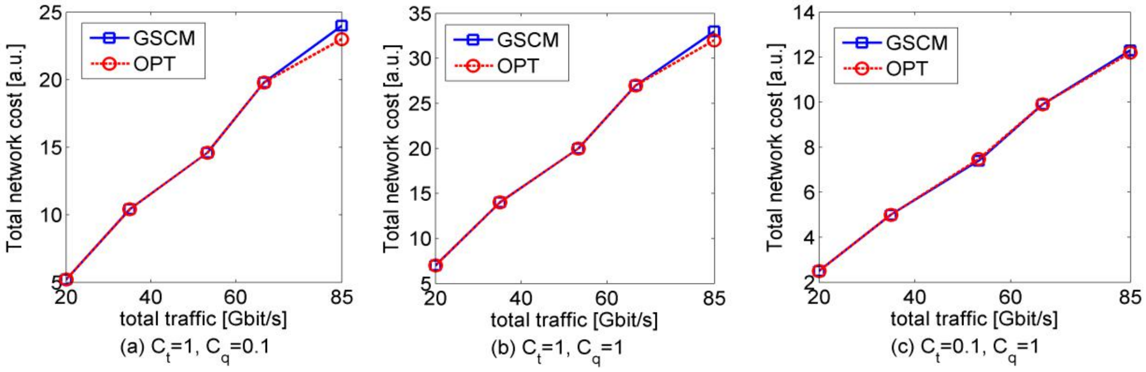
To validate the GSCM algorithm, we compare the optimal solution (OPT) calculated by the ILP model, and the solution found by GSCM algorithm, in a randomly generated Scenario 3 (to test the GSCM performance in the most general case). The results are presented in Fig. 6. The simulations are performed for different value of the cost ratio between the cost of channel interface and the cost of channel. The following cost assumptions are considered: 1) $C_t = 1, C_q = 0.1$, 2) $C_t = 1, C_q = 1$, and 3) $C_t = 0.1, C_q = 1$. The results suggest that GSCM obtains excellent performances on the selected random traffic scenario, and its results are very close to the optimal solution. The highest discrepancy from the optimal solution was less than 5%, and this performance was not affected by the change of the cost ratio of channel interface and channels.

We have also compared the solution found by the GSCM and ILP model on the hub-and-spoke Scenario 2, that is centralized, and 0% error of the GSCM (w.r.t. the optimal solution) has been observed. Such good results of GSCM algorithm can be explained by the fact that the channels for transport of the isochronous traffic need to be allocated per source, to preserve the "continuity" of the channel, as previously discussed, which allows to eliminate or to limit the jitter. Because of this, different sources are not allowed to use the same channel over the same links, which allows good results to the heuristics based on the greedy approach in optimizing the source cost.

C. *The results obtained by using the GSCM algorithm*

The proposed GSCM is scalable. Indeed, its complexity is limited to $O(n^{3}{}_{D_{max}})$, i.e. this is a polynomial time algorithm. In the current section, we exploit the algorithm GSCM to compare the cost and energy consumption of N-GREEN and Ethernet for a greater and a realistic size of the N-GREEN network of up to 10 nodes.

The simulations are performed for Scenario 4, and for different rings sizes. The transponder cost comparison between N-GREEN and Ethernet is presented in Fig. 7. From the figure, we can see that the savings of N-GREEN w.r.t. Ethernet increase with the increase of ring size. When the network is sufficiently loaded, even if the ring size is not large ($n = 6$), N-GREEN network costs much less than Ethernet, measured in transponder cost. If the integration technology used for production of the WDM-TRX transponders enables the cost reduction of these devices of $\delta \leq 3$, the potential savings of costs go to more than 6 times, obtained for $n = 10$. For $\delta \leq 6$,



(a) C_t=1, C_q=0.1 (b) C_t=1, C_q=1 (c) C_t=0.1, C_q=1

Fig. 6. The validation of the algorithm GSCM in Scenario 3. Results show high accuracy of the GSCM algorithm.

the savings are more than 3 times (result not shown in the figure), etc.

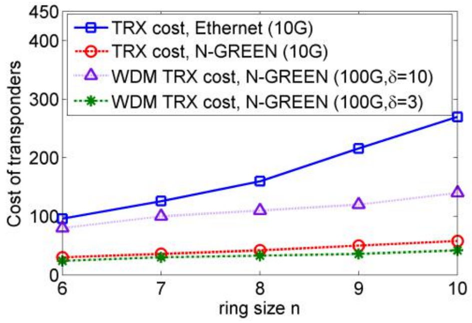


Fig. 7. Transponder cost for different ring sizes in Scenario 4 (results of GSCM algorithm). Results show high potential savings of N-GREEN vs Ethernet in the cost of transponders (more than 3 times when WDM-TRX are used)

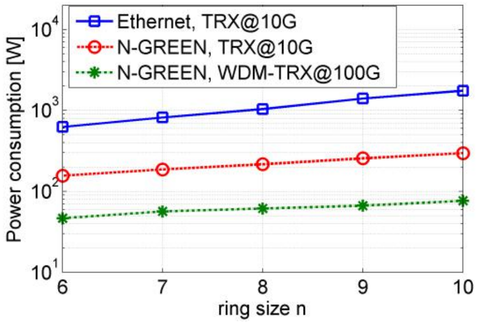


Fig. 8. Power consumption of N-GREEN vs Ethernet in Scenario 4 (results of GSCM algorithm). Results show high potential savings of NGREEN vs Ethernet (up to 10 times) in power consumption.

Fig. 8 shows the power consumption performance of N-GREEN and Ethernet, in Scenario 4. For Ethernet, we suppose the consumption of each SFP+ 10 Gbit/s TRX to be 1.5 W [6] and the switching fabric consumption to be ≈5W/port (calculated from [7]). For N-GREEN, the consumption of SOA is 1.5W, while 5W is the estimated consumption of the electronic part of the node. The consumption of WDM-TRX is 5W [8]. From Fig. 8 we can see that N-GREEN achieves significant savings (up to 10 times) in the power consumption, when compared with Ethernet (also valid for large switches).

VII. Conclusion

A cost optimized and deterministic scheduler for the N-GREEN WSADM technology, supporting the time-sensitive CPRI traffic with zero jitter has been proposed. Then, an Integer Linear Program and a highly performant heuristic have been provided as tools enabling to calculate this scheduling. The benchmarking studies have been performed and have shown that significant cost and energy consumption reductions could be obtained when compared to a same capacity state-of-the-art Ethernet technology. The WDM technology proposed in the N-GREEN project is positioning as a highly competitive solution for a future generation of Xhaul networks.

Acknowledgment

We thank François Taburet for the insightful exchanges about the topic. We thank the French National Agency of Research (ANR) for partial funding of the N-GREEN project.

References

[1] D. Chitimalla et al., "5G fronthaul-latency and jitter studies of CPRI over ethernet," in IEEE/OSA JOCN, Feb. 2017.

[2] D. Chiaroni and B. Uscumlic, "Potential of WDM Packets", invited paper, ONDM 2017, Budapest, Hungary, 2017.

[3] A. Triki, A. Gravey, P. Gravey, M. Morvan, "Long-Term CAPEX Evolution for Slotted Optical Packet Switching in a Metropolitan Network". ONDM 2017, Budapest, Hungary, 2017.

[4] Xilinx All Programmable, XAPP1252 (v1.1) November 17, 2016, "Burst-Mode Clock Data Recovery with GTH and GTY Transceivers", Author: Edward Lee and Caleb Leung, https://www.xilinx.com/support/documentation/application_notes/xapp1252-burst-clk-data-recovery.pdf, accessed on Oct. 29, 2017.

[5] Cost of 10G DWDM SFP+ Transceivers (found at https://www.cozlink.com/), https://www.cozlink.com/10g-dwdm-sfp-transceivers-c322-323-zh381, accessed on Oct. 29, 2017.

[6] Finisar, Product Specification, 10Gb/s, 40km Single Mode, Multi-Rate SFP+ Transceiver, FTLX1672D3BTL, https://www.finisar.com/sites/default/files/downloads/finisar_ftlx1672d3btl_10gbase-er_40km_sfp_optical_transceiver_product_specb1.pdf, accessed on Oct. 29, 2017.

[7] Cisco Catalyst 2960-S Series Switches Data Sheet, https://www.cisco.com/c/en/us/products/collateral/switches/catalyst-2960-s-series-switches/data_sheet_c78-726680.html, accessed on Oct. 29, 2017.

[8] Cisco CPAK 100GBASE Modules Data Sheet, https://www.cisco.com/c/en/us/products/collateral/interfaces-modules/transceiver-modules/data_sheet_c78-728110.html, accessed on Oct. 29, 201

Design Methodologies and Algorithms for Survivable C-RAN

Bahare M. Khorsandi, Federico Tonini, Carla Raffaelli
DEI, University of Bologna
Viale Risorgimento 2, 40136 Bologna, Italy
email: {bahare.masood, f.tonini, carla.raffaelli}@unibo.it

Abstract— **In centralized/cloud radio access networks (C-RANs), baseband units (BBUs) are decoupled from remote radio units (RRUs) and placed in BBU hotels. In this way baseband processing resources can be shared among RRUs, providing opportunities for radio coordination and cost/energy savings. However, the failure of a BBU hotel can affect a large number of RRUs creating severe outages in the radio segment. For this reason, the design of a resilient C-RAN is imperative. In this paper, an extension of the facility location problem (FLP) is proposed to find the placement of BBU hotels that guarantees survivability against single hotel failure while the delay is minimized. Different strategies are proposed based on heuristic and integer linear programming (ILP) to solve the survivable BBU location problem and optimizing the sharing of backup resources. The results compare the proposed methodologies in terms of the costs of the BBU placement by referring to different network topologies. The heuristic algorithm is shown to find solutions close to those obtained by the ILP, although evidencing different contributions that are suitably discussed.**

Index Terms—**C-RAN, Fronthaul, Resiliency, Facility Location, ILP, Heuristics.**

I. Introduction

Centralized radio access network (C-RAN) was originally introduced to accommodate growth in mobile networking [1]. As opposed to the traditional distributed access networks, where the radio and baseband processing functions are performed at the base station (BS) sites, C-RAN decouples baseband units (BBUs) from BS sites and place them in centralized locations, called BBU hotels. BBUs, that performs baseband processing functions, are connected to remote radio units (RRUs), performing radio processing at BS sites, through the so called fronthaul segment [2], typically based on the common public radio interface (CPRI) [3].

C-RAN introduces considerable benefits compared to the distributed access network, especially when coupled with Network Function Virtualization (NFV) and Sofwtare Defined Networking (SDN), that enable the cloud RAN concept and lead to enhanced flexibility and effectiveness in support of energy and cost reduction, advanced coordination techniques and baseband function virtualization [2]. Despite these advantages, C-RAN introduces many challenges, like the deployment of a reliable C-RAN capable of meeting strict capacity and delay requirements for a large number of cells in wavelength division multiplexing (WDM) optical network. The failure of a BBU hotel can strongly impact the network performance, resulting in service interruption for a potentially large number of mobile users and devices.

Network resiliency against failures is one of the well-established research area for WDM optical networks. The work in [4] presents a routing algorithm for a survivable all-optical mesh topology based on WDM. The authors introduce three primary and backup route computation mechanisms that aim at improving the overall network performance. However, strict latency and capacity requirements, like the one imposed by fronthaul links, are not considered. In [5], authors present a path and link restoration technique for link failures, but protection against node failures is not investigated. The problem of finding optimal location for network functionalities, such as baseband processing functions, is investigated in [6] and [7], while the assignment of BBU functionalities in C-RAN over WDM networks is discussed in [8], but all of the above studies do not consider protection against failures. In [9], the authors deal with protection in traditional distributed radio access networks, but no considerations are made regarding centralized architectures. A previous work proposed a resilient BBU hotel placement against single BBU hotel failure [10]. The approach is based on heuristic with constraints on starting point and maximum distance between each BBU hotel and RRU pair. The results are compared with the case of no protection and show that by adding only 30% more BBU hotels, the resiliency can be guaranteed.

In this work, the classical facility (or node) location problem (FLP) presented in [6] and [7] is extended by introducing the concept of resiliency against single BBU hotel failure. Different design methodologies for survivable C-RAN architectures based on heuristic and an integer linear programming (ILP) are proposed. The main objective of the study is to find the optimal placement for the BBU hotels in order to have protected service for RRUs while minimizing the total distance between RRUs and BBUs. The minimization of backup BBUs and the related deployment are also discussed.

The remainder of this work is organized as follows: in section II, the reference architecture and problem definition are introduced. Section III provides algorithms based on different methodologies to solve the problem. Numerical results obtained for different topologies are described in section IV, while section V concludes the paper.

978-3-903176-07-2 © 2018 IFIP

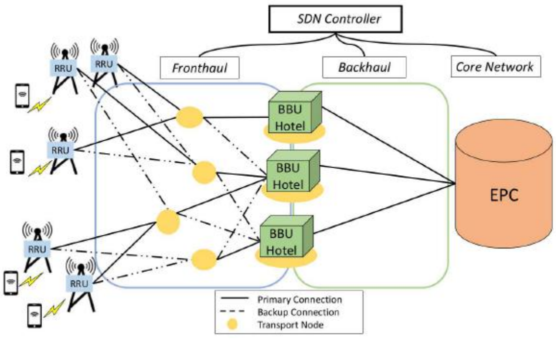


Fig. 1. C-RAN architecture.

II. Reference Architecture and Problem Definition

A C-RAN architecture is shown in figure 1. It is an architecture where a set of RRUs in an area is divided into groups and connected to different nodes of the transport network, called transport nodes. Each transport node represents also a potential BBU hotel site and is connected to one or more transport nodes by means of fiber cables, creating the fronthaul network. A BBU hotel contains several BBUs, each serving a RRU. All RRUs connected to the same transport node have their BBUs hosted in the same BBU hotel for the mitigation of interference in the area. In addition to the fronthaul, another network segment, the backhaul, provides connectivity between BBU hotels and the mobile core network, i.e., the evolved packet core (EPC).

The survivable fronthaul design problem addressed in this paper is defined as follows:

- **Given:** the physical topology of the WDM mesh transport network, the number of RRUs connected to each transport node, the potential location and the cost of activating a new BBU hotel, and the cost of connecting RRUs in transport nodes to BBU hotels.
- **Find:** the minimum cost solution that minimizes the BBU hotel activations and distance between each pair of BBU hotels and RRUs in transport nodes. The solution must ensure that each RRU is always connected to a BBU, also when a single hotel failure occurs.

In the following, some useful parameters and variables are defined, while the notation used throughout the paper is summarized in Table I.

The activation cost of BBU hotels needed to provide full coverage and resiliency of the target area is calculated using the following formula:

$$C_B = \sum_{i=1}^{s} B_i \beta_i \tag{1}$$

where B_i is a boolean variable equal to 1 when the node is set as a BBU hotel, that is when it hosts BBU functionalities related to one or more RRUs. β_i is a parameter associated to the activation cost for a BBU hotel in node i.

TABLE I
Notations used in the different procedures

	Parameters:
S	Set of transport nodes, $\lvert S \rvert = s$.
H	$s \times s$ matrix. h_{ij} is the distance in hops between nodes i and j computed with the shortest path.
α	Weight of the hops in the cost function F.
β_i	Weight of the active BBU hotel i in the cost function F.
γ	Weight of the BBU hotel ports in the cost function G.
	Variables:
B_i	1 if node $i \in S$ hosts a BBU hotel, 0 otherwise.
p_{ij}	1 if BBU hotel i is assigned as primary for RRUs at node j $i, j \in S$, 0 otherwise.
b_{ij}	1 if BBU hotel i is assigned as backup for RRUs at node j $i, j \in S$, 0 otherwise.
x_i	Number of BBU ports required at hotel site i for primary purposes.
y_i	Number of BBU ports required at hotel site i for backup purposes.
W	Average number of wavelengths per link.

In order to provide reliability against single BBU hotel failure, it is sufficient to ensure that each RRU is connected to two BBU ports placed in different BBU hotels, one in the primary and one in the backup hotel. The overall distance between BBU hotels and RRUs connecting to the transport nodes in the network, considering both primary and backup hotels, is denoted as D_H:

$$D_H = \sum_{i=1}^{s} \sum_{j=1}^{s} p_{ij} h_{ij} + \sum_{i=1}^{s} \sum_{j=1}^{s} b_{ij} h_{ij} \tag{2}$$

where p_{ij} and b_{ij} are boolean variables that indicates if hotel i is assigned as primary or a backup for the group of RRUs at transport node j. h_{ij} represents the distance, in hops, between transport node i and j computed solving the shortest path problem. By multiplying (2) by the parameter α, the total cost for the distance is achieved:

$$C_H = D_H \alpha \tag{3}$$

Finally, to solve the problem, the proper number of BBU ports must be allocated in each hotel. The total number of primary and backup BBU ports, and the related cost, are calculated according to the following formulas:

$$N = \sum_{i=1}^{s} x_i + \sum_{i=1}^{s} y_i = N_P + N_B \tag{4}$$

$$C_P = N \gamma \tag{5}$$

where N_P and N_B are the total number of primary and backup ports respectively. C_P is the contribution of the total number of ports in each hotel multiplied by the cost parameter γ associated to each port. Since the protection requires that each RRU is connected to two different BBU hotels, the total number of ports should be twice the number of RRUs, and consequently the value for C_P can be fixed. However, only N_P is fixed, while N_B can be reduced. In fact, if exist RRUs which

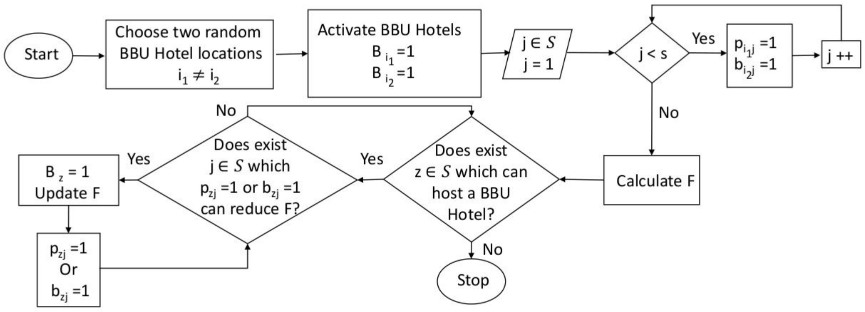

Start
Choose two random BBU Hotel locations $i_1 \neq i_2$
Activate BBU Hotels $B_{i_1}=1$ $B_{i_2}=1$
$j \in S$ $j = 1$
$j < s$
Yes
$p_{i_1j}=1$ $b_{i_2j}=1$
j ++
No
Calculate F
Does exist $z \in S$ which can host a BBU Hotel?
Yes
No
Stop
Does exist $j \in S$ which $p_{zj}=1$ or $b_{zj}=1$ can reduce F?
Yes
No
$B_z = 1$ Update F
$p_{zj}=1$ Or $b_{zj}=1$

Fig. 2. Flowchart of the BBU hotel placement of heuristic solution.

have separate primary BBU hotels, they can share backup ports due to the single failure assumption done in this work. By sharing the backup ports among RRUs the value for C_P can be reduced, and further cost saving can be achieved.

III. Design Methodologies

In the following, two solutions for survivable fronthaul design are presented. First the problem is solved by the heuristic and in the next subsection an ILP formulation is introduced for comparison.

A. Heuristic

The proposed algorithm is based on the FLP presented in [7], which is applied to networking contexts to find optimal locations for network functions, given a set of possible nodes, under cost constraints. The FLP is extended here by considering also the location of backup functions, in addition to primary functions, while choosing the BBU hotels within the set of transport nodes in the fronthaul network. In the proposed approach, the overall cost of deploying BBU hotels and overall distance, in hops, between BBU hotels and RRUs is minimum, even though it is not guaranteed that a RRU is connected to either a primary or a backup BBU hotel within a given distance.

The heuristic aims at connecting s transport nodes, each containing a given amount of RRUs, through a list of possible BBU hotel locations so that the total cost F is minimum. Let us introduce the total cost F, given by the sum of the cost of activating a new BBU hotel (C_B) and the overall cost of connecting RRUs to BBU hotels (C_H) as follows:

$$Minimize\, F = C_B + C_H \tag{6}$$

The flowchart of the proposed algorithm is reported in figure 2. As an input to the algorithm, it is given an $s \times s$ matrix H which contains the information about the distance computed according to shortest path between each pair of nodes in the network. Also other parameters, α and β_i, are given. These two parameters are related to the cost of distance in hops and activating a new BBU hotel, respectively. The algorithm starts by randomly choosing two candidate nodes for hosting BBU hotels, one for primary $i_1 \in S$ and the other one for backup $i_2 \in S$. In order to provide resiliency, these two locations must be different. After activating new BBU hotels at nodes i_1 and i_2, all RRUs at node $j \in S$ are connected to these two hotels, one as a primary ($p_{i_1j} = 1$) and the other one as a backup BBU hotel ($b_{i_2j} = 1$). The total cost F of the initial solution is then computed and used as a reference value. The aim of the rest of the procedure is to reduce the value of F by adding further BBU hotels in order to reduce the contribution of C_H.

The search for a new BBU hotel is performed in the following way. A new location $z \in S$, which is not hosting a BBU hotel, is selected and a new BBU hotel is activated in z. The RRUs involved in the cost reduction, i.e., the ones that can reduce C_H, are then disconnected from their former BBU hotel i_1 or i_2 and connected to the new BBU hotel z. In all steps of the procedure, all RRUs are always connected to two different BBU hotels. The search for a new hotel is repeated until no improvement in F is experienced, and the last solution is considered to be the best solution to the resilient BBU hotel placement, meaning that all RRUs are connected to two different BBU hotels and the obtained cost F is the best combination of the total cost for activating BBU hotels C_B and the cost related to connection C_H.

Once the BBU hotel placement is performed, another procedure is performed to investigate further cost reduction by sharing backup BBU port. BBU hotel port sharing is allowed if and only if two RRUs, namely j_1 and j_2, share the same backup BBU hotel i' and are assigned to different primary BBU hotels. Only in this case, j_1 and j_2 can share the backup ports in BBU hotel i'. This procedure is repeated for all shared backup BBU hotels with the above property.

B. ILP Optimization

The core of our problem is based on the ILP formulation of the FLP introduced in [7]. The formulation in [7] has been modified in order to provide protection, by means of backup hotels, and to include the effects of BBU hotel ports. The problem is here formulated in such a way that, by properly tuning the parameter of the objective function, BBU ports can be minimized while solving the survivable fronthaul design problem.

Additional parameters:

- r_j number of RRUs at site j.
- M a large number.

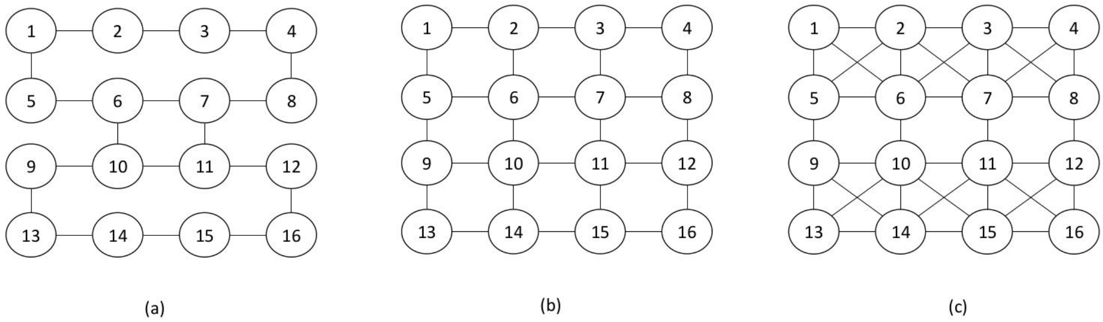


Fig. 3. The reference network topologies, (a) network A with connectivity $N_A = 2.25$, (b) network B with connectivity $N_B = 3$ and (c) network C with connectivity $N_C = 4.5$.

Additional variables:

- $c_{j,i,i'} = 1$ if source j is using destination i as primary and i' as backup hotel site; 0 otherwise.

Objective function:

$$Minimize\ G = C_B + C_H + C_P \tag{7}$$

The multi-objective function (7) is composed of three members. The first term takes into account the activation cost of each hotel (C_B). The second term accounts for the cost to connect RRUs to BBU hotels, both primary and backup (C_H) while the third term accounts for the cost of BBU ports required in each hotel (C_P).

The problem is subject to the following constraints:

$$\sum_{i \in S} p_{i,j} = 1, \forall j \in S \tag{8}$$

$$\sum_{i \in S} b_{i,j} = 1, \forall j \in S \tag{9}$$

$$p_{i,j} + b_{i,j} \leq 1, \forall i,j \in S \tag{10}$$

$$x_{i,j} \geq \sum_{i \in S} p_{i,j} r_i, \forall i \in S \tag{11}$$

$$c_{j,i,i'} \geq p_{j,i} + b_{j,i'} - 1, \forall i,j \in S, i' \in S - \{i\} \tag{12}$$

$$y_{i'} \geq \sum_{j \in S} c_{j,i,i'} r_j, \forall i \in S, i' \in S - \{i\} \tag{13}$$

$$B_i \cdot M \geq \sum_{j \in S} p_{i,j} + b_{i,j}, \forall i \in S \tag{14}$$

Constraints (8) and (9) ensure that there is one primary and one backup hotel for each RRU. Constraint (10) imposes primary and backup hotels to be disjoint. Constraint (11) counts the number of BBU ports to be installed in each primary hotel. Constraint (12) tells if a primary hotel is in common to a backup hotel for each source and is used in constraint (13) to ensure that there are enough BBU ports in each backup hotel. These two constraints, along with (7), allow to minimize the number of ports in each backup hotel. In fact, the number of BBU ports required at each backup hotel equals the largest number of RRUs that shares the same primary hotel. Finally, constraint (14) activates hotels (i.e., tells if the hotel is a primary and/or backup for RRUs).

IV. Use case Settings and Numerical Results

A. Reference Scenarios

This section presents the analysis of survivable fronthaul in C-RAN to evaluate the strategies proposed in Section III and applied to different scenarios. The reference topologies of the optical transport network used in the performance assessment are presented in figure 3. Three metro/aggregation networks are considered with 16 nodes each but with different levels of connectivity. The connectivity N_i for network i is defined as follows

$$N_i = \frac{\sum_{i=1}^{s} NO_i}{s} \tag{15}$$

where NO_i is the number of optical interfaces in node i and s is the total number of nodes, 16 for all networks in this evaluation.

In all the topologies each node represents a cell site, assumed to serve a value of the upstream traffic equal to 10 RRUs connected to the node, each one requiring two lightpaths, i.e., one connecting the RRU to the primary and one connecting the same RRU to the backup BBU hotel. Each edge in the graph represents a bidirectional fiber connection, all with the same length. The results discussed in this section are obtained using a Java-based simulator and compared with the optimal solution from ILP, obtained using CPLEX commercial tool [11]. The results from the heuristic are averaged over all the possible combinations of BBU hotels pairs that can be used as a starting point. Among the solutions, the maximum observed deviation from the average is 22% which shows the limited impact of the starting point on the results and allows the algorithm to start by random locations. In all the graphs reporting F and G, the results are normalized with respect to α (that was constant) and are reported in each case. All β_i were considered constant and equal to β. The following parameters are used:

$$R = \frac{\beta}{\alpha} \quad , \quad Q = \frac{\gamma}{\alpha} \tag{16}$$

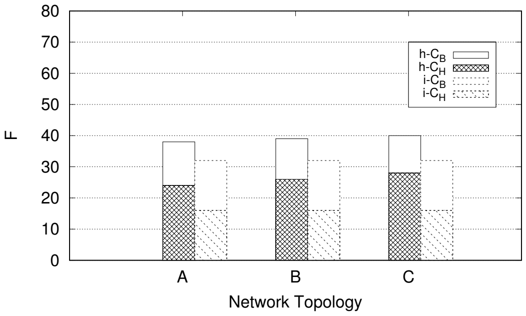


Fig. 4. Total cost F, normalized with respect to α, for ILP (i) and heuristic (h), representing the contributions of the BBU hotel activation cost C_B and the overall distance between each pair of RRUs and BBU hotels C_H, in networks A, B and C when $R = 1$.

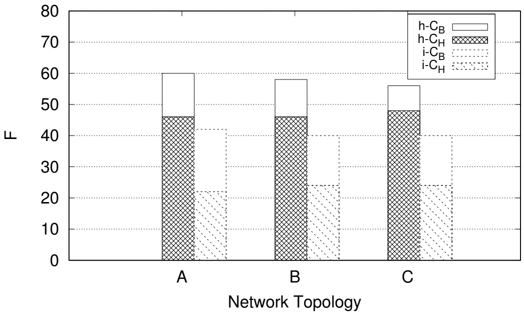


Fig. 5. Total cost F, normalized with respect to α, for ILP (i) and heuristic (h), representing the contributions of the BBU hotel activation cost C_B and the overall distance between each pair of RRUs and BBU hotels C_H, in networks A, B and C when $R = 2$.

B. Numerical Results

Figure 4 reports the total cost of the survivable fronthaul design solution (i.e., the cost function F). In the figure, the two contributions to F are shown for each network when $R = 1$, and the total cost is normalized with respect to α. The cost obtained with the heuristic is compared to the one of the ILP when $\gamma = 0$, so that F has the same meaning as G. The total cost is lower for the ILP, with a different contributions of BBU hotels and distance. While the ILP cost is constant with respect to different network connectivities, the cost of the heuristic is slightly higher when the network connectivity is higher. The reason is that the heuristic is able to activate less BBU hotels than the ILP, which causes the number of hops to grow, and results in an increased overall cost. Similarly to the previous figure, figures 5 and 6 show the total cost function F, normalized respect to α when R equals 2 and 10, respectively. By increasing R, the hotel activation cost becomes more relevant in F, therefore the number of selected hotels decreases when R increases. For $R = 2$ the contribution of the BBU hotels to F is less than in the case $R = 1$. When $R = 10$, the number of active BBU hotels keep decreasing but their contribution to the total cost becomes higher than in the case $R = 2$, due to the large R factor. As a final note, the heuristic provides a good approximation of the ILP when the activation cost and the distance have similar weight in F $(R = 1)$ and when the activation cost is much more relevant than the distance $(R = 10)$. In the case $R = 2$ instead, the heuristic solution is up to 40% more expensive then the ILP.

The number of BBU ports, that is the number of functional interfaces to serve the related RRUs, is calculated based on the number and location of BBU hotels. The previous results, obtained using F or G with $\gamma = 0$, do not include any consideration on the number of ports, not considered so far. In order to compare the results of the heuristic and ILP, the latter has has been run once again to derive the minimum number of BBU hotel ports. α and β were all set to zero, γ was set to 1 and the hotel placement previously obtained was introduced in the ILP model as additional constraint, in order to set the position of the BBU hotels. The overall number of backup ports obtained from the modified ILP is compared to the heuristic one, averaged over all the initial cases, and is reported in figures 7 and 8, for the three network topologies when R equal to 1 and 10, respectively. Since the total number of primary BBU hotel ports is fixed and equal to the number of RRUs, it is not included in these figures.

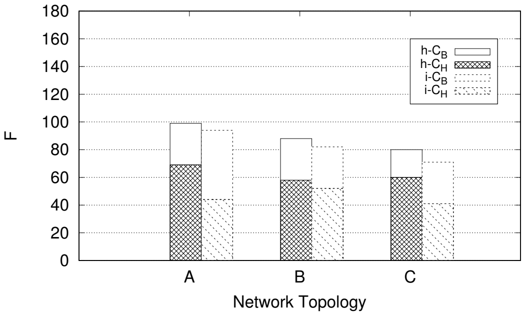


Fig. 6. Total cost F, normalized with respect to α, for ILP (i) and heuristic (h), representing the contributions of the BBU hotel activation cost C_B and the overall distance between each pair of RRUs and BBU hotels C_H, in networks A, B and C when $R = 10$.

Figure 7 shows that the number of backup BBU hotel ports required by the ILP is lower than the heuristic one. In the case of $R = 1$, both ILP and heuristic have a large number of active BBU hotels, and since this number is higher for the ILP, ILP results more efficient in sharing BBU hotel ports. By increasing the network connectivity, the ILP easily assigns primary and backup BBU hotels such that the sharing of backup ports results higher than with the heuristic that, instead, assigns primary and backup hotels based only on F, and therefore is not aware of their impact on the number of shared backup ports.

Figure 8 shows that the sharing of BBU hotel ports is extremely difficult for the heuristic when R is high and the number of active hotels is very low. The total number of ports is high independently of the connectivity due to the fact that the solution obtained with the heuristic, averaged over all possible starting nodes, requires just two or three hotels to be active. The ILP instead, finds solutions with slightly more

TABLE II
THE EFFECTS OF Q ON THE COST COMPONENTS OF THE OBJECTIVE FUNCTION G FOR THE DIFFERENT NETWORK TOPOLOGIES ($R = 2$).

	Network A							Network B							Network C						
Q	C_B	C_H	F	N_B	C_P	G	W	C_B	C_H	F	N_B	C_P	G	W	C_B	C_H	F	N_B	C_P	G	W
0	20	22	42	100	0	42	12.2	16	24	40	80	0	40	10	16	24	40	60	0	40	6.7
0.001	20	22	42	100	0.1	42.1	12.2	16	24	40	80	0.08	40.08	10	14	26	40	50	0.05	40.05	6.9
0.1	22	21	43	80	8	51	11.7	18	23	41	70	7	48	9.6	14	26	40	50	5	45	6.9

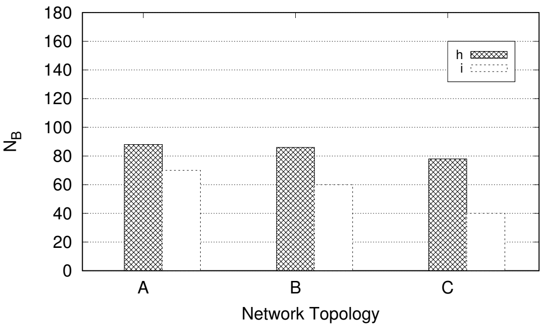


Fig. 7. Total number of backup ports N_B for ILP (i) and heuristic (h) in networks A, B and C with $R = 1$.

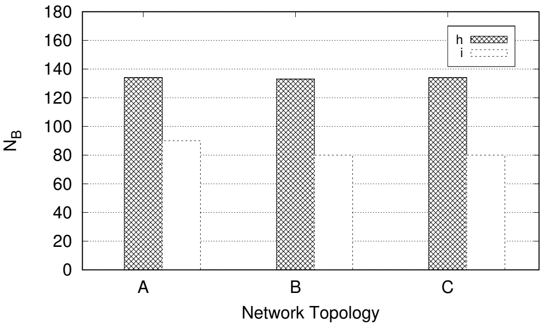


Fig. 8. Total number of backup ports N_B for ILP (i) and heuristic (h) in networks A, B and C with $R = 10$.

active hotels, and therefore can limit the number of BBU hotel ports to lower values.

In order to see the effects of γ on the placement, the value of parameter Q is varied. Table II shows the different values for F and G in the three networks when Q is equal to 0, 0.001 and 0.1, while R is considered constant and equal to 2. As expected, the total cost in each network increases by increasing Q, due to the cost introduced by the ports. For these values of Q, the sum of activation and distance cost are almost the same in the three cases, while their contribution changes. In fact, there may be solutions employing different number of hotels and that leads to have slightly different cost like the case of $Q = 0.1$. The impact of γ on the cost is therefore to select the solution, among solutions with the same cost (measured by F), that minimizes also the total number of ports. Table II also shows the average number of wavelengths per link without considering wavelength continuity. It is possible to notice how the required wavelengths per link decrease when the network connectivity increase, due to the higher number of available links to connect transport nodes. In conclusion, when the contribution of the BBU hotel ports is considerably less relevant with respect to the activation and distance, which well represent a real case scenario, it is safe to neglect the contribution of the BBU hotel ports in a first computational phase. Then, when the hotels to activate are selected and the delay is minimized, a dedicated minimization can be performed to limit the number of BBU hotel ports.

V. CONCLUSION

The paper presents a survivable fronthaul design in C-RAN. Two methodologies have been proposed and compared in terms of relevant cost parameters, namely the number of BBU hotels, overall distance between BBU hotels and RRUs and BBU hotel ports. The different contributions to cost, calculated by heuristics and ILP have been discussed, evidencing the influence of different cost weights on results. The methodologies has been tested against different network topologies of the same size characterized by different connectivity level, showing limited impact on final costs.

REFERENCES

[1] China Mobile. 'C-RAN: the road towards green RAN'. White Paper, ver.2 2011.

[2] A. Checko, H. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. Berger, L. Dittmann, 'Cloud RAN for Mobile Networks. A Technology Overview', IEEE Communications Surveys and Tutorials, 17(1), pp.405-426, 2015.

[3] CPRI, 'Interface specification' ver.7.0, 2015.

[4] S. Ramamurthy, L. Sahasrabuddhe, B. Mukherjee, 'Survivable WDM mesh networks.', IEEE/OSA Journal of Lightwave Technology, 21(4), pp.870-883, 2013.

[5] S. Gowda, K. M. Sivalingam. 'Protection mechanisms for optical WDM networks based on wavelength converter multiplexing and backup path relocation techniques'. In Proc. IEEE Conference of Computer and Communications (INFOCOM), 2003, San Francisco, USA, Vol. 1, pp. 12-21.

[6] L. Tang, C. Zhu, Z. Lin, J. Shi, W. Zhang, 'Reliable Facility Location Problem with Facility Protection', PLOS ONE, 11(9), p.e0161532, 2016.

[7] M. Pioro, D. Medhi, 'Routing, Flow, and Capacity Design in Communication and Computer Networks' Saint Louis: Elsevier Science, 2014.

[8] F. Musumeci, C. Bellanzon, N. Carapellese, M. Tornatore, A. Pattavina, S. Gosselin, 'Optimal BBU Placement for 5G C-RAN Deployment Over WDM Aggregation Networks', IEEE/OSA Journal of Lightwave Technology, 34(8), pp.1963-1970, 2016.

[9] A. Douik, H. Dahrouj, T. Y. Al-Naffouri, M. Alouini. 'Resilient backhaul network design using hybrid radio/free-space optical technology'. In Proc. IEEE International Conference on Communications (ICC), 2016, Kuala Lumpur, Malaysia.

[10] B. M. Khorsandi, C. Raffaelli, M. Fiorani, L. Wosinska, P. Monti. 'Survivable BBU Hotel placement in a C-RAN with an Optical WDM Transport', In Proc. International Conference Design of Reliable Communication Networks (DRCN), 2017, Munich, Germany.

[11] IBM ILOG CPLEX Optimization Studio ver.12.6.3.

Cognitive Zone-Based Spectrum Assignment Algorithm for Elastic Optical Networks

Rodrigo Stange Tessinari[†], Didier Colle[*], Anilton Salles Garcia[†]
[†] *LabTel, Federal University of Espírito Santo, Brazil*
[*] *IMEC, Ghent University, Belgium*
Email: stange@inf.ufes.br

***Abstract*—In this work, we present the Cognitive Zone-Based spectrum assignment algorithm (CZB). Our algorithm is capable of observing the network traffic and acquiring information regarding the network services, thus using it to calibrate the division of the spectrum into partitions. To achieve this, the CZB algorithm uses an upgraded version of the Static Zone-Based spectrum assignment algorithm. Our results show that CZB algorithm can indeed achieve its objective, and even improve the fairness under certain conditions. Simulations show that CZB algorithms can achieve up to 10 times better fairness under certain conditions when compared to the Spectrum Sharing First-Fit algorithm.**

***Index Terms*—Elastic Optical Networks, Fairness, Spectrum Management, Spectrum Assignment.**

I. Introduction

According to Cisco Visual Networking Index Report [1], the traffic from wireless and mobile devices will account for more than 63 percent of total IP traffic by 2021, and roughly half of that will be from Smartphones. This new traffic characteristic era represents a significant challenge to be faced by service providers since supporting this highly dynamic traffic demands a flexible and agile optical core networks [2].

In the last few years, the Elastic Optical Networks (EON) has emerged as a promising technology to fulfill this niche due to its flexibility and higher spectral efficiency [3], [4]. This added flexibility and efficiency come at the price of increased complexity and new hurdles, such as the spectrum fragmentation and unfairness among the network services [5], [6]. The unfairness problem can be especially worsened with the extra mobile traffic expected in the years to come.

Previous works proposed to tackle the unfairness problem by using a myriad of spectrum management techniques, mostly involving dividing the spectrum into partitions. The division of the spectrum is usually made following specific criteria and are classified as Dedicated Partition (DP) or Shared Partition (SP). The DP schemes work in a restrictive manner, in which the services are obliged to fit the designed partition [6], [7], whereas the SP techniques are priority-based, allowing services to be accepted outside the ideal partition [8], [9]. Some schemes work in a mixed mindset with restrictive partitions but using a spectrum range that can be shared [10]. The more complex the spectrum management technique is, more information from the network is needed to establish and manage the partitions. Most methods need knowledge about the services that may use the network [7], [9], [10], the ratio between them [6], and even the blocking probabilities [6], [8]. All information is made available as an input for the allocation algorithm, also known as Routing and Spectrum Assignment (RSA) algorithm, responsible for accepting or blocking new requests trying to access the network.

These methods require an increasing number of inputs and features to increase the accuracy of the obtained results. In this sense, an approach that reduces the complexity required is desirable. On the other hand, we find cognitive methods that can infer results from a reduced set of inputs with reduced or negligible losses. In the telecommunications context, the word cognitive evokes the ability to observe and to extract information from the network conditions, and then to use this information in a useful manner [11].

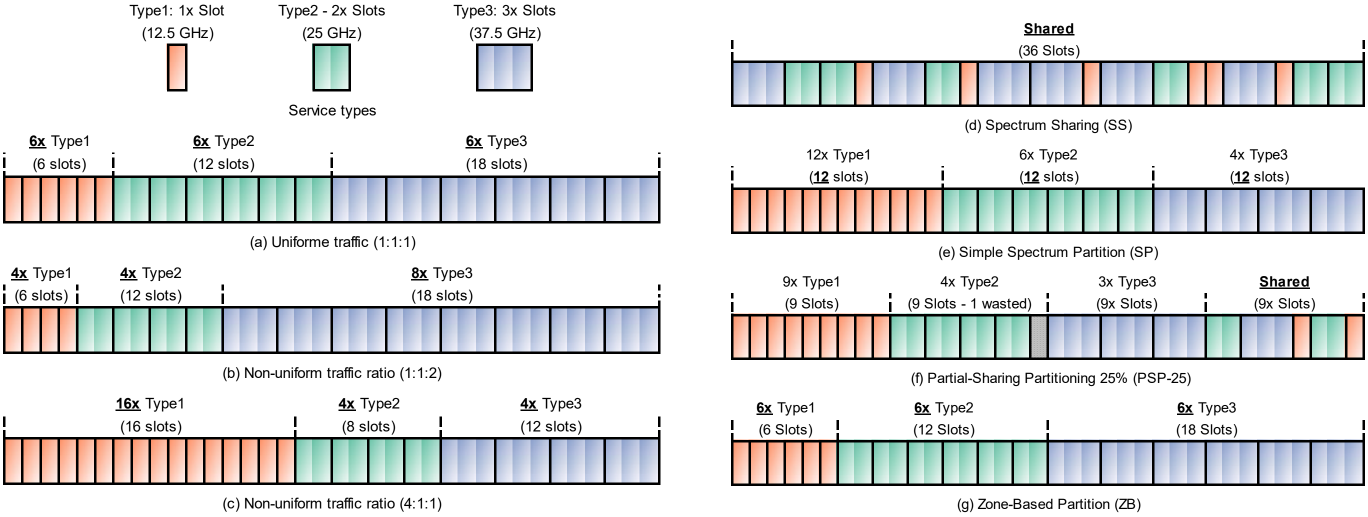
In this context, we present the Cognitive Zone-Based (CZB) spectrum assignment algorithm. Our technique is capable of observing the network traffic and acquire information regarding the services using the network. Using the acquired data, CZB infers the traffic ratio of the services and uses it to calibrate the spectrum division into partitions. Since CZB is based on an improved version of previous work [7], in this paper we also present the updates we developed in our previous Zone-Based algorithm.

The remainder of this paper is divided as follows: Section II presents the updates we made in our previous method, now called Static Zone-Based (SZB) assignment algorithm, whereas Section III shows the new CZB version of it. Section IV presents the tests and rhe results we obtained using both techniques. Finally, Section V concludes this paper.

II. An Upgrade to the Static Zone-Based Spectrum Assignment Algorithm

In a dynamic network scenario, incoming requests are established and released in an entirely random fashion. This randomness induces spectral resources to be highly fragmented, and consequently, “gaps” are unavoidably introduced leading to the so-called intra-link fragmentation, thus degrading spectrum utilization efficiency. Due to the contiguity constraint, the more fragmented is the spectrum, the harder is for new connections to be established. In a heterogeneous environment, more spectrally demanding services suffer from an increased difficulty to get requests accepted by the network when comparing to less demanding services, thus the blocking ratio is proportional to the number of resources requested. In

978-3-903176-07-2 © 2018 IFIP



(a) Uniforme traffic (1:1:1)

(b) Non-uniform traffic ratio (1:1:2)

(c) Non-uniform traffic ratio (4:1:1)

(d) Spectrum Sharing (SS)

(e) Simple Spectrum Partition (SP)

(f) Partial-Sharing Partitioning 25% (PSP-25)

(g) Zone-Based Partition (ZB)

Fig. 1. Example of zone divisions configured by the proposed Static Zone-Based algorithm according to different traffic ratios: (a) Uniform traffic (1:1:1), (b) Non-uniform pattern (1:1:2), and (c) Non-uniform pattern (4:1:1). Comparison between (d) Spectrum Sharing (SS), (e) Simple Spectrum Partition (SP), (f) Partial-Sharing Partitioning 25% (PSP-25) [9], and (g) Zone-Based partitioning (ZB) [7].

this sense, the disparity of the services blocking ratios is what we call unfairness.

To mitigate the unfairness effect, in previous work [7] we proposed the first version of the Zone-Based spectrum assignment algorithm, which idea is to divide the spectrum transforming the expected heterogeneous environment in a set of smaller and homogeneous environments (i.e., partitions or zones). It pays particular attention to partition delimitation, ensuring coexistence of similar services within each partition only and ensuring each partition has the same capacity (i.e., can accommodate the same maximum number of connections at a given time). Therefore, homogeneity is guaranteed as each partition only supports connections with the same size. Such homogeneity mitigates the effects of intra-link fragmentation. Although the fragmentation is not entirely removed, this division ensures that every single fragment of the spectrum available has enough space to accommodate at least one connection, therefore not being a problem anymore. The first version of our solution achieves its objective assuming a uniform traffic pattern. Furthermore, the focus of this section is presenting an upgrade, allowing the technique to handle non-uniform traffic.

The main change of this updated version of our technique is taking the traffic pattern into account during the zone delimitation. The idea is to maintain the same maximum number of connections for each possible service type, adjusted by the traffic pattern. For example, supposing three service types are coexisting in the same network and uniform traffic; the spectrum would be divided into three zones, holding C_{max} connections each. In total, 3 C_{max} connections could exist at the same time, C_{max} connections per service type. This concept is illustrated in Figure 1. In Figure 1 (a) the zones are configured based on uniform traffic with the ratio (1:1:1) between services types, implying in the same maximum number of connections accommodated within each zone simultaneously (i.e., $C_{max} = 6$ for each service type). Figure 1 (b) illustrates how the zones are influenced by non-uniform traffic, the ratio 1:1:2 results in doubling *Type3* connections, whereas the proportion (4:1:1) implies in four times more *Type1* connections (c).

To describe the zones division we introduce the following notation:

B_{max} : Total bandwidth available in each link (in *slots*).
St : Set of all possible services, $St = \{St_1, ..., St_n\}$ (in *slots*).
Tr : Set of service traffic ratio, $Tr = \{Tr_1, ..., Tr_n\}$.
C_{max} : The maximum number of connections allowed within each zone (before traffic ratio compensation).
Zc : Set of zones capacities, $Zc = \{Zc_1, ..., Zc_n\}$ (in slots)

The following equations define C_{max} and Zc:

$$C_{max} = \left\lfloor \frac{B_{max}}{\sum_{i=1}^{n} St_i \, Tr_i} \right\rfloor \tag{1}$$

$$Zc_i = C_{max} \, St_i \, Tr_i \tag{2}$$

Using Figure 1 (c) values as example, $B_{max} = 36$, $St = \{1, 2, 3\}$, and $Tr = \{4, 1, 1\}$, Equation (1) gives $C_{max} = 4$, and Equation (2) gives $Zc = \{16, 8, 12\}$. The value $C_{max} \, Tr_i$ gives the maximum number of connections that can coexist simultaneously at any given time for a service type St_i whereas the Zc_i represents the number of slots needed by the zone to support those connections.

Figure 1 also illustrates how the Zone-Based spectrum assignment technique compares to other related work under uniform traffic pattern. From top to bottom, first, Figure 1 (d) shows the Spectrum Sharing (SS), i.e., if connections are allocated without any spectrum management. Next, (e) shows the most straightforward way to manage spectrum, by dividing it into partitions with the same number of slots. It is called Simple Spectrum Partition (SP). The third algorithm presented in (f) is the Partial Sharing Partitioning 25% (PSP-25), that uses a shared zone that can be used as an “overflow zone” [9]. Finally, our Zone-Based method (g), which fixes the same maximum number of connections within each partition [7].

Although our Static Zone-Based algorithm (SZB) was developed to work using as less information as possible, it still needs information regarding the traffic expected in the network, more specifically, the types of services and their ratio. As shown in Section IV, when this data is available, SZB performs well and increases the fairness among the different services within the network. For the cases in which this information regarding the traffic is not available, in the next section, we present the cognitive version of our algorithm.

III. Cognitive Zone-Based Spectrum Assignment Algorithm

Spectrum partitioning techniques are *highly dependent* on the traffic pattern in the network. In cases where the nature of the network traffic is known beforehand, and the traffic does not change, it is possible to feed a partitioning algorithm with the required information it needs to work correctly. The question that naturally comes next is “what if the traffic pattern is unknown or varies over time?” In both cases, “static” algorithms (including SZB) would not be capable of establishing proper zones boundaries and would not perform according to the expected.

In this context, we present the Cognitive Zone-Based (CZB) spectrum assignment algorithm. Our technique is capable of monitoring the network requisitions received by the network controller, acquiring information regarding the services using the network. With this data, CZB infers a compatible distribution that fits the received traffic, and then using the ratio between the types of services, calls an instance of the Static Zone-Based (SZB) algorithm, providing the services types using the network (St) and their ratio (Tr) as input.

At first, it is assumed that no information regarding the network traffic is known during the algorithm initialization. Therefore, the first step of the CZB algorithm is to acquire data regarding the traffic received. However, the network can not afford to reject requisitions while waiting for the calibration of the algorithm, thus needing to attend the arriving requests. Consequently, the use of an auxiliary algorithm is necessary during this “initialization phase.”

Although our method allows any spectrum assignment algorithm to be used as auxiliary algorithm during the initialization phase, after dozens of simulations performed, empirical results induced us to the conclusion that the combination of Spectrum Sharing and First-Fit algorithms (SS_FF) is a good solution, especially under lower loads. Furthermore, the First-Fit algorithm is easy to implement and has low complexity. Suppose that a network is starting its operation, and there are plenty of resources available. If the network load is low and no requests are being blocked, there is no need to use more robust algorithms, and the SS_FF is enough. Therefore it makes sense to use the Spectrum Share First-Fit algorithm as the auxiliary algorithm throughout the initialization of the CZB algorithm.

The Cognitive Zone-Based algorithm needs three inputs to work properly: the total number of slots available in each link (B_{max}), as described in Section II; the blocking threshold (blk_thr); and the size of its receiving window ($window\_size$). The blk_thr input is an integer representing the minimum number of requests that must be blocked to trigger a change in the partitions. Additionally, CZB has a “receiving window”, an array of size $window\_size$ that stores data regarding the arriving requests. Figure 2 shows how the algorithm works.

As requests arrive in the network, the CZB algorithm stores in its receiving window the number of slots (*num_slots*) needed to fulfill each request. After *window_size* requisitions, the algorithm verifies if the network rejected at least blk_thr of the latest $window\_size$ requests. If not, the CZB calls the active Spectrum Assigned (SA) algorithm (SZB or the auxiliary algorithm), that returns the slot index (if available) to allocate the request. This situation happens if current zones are good enough or if the traffic load is low.

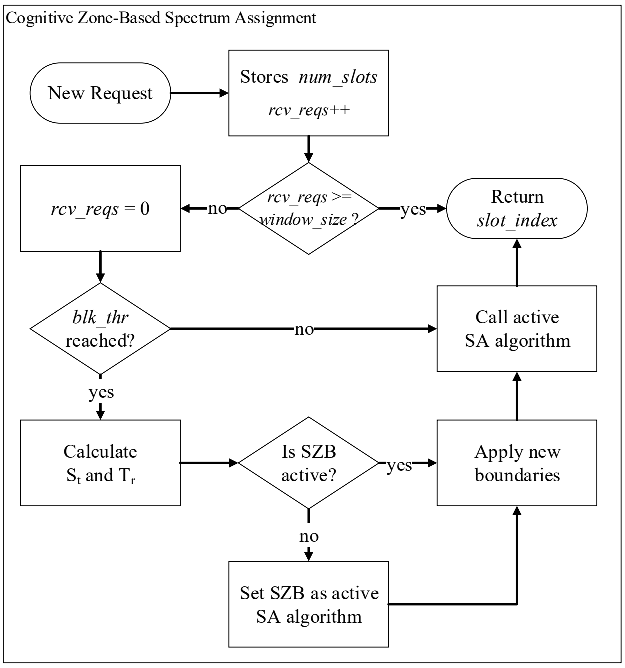


Fig. 2. Cognitive Zone-Based spectrum assignment algorithm.

If at least blk_thr requests were rejected, CZB reads the stored data of all $window\_size$ previous requisitions and estimates the ratio between the number of slots (num_slots) held. The estimation is done by accounting how many times each *num_slots* was requested; then normalizing by the lowest value, rounding the results to the nearest integer. The result of this step is a list of current network services (St) and their relative ratio (Tr). After the estimation of the traffic ratio, CZB verifies if SZB is the active SA algorithm, setting it as active if it is not. Next, SZB is called, passing St and Tr as input. The Static Zone-Based algorithm then uses the estimated traffic ratio and delimits the new boundaries of the zones, according to Equations (1) and (2). Finally, the active SA algorithm is called (now SZB) and the resulting slot index is returned as the result of the whole process.

It is important to notice that the bigger the receiving window size (*window_size*) is, the higher is the chance of estimating the traffic ratio correctly. However, it takes longer to the network to adapt the zones to the newly detected traffic pattern. There is an opportunity cost involved, a trade-off between precision and requests missed because of a wrong zone delimitation. In the next section, we present the simulations performed and the results obtained using the Cognitive Zone-Based algorithm.

IV. Simulation and Results

As Equations (1) and (2) show, both SZB and CZB, are dependent on the traffic transported by the network. Therefore, it is interesting to evaluate how the algorithm responds to different traffic patterns. The objectives of the proposed tests are to evaluate if CZB can detect the right traffic and adapt correctly to it, and also evaluate how SZB and CZB perform concerning fairness.

We propose two tests scenarios: test T1 uses a non-uniform traffic ratio between services, increasing the ratio of most demanding services (i.e., 400 Gbps and 1 Tbps), whereas test T2 utilizes a non-uniform traffic ratio between services increasing the proportion of less demanding services (i.e., 40 Gbps and 100 Gbps). Since in previous work [7] we already tested the SZB algorithm performance under uniform traffic ratio, we omit it in this paper.

All simulations were performed using the ElasticO++ framework [12], and the NSFNET topology, composed of 14 nodes and 21 bidirectional links [4]. A dynamic network operation scenario is simulated following the Erlang model with new requests arriving at λ Poisson rate and exponential holding time (with a normalized mean of $1/\mu = 1$). Network load is given by $\rho = \lambda/\mu = \lambda$ (Erlang). In each simulation run, $1x10^5$ requests are generated, and each chart point is represented by the average of 30 runs with different random seeds. Error bars represent the standard deviation.

Each new request is composed of a source, a destination, and a bitrate requirement. In all tests, four different service types are allowed in the network with bitrates of 40 Gbps, 100 Gbps, 400 Gbps, and 1 Tbps. Those bitrates are translated to the set of service types $St = \{3, 4, 7, 16\}$ (in slots) after applying an implementation of the DP-QPSK modulation format and considering a 10 GHz guard band, according to the Table 1 in [3]. Moreover, each test follows a distinct set of traffic ratio between services. Tests T1 and T2 utilize a non-uniform ratio of $Tr = \{1, 2, 3, 5\}$ and $Tr = \{5, 3, 2, 1\}$ respectively.

In each test, four Routing, Modulation, and Spectrum Assignment (RMSA) algorithms are compared. Each RMSA algorithm is composed of four parts: a routing algorithm, a modulation scheme, a spectrum management technique, and a spectrum assignment algorithm as described in [13]. Since the focus of this work is the spectrum management and assignment, and as an effort to reduce the number of variables, each RMSA algorithm tested shares the same routing algorithm and the modulation scheme, thus reducing this test to a SA (Spectrum Assignment) problem. The routing algorithm chosen is the Yen's K-Shortest Paths [14], using $K = 1$ for simplicity. Link lengths are also not considered in this test. Therefore, Yen's algorithm selects the shortest route by evaluating the number of hops. Finally, regarding the modulation scheme, it is assumed that DP-QPSK modulation can be assigned to all connections with no physical layer problems. The compared algorithms are:

- SS_FF: Spectrum Sharing First-Fit [15].
- SZB: Static Zone-Based.
- CZB27k and CZB45k: Cognitive Zone-Based with 27k and 45k sized receiving windows respectively. Both CZB27k and CZB45k use $blk\_thr = 100$.

Four metrics are used to evaluate algorithms performance in following tests: requests blocked rate (RBR), bitrate blocked rate (BBR), and two fairness metrics (RBR_{Sti} and RBR_{diff}). RBR is defined as $RBR = R_b/R_t$, where R_b represents the number of requests blocked at the end of the simulation and R_t is the total number of requisitions generated. Likewise, the BBR is given by $BBR = B_b/B_t$, where B_b is the total bitrate blocked, and B_t is the total bitrate requested. The first fairness metric is a comparison of the service requests blocked rates among all service types, and is defined as $RBR_{Sti} = R_{bSti}/R_t$, where R_{bSti} stands for the number of requests blocked of service type St_i and R_t is the total number of requisitions generated. The greater are the differences between the blocked rates the more unfair is the algorithm in the scenario analyzed. Finally, the second fairness metric reflects the difference between the maximum and minimum service requests blocked rates, obtained through: $RBR_{diff} = max(RBR_{Sti}) - min(RBR_{Sti})$.

A. Test T1 - Traffic Ratio (1:2:3:5)

The test T1 is set to use a non-uniform traffic ratio between service types, prioritizing the arrival of heavier demands. The traffic ratio used is $Tr = \{1, 2, 3, 5\}$. Using Tr values with $B_{max} = 336$ and $St = \{3, 4, 7, 16\}$, Equation (1) gives $C_{max} = 3$. The value $B_{max} = 336$ was chosen to prevent non-integer C_{max} values, that would imply in spectrum wasted by SZB and consequently by CZB, making this comparison less fair since SS_FF would have extra resources.

T1 results are presented in Figure 3 and are organized in the following manner: (a) and (b) show the request blocked rates (RBR) and bitrate blocked rates (BBR) respectively. The request blocked ratios (RBR_{Sti}) for each service type are shown in Figure 3 (c) SS_FF, (d) CZB27k, and (e) SZB; whereas (f) shows the difference between the maximum and minimum service requests blocked rates (RBR_{diff}).

In Figure 3 (a), the SS_FF algorithm presents the best performance concerning RBR, followed by CZB45k and CZB27k, and at last by SZB algorithm. Those results are expected since Dedicated Partition algorithms tend to block more due to the restrictive zone allocation policy. This extra blocking happens in situations where one zone is entirely occupied due to a burst of requisitions, while other zones may still have free resources. As SZB zone assignment is restrictive, requests can be blocked even when there are available resources in the network. It is interesting to see that before 125 Erlang, neither the CZB algorithms nor SS_FF experience requests blocked. In fact,

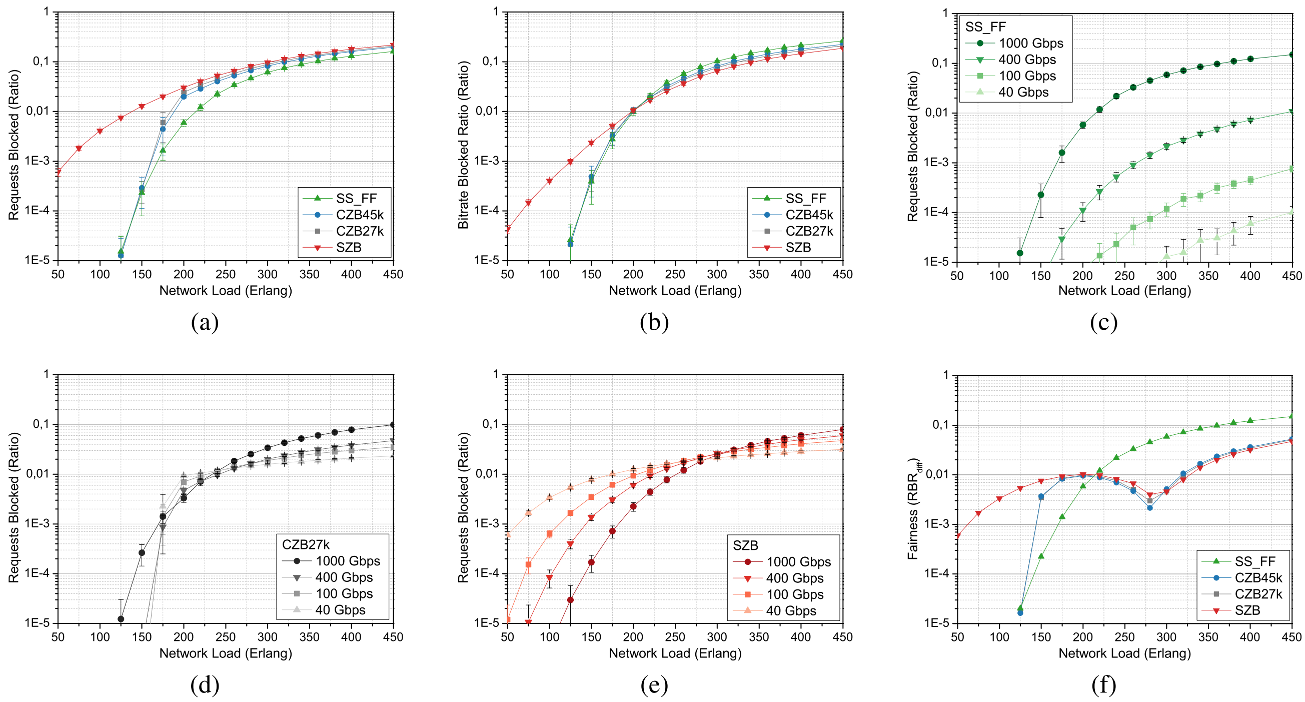


Fig. 3. Test T1 - traffic ratio (1:2:3:5) results: (a) Requests blocked ratio (RBR). (b) Bitrate blocked ratio (BBR). Requests blocked ratio distinguish per Service Type (RBR_{Sti}): (c) Spectrum Sharing First-Fit (SS_FF); (d) Cognitive Zone-Based (CZB27k). (e) Static Zone-Based (SZB); (f) Fairness comparison between the four tested algorithms (RBR_{diff}).

this is due to SS_FF being chosen as the auxiliary algorithm during CZB initialization phase. Since not enough requests were blocked (less then blk_thr), the SZB algorithm was not activated (Figure 2). For network loads of 125 Erlang and higher, the CZB algorithms achieve results between SS_FF and SZB, tending to the SZB results.

Similar behavior is observed in the BBR results (b). The SZB algorithm performs worst than SS_FF under light loads but starts performing better after ≈220 Erlang. We believe two reasons justify this behavior: the heavier traffic pattern and the fairness. As the strongest point of the SZB is to prevent unfairness, it enables a higher number of more demanding services to be accepted at the expense of less demanding services, thus reducing the total bitrate blocked. Once more, CZB results range between SZB and SS_FF curves. It should be noted that CZB improves SZB results under lower loads.

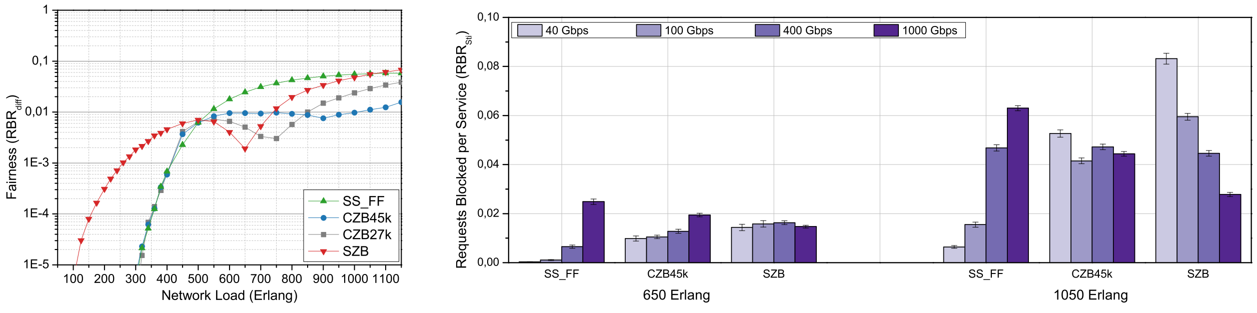
Regarding fairness, Figure 3 (c) presents the results of the Spectrum Sharing First Fit (SS_FF) algorithm. As SS_FF does not apply any spectrum management method, the higher is the bitrate requirement of the requisition, the higher is the probability of it being blocked. Comparing the SS_FF results with the results of CZB27k (d) and SZB (e), it is clear how the curves are more distant, thus indicating increased levels of unfairness. The differences between the maximum and minimum services blocked ratios (RBR_{diff}) are plotted in (f). It can be seen that for loads smaller than 125 Erlang, both CZB algorithms and SS_FF obtain same fairness levels. Between 125 and 200 Erlang, it is possible to see the CZB transition between SS_FF and SZB. In this same range, the SS_FF algorithm obtains the best results. We believe this happens due to SS_FF overall lower RBR at this load (≈ 0.6% of blocked requests). Above ≈220 Erlang (≈ 1.2% of blocked requests), the other algorithms start to outperform the SS_FF in this fairness metric. It is also noticeable that between 220 and 300 Erlang, both CZB algorithms achieve better results than SZB and SS_FF, up to 10 times better fairness under 260 Erlang (≈ 3.4% of blocking ratio) when compared to SS_FF.

Finally, it is interesting to notice that the obtained results for CZB45k are more similar to the SS_FF results than the CZB27k. That is explained by the receiving window size of the algorithms. Since CZB45k has a bigger receiving window size, it takes longer to switch from the auxiliary algorithm (SS_FF) to SZB. Moreover, it is intriguing to think CZB as a combination of SS_FF and SZB, or, in a broader sense, as a combination of the auxiliary algorithm and SZB. This combination is especially impactful depending on the ratio between the total number of requests of the simulation and the receiving window.

B. *Test T2 - Traffic Ratio (5:3:2:1)*

In Test T2 we simulate a scenario with a non-uniform traffic ratio between service types, prioritizing the arrival of lighter demands, i.e., $Tr = \{5, 3, 2, 1\}$. We use $B_{max} = 342$ to obtain an integer value for C_{max} ($C_{max} = 6$). Since the smallest services are the most abundant in the network, it is expected to SP_FF algorithm obtain better fairness results when comparing to the previous test. We omitted some charts since the results follow a similar behavior from the observed on test T1.

The fairness metric RBR_{diff} is shown in Figure 4 (a). It is interesting to note that since SS_FF is more suited to the Tr



(a) (b)

Fig. 4. Test T2 - traffic ratio (5:3:2:1) results: (a) Fairness comparison between the four tested algorithms (RBR_{diff}). (b) Requests blocked ratio distinguish per Service Type (RBR_{Sti}) under 650 Erlang and 1050 Erlang load.

used in this test, and since CZB is a combination of SS_FF and SZB, from ≈750 Erlang, both CZB27k and CZB45k start to outperform SZB in fairness. Figure 4 (b) shows two points of interest of (a): the 650 Erlang load in which the SZB has its local minimum value, and the 1050 Erlang load where SS_FF obtains a better result then SZB.

Two points in Figure 4 are worth a mention. First, in (b) the ascending trend observed in the SS_FF algorithm under 1050 Erlang is counterbalanced by the descending trend of the SZB algorithm, culminating in the fairer result obtained by CZB45k. This observation reinforces our conclusion that CZB is a combination of those algorithms. Second, even though the SS_FF result in (a) matches the SZB result under higher loads, a more in-depth analysis reveals the differences among the services blocked ratio are more evenly distributed in SZB algorithm than in SS_FF, Figure 4 (b) 1050 Erlang. However, this result compels us to investigate the fairness in non-uniform situations further and consider novel ways to establish the partitions other than following the traffic ratio linearly.

V. Conclusion

In this work, we present the Cognitive Zone-Based (CZB) spectrum assignment algorithm. Our algorithm is capable of observing the network traffic and acquiring information regarding the network services, thus using it to calibrate the division of the spectrum into partitions. To achieve such a partitioning, the CZB algorithm uses an upgraded version of the Static Zone-Based spectrum assignment algorithm [7]. Our results show that CZB algorithm can indeed achieve its objective, and even improve the fairness under certain conditions. The fairness results are presented in Section IV, and as Figure 3 (f) shows, between 220 and 300 Erlang, CZB algorithms obtain the best results, achieving up to 10 times better fairness under 260 Erlang load (≈ 3.4% of blocking ratio) when compared to the Spectrum Sharing First-Fit algorithm (SS_FF). Those results are confirmed in test T2 (Figure 4) in which CZB algorithms also achieves the best fairness results under higher loads (after 700 Erlang load).

Acknowledgment

This work was supported in part by FAPES and by CAPES Foundation, Ministry of Education of Brazil.

References

[1] C. V. N. Index, "Forecast and methodology, 2016–2021," *Technical Report, Cisco, Tech. Rep.*, 2017. [Online]. Available: "https://www.cisco.com/c/dam/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.pdf"

[2] R. Wang, S. Bidkar, R. Nejabati, and D. Simeonidou, "Load-aware nonlinearity estimation for efficient resource allocation in elastic optical networks," in *International Conference on Optical Networking Design and Modeling (ONDM)*, 2017, pp. 1–6.

[3] O. Gerstel, M. Jinno, A. Lord, and S. B. Yoo, "Elastic optical networking: A new dawn for the optical layer?" *IEEE Communications Magazine*, vol. 50, no. 2, pp. s12–s20, 2012.

[4] R. Wang and B. Mukherjee, "Provisioning in elastic optical networks with non-disruptive defragmentation," in *International Conference on Optical Networking Design and Modeling (ONDM)*, 2013, pp. 223–228.

[5] S. Talebi, F. Alam, I. Katib, M. Khamis, R. Salama, and G. N. Rouskas, "Spectrum management techniques for elastic optical networks: A survey," *Optical Switching and Networking (OSN)*, vol. 13, pp. 34–48, 2014.

[6] R. Wang and B. Mukherjee, "Spectrum management in heterogeneous bandwidth optical networks," *Optical Switching and Networking (OSN)*, vol. 11, pp. 83–91, 2014.

[7] R. S. Tessinari, B. Puype, D. Colle, and A. S. Garcia, "Zone based spectrum assignment in elastic optical networks: A fairness approach," in *Opto-Electronics and Communications Conference (OECC)*, 2015, pp. 1–3.

[8] N. Hara and T. Takahashi, "A study on dynamic spectrum assignment for fairness in elastic optical path networks," in *International Conference on Signal Processing and Communication Systems (ICSPCS)*, 2014, pp. 1–7.

[9] R. A. Scaraficci and N. L. Da Fonseca, "Alternative routing and zone-based spectrum assignment algorithm for flexgrid optical networks," in *IEEE International Conference on Communications (ICC)*, 2014, pp. 3295–3300.

[10] J.-L. Izquierdo-Zaragoza, P. Pavon-Marino, and M.-V. Bueno-Delgado, "Distance-adaptive online rsa algorithms for heterogeneous flex-grid networks," in *International Conference on Optical Network Design and Modeling (ONDM)*, 2014, pp. 204–209.

[11] R. W. Thomas, D. H. Friend, L. A. DaSilva, and A. B. MacKenzie, "Cognitive networks," in *Cognitive radio, software defined radio, and adaptive wireless systems*, 2007, pp. 17–41.

[12] R. S. Tessinari, B. Puype, D. Colle, and A. S. Garcia, "ElasticO++: An elastic optical network simulation framework for OMNeT++," *Optical Switching and Networking (OSN)*, vol. 22, pp. 95–104, 2016.

[13] R. S. Tessinari, D. Colle, and A. S. Garcia, "A defragmentation-ready simulation framework for elastic optical networks," *Journal of Communication and Information Systems (JCIS)*, vol. 32, no. 1, 2017.

[14] J. Y. Yen, "Finding the k shortest loopless paths in a network," *Management Science*, vol. 17, no. 11, pp. 712–716, 1971.

[15] M. Jinno, B. Kozicki, H. Takara, A. Watanabe, Y. Sone, T. Tanaka, and A. Hirano, "Distance-adaptive spectrum resource allocation in spectrum-sliced elastic optical path network," *IEEE Communications Magazine*, vol. 48, no. 8, pp. 138–145, 2010.

Control Plane Robustness in Software-Defined Optical Networks under Targeted Fiber Cuts

Jing Zhu[*,†], Carlos Natalino[†], Lena Wosinska[†], Marija Furdek[†], Zuqing Zhu[*]
[*]School of Information Science and Technology, University of Science and Technology of China, Hefei, China
[†]Optical Networks Laboratory (ONLab), KTH Royal Institute of Technology, Sweden
[*]Email: zhujinng@mail.ustc.edu.cn, {carlosns, wosinska, marifur}@kth.se, zqzhu@ieee.org

***Abstract*—The Software-Defined Optical Networking (SDON) paradigm enables programmable, adaptive and application-aware backbone networks. However, aside from the manifold advantages, the centralized Network Control and Management in SDONs also gives rise to a number of security concerns at different network layers. As communication between the control and the data plane devices in an SDON utilizes the common optical fiber infrastructure, it can be subject to various targeted attacks aimed at disabling the underlying optical network infrastructure and disrupting the services running in the network.**

In this work, we focus on the threats from targeted fiber cuts to the control plane (CP) robustness in an SDON under different link cut attack scenarios with diverse damaging potential, modeled through a newly defined link criticality measure based on the routing of control paths. To quantify the robustness of a particular CP realization, we propose a metric called Average Control Plane Connectivity (ACPC) and analyze the CP robustness for a varying number of controller instances in master/slave configuration. Simulation results indicate that CP enhancements in terms of controller addition do not necessarily yield linear improvements in CP robustness but require tailored CP design strategies.

***Index Terms*—Control plane robustness, Physical-layer security, Software-defined optical networks, Targeted fiber cuts.**

I. Introduction

Optical backbone networks are critical communication infrastructure supporting a variety of vital network services. In order to enable programmable, scalable and flexible network control and management (NC&M), Software-Defined Networking (SDN) has been proposed to decouple the network control and data planes (CP and DP), such that the NC&M tasks are handled by logically centralized controllers while the DP devices only take care of packet forwarding/data transmission [1], [2]. Hence, implementing Software-Defined Optical Networks (SDONs) enables flexible and programmable optical backbone networks, and significantly shortens the time-to-market of new services [3], [4]. Similar to its packet-based counterparts, the CP of an SDON uses centralized controllers to collect the statuses and configure the operation of DP devices (*e.g.*, optical transponders and switches) [5].

One of the essential aspects in SDON planning is the CP design [6]. As each fiber link in an SDON can carry Tb/s traffic, a well-designed CP should be able to simultaneously satisfy the requirements on low communication latency and high reliability of the control channels [7]. In general, the CP comprises one or multiple controller instances and each of them controls a subset of DP devices. Each DP device can connect to multiple controller instances, typically two, with one serving as master and the other as slave (Fig. 1). Several studies have addressed resilient SDN control plane design [6], [8]–[11]. Nevertheless, all of them considered CP disruptions due to random failures only, whereas the damage caused by deliberate attacks has not been investigated yet.

Optical networks are subject to physical-layer vulnerabilities which can be leveraged by malicious users to launch attacks aimed at service disruption [12]. In SDON, such attacks can affect not only the data plane communication, but may seriously disrupt the control plane as well. As the network 'brain', the robustness of the control plane is an important prerequisite for robust SDON deployment. The damaging potential of attacks can be boosted by design of attack techniques, e.g., by targeting the most critical components. In particular, we focus on deliberate fiber cut attacks where an attacker cuts the most critical links in an effort to maximize the CP communication disruption. Targeted fiber cuts have a larger disruptive effect than random failures [13], and are more challenging to address through careful network design.

In our previous work [14], we have investigated the robustness of data plane communication to targeted link cuts. In this paper, we consider the threats from targeted fiber cuts to the control plane and evaluate the CP robustness in an SDON from the perspectives of connectivity and transmission distance. Our evaluation is based on two newly proposed metrics: *(i)* a link criticality measure that quantifies the importance of links to support the CP connections and *(ii)* the Average Control Plane Connectivity (ACPC) that evaluates the robustness of a specific CP realization (i.e., the controller placement and the routing of control channels over the optical fiber topology). We consider

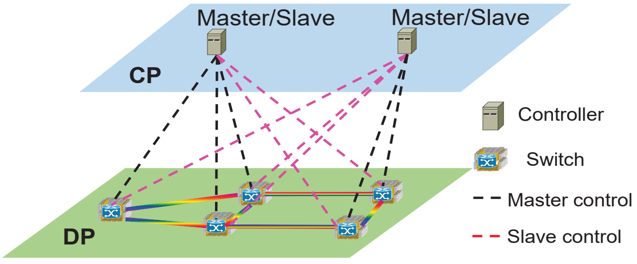


Fig. 1. An example of a software-defined optical network.

978-3-903176-07-2 © 2018 IFIP

two attack scenarios: one, where the attacker is not aware of the CP realization and, thus, uses general knowledge of the topology to select the targeted links to cut; and the other, where the attacker is aware of the CP realization and, thus, selects the most critical fibers to cut. Extensive simulation experiments are conducted for three realistic backbone topologies, where we analyze the CP robustness depending on the number of controller instances in the network and assess whether adding master/slave controller configuration to the switches can enhance the CP robustness. Results show that adding controller instances or considering master/slave configuration might not always lead to an increase in CP robustness, especially when the knowledge of the CP realization is available to the attacker.

The remainder of the paper is organized as follows. Section II reviews the related work. The proposed control plane connectivity measures are presented in Section III. Sections IV and V analyze network performance in the two considered attack scenarios, while Section VI concludes the paper.

II. Related Work

Since the inception of SDN, there have been intensive efforts on control plane design. The fundamental problem of CP design, i.e., how many controllers to deploy and where to place them, has been addressed in [15]. A comprehensive survey on fault management in SDN can be found in [16]. Control plane resiliency was investigated under various failure scenarios in [6], [8]–[11]. In [8], the authors proposed a method for controller placement aimed at maximizing the number of protected SDN switches. The work in [9] compared several controller placement schemes in terms of CP connectivity. The study in [10] considered failures of fiber links, switches and controllers, and designed an algorithm for Pareto-optimal controller placement with load balancing. Resilience from cascading controller failures was addressed in [11], by designing several algorithms to balance and redistribute the load among controllers. In [6], a survivable CP establishment scheme was proposed to protect SDONs against single node failures, utilizing a mutual backup model for the controllers. However, these studies did not consider failure scenarios caused by malicious man-made attacks.

Robustness of large-scale network topologies in the presence of targeted attacks was evaluated in [17]. Santos *et al.* [18] investigated the identification of critical nodes in a telecommunication network, *i.e.*, nodes whose removal would minimize the network connectivity. The work in [14] studied the robustness of optical content delivery networks in the presence of targeted fiber cuts, gauged by average content accessibility. As the aforementioned investigations only addressed survivability issues concerning the data plane, they cannot be directly mapped to assess the control plane robustness in SDONs. Attacks aimed at disabling control plane elements were investigated in [19], where the authors proposed a cost-efficient controller assignment algorithm to protect an SDN with multiple controllers from Byzantine attacks targeting controllers and control channels. They assumed that the attacker has complete knowledge about the CP realization, *i.e.*, the controller location and connectivity. The assumption of complete CP realization knowledge might not always be applicable because network operators typically try to prevent disclosing operational details. In this paper, we consider both cases, *i.e.*, the scenario where the attacker is aware only of the network topology, and the case where the attacker is also aware of the CP realization.

III. Control Plane Connectivity Measures

We consider a backbone SDON with topology modeled as a graph $G(V, E)$, where V denotes the set of nodes hosting switching elements, and E the set of undirected fiber links. We assume that the CP and DP of the SDON are supported by the same physical infrastructure, which means that the controllers are co-located with the optical switches, while the control channels share fiber links with data plane connections (i.e., in-band control). There are $|U|$ controller instances in the SDON, and the set U ($U \subset V$) represents their locations. To realize CP resiliency, each controller manages several optical switches, and each switch may connect to one or two controller instances, i.e., one master and one slave [7]. To reduce the control latency, each optical switch is assumed to connect to the physically closest controller instances.

In a targeted fiber link cut attack, the attacker deliberately chooses certain fiber links to cut. Link selection can be guided by different policies, depending on the knowledge and the aim of the attacker. However, a generalized strategy of an attacker would be to try to maximize the level of resulting disruption at a minimal effort, by selecting the links deemed most critical. Link criticality can be evaluated according to different criteria, such as topological properties or the number of carried connections. Link cuts can be performed simultaneously or sequentially. In this paper, we examine the former approach of simultaneously cutting a certain number of the most critical links, under different criticality considerations. If the set of intact fiber links upon an attack is denoted with E', the attack intensity can be quantified by the link cut ratio r:

$$r = \frac{|E| - |E'|}{|E|}. \tag{1}$$

Note that the targeted fiber cuts can disrupt the connectivity between switches and controllers, among the switches, and among the controllers. We focus on the case where the connectivity between switches and controllers is disrupted, which affects CP robustness in the SDON, i.e., the survivability of the control channels [6]. Here, we assume that the connectivity between a switch and its controller is lost if no path exists between them in $G(V, E')$ after the attack.

The following notation is used throughout the paper to assist CP robustness evaluation in SDONs.

- $x_{u,v}$: boolean variable that equals 1 if the optical switch at node v connects to the controller at node u, and 0 otherwise.
- $P_{u,v}$: the shortest path between the controller at node u and the optical switch at node v before the attack.

- $z_{u,v,e}$: boolean variable that equals 1 if link e is traversed by $P_{u,v}$, and 0 otherwise.
- $y_{u,v,r}$: boolean variable that equals 0 if, after an attack with cut ratio r, the connectivity between the optical switch at node v and the controller at node u is lost, and 1 otherwise.
- $P_{u,v,r}$: the shortest path between the controller at node u and the optical switch at node v after an attack with cut ratio r.
- $d_{u,v,r}$: the transmission distance of path $P_{u,v,r}$.

Using these notations, we define three metrics to measure link criticality with respect to the control plane, and to evaluate the CP robustness upon an attack with cut ratio r.

1) Link Criticality (L_c)

If the attacker is aware of the CP realization, the cut fiber links can be selected according to their importance to the CP. So far, there are no metrics that define the criticality of a link based on its importance to the CP. Therefore, we define link criticality L_c metric to quantify the importance of each link in the network proportionally to the number of traversing control channels. Formally, the metric is defined as:

$$L_c(e) = \sum_{u \in U, v \in V} x_{u,v} \cdot z_{u,v,e}. \quad (2)$$

2) Average Control Plane Connectivity (ACPC)

The ACPC quantifies the portion of network switches that can still connect to any of their controller instances (master or slave) after an attack. Formally, the ACPC upon an attack with cut ratio r can be calculated as:

$$ACPC(r) = \frac{\sum\limits_{u \in U, v \in V} x_{u,v} \cdot y_{u,v,r}}{|V|}. \quad (3)$$

3) Average Transmission Distance (ATD)

Besides connectivity, the latency of control channels is also a critical enabler of the efficient operation of an SDON. In optical networks, a significant portion of latency is related to the propagation of the optical signal in the fiber. Hence, transmission distance is a major factor for the latency. We define the ATD as:

$$ATD(r) = \frac{\sum\limits_{u \in U, v \in V} d_{u,v,r} \cdot x_{u,v} \cdot y_{u,v,r}}{|V|}. \quad (4)$$

Note that ATD is computed only for working control paths, i.e., those disrupted by the attack are not taken into account.

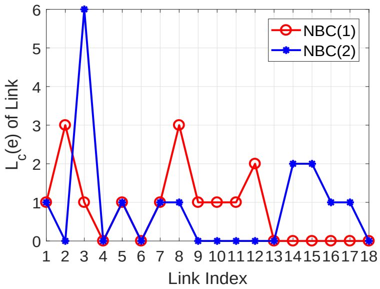


Fig. 3. $L_c(e)$ in Sprint topology with controllers placed according to NBC.

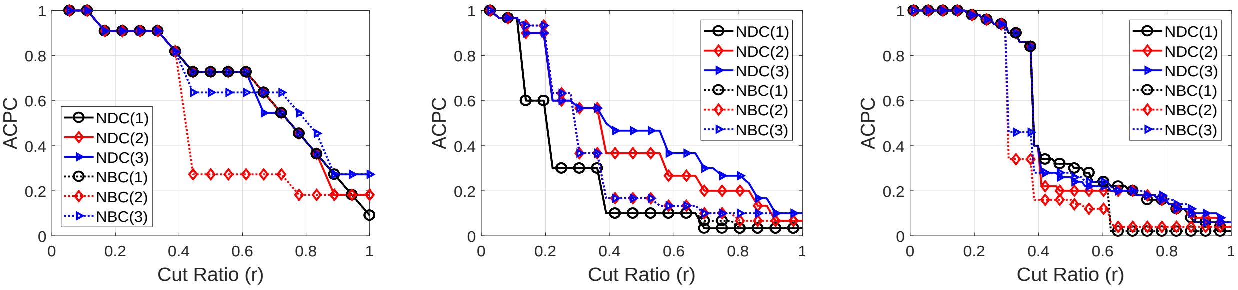
IV. Attack Scenario with No CP Realization Knowledge

Our simulation experiments are carried out using a custom-built Java-based tool that leverages GraphStream [20] for graph manipulation. We consider three realistic topologies whose characteristics are summarized in Table I. We consider two controller placement schemes, i.e., the Node Degree Centrality (NDC) and the Node Betweenness Centrality (NBC). The NDC scheme places the controller instances at the nodes with higher nodal degree. The NBC scheme places the controller instances at the nodes with higher node betweenness centrality, which refers to the number of all-node-pairs shortest paths traversing a node [21]. We first analyze how the number of controller instances in an SDON affects the CP robustness in the case where each optical switch only connects to its master controller (i.e., no slave controller is used). Then, we investigate whether considering master/slave controller configuration improves the CP robustness.

This section considers the less sophisticated attack scenario denoted as unavailable knowledge scenario (UKS) where the attacker has the knowledge of the physical network topology ($G(V, E)$, but does not know the details of the CP realization. According to [21], one effective scheme for selecting the most critical links is utilizing the link betweenness centrality, which

TABLE I
Topology Characteristics

Topology	Nodes	Links	Degree (± Deviation)	Diameter (hops)
Sprint [22]	11	18	3.27 (± 1.42)	4
USNET [23]	30	36	2.4 (± 0.6)	11
Germany [24]	50	88	3.5 (± 1.04)	9



(a) Sprint (b) USNET (c) Germany

Fig. 2. Average Control Plane Connectivity in the UKS scenario with single switch-controller assignment.

TABLE II
LINKS TO BE CUT WITH $r = 0.5$ IN SPRINT (CP CRITICAL LINKS IN RED)

Scenario	Link Index
UKS(1)	[1, 2, 3, 5, 8, 12, 14, 15, 17]
UKS(2)	[1, 2, 3, 5, 8, 12, 14, 15, 17]

is defined as the number of the shortest paths between all node pairs that traverse a specific link. Hence, in UKS, we assume that the attacker aims at maximizing the disruption potential of the attack by targeting the fiber links with higher link betweenness centrality.

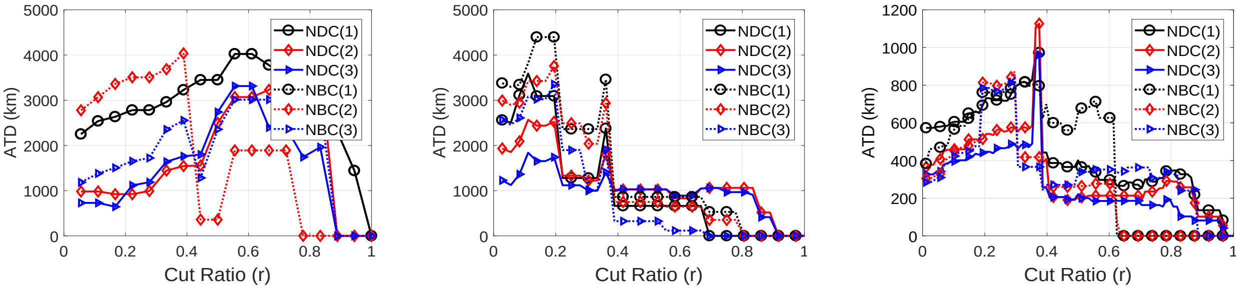
Fig. 2 shows the ACPC in the UKS scenario with single switch-controller assignment. It means that each switch is statically assigned to one (the closest) controller, and does not connect to any other controller even in the presence of attacks. Here, the curves in each plot correspond to a controller placement scheme with a certain number of controllers, e.g., "NDC(1)" represents the case where the SDON has one controller placed according to the NDC scheme. We observe that for a given number of controllers and a placement scheme, ACPC decreases for higher cut ratio r until it reaches the minimum, where the controller(s) are reachable only by its local optical switch(es) placed at the same node. However, there is a large variation in the impact of link cuts depending on the network topology. For instance, in USNET, when there is one controller, a drastic ACPC decrease occurs at around $r = 0.2$, while for Sprint and Germany the ACPC it does not drop significantly until about $r = 0.4$. The lower connectivity of USNET (as listed in Table I) makes this topology more vulnerable to targeted fiber cuts.

Interestingly, **in the UKS scenario with static single switch-controller assignment, a larger number of controllers does not guarantee a higher ACPC, and in some cases, ACPC can degrade with the number of controllers.** For example, in Fig. 2(a), when up to 7 links are cut ($r \leq 0.39$), the ACPC results are the same regardless of the number of controllers for both placement strategies. When we have r within $[0.44, 0.61]$, the ACPC for NBC(1) is higher than that for NBC(2). The same phenomenon can be observed by comparing the ACPC results for NDC(1) and NDC(3) at $r = 0.67$. These situations occur because when a controller is added to the network, the routing of control paths changes significantly. The control channels tend to be distributed more evenly over the links, which makes targeted attacks based on link betweenness centrality more effective.

To verify our analysis above,in Fig. 3 we show the link criticality w.r.t control plane $L_c(e)$ in the Sprint network for the scenarios that place 1 and 2 controllers with the NBC scheme. We also list the links that are selected by the link betweenness centrality with $r = 0.5$ in the UKS case in Table II. By cross-referencing the results in Fig. 3 and Table II, we find that the link betweenness centrality selects 6 and 7 links truly critical for the control plane in the two scenarios, respectively. Hence, **placing more controllers in an SDON that statically assigns single controllers does not necessarily improve the SDON robustness**.

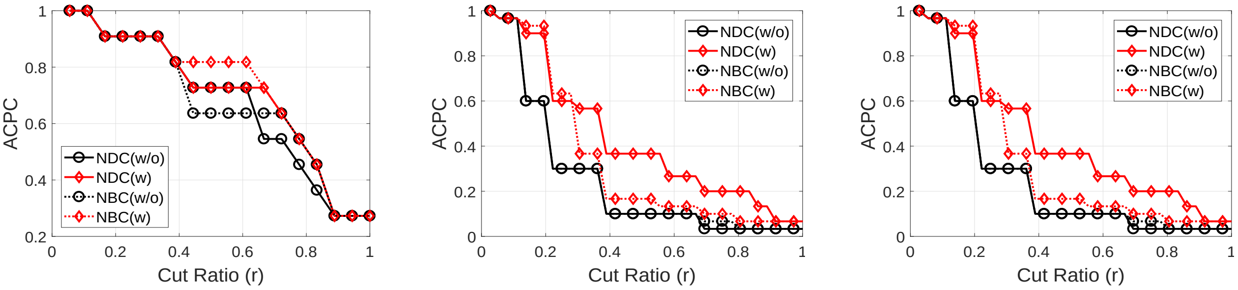
The ATD values for UKS with single switch-controller assignment are plotted in Fig. 4. A general observation is that CP needs to use longer paths as links are cut, leading to an increase in ATD. Recall that only working control paths are accounted. By ignoring the disrupted control paths, it is possible to measure the ATD for the control paths that remain connected. The drops in ATD shown in Fig. 4 are associated with drops in ACPC for the same cut ratio, i.e., cutting links tends to disrupt control path of the farthest switch(es), which leads to a decrease in the ATD for the remaining working control paths. For instance, in Fig. 4(a), when r increases from 0.06 to 0.11, the value of ACPC is 1 although there is an increase in ATD. Nevertheless, when r changes from 0.39 to 0.44, ATD for both NBC(2) and NBC(3) decreases due to a drop in ACPC, which accounts for the fact that the topology is no longer fully connected, and thus the surviving control channels can only take relatively shorter paths.

We also analyze whether CP robustness can be improved by considering a master/slave controller configuration for each optical switch. Fig. 5 shows the ACPC for the cases with single or master/slave switch-controller assignment. The number of controller instances placed in the network is set to 3. In the master/slave controller configuration, two controllers are assigned to each optical switch. Every controller instance can simultaneously act as the master for one set of switches and the slave for another set. Results indicate that considering master/slave controller assignment tends to increase the ACPC. However, such benefits are observed at different cut ratios depending on the network topology. These results suggest



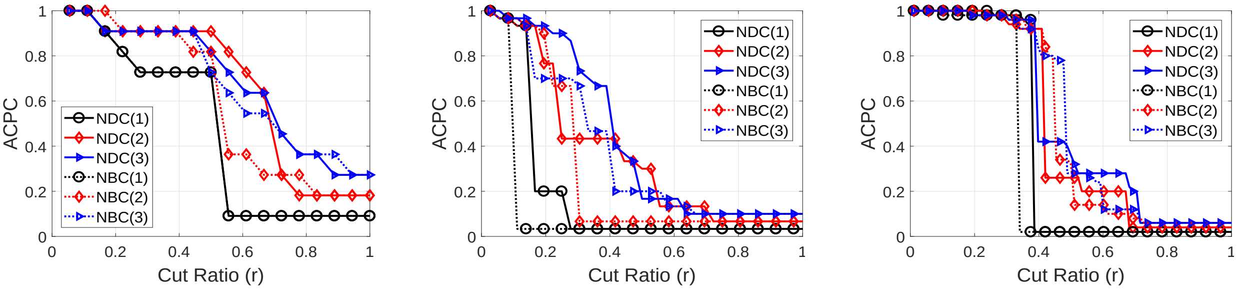
(a) Sprint (b) USNET (c) Germany

Fig. 4. Average Transmission Distance in the UKS scenario with single switch-controller assignment.



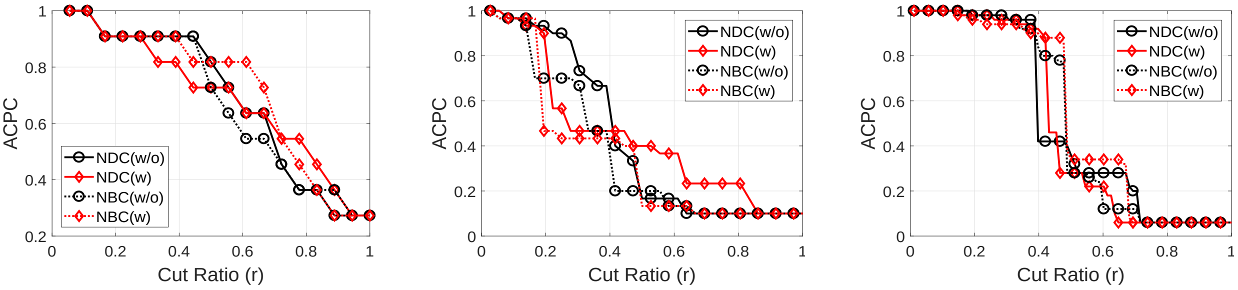
(a) Sprint (b) USNET (c) Germany

Fig. 5. Comparison of the ACPC for UKS with single and master/slave switch-controller assignment (3 controllers in the SDON).



(a) Sprint (b) USNET (c) Germany

Fig. 6. Results on ACPC for AKS with single switch-controller assignment.



(a) Sprint (b) USNET (c) Germany

Fig. 7. ACPC for the AKS scenario with single and master/slave switch-controller assignment (3 controllers in the SDON).

that **by considering master/slave controller configuration in UKS, the ACPC can be enhanced**. This can be easily understood since in UKS, the importance of links targeted by the attack is independent of the existence of slave controllers.

V. Attack Scenario with Full CP Realization Knowledge

In this section, we analyze the available knowledge scenario (AKS), where we assume that the attacker knows the details of the CP realization and is able to calculate $L_c(e)$ and selects the $\lfloor r \cdot |E| \rfloor$ links with the highest value of $L_c(e)$ to cut. Apart from the link selection strategy, this experiment follows the same setup as that in Section IV.

Fig. 6 shows the obtained ACPC for AKS with statically assigned single controller. In AKS, the general trend of ACPC with respect to r is similar to that of the UKS scenario. When comparing the curves for different number of controllers, it can be observed that **adding more controllers does not always improve the ACPC**. However, gains are visible in most cases. For instance, for Sprint and USNET, higher gains are observed when moving from 1 to 2 controller instances. Further addition of controllers still provides gains, but less pronounced. For example, the results in Fig. 6(a) indicate that for $r = 0.17$ and the NBC placement scheme, the ACPC decreases when the number of controllers increases from 2 to 3 (compare the curves of NBC(2) and NBC(3)). Moreover, in Fig. 6(a), the ACPC obtained for NBC(2) and NBC(3) is the same when r changes within $[0.22, 0.39]$, while for $r = 0.5$, the ACPC for NBC(2) is higher than that of NBC(3). This phenomenon can be explained as follows. When more controllers are placed in the SDON, the control channels of switches to different controllers may traverse the same links. Hence, when these links are cut, multiple control channels are interrupted. In Fig. 6, we observe that this phenomenon occurs more frequently in Sprint and Germany than in USNET. This is because they have larger deviations on nodal degree, which makes link sharing among control channels more common.

The ATD for AKS follows similar trends as in the UKS case, and is omitted for conciseness. Fig. 7 shows the ACPC for scenarios with single and master/slave switch-controller assignment when there are 3 controllers in the SDON. The results indicate that **considering master/slave controller configuration might not improve ACPC if the attacker has the knowledge of the CP realization**. At certain values of r, adding slave controllers can even degrade ACPC. For instance, when r ranges within $[0.06, 0.28]$ in Fig. 7(a), there is no improvement of ACPC for both controller placement schemes after considering a master/slave controller configuration. This can be explained by the fact that considering master/slave controller configuration generates more control channels and in turn makes certain links more vulnerable to targeted fiber cuts by increasing their $L_c(e)$.

VI. Conclusion

The paper considers the threats from targeted fiber cuts and evaluates control plane robustness in SDONs in terms of Average Control Plane Connectivity (ACPC) and Average Transmission Distance (ATD). Two attack scenarios were considered with different extents of control plane realization knowledge available to the attacker, and the impact of the number of controller instances to CP robustness was assessed. Moreover, two controller assignment configurations were considered: single or master/slave switch-controller assignment. For attacks with unknown CP realization and single controller assignment, adding more controllers does not guarantee an increase in ACPC, but adopting master/slave controller configuration benefits the CP robustness. When the attacker has the CP realization details, considering master/slave configuration or adding more controllers does not ensure improved ACPC. The extensive simulation results indicate strong necessity to protect the information related to the CP realization.

Acknowledgments

This work was supported in part by the NSFC Project 61701472, CAS Key Project (QYZDY-SSW-JSC003), NGB-WMCN Key Project (2017ZX03001019-004), China Postdoctoral Science Foundation (2016M602031), and Fundamental Research Funds for the Central Universities (WK2100060021).

C. Natalino, L. Wosinska and M. Furdek were supported in part by the RESyST project funded by the Unity through Knowledge Fund of the Croatian Ministry of Science, grant agreement no. 1/17, and the COST Action 15127 RECODIS.

References

[1] D. Kreutz, F. M. V. Ramos, P. E. Verłssimo, C. E. Rothenberg, S. Azodolmolky, and S. Uhlig, "Software-defined networking: A comprehensive survey," *Proc. IEEE*, vol. 103, pp. 14–76, Jan. 2015.

[2] S. Li, D. Hu, W. Fang, S. Ma, C. Chen, H. Huang, and Z. Zhu, "Protocol oblivious forwarding (POF): Software-defined networking with enhanced programmability," *IEEE Netw.*, vol. 31, pp. 12–20, Mar./Apr. 2017.

[3] Z. Zhu, C. Chen, S. Ma, L. Liu, X. Feng, and S. Yoo, "Demonstration of cooperative resource allocation in an OpenFlow-controlled multidomain and multinational SD-EON testbed," *J. Lightw. Technol.*, vol. 33, pp. 1508–1514, Apr. 2015.

[4] C. Chen, X. Chen, M. Zhang, S. Ma, Y. Shao, S. Li, M. S. Suleiman, and Z. Zhu, "Demonstrations of efficient online spectrum defragmentation in software-defined elastic optical networks," *J. Lightw. Technol.*, vol. 32, pp. 4701–4711, Dec. 2014.

[5] Z. Zhu, X. Chen, C. Chen, S. Ma, M. Zhang, L. Liu, and S. Yoo, "OpenFlow-assisted online defragmentation in single-/multi-domain software-defined elastic optical networks," *J. Opt. Commun. Netw.*, vol. 7, pp. A7–A15, Jan. 2015.

[6] B. Zhao, X. Chen, J. Zhu, and Z. Zhu, "Survivable control plane establishment with live control service backup and migration in SD-EONs," *J. Opt. Commun. Netw.*, vol. 8, pp. 371–381, Jun. 2016.

[7] X. Chen, B. Zhao, S. Ma, C. Chen, D. Hu, W. Zhou, and Z. Zhu, "Leveraging master-slave openflow controller arrangement to improve control plane resiliency in SD-EONs," *Opt. Express*, vol. 23, pp. 7550–7558, Mar. 2015.

[8] N. Beheshti and Y. Zhang, "Fast failover for control traffic in software-defined networks," in *Proc. of GLOBECOM*, Dec. 2012, pp. 2665–2670.

[9] Y. Hu, W. Wendong, X. Gong, X. Que, and C. Shiduan, "Reliability-aware controller placement for software-defined networks," in *Proc. of IFIP/IEEE IM*, May 2013, pp. 672–675.

[10] D. Hock, M. Hartmann, S. Gebert, M. Jarschel, T. Zinner, and P. Tran-Gia, "Pareto-optimal resilient controller placement in SDN-based core networks," in *Proc. of ITC*, Sept. 2013, pp. 1–9.

[11] G. Yao, J. Bi, and L. Guo, "On the cascading failures of multi-controllers in software defined networks," in *Proc. of ICNP*, Oct. 2013, pp. 1–2.

[12] N. Skorin-Kapov, M. Furdek, S. Zsigmond, and L. Wosinska, "Physical-layer security in evolving optical networks," *IEEE Commun. Mag.*, vol. 54, pp. 110–117, Aug. 2016.

[13] R. Albert, H. Jeong, and A. Barabasi, "Error and attack tolerance of complex networks," *Nature*, vol. 406, pp. 378–382, Jul. 2000.

[14] C. Natalino, A. Yayimli, L. Wosinska, and M. Furdek, "Content accessibility in optical cloud networks under targeted link cuts," in *Proc. of ONDM*, May 2017, pp. 1–6.

[15] B. Heller, R. Sherwood, and N. McKeown, "The controller placement problem," in *Prof. of HotSDN*, Aug. 2012, pp. 7–12.

[16] P. Fonseca and E. Mota, "A survey on fault management in software-defined networks," *IEEE Commun. Surveys Tut.*, vol. 19, pp. 2284–2321, Fourth Quarter 2017.

[17] S. Iyer, T. Killingback, B. Sundaram, and Z. Wang, "Attack robustness and centrality of complex networks," *PLoS ONE*, vol. 8, pp. 1–17, Apr. 2013.

[18] D. Santos, A. Sousa, and P. Monteiro, "Compact models for critical node detection in telecommunication networks," in *Proc. of INOC*, Feb. 2017, pp. 1–10.

[19] H. Li, P. Li, S. Guo, and A. Nayak, "Byzantine-resilient secure software-defined networks with multiple controllers in cloud," *IEEE Trans. Cloud Comput.*, vol. 2, pp. 436–447, Oct. 2014.

[20] Y. Pign, A. Dutot, F. Guinand, and D. Olivier, "Graphstream: A tool for bridging the gap between complex systems and dynamic graphs," in *Proc. of ECCS*, Sep. 2007, pp. 1–10.

[21] D. Rueda, E. Calle, and J. Marzo, "Robustness comparison of 15 real telecommunication networks: Structural and centrality measurements," *J. Netw. Syst. Manag.*, vol. 25, pp. 269–289, Apr. 2017.

[22] S. Knight, H. X. Nguyen, N. Falkner, R. Bowden, and M. Roughan, "The internet topology zoo," *IEEE J. Sel. Areas Commun.*, vol. 29, pp. 1765–1775, Oct. 2011.

[23] J. Simmons, *Optical Network Design and Planning*, 2nd ed. Springer, 2014.

[24] S. Orlowski, M. Pióro, A. Tomaszewski, and R. Wessäly, "Sndlib 1.0: Survivable network design library," in *Proc. of INOC*, Apr. 2007, pp. 1–11.

Fully Automated Peer Service Orchestration of Cloud and Network Resources Using ACTN and CSO

Ricard Vilalta, Ramon Casellas,
Ricardo Martínez, Raul Muñoz
Centre Tecnològic de Telecomunicacions
de Catalunya (CTTC/CERCA)
Email: ricard.vilalta@cttc.es

Young Lee, Haomian Zheng,
Yi Lin
Huawei

Victor López, Luis Miguel Contreras
Telefónica I+D/Global CTO

Abstract—**Current applications, such as on-line gaming or media content delivery, require dynamic bandwidth allocation and a tight integration between the network resources and computing resources. Service orchestration faces the challenge of combining and controlling the resources of these different stratums and optimizing them.**

This paper proposes the fully automated establishment of a network service using a peer inter-CSO interface in ACTN. The underlying network resources have been abstracted and virtualized in order to provide a network slice. We present the CSO/ACTN architecture and detail the main components. The system is implemented and demonstrated in an experimental testbed, where we characterize the setup delay of a virtual deep packet inspection service across several network domains using the proposed peer interface.

I. Introduction

Cloud computing is able to provide a variety of services (such as on-line gaming, media content delivery, network slicing). These services allow end-users to access large pools of compute and storage resources, enabling various application services (e.g., Video Caching, VM mobility, media content delivery, IoT, etc.). Cloud computing services is one of the faster emerging businesses for Internet Server Providers (ISP).

Data centers (DC) provide the physical and virtual infrastructure in which cloud computing applications are deployed and chained into end-services. The proposed services might be also aligned with Network Function Virtualization (NFV) services. Since the DCs used to provide services may be distributed geographically around a set of interconnected networks, service deployment can affect on the state of the network resources. Conversely the capabilities and current state of the network can have a major impact on the service performance; DCs have been spread geographically to reduce latency to the end user, and that has led into an exponential growth on the inter-datacenter traffic [1]. Consequently, DC interconnection is one of the major problems that service providers have to face, along the need to adapt the actual rigid and fixed transport networks, enabling them with the flexibility provided with the Software Defined Networking (SDN) architecture.

SDN is the solution to improve network programmability, including the dynamic allocation of the network resources. SDN proposes a centralized architecture where the control entity (SDN controller) is responsible for providing an abstraction of network resources through programmable Application Programming Interfaces (APIs). One of the main benefits of this architecture resides on the ability to perform control and management tasks of different network forwarding technologies such as packet/flow switches, circuit switching and optical wavelength switched transport technologies. In the IETF new data models are currently under definition, in order to allow the integration of abstracted traffic enginery (TE) information in the description of network resources [2] and in the allocation of these resources [3].

Moreover, Abstraction and Control of Traffic Engineered networks (ACTN) enables virtual network operations using abstraction and virtualization mechanisms [4]. It allows customers to request a virtual network over operators transport networks, which are often multi-layer and multi-domain TE networks. This virtual network is presented as an abstract topology to customers, in such a way that they can use this abstracted topology to offer applications over its virtual network. Therefore, ACTN enables multi-tenant virtual network services with flexibility and dynamicity. Along ACTN, and in order to deal with the joint allocation of DC and network resources, Cross Stratum Optimization (CSO) [5] involves a cooperation between the Application Stratum and Network Stratum to efficiently utilize cloud and network resources and provide for overall application level quality of service.

In this paper we present a possible CSO architecture involving multiple CSO service orchestrators, which interact in a distributed (peer) model with the objective of provisioning an End-to-End (E2E) service over multiple administrative network and cloud domains. A virtual Deep Packet Inspection (vDPI) service is provided as our use case, and we measure the setup delay in deploying the service using the cloud computing platform of the ADRENALINE testbed [6].

978-3-903176-07-2 © 2018 IFIP

II. State of the Art

In this section we will present the current state of the art of both the ACTN and CSO functional architectures, setting the basis for highlighting their complementarity and how they can interact and be used to fulfill our purpose of dynamic service establishment over peer-CSO service orchestrators using ACTN.

A. *Abstraction and Control of Traffic Engineered Networks*

ACTN provides an architecture for the virtualization of TE networks. The architecture defines multiple functional entities (controllers) according to their main functions and roles. In this setting, the types of controller defined in the ACTN architecture are shown in Fig. 1 and are as follows:

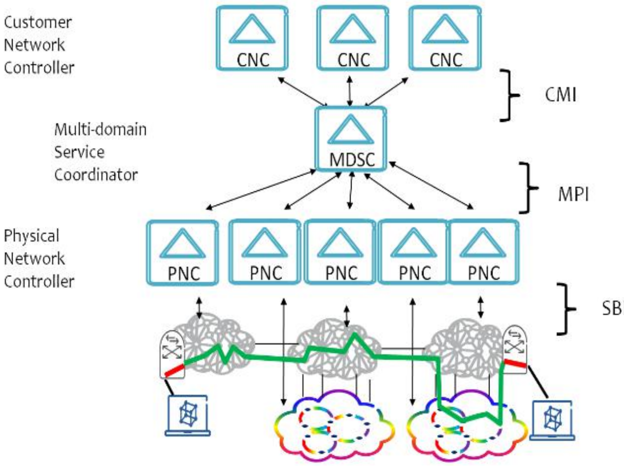


Fig. 1. ACTN architecture

1) Customer Network Controller (CNC): A Virtual Network is requested by the Customer Network Controller via the CNC-MDSC Interface (CMI). As the Customer Network Controller directly interfaces to the applications, it is able to understand multiple application requirements and their needs. It is assumed that the CNC and the MDSC have a common knowledge of the end-point interfaces based on their business negotiations prior to service instantiation.

2) Multi Domain Service Coordinator (MDSC): The Multi Domain Service Coordinator (MDSC) sits between the CNC that issues VN requests and the Physical Network Controllers (PNCs) that manage the physical network resources. The MDSC is the responsible for the following functions: multi-domain coordination, virtualization/abstraction, customer mapping/translation, and virtual service coordination. Multi-domain coordination and virtualization/abstraction are referred to as network control/coordination functions. While customer mapping/translation and virtual service coordination are referred to as service control/coordination functions. The key point of the MDSC is detaching the network and service control from underlying technology to help the customer express the network as desired by business needs. The MDSC provides the deployment and control of the right technology to meet business criteria. In essence it controls and manages the primitives to achieve functionalities as desired by the CNC. A hierarchy of MDSCs can be foreseen for scalability and administrative choices.

3) Physical Network Controller (PNC): The Physical Network Controller (PNC) refers to a standard SDN controller, which allows configuration of the network elements, monitoring the topology (physical or virtual) of the network, and passing information about the topology (either raw or abstracted) to the MDSC.

Moreover, a data model for the control of a data network has also been proposed in [7]. This data model can be applied as the NBI of a MDSC. It provides the basic elements for Create/Read/Update/Delete (CRUD) operations over ACTN VNs. It also describes two main ACTN elements:

- VN member: The VN can be understood as set of end-to-end tunnels from a customer point of view, where each tunnel is known as a VN member. Each VN member might be formed by recursive abstraction of paths in underlying networks.
- Access point (AP): An access point provides confidentiality between the customer and the provider. It is a logical identifier shared between the customer and the provider, used to map the end points of the border node in both the customer and the provider network. A set of APs are used by the customer when requesting VN service to the provider.

B. *Cross Stratum Optimization*

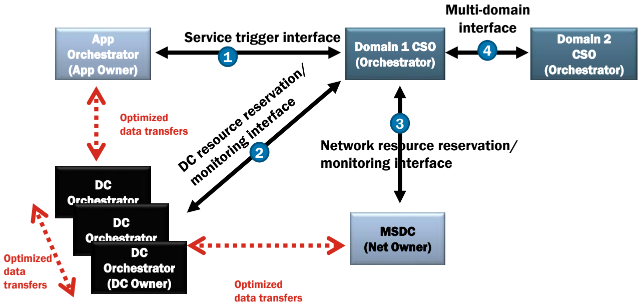


Fig. 2. CSO architecture

Cross Stratum Optimization (CSO) involves a cooperation between the Application Stratum and Network Stratum in order to optimize cloud and network resources and provide for overall application level quality of service.

The *Application stratum* is the functional grouping which considers application resources and the control and management of these resources. These application resources are used along with network services to provide an application service to customers. Application resources are non-network resources critical to achieving the application service functionality. Examples of application resources include: caches, mirrors, application specific servers, content, large data sets, and computing and storage power. Application service is a networked application offered to a variety of clients. Several application service examples are: server backup, VM migration, video cache, virtual network on-demand, 5G network slicing. The entity responsible for application stratum control and management of its resources is referred to as application orchestrator.

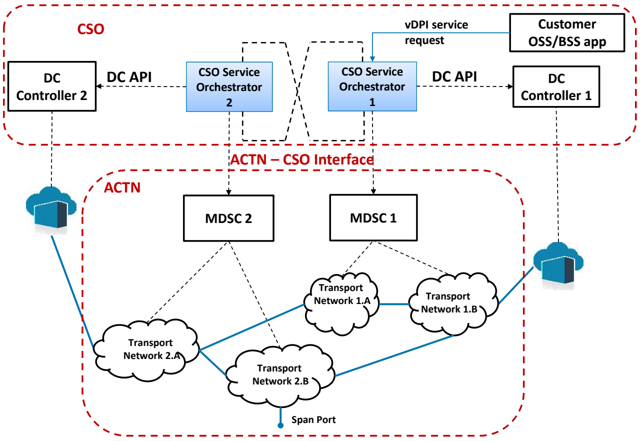


Fig. 3. Proposed ACTN-CSO architecture

The *Network stratum* is the functional grouping which includes network resources and the control and management of these resources providing transport of data between customers and application sources. Network resources are resources such as bandwidth, links, paths, path processing (creation, deletion, and management), network databases, path computation, admission control, and resource reservation. There are different types of network stratum controllers/orchestrators.

Figure 2 summarizes the CSO architecture, where the data center orchestrator provides its data center network resource abstraction pertaining to the applications to the CSO. Since application services may use resources in multiple data centers via data center interconnection, each data center orchestrator involving in the application service should provide its resource abstraction to the CSO so that the CSO would be able to compute the optimal resource sequence and path meeting the service objective.

The Wide Are Network (WAN) SDN controller(s) is(are) also involved for an end-to-end service instantiation and its life cycle operation that may traverse multiple data centers dispersed in multiple domains. WAN SDN controllers may also comprise multiple hierarchical SDN controllers, each of which is responsible for domain control of IP or optical networks in multiple domain networks.

The offered interfaces are the following:

- CSO interface type 1 is referred to as the Service Trigger Interface. This interface shall be able to describe service requirements, taking into consideration DC and network requisites.
- CSO interface type 2 is referred to as the DC Resource Reservation and Monitoring Interface. This interface on a high level abstracts a DC level resource abstraction as well as a host/server level resource abstraction needed per application.
- CSO interface type 3 is referred to as the Network Resource Reservation and Monitoring Interface. This interface should provide several functionalities such as: a) abstraction of the network resource information of the operators transport networks providing DC interconnection; b) service connection reservation request; c) monitoring data and measurement pertaining to the service connection among the key requirements.
- CSO interface type 4 is referred to as multi-domain CSO interface.

III. Proposed ACTN/CSO Joint Architecture

Figure 3 shows the proposed architecture, which is able to jointly orchestrate IT and network resources. It consists of four main building blocks: the customer application, CSO service orchestrator, DC controller and Multi-domain Service Coordinator (MDSC). The CSO service orchestrator offers the NorthBound Interface (NBI) for the dynamic service deployment, considering the necessary joint orchestration of IT and network resources.

Coupling the ACTN with the CSO, an end-to-end automated orchestration of services/applications is made possible with the joint optimization of cloud and network resources the applications consume. The MDCS of the ACTN architecture acts as a network orchestrator of multi-layer and multi-domain networks. A CSO Orchestrator is able to request virtual networks to each MDSC. The CSO is the entity that has the application requirements knowledge and the necessary cloud/DC resources associated with the application. A CSO in one operator domain interacts with another operator domains CSO as the applications are provided across multiple operator domains.

The CSO service orchestrator is responsible for peer coordination with other CSO service orchestrators in other administrative domains. Moreover, each CSO service orchestrator is responsible for its domain network and cloud resources. In the following subsections, CSO service orchestrator is detailed.

The DC controller is responsible for the creation/migration/deletion of VM instances (computing service), disk images storage (image service), and the management of the VM network interfaces (networking service). The computing service (e.g., Nova in OpenStack) is responsible for the management of the VM into the compute hosts (i.e., hypervisors). A compute service agent is running in each host and controls the computing hypervisor (e.g., KVM) responsible of the creation/deletion of the VMs. The image service (e.g., Glance in OpenStack) handles the disk images which are used as templates for VM file systems; it also operates in a centralized manner by maintaining a copy of all the disk images in the DC controller. An image-service agent is running in each host to request the download of images when a new VM instantiation requires it. It also permits to create new images of the currently working instances, a process known as snapshotting, which is used for VM migration. Finally, the connectivity between VMs and virtual switches inside the hosts is managed by the networking service (e.g., Neutron in OpenStack). It creates the virtual interfaces, attaches them into the virtual switches, such as OpenVSwitch, and it offers a DHCP service for the VMs to get the assigned IP address. The CSO service orchestrator controls the DC controller through a RESTful API (e.g., OpenStack API), which is used to trigger the DC controller actions and to get the necessary information about the running VM instances.

The MDSC is introduced in order to support end-to-end connectivity by orchestrating the different network technologies or control domains. In the proposed network architecture, the inter-DC network is controlled using IETF TEAS topology [2] and tunnel [3] data models using RESTconf protocol [8]. We have based the MDSC architecture on the IETF ABNO [9]. The ABNO is the IETF reference architecture for SDN controllers, where it reuses standardized components, such as path computation element (PCE). The Topology Server, included in the ABNO, is the component responsible of gathering the network topology from each control domain and building the whole network topology which it is stored in the Traffic Engineering Databased (TED). The TED includes all the information about network links and nodes, and it is used by the PCE for calculating routes across the network. The Virtual Network Topology Manager (VNTM) is the responsible for managing the multi-layer provisioning. In the proposed architecture, the VNTM will arrange the set-up of an optical connection and offer it as a L2 logical link to satisfy the L2 connectivity demand. The Provisioning Manager implements the different provisioning interfaces to push the forwarding rules and the establishment of segments into the data plane. Flow server stores the connections established in the network into a FlowDB. Finally the Network Orchestration Controller handles all the processes involved inside the MDSC to satisfy the provisioning of end-to-end connectivity.

A. Customer Network Service and Inter-CSO Peering model

Figure 4 shows the customer view offered from CSO service orchestrator 1 to the Customer application. A virtual network is provisioned using ACTN, through requesting several VN members. Each VN member is an e2e tunnel from the customer perspective. It can be observed that the served VN is an abstract construct on top of Transport Network 2.B and 1.B, each from a different domain and controlled by a different peer CSO service orchestrator. The costumer application only has the necessary information regarding the deployed application and the VN interconnecting it. In this case the customer application is a vDPI, where a span port (where the traffic to be analyzed is captured) is interconnected to the vDPI application running in DC1.

B. Proposed CSO service orchestrator architecture

Figure 5 shows the internal architecture for CSO service orchestrator. It can be observed that the main component is the internal orchestrator, which is the responsible for providing the necessary workflows in order to deploy, modify and tear down services. On top, four NBIs are offered:

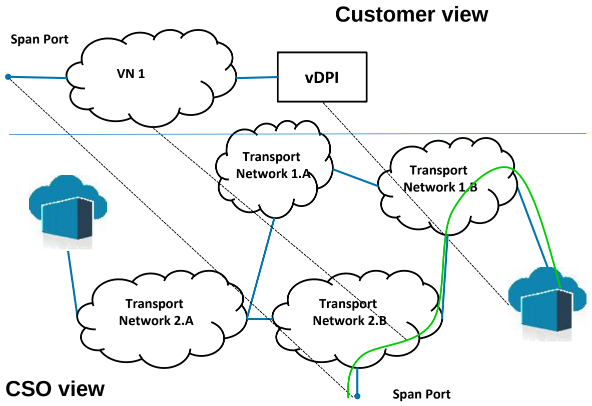


Fig. 4. Customer's view from application perspective

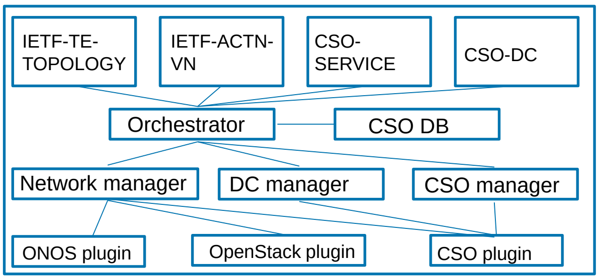


Fig. 5. Proposed CSO service orchestrator architecture

1) CSO Service: This interface is offered for the end-user. It provides network intents and compute intents in order to scale in/out the requested service, which might include several applications.

2) CSO-DC: This interface is used to provide computing and storage resources to end-user and to other CSO service orchestrators. It provides an abstraction to OpenStack API, in order to offer a generalized compute and storage service, which might be offered using different underlying controllers.

3) ACTN VN: It provides VN resources to end user and to other CSO service orchestrators. A list of access points is provided, which represent the possible end points of the VN members.

4) TE Topology: A more detailed view of network topology might be obtained using this interface by a peer CSO in order to perform better network resource optimization.

The CSO database (DB) is used to keep synchronized all the information related to peer CSOs, as well as the information regarding the underlying controlled DCs and MDSCs. Finally, the Network, DC, and CSO manager are the responsible for each of the underlying resources, and they use the necessary plugins (to consume the NBI defined for each OSS, such as

ONOS and OpenStack).

C. Resource allocation algorithm which considers Compute and Network resources

A heuristic algorithm for the CSO has been developed in is further explained in [10]. Once the CSO receives a request, it allocates the necessary compute and networking resources by using the previously recovered information. The CSO heuristic algorithm receives the following inputs: Network Service requested (including compute requirements, and network interconnections and span ports). The algorithm outputs are the VM allocation per DC, and the necessary VM interconnections though transport networks.

D. Fully Automated Service workflows

Figure 6 shows the information recovery workflow from the CSO service orchestrator 1 (CSO1) perspective in order to discover all the cloud and network available resources in a peer hierarchical approach as has been presented in [11]. Four steps are included in this process:

1) Discovery of underlying DC resources using OpenStack API [12] towards DC controller.
2) Discovery of underlying Network resources using TE Topology towards MDSC1.
3) Discovery of peer CSO service orchestrator 2 (CSO2) DC resources using CSO DC interface. CSO2 will initiate discovery of underlying DC resources.
4) Discovery of available access points in CSO2 in order to provide network resources.

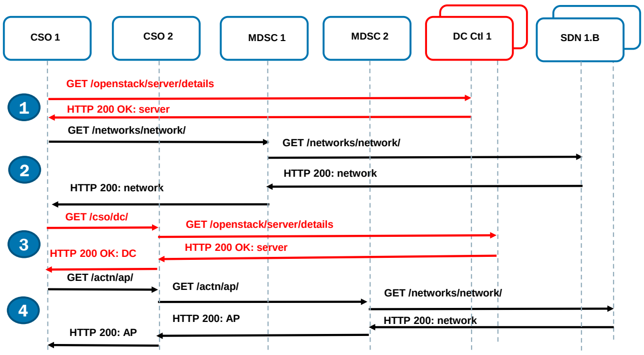


Fig. 6. Message exchange for inititializing CSO1 and CSO2

Figure 7 show the message exchange in order to dynamically deploy a service, such as the depicted in the reference scenario (Fig. 3). It consists of the following steps:

1) A customer requests a service to CSO1.
2) The resource allocation algorithm is triggered and allocates the necessary VM in DC1.
3) It also computes the necessary virtual network members, and it triggers the virtual network creation. CSO1 requests to CSO2 a virtual network, in order to reach the requested span port. CSO2 forwards the request to the MDSC under its domain.
4) A virtual network is directly requested to MDSC1. Each request triggers the creation of a TE tunnel.
5) Finally, the CSO1 offers the resource view to the customer, as seen in Figure 4.

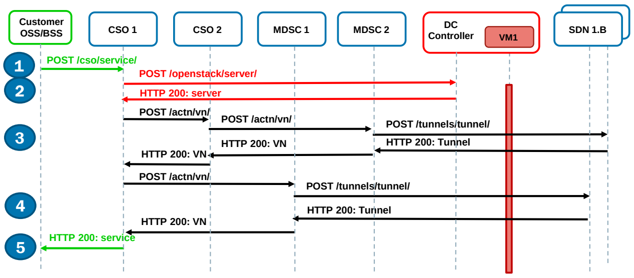


Fig. 7. Message exchange for provisioning service

IV. Experimental Demonstration

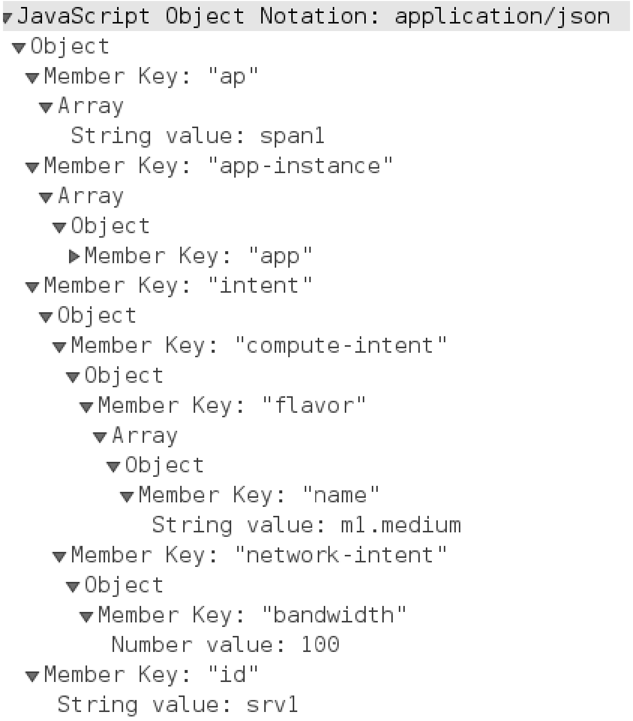
The proposed architecture has been validated in the Cloud Computing platform of the ADRENALINE Testbed. The OpenStack Havanna release has been deployed into five physical servers with 2 x Intel Xeon E5-2420 and 32GB RAM each, one dedicated to the cloud controller and the other four as compute pool (hypervisors) for VM instantiation [13]. The network resources have been emulated using mininet and ONOS, on top of which we have included an agent to provide the necessary TE topology and ACTN VN interfaces. The CSO service orchestrator has been developed using python and its interfaces how been generated using swagger codegen.

In Figure 8, a JSON object of the requested service can be observed. The span port is included in the access point array, as well as the network and compute intents (including bandwidth and flavor, respectively). The application to be deployed (in this case vDPI) is described in app-instance array.

Figure 9 shows the HTTP conversation between CSO1 and CSO2 service orchestrators at their initialization. Firstly, it can be observed how CSO1 loads the information from underlying DC controller 1 (OPS1). Secondly, CSO1 requests to MDSC1 the topological network information. Thirdly, DC status of resources is requested to CSO2. Finally, access points are requested.

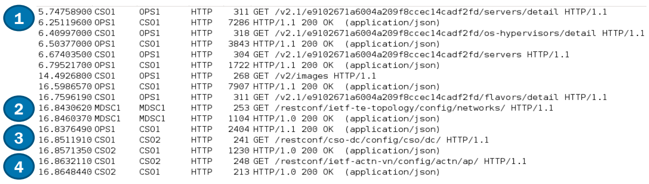
Figure 10 shows the capture HTTP message exchange between CSO1 and CSO2 in order to provision a service requested by client. Firstly, we can observe the requested service. Secondly, A VM is deployed in order to support the requested application. CSO1 checks the status of the VM until the VM is fully available. Thirdly, CSO1 request to MDSC1 the creation of a VN. Fourthly, CSO1 requests CSO2 the creation of a VN, considering the span access point. Finally, the service is provided to the client.

It can be observed in Figure 10 that the setup delay to provision a service is of 14.7 seconds, being the main delay contributor, the span of the VM to deploy the requested application. The necessary networking resources have been allocated in 132 milliseconds.



```
▼JavaScript Object Notation: application/json
 ▼Object
  ▼Member Key: "ap"
   ▼Array
     String value: span1
  ▼Member Key: "app-instance"
   ▼Array
    ▼Object
     ▶Member Key: "app"
  ▼Member Key: "intent"
   ▼Object
    ▼Member Key: "compute-intent"
     ▼Object
      ▼Member Key: "flavor"
       ▼Array
        ▼Object
         ▼Member Key: "name"
           String value: m1.medium
    ▼Member Key: "network-intent"
     ▼Object
      ▼Member Key: "bandwidth"
        Number value: 100
  ▼Member Key: "id"
    String value: srv1
```

Fig. 8. Wireshark capture of requested service JSON object



```
5.74758900 CS01    OPS1    HTTP   311 GET /v2.1/e9102671a6004a209f8ccec14cadf2fd/servers/detail HTTP/1.1
6.25119600 OPS1    CS01    HTTP  7286 HTTP/1.1 200 OK  (application/json)
6.40997000 CS01    OPS1    HTTP   318 GET /v2.1/e9102671a6004a209f8ccec14cadf2fd/os-hypervisors/detail HTTP/1.1
6.50377000 OPS1    CS01    HTTP  3843 HTTP/1.1 200 OK  (application/json)
6.67403500 CS01    OPS1    HTTP   304 GET /v2.1/e9102671a6004a209f8ccec14cadf2fd/servers HTTP/1.1
6.79521700 OPS1    CS01    HTTP  1722 HTTP/1.1 200 OK  (application/json)
14.4926800 CS01    OPS1    HTTP   268 GET /v2/images HTTP/1.1
16.5986570 OPS1    CS01    HTTP  7907 HTTP/1.1 200 OK  (application/json)
16.7596190 CS01    OPS1    HTTP   311 GET /v2.1/e9102671a6004a209f8ccec14cadf2fd/flavors/detail HTTP/1.1
16.8430620 MDSC1   MDSC1   HTTP   253 GET /restconf/ietf-te-topology/config/networks/ HTTP/1.1
16.8460370 MDSC1   MDSC1   HTTP  1104 HTTP/1.0 200 OK  (application/json)
16.8376490 OPS1    CS01    HTTP  2404 HTTP/1.1 200 OK  (application/json)
16.8511910 CS01    CS02    HTTP   241 GET /restconf/cso-dc/config/cso/dc/ HTTP/1.1
16.8571350 CS02    CS01    HTTP  1230 HTTP/1.0 200 OK  (application/json)
16.8632110 CS01    CS02    HTTP   248 GET /restconf/ietf-actn-vn/config/actn/ap/ HTTP/1.1
16.8648440 CS02    CS01    HTTP   213 HTTP/1.0 200 OK  (application/json)
```

Fig. 9. Captured wireshark message exchange for inititializing CSO1 and CSO2

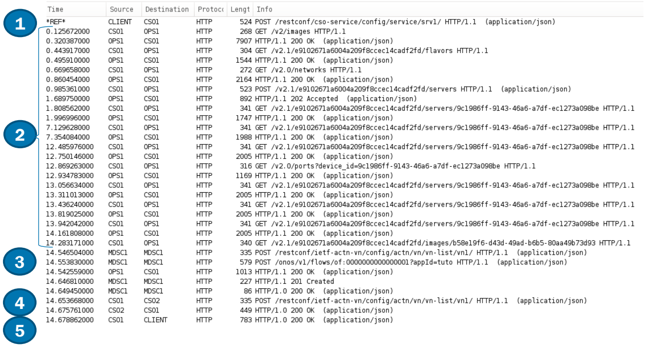
V. Conclusion

Several innovations are presented in this paper. An inter-CSO peer interface has been introduced, by providing descriptions to consider DC/cloud resources and VN resources. Internal CSO service orchestrator has been described. Also, SDN controller support for TEAS Topology and ACTN VN interfaces has been demonstrated. Finally, a proof of concept has been experimentally demonstrated in order to validate the proposed architecture.

As further work, scalability of the proposed architecture should be addressed, in terms of necessary setup requests, as well as scale in/down mechanisms for the requested resources.

Acknowledgments

This research has been partly funded by EC H2020 5GPPP 5GTANGO (H2020-ICT-2016-2 761493) and Spanish MINECO project DESTELLO (TEC2015-69256-R).



```
Time          Source   Destination  Protoc  Lengt  Info
*REF*         CLIENT   CS01         HTTP     524 POST /restconf/cso-service/config/service/srv1/ HTTP/1.1  (application/json)
0.125672000   CS01     OPS1         HTTP     268 GET /v2/images HTTP/1.1
0.320387000   OPS1     CS01         HTTP    7907 HTTP/1.1 200 OK  (application/json)
0.443917000   CS01     OPS1         HTTP     304 GET /v2.1/e9102671a6004a209f8ccec14cadf2fd/flavors HTTP/1.1
0.495910000   OPS1     CS01         HTTP    1544 HTTP/1.1 200 OK  (application/json)
0.669658000   CS01     OPS1         HTTP     272 GET /v2.0/networks HTTP/1.1
0.860454000   OPS1     CS01         HTTP    2164 HTTP/1.1 200 OK  (application/json)
0.985361000   CS01     OPS1         HTTP     523 POST /v2.1/e9102671a6004a209f8ccec14cadf2fd/servers HTTP/1.1  (application/json)
1.689750000   OPS1     CS01         HTTP     892 HTTP/1.1 202 Accepted  (application/json)
1.808562000   CS01     OPS1         HTTP     341 GET /v2.1/e9102671a6004a209f8ccec14cadf2fd/servers/9c1986ff-9143-46a6-a7df-ec1273a098be HTTP/1.1
1.996996000   OPS1     CS01         HTTP    1747 HTTP/1.1 200 OK  (application/json)
7.129628000   CS01     OPS1         HTTP     341 GET /v2.1/e9102671a6004a209f8ccec14cadf2fd/servers/9c1986ff-9143-46a6-a7df-ec1273a098be HTTP/1.1
7.354084000   OPS1     CS01         HTTP    1988 HTTP/1.1 200 OK  (application/json)
12.485976000  CS01     OPS1         HTTP     341 GET /v2.1/e9102671a6004a209f8ccec14cadf2fd/servers/9c1986ff-9143-46a6-a7df-ec1273a098be HTTP/1.1
12.750146000  OPS1     CS01         HTTP    2005 HTTP/1.1 200 OK  (application/json)
12.869263000  CS01     OPS1         HTTP     316 GET /v2.0/ports?device_id=9c1986ff-9143-46a6-a7df-ec1273a098be HTTP/1.1
12.934783000  OPS1     CS01         HTTP    1169 HTTP/1.1 200 OK  (application/json)
13.056634000  CS01     OPS1         HTTP     341 GET /v2.1/e9102671a6004a209f8ccec14cadf2fd/servers/9c1986ff-9143-46a6-a7df-ec1273a098be HTTP/1.1
13.311013000  OPS1     CS01         HTTP    2005 HTTP/1.1 200 OK  (application/json)
13.436240000  CS01     OPS1         HTTP     341 GET /v2.1/e9102671a6004a209f8ccec14cadf2fd/servers/9c1986ff-9143-46a6-a7df-ec1273a098be HTTP/1.1
13.819025000  OPS1     CS01         HTTP    2005 HTTP/1.1 200 OK  (application/json)
13.942042000  CS01     OPS1         HTTP     341 GET /v2.1/e9102671a6004a209f8ccec14cadf2fd/servers/9c1986ff-9143-46a6-a7df-ec1273a098be HTTP/1.1
14.161808000  OPS1     CS01         HTTP    2005 HTTP/1.1 200 OK  (application/json)
14.283171000  CS01     OPS1         HTTP     340 GET /v2.1/e9102671a6004a209f8ccec14cadf2fd/images/b58e19f6-d43d-49ad-b6b5-80aa49b73d93 HTTP/1.1
14.546504000  MDSC1    MDSC1        HTTP     335 POST /restconf/ietf-actn-vn/config/actn/vn/vn-list/vn1/ HTTP/1.1  (application/json)
14.553830000  MDSC1    MDSC1        HTTP     579 POST /onos/v1/flows/of:0000000000000001?appId=tuto HTTP/1.1  (application/json)
14.542559000  OPS1     CS01         HTTP    1013 HTTP/1.1 200 OK  (application/json)
14.646810000  MDSC1    MDSC1        HTTP     227 HTTP/1.1 201 Created
14.649450000  MDSC1    MDSC1        HTTP      86 HTTP/1.0 200 OK  (application/json)
14.653668000  CS01     CS02         HTTP     335 POST /restconf/ietf-actn-vn/config/actn/vn/vn-list/vn1/ HTTP/1.1  (application/json)
14.675761000  CS02     CS01         HTTP     449 HTTP/1.0 200 OK  (application/json)
14.678862000  CS01     CLIENT       HTTP     783 HTTP/1.0 200 OK  (application/json)
```

Fig. 10. Captured wireshark message exchange for provisioning service

References

[1] Cisco Global Cloud Index: Forecast and Methodology, 20152020, White Paper, 2016.

[2] X. Liu, et al. ”YANG Data Model for TE Topologies.”, draft-ietf-teas-yang-te-topo-06, IETF, Oct. 2016.

[3] T. Saad (ed), et al. A YANG Data Model for Traffic Engineering Tunnels and Interfaces, draft-ietf-teas-yang-te-05, IETF, Oct. 2016.

[4] D. Ceccarelli and Y. Lee (ed.), Framework for Abstraction and Control of Traffic Engineered Networks, IETF draft, October 2016.

[5] Mapping Cross Stratum Orchestration (CSO) to the SDN architecture, TR-528, ONF, 2016.

[6] R. Vilalta, A. Mayoral, R. Casellas, R. Martínez, R. Muñoz, Experimental Demonstration of Distributed Multi-tenant Cloud/Fog and Heterogeneous SDN/NFV Orchestration for 5G Services , in Proceedings of European Conference on Networks and Communications (EuCnC), June 27-30 2016, Athens (Greece).

[7] Y. Lee (ed.), A Yang Data Model for ACTN VN Operation, IETF draft, October 2016.

[8] A. Bierman et al., RESTCONF Protocol, IETF draft-ietf-netconf-restconf-18, 2016.

[9] A. Aguado, V. López, J. Marhuenda, O. González de Dios, and J. P. Fernández-Palacios, ABNO: A feasible SDN approach for multivendor IP and optical networks [invited], Journal of Optical Communications and Networking, vol. 7, num. 2, A356-A362, 2015.

[10] A. Mayoral, R. Muñoz, R. Vilalta, R. Casellas, R. Martínez, V. López, Need for a Transport API in 5G for Global Orchestration of Cloud and Networks Through a Virtualized Infrastructure Manager and Planner [Invited], IEEE/OSA Journal of Optical Communications and Networking, vol. 9, num. 1, A55-A62, 2017.

[11] R. Vilalta et al., Peer SDN Orchestration End-to-End Connectivity Service Provisioning Through Multiple Administrative Domains, ECOC16.

[12] OpenStack API. http://developer.openstack.org/api-guide/quick-start/

[13] A. Mayoral, R. Vilalta, R. Muñoz, R. Casellas, R. Martínez, SDN orchestration architectures and their integration with cloud computing applications, Optical Switching and Networking, 2016, Elsevier.

AI-Assisted Resource Advertising and Pricing to Realize Distributed Tenant-Driven Virtual Network Slicing in Inter-DC Optical Networks

Wei Lu, Hongqiang Fang, Zuqing Zhu[†]
School of Information Science and Technology, University of Science and Technology of China, Hefei, China
[†]Email: {zqzhu}@ieee.org

***Abstract*—We propose a novel artificial intelligence (AI) assisted framework to realize virtual network (VNT) slicing in an inter-datacenter optical network (IDCON), where the infrastructure provider (InP) performs resource advertising and pricing based on deep reinforcement learning (DRL) and grants the virtual network embedding (VNE) schemes calculated distributedly by the tenants. Simulation results confirm that compared with the traditional centralized VNT slicing framework, our proposal can not only make the InP more profitable but also relieve its computation complexity effectively.**

***Index Terms*—Inter-DC optical networks, Virtual network slicing, Knowledge-defined networking, Artificial intelligence.**

I. Introduction

Recently, the omnipresent requirements of cloud computing are demanding an unprecedented amount of data to be transferred among datacenters (DCs) [1]. Therefore, the architecture of inter-DC optical networks (IDCONs) [2] and the network virtualization schemes in them [3] have received intensive research interests. With network virtualization, service providers (SPs) (*i.e.*, tenants) are allowed to lease substrate network (SNT) resources from an infrastructure provider (InP) and build various virtual networks (VNTs) in a “pay as you use” manner [4, 5]. This is extremely useful in an IDCON, since the InP can allocate bandwidth and IT resources dynamically and adaptively to slice VNTs for the tenants and help them satisfy the time-varying and diversified demands from their services [6]. Hence, a win-win situation can be achieved, *i.e.*, the InP’s substrate resource utilization can be improved and the tenants’ time-to-market can be reduced.

Note that, for VNT slicing, the InP of an IDCON usually needs to 1) select a substrate DC node to host each virtual node (VN) of the VNT for satisfying the IT resource requirement (*i.e.*, the node mapping), and 2) reserve sufficient optical spectra on a substrate path to carry each virtual link (VL) between a VN pair for satisfying the bandwidth requirement (*i.e.*, the link mapping), which is also known as virtual network embedding (VNE) [7]. Previously, the problem of VNE has already been studied intensively in various network scenarios and with different optimization objectives [7–10], and related network system prototypes have been experimentally demonstrated in [11–13]. However, all these previous investigations assumed that the InP is in charge of VNT slicing solely without any involvement of the tenants, and it calculates VNE schemes based on current network status without the intelligence for forecasting. This would not only complicate the network control and management (NC&M) of the InP but also limit the cost-effectiveness of VNT slicing. For instance, it is known that the energy efficiency or the cost-effectiveness of an IDCON can be improved with network-wide resource consolidation, *i.e.*, consolidating computing tasks on fewer DCs and grooming inter-DC traffic to fewer fiber links [14, 15]. Nevertheless, in case of VNT slicing, the InP can hardly realize the most effective resource consolidation, if it cannot directly forecast future VNT requests from the tenants or indirectly affect their behaviors on submitting VNT requests.

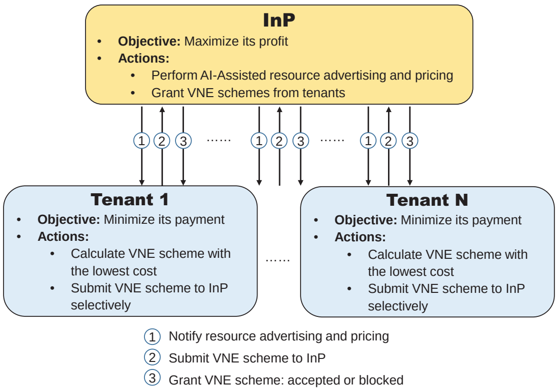

InP
Objective: Maximize its profit
Actions:
Perform AI-Assisted resource advertising and pricing
Grant VNE schemes from tenants
1 2 3 ...... 1 2 3 ...... 1 2 3
Tenant 1
Objective: Minimize its payment
Actions:
Calculate VNE scheme with the lowest cost
Submit VNE scheme to InP selectively
......
Tenant N
Objective: Minimize its payment
Actions:
Calculate VNE scheme with the lowest cost
Submit VNE scheme to InP selectively
1 Notify resource advertising and pricing
2 Submit VNE scheme to InP
3 Grant VNE scheme: accepted or blocked

Fig. 1. Proposed framework for distributed tenant-driven VNT slicing.

The aforementioned issue with existing approaches motivates us to revisit the problem of VNT slicing in IDCONs. Specifically, inspired by the idea of knowledge-defined networking (KDN) [16], we propose to add three new mechanisms into the framework of VNT slicing to make it operate in a distributed tenant-driven manner and more profitable:

- The InP performs resource advertising and pricing to tell the tenants about the DCs and fiber links that can be used to embed their VNTs and the cost of using the corresponding IT and bandwidth resources[1].

[1] Here, the advertised resources might not be all the available ones in the IDCON. For the purpose of resource consolidation, the InP may choose to hide certain resources from advertising.

978-3-903176-07-2 © 2018 IFIP

- Based on the advertisement from the InP, each tenant distributedly calculates the VNE scheme for its VNT with the lowest cost, and determines whether the price is affordable. If yes, it will submit the scheme to the InP.
- The InP collects all the requests from the tenants, grants them based on current network status, calculates the profit from the VNT slicing, and feeds all the information into an artificial intelligence (AI) module based on deep learning to obtain the strategy of the next round of resource advertising and pricing for maximizing its profit.

As shown in Fig. 1, with the new mechanisms, VNT slicing is realized in a distributed and thus much more time-efficient way, and the InP would not be directly involved in the computation of VNE schemes. Hence, the InP's intelligence lies in being able to maximize the profit of VNT slicing by leveraging the AI-assisted resource advising and pricing. In this work, based on the framework in Fig. 1, we first lay out the network model and design an integer linear programming (ILP) model for each tenant to distributedly calculate the VNE scheme for its VNT with the lowest cost. Then, we study how to perform AI-assisted resource advertising and pricing in the InP for profit maximization. Specifically, we design a deep reinforcement learning (DRL) based algorithm to help the InP learn the relation between the strategy of resource advertising and pricing and the profit from VNT slicing. In other words, the DRL-based algorithm enables the InP to analyze the tenants' behaviors on distributed VNE computation for making wise decisions on resource advertising and pricing.

The rest of the paper is organized as follows. We formulate the problem in Section II. The DRL-based resource advertising and pricing algorithm is proposed in Section III. Section IV evaluates the performance of our proposal. Finally, we summarize the paper in Section V.

II. Problem Formulation

A. Network Model of IDCON

We model the topology of an IDCON as $G(V,E)$, where V and E denotes the sets of nodes and fiber links in it, respectively. Note that, there are actually two types of nodes in the IDCON, as shown in Fig. 2(a). Each of the first type ones consists of a local DC and an optical switch (OXC), which is referred to as an edge node and included in set V^E. The second type ones are intermediate nodes, each of which only includes an OXC and is included in set V^I. Apparently, we have $V^E \cap V^I = \emptyset$ and $V^E \cup V^I = V$. In the IDCON, each DC offers IT resources and each fiber link provides bandwidth, for VNT slicing. To facilitate distributed tenant-driven VNT slicing, the InP needs to perform resource advertising and pricing periodically. An example of the resource advertising is illustrated in Fig. 2(b), where for cost saving, the InP only turns on partial of the network elements in the IDCON and advertises the resources on them. Meanwhile, in order to maximize its profit and encourage the tenants to use substrate resources in a balanced manner, the InP needs to price the advertised resources properly. In the next section, we will design a DRL-based algorithm to help the InP achieve this.

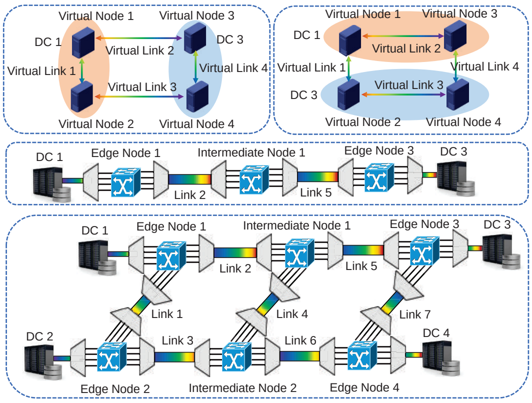


Fig. 2. Example on distributed tenant-driven VNT slicing, (a) IDCON, (b) Resource advertisement from InP, and (c) VNE schemes computed by tenants.

TABLE I
Notations for Resource Cost Model

Notation	Explanation
IDCON:	
v_i^E	the i-th edge node in V^E
v_i^I	the i-th intermediate node in V^I
$E(v)$	the set of links that connect to node $v \in V$
R_i^c	the amount of available IT resources in the DC of v_i^E
R_e^b	the available bandwidth on a fiber link $e \in E$
Edge Node $v_i^E \in V^E$:	
$\dot{C}_{o,i}^E$	the base cost of the OXC in v_i^E if it is working
$C_{o,i}^E$	the unit cost of switching capacity of the OXC in v_i^E
$\dot{C}_{d,i}^E$	the base cost of the DC in v_i^E if it is working
$C_{d,i}^E$	the unit cost of IT resources in the DC in v_i^E
Intermediate Node $v_i^I \in V^I$:	
$\dot{C}_{o,i}^I$	the base cost of the OXC in v_i^I if it is working
$C_{o,i}^I$	the unit cost of switching capacity of the OXC in v_i^I
Fiber Link $e \in E$:	
$\dot{C}_e$	the base cost of e if it is active with traffic
C_e	the unit cost of bandwidth usage on e
$\tilde{C}_e$	the merged unit cost of bandwidth usage on e

To assist the resource advertising and pricing, we define a few notations for the cost model of resources, which are listed in Table I. Here, for each network element in the IDCON (*i.e.*, a DC, an OXC or a fiber link), we assume that the cost of using it consists of a static component (*i.e.*, the base cost of turning it on) and a dynamic component (*i.e.*, the one that increases linearly with the actual resource usage on it). Note that, since the data transmission on fiber link e uses both the link and the two OXCs in its end-nodes, we combine the unit costs of bandwidth usage on them to get the merged unit cost $\tilde{C}_e$ as

$$\tilde{C}_e = C_e + \sum_{\{v_i^E:\ e \in E(v_i^E)\}} C_{o,i}^E + \sum_{\{v_i^I:\ e \in E(v_i^I)\}} C_{o,i}^I. \quad (1)$$

With this cost model, the InP needs to determine its strategy of resource advertising and pricing for profit maximization, and the strategy can be denoted with the variables defined in Table II. Here, for simplicity, we also get the merged unit price of bandwidth usage on fiber link e, which is

$$\tilde{P}_e = P_e + \sum_{\{v_i^E:\ e \in E(v_i^E)\}} P_{o,i}^E + \sum_{\{v_i^I:\ e \in E(v_i^I)\}} P_{o,i}^I. \tag{2}$$

TABLE II
VARIABLES DEFINED FOR RESOURCE ADVERTISING AND PRICING

Variable	Definition
$x_{o,i}^E$	Boolean variable that equals 1 if the OXC in edge node v_i^E is advertised (*i.e.*, turned on) by the InP, and 0 otherwise.
$x_{d,i}^E$	Boolean variable that equals 1 if the DC in edge node v_i^E is advertised by the InP, and 0 otherwise.
$x_{o,i}^I$	Boolean variable that equals 1 if the OXC in intermediate node v_i^I is advertised by the InP, and 0 otherwise.
y_e	Boolean variable that equals 1 if fiber link e is advertised by the InP, and 0 otherwise.
$P_{o,i}^E$	Positive real variable that represents the unit price of switching capacity of the OXC in edge node v_i^E.
$P_{d,i}^E$	Positive real variable that represents the unit price of IT resources in the DC in edge node v_i^E.
$P_{o,i}^I$	Positive real variable that represents the unit price of switching capacity of the OXC in intermediate node v_i^I.
P_e	Positive real variable that represents the unit price of bandwidth usage on fiber link e.
$\tilde{P}_e$	Positive real variable that represents the merged unit price of bandwidth usage on fiber link e.

B. Distributed Tenant-driven VNT Slicing

We assume that there are K pending VNT requests from the tenants. The k-th VNT request can be represented as $G_k^r(V_k^r, E_k^r, \hat{P}_k^r)$, where V_k^r and E_k^r are the sets of virtual nodes (VNs) and virtual links (VLs), respectively, and $\hat{P}_k^r$ is the highest cost that the tenant can afford. Here, each VN $v_{k,i}^r \in V_k^r$ has an IT resource requirement of $R_{k,i}^r$, and it should be mapped onto an edge node in V^E with sufficient IT resources in its DC. Note that, for node mapping, a tenant may have a location constraint from its services, *i.e.*, its VNs should only be mapped onto a subset of edge nodes in the IDCON to ensure certain access latency and/or coverage of its services [10]. We denote the subset of the edge nodes that VN $v_{k,i}^r$ can be mapped onto as $V_{k,i}^E$ and have $V_{k,i}^E \subseteq V^E$. Each VL $e \in E_k^r$ has a bandwidth requirement of $R_{k,e}^r$, and it should be mapped onto a substrate path with sufficient bandwidth.

Based on the resource advertisement from the InP, the tenant calculates the VNE scheme for its VNT request with the lowest cost, as shown in Fig. 2(c). This can be done by leveraging the ILP model listed in Table III. Note that, in Eq. (9), $E(v)^-$ and $E(v)^+$ mean the sets of egress and ingress links to node v, respectively. After obtaining the VNE scheme by solving the ILP, the tenant checks whether the scheme's cost is affordable (*i.e.*, not exceeding $\hat{P}_k^r$). If yes, the tenant will submit the VNE scheme and the corresponding payment to the InP. Otherwise, it will cancel its VNT request temporarily.

For the VNT requests submitted to the InP, we denote the set of their indices as K'. Then, based on the payments from the tenants and the corresponding resource costs, the InP calculates the profit from each VNT request (*i.e.*, payment minus total resource cost), sort the requests in descending order of the profits from them, and grant them one-by-one in sorted order. Note that, in this process, a VNT request can be blocked due to insufficient resources in the IDCON. Hence, the granted VNT requests may be a subset of the submitted ones, and the set of their indices can be denoted as K''. Finally, with K'', the InP can calculate its profit from this round of VNT slicing, which is denoted as $\mathcal{P}$. According to Table II, the strategy of resource advertising and pricing can be represented with the advertisement matrix $\mathbf{A} = [\{x_{o,i}^E\}, \{x_{d,i}^E\}, \{x_{o,i}^I\}, \{y_e\}]$ and the price matrix $\mathbf{P} = [\{P_{o,i}^E\}, \{P_{d,i}^E\}, \{P_{o,i}^I\}, \{P_e\}, \{\tilde{P}_e\}]$. We can see that the profit $\mathcal{P}$ is actually a function of $\mathbf{A}$ and $\mathbf{P}$, *i.e.*, $\mathcal{P} = f(\mathbf{A}, \mathbf{P})$. In the next section, we will design a DRL-based algorithm to let the InP learn $\mathcal{P} = f(\mathbf{A}, \mathbf{P})$ intelligently.

TABLE III
ILP MODEL FOR TENANT TO CALCULATE VNE SCHEME OF THE k-TH VNT REQUEST

Variable	Definition
$x_{i,i'}$	Boolean variable that equals 1 if the i-th VN $v_{k,i}^r$ in V_k^r is mapped onto the i'-th edge node $v_{i'}^E$ in V^E, and 0 otherwise.
$y_{e,e'}$	Boolean variable that equals 1 if VL $e \in E_k^r$ goes through fiber link $e' \in E$, and 0 otherwise.

Objective:

$$\text{Minimize} \left(\sum_{v_{k,i}^r \in V_{k,i}^E} \sum_{v_{i'}^E \in V^E} P_{d,i'}^E \cdot x_{i,i'} \cdot R_{k,i}^r + \sum_{e \in E_k^r} \sum_{e' \in E} \tilde{P}_{e'} \cdot y_{e,e'} \cdot R_{k,e}^r \right). \tag{3}$$

Node Mapping Constraints:

$$x_{i,i'} \leq x_{o,i'}^E, \quad \forall v_{k,i}^r \in V_k^r,\ v_{i'}^E \in V_{k,i}^E, \tag{4}$$

$$x_{i,i'} = 0, \quad \forall v_{k,i}^r \in V_k^r,\ v_{i'}^E \notin V_{k,i}^E, \tag{5}$$

$$\sum_{v_{i'}^E \in V_{k,i}^E} x_{i,i'} = 1, \quad \forall v_{k,i}^r \in V_k^r, \tag{6}$$

$$x_{o,i'}^E + x_{i,i'} - 1 \leq x_{d,i'}^E, \quad \forall v_{k,i}^r \in V_k^r,\ v_{i'}^E \in V_{k,i}^E. \tag{7}$$

Link Mapping Constraints:

$$y_{e,e'} \leq y_{e'}, \quad \forall e \in E_k^r,\ e' \in E, \tag{8}$$

$$\begin{aligned} &\sum_{e' \in E(v_{i'}^E)^-} y_{e,e'} - \sum_{e' \in E(v_{i'}^E)^+} y_{e,e'} = x_{i,i'} - x_{j,i'}, \\ &\{e : e = (v_{k,i}^r, v_{k,j}^r), e \in E_k^r\},\ \forall v_{i'}^E \in V^E. \end{aligned} \tag{9}$$

Resource Constraints:

$$\sum_{v_{k,i}^r \in V_k^r} x_{i,i'} \cdot R_{k,i}^r \leq R_{i'}^c, \quad \forall v_{i'}^E \in V^E, \tag{10}$$

$$\sum_{e \in E_k^r} y_{e,e'} \cdot R_{k,e}^r \leq R_{e'}^b, \quad \forall e' \in E. \tag{11}$$

III. AI-ASSISTED RESOURCE ADVERTISING AND PRICING

We first design an evaluate function $\widehat{Q}(\cdot)$ that can rank network elements in the IDCON to get the advertisement matrix

$\mathbf{A}$, and then propose a DRL-based algorithm to parameterize $\widehat{Q}(\cdot)$ such that the price matrix $\mathbf{P}$ can be learned iteratively.

A. Design of Evaluation Function $\widehat{Q}(\cdot)$

The evaluation function $\widehat{Q}(\cdot)$ should be able to rank network elements in the IDCON such that if the InP turns down them in the sorted order (*i.e.*, maximizing $\widehat{Q}(\cdot)$ each time), its profit can be maximized. We formulate $\widehat{Q}(\cdot)$ as $\widehat{Q}(\mathbf{A}, n_d; \Theta, \mathbf{P})$, which is a function of $\mathbf{A}$ and n_d with parameters Θ and $\mathbf{P}$, and n_d is the network element (*i.e.*, a DC, an OXC or a fiber link) to be shut down and removed from the upcoming advertisement. Supposing that $\widehat{Q}(\cdot)$ has already been parameterized with known Θ and $\mathbf{P}$, the InP can use the simple procedure in *Algorithm* 1 to obtain the advertisement matrix $\mathbf{A}$.

Algorithm 1: Determining Advertisement Matrix $\mathbf{A}$ with Evaluation Function $\widehat{Q}(\mathbf{A}, n_d; \Theta, \mathbf{P})$

```
initialize $\mathbf{A}$ as turning on all the elements in $G(V, E)$;
calculate the InP's profit $\mathcal{P}$ based on $\mathbf{A}$ and $\mathbf{P}$ with the
    approach in Section II;
$\mathcal{P}' = 0$;
while $\mathcal{P} > \mathcal{P}'$ do
    $\mathcal{P}' = \mathcal{P}$, $n_d = \underset{n_d \in \mathbf{A}}{\operatorname{argmax}} \widehat{Q}(\mathbf{A}, n_d; \Theta, \mathbf{P})$;
    shut down $n_d$ and update $\mathbf{A}$ accordingly;
    calculate the InP's profit $\mathcal{P}$ based on $\mathbf{A}$ and $\mathbf{P}$;
end
```

We design a recursive structure [17] for $\widehat{Q}(\cdot)$ to capture the features of each network element, by considering both the characteristics of the IDCON's topology $G(V, E)$ and the element's relation with other elements in the IDCON. Specifically, at the t-th recursion, the features of n_d are represented by a $(2|V^E|+|V^I|+|E|)$-dimensional vector $\varpi^{(t)}_{n_d}$, and the recursive relations are defined as follows.

$$\varpi^{(t)}_{n_d} = \begin{cases} f_0(\theta_1 \cdot x^E_{o,i} + \theta_2 \cdot A_1 + \theta_3 \cdot f_0(B_1)), & n_d \text{ is the OXC in } v^E_i, \\ f_0(\theta_1 \cdot x^E_{d,i} + \theta_2 \cdot \varpi^{(t-1)}_{O(v^E_i)}), & n_d \text{ is the DC in } v^E_i, \\ f_0(\theta_1 \cdot x^I_{o,i} + \theta_2 \cdot A_2 + \theta_3 \cdot f_0(B_2)), & n_d \text{ is the OXC in } v^I_i, \\ f_0(\theta_1 \cdot y_e + \theta_2 \cdot A_3 + \theta_3 \cdot f_0(B_3)), & n_d \text{ is fiber link } e, \end{cases}$$

where $f_0(x) = x$ if $x \geq 0$, and 0 otherwise, and the parameters $\{A_m, B_m : m \in [1, 3]\}$ are calculated as follows

$$\begin{cases} A_1 = \varpi^{(t-1)}_{D(v^E_i)} + \sum\limits_{n_d \in N(v^E_i)} \varpi^{(t-1)}_{n_d} + \sum\limits_{n_d \in E(v^E_i)} \varpi^{(t-1)}_{n_d}, \\ B_1 = \sum\limits_{v^E_j \in N(v^E_i)} \theta_4 \cdot \left(P^E_{o,j} + P^E_{d,j}\right) + \sum\limits_{v^I_j \in N(v^E_i)} \theta_5 \cdot P^I_{o,j} + \sum\limits_{e \in E(v^E_i)} \theta_6 \cdot P_e, \\ A_2 = \sum\limits_{n_d \in N(v^I_i)} \varpi^{(t-1)}_{n_d} + \sum\limits_{n_d \in E(v^I_i)} \varpi^{(t-1)}_{n_d}, \\ B_2 = \sum\limits_{v^E_j \in N(v^I_i)} \theta_4 \cdot \left(P^E_{o,j} + P^E_{d,j}\right) + \sum\limits_{v^I_j \in N(v^I_i)} \theta_5 \cdot P^I_{o,j} + \sum\limits_{e \in E(v^I_i)} \theta_6 \cdot P_e, \\ A_3 = \sum\limits_{\{v^E_i :\, e \in E(v^E_i)\}} \varpi^{(t-1)}_{O(v^E_i)} + \sum\limits_{\{v^I_i :\, e \in E(v^I_i)\}} \varpi^{(t-1)}_{O(v^I_i)}, \\ B_3 = \sum\limits_{\{v^E_i :\, e \in E(v^E_i)\}} \theta_4 \cdot \left(P^E_{o,i} + P^E_{d,i}\right) + \sum\limits_{\{v^I_i :\, e \in E(v^I_i)\}} \theta_5 \cdot P^I_{o,i}, \end{cases}$$

where $N(v)$ returns the set of OXCs in adjacent nodes of node v, and $D(v)$ and $O(v)$ return the DC and OXC in node v, respectively. With T recursions, the features of each network element are spread to those that are T hops away from it. Then, the evaluation function $\widehat{Q}(\mathbf{A}, n_d; \Theta, \mathbf{P})$ can be formulated as

$$\widehat{Q}(\mathbf{A}, n_d; \Theta, \mathbf{P}) = \theta_7^\top \cdot f_0\left(\left[\theta_8 \cdot \sum_{n'_d \in G} \varpi^{(T)}_{n'_d},\ \theta_9 \cdot \varpi^{(T)}_{n_d}\right]\right), \quad (12)$$

where $\Theta = \{\theta_i : i \in [1, 9]\}$.

B. DRL-based Algorithm to Parameterize $\widehat{Q}(\cdot)$

We propose a DRL-based algorithm with the following principle to parameterize $\widehat{Q}(\cdot)$, *i.e.*, determining Θ and $\mathbf{P}$.

- *States*: each state corresponds to a feasible $\mathbf{A}$.
- *Actions*: an action is to shut down one network element n_d at the current state $\mathbf{A}$.
- *Rewards*: the reward of an action at the current state $\mathbf{A}$ is calculated as:

  $$f_r(\mathbf{A}, n_d) = f(\mathbf{A}/n_d, \mathbf{P}) - f(\mathbf{A}, \mathbf{P}), \quad (13)$$

  where $f(\mathbf{A}, \mathbf{P})$ calculates the InP's profit, and $\mathbf{A}/n_d$ means to shut down n_d at state $\mathbf{A}$.

Based on Eq. (13), we define an n-step-forward function

$$y = \sum_{i=0}^{n-1} f_r(\mathbf{A}^{(t+i)}, n_d^{(t+i)}) + \beta \cdot \max_{n_d} \left[\widehat{Q}(\mathbf{A}^{(t+n)}, n_d; \Theta, \mathbf{P})\right], \quad (14)$$

where t is the index of the current iteration, $\mathbf{A}^{(t+i)}$ and $n_d^{(t+i)}$ are the state and action at the $(t+i)$-th iteration, respectively, and β is a constant coefficient. Then, in the DRL, we try to minimize the squared regression loss defined as

$$\left[y - \widehat{Q}(\mathbf{A}^{(t)}, n_d^{(t)}; \Theta, \mathbf{P})\right]^2. \quad (15)$$

Algorithm 2 shows the procedure of the proposed DRL-based algorithm. In each round of training, we first create two sets Γ and Δ (*Lines* 2-3). The former is to store all the valid training samples, and the latter is the training set with a fixed size for an iteration. *Line* 4 initializes $\mathbf{A}^{(1)}$, and the for-loop covering *Lines* 5-21 tries to shut down a network element in each iteration. Here, to diversify the training samples, we generate a random number $\epsilon \in [0, 1]$ (*Line* 6), and test whether it is smaller than a preset threshold T_h. If yes, the action $n_d^{(t)}$ is randomly selected within $\mathbf{A}^{(t)}$ (*Line* 8). Otherwise, the action is determined according to the policy in *Line* 10. Then, we get $\mathbf{A}^{(t+1)}$ accordingly (*Line* 12) and calculate the corresponding reward in *Line* 13. Due to the n-step-forward function in Eq. (14), only when the iteration number is larger than n, $\{\mathbf{A}^{(t-n)}, n_d^{(t-n)}, f_r(\mathbf{A}^{(t-n)}, n_d^{(t-n)})\}$ becomes a valid sample. Hence, it is added into Γ in *Line* 15. Once there are more than $|\Delta|$ samples in Γ (*Line* 16), the training set Δ can be formed by selecting $|\Delta|$ samples from Γ randomly (*Line* 17), and then the values of $\{\Theta, \mathbf{P}\}$ are updated by performing stochastic gradient descent (SGD) over Eq. (15) for Δ (*Line* 18). Finally, after M rounds of training, *Algorithm* 2 determines and returns the values of $\{\Theta, \mathbf{P}\}$.

Algorithm 2: DRL-based Algorithm to Parameterize $\widehat{Q}(\cdot)$

```
for each round $j \in [1, M]$ do
    create a set $\Gamma$ to store valid training samples;
    create a training set $\Delta$ with a fixed size;
    initialize $\mathbf{A}^{(1)}$ as turning on all elements in $G(V, E)$;
    for each iteration $t \in [1, 2|V^E| + |V^I| + |E|]$ do
        generate a random real number $\epsilon \in [0, 1]$;
        if $\epsilon \leq T_h$ then
            select $n_d^{(t)}$ randomly in $\mathbf{A}^{(t)}$;
        else
            $n_d^{(t)} = \underset{n_d \in \mathbf{A}^{(t-1)}}{\text{argmax}}\ \widehat{Q}(\mathbf{A}^{(t)}, n_d; \Theta, \mathbf{P})$;
        end
        get $\mathbf{A}^{(t+1)}$ from $\mathbf{A}^{(t)}$ by removing $n_d^{(t)}$;
        calculate $f_r(\mathbf{A}^{(t)}, n_d^{(t)})$ with Eq. (13);
        if $t \geq (n + 1)$ then
            $\{\mathbf{A}^{(t-n)}, n_d^{(t-n)}, f_r(\mathbf{A}^{(t-n)}, n_d^{(t-n)})\} \rightarrow \Gamma$;
            if $|\Gamma| \geq |\Delta|$ then
                select $|\Delta|$ samples from $\Gamma$ randomly to
                  form $\Delta$;
                update $\{\Theta, \mathbf{P}\}$ by performing SGD over
                  Eq. (15) for $\Delta$;
            end
        end
    end
end
return $\{\Theta, \mathbf{P}\}$;
```

IV. Performance Evaluation

We conduct numerical simulations to evaluate the proposed framework with DRL-based resource advertising and pricing, which run on a computer with 4.0 GHz Inter Core i7-6700K CPU, 16 GB RAM and 11 GB NVIDIA GTX 1080Ti GPU. The DRL-based algorithm is implemented with TensorFlow 1.4.1. The topology of the IDCON can take either the 8-node or the NSFNET topologies in Fig. 3. The cost of resources are uniformly distributed within $[10, 30]$ units in both topologies. Here, unit stands for a general currency unit. We generate each VNT request in a way as: 1) the number of VNs is uniformly distributed within $[1, 5]$, 2) the subset of edge nodes that each VN is location-constrained within is randomly selected, 3) each VN pair is connected by a VL with a probability of 0.6, and 4) the highest cost that a tenant can afford is linearly proportional to the total number of VNs and VLs, with a slope uniformly distributed within $[16, 116]$ and $[50, 150]$ units in the 8-node and NSFNET topologies, respectively.

When training the proposed DRL-based algorithm, we use $M = 200$ as the maximum number of training rounds, set the number of VNT requests in each round as uniformly distributed within $[5, 45]$, and have $|\Delta| = 5$ as the number of training samples. After the training is done, we compare the proposed DRL-based algorithm with a centralized benchmark. The benchmark prices resources according to a normal distribution with mean μ and standard deviation σ equal to $\{65.7, 26.8\}$ and $\{123, 37.4\}$ for the 8-node and NSFNET topologies, respectively, and performs resource advertising in a greedy manner. Specifically, in the benchmark, by estimating the network element that can be shut down to bring in the maximum profit gain, the InP removes selected network elements one-by-one from the upcoming resource advertisement until its profit is maximized, and the whole process does not consider any inputs from the tenants'. The simulations average the results from 5 independent runs to get each data point.

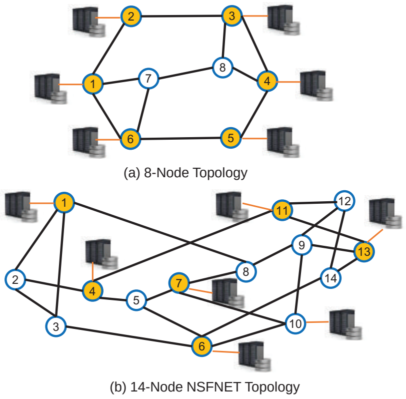
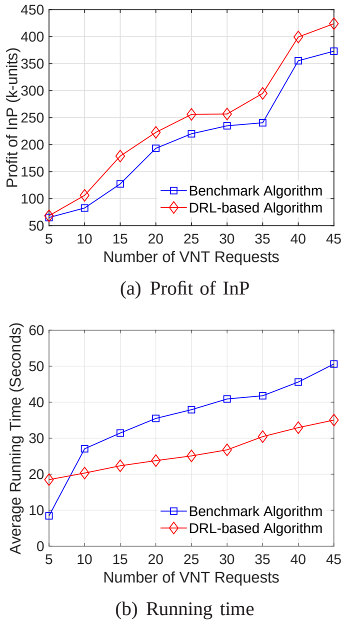


Fig. 3. IDCON topologies used in simulations.

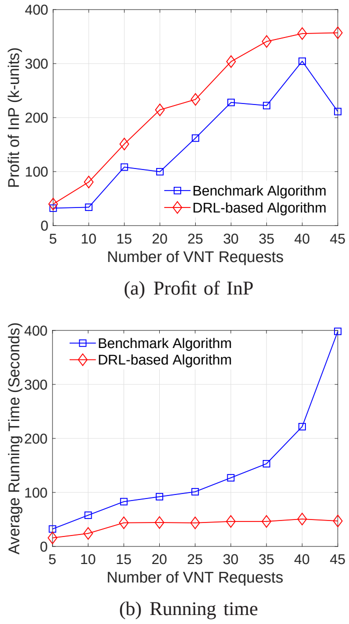


(a) Profit of InP

(b) Running time

Fig. 4. Results in 8-node topology.

Fig. 4 shows the results on the InP's profit and the algorithms' average running time for the 8-node topology. We can see that the proposed DRL-based algorithm achieves $\sim 14.56\%$



(a) Profit of InP

(b) Running time

Fig. 5. Results in the 14-node NSFNET topology.

more profit than the benchmark in all the cases. Meanwhile, the DRL-based algorithm also consumes less running time than the benchmark. Here, to achieve fair comparisons, the running time of the DRL-based algorithm includes the training time. This is because with the benchmark, the InP has to calculate the VNE schemes for the VNT requests and determine the resource advertising scheme in a centralized manner, while with the DRL-based algorithm, it only needs to perform network advertising according to the trained evaluation function and grant the VNE schemes from the tenants based on resource availability. Hence, the results in Fig. 4 verify that our proposed framework can not only make the InP more profitable but also relieve its computation complexity effectively.

Fig. 5 illustrates the results on the InP's profit and average running time in the NSFNET topology. We observe the similar trends as in Fig. 4. Actually, as the IDCON's size is larger, the proposed DRL-based algorithm achieves a larger profit increase over the benchmark, *i.e.*, $\sim 58.40\%$ on average. Since with the benchmark, the InP does not consider the tenants' inputs when determining resource advertising and pricing schemes, its profit in Fig. 5(a) does not exhibit a stable trend. Moreover, the centralized scheme of the benchmark does not scale well with the problem's size, and thus its running time increases exponentially with the number of VNT requests in Fig. 5(b). This makes our proposed framework's advantage on reduced time complexity much more significant.

V. Conclusion

We proposed a novel framework to realize VNT slicing in an IDCON, where the InP performs AI-assisted resource advertising and pricing and grants the VNE schemes calculated distributedly by the tenants. Then, for the InP, we designed a DRL-based resource advertising and pricing algorithm for profit maximization. Simulation results confirmed that compared with the traditional centralized VNT slicing framework, our proposal can not only make the InP more profitable but also relieve its computation complexity effectively. In the future, we need to further study both the timing and methods for predicting the tenant demands and their affordable prices accurately.

Acknowledgments

This work was supported in part by the NSFC Projects 61701472 and 61771445, CAS Key Project (QYZDY-SSW-JSC003), NGBWMCN Key Project (2017ZX03001019-004), China Postdoctoral Science Foundation (2016M602031), and Research Funds for Central Universities (WK2100060021).

References

[1] "Cisco global cloud index: Forecast and methodology, 2016-2021," 2017, [Online]. Available: https://www.cisco.com/c/en/us/solutions/service-provider/visual-networking-index-vni/index.html.

[2] P. Lu *et al.*, "Highly-efficient data migration and backup for big data applications in elastic optical inter-datacenter networks," *IEEE Netw.*, vol. 29, pp. 36–42, Sept./Oct. 2015.

[3] H. Jiang, Y. Wang, L. Gong, and Z. Zhu, "Availability-aware survivable virtual network embedding in optical datacenter networks," *J. Opt. Commun. Netw.*, vol. 7, pp. 1160–1171, Dec. 2015.

[4] M. Chowdhury and R. Boutaba, "Network virtualization: State of the art and research challenges," *IEEE Commun. Mag.*, vol. 47, pp. 20–26, Jul. 2009.

[5] L. Gong, Y. Wen, Z. Zhu, and T. Lee, "Toward profit-seeking virtual network embedding algorithm via global resource capacity," in *Proc. of INFOCOM 2014*, pp. 1–9, Apr. 2014.

[6] M. Bari *et al.*, "Data center network virtualization: A survey," *IEEE Commun. Surveys Tuts.*, vol. 15, pp. 909–928, Second Quarter 2013.

[7] L. Gong and Z. Zhu, "Virtual optical network embedding (VONE) over elastic optical networks," *J. Lightw. Technol.*, vol. 32, pp. 450–460, Feb. 2014.

[8] X. Cheng *et al.*, "Virtual network embedding through topology-aware node ranking," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 41, pp. 38–47, April 2011.

[9] M. Chowdhury, M. Rahman, and R. Boutaba, "Vineyard: Virtual network embedding algorithms with coordinated node and link mapping," *IEEE/ACM Trans. Netw.*, vol. 20, pp. 206–219, Feb. 2012.

[10] L. Gong, H. Jiang, Y. Wang, and Z. Zhu, "Novel location-constrained virtual network embedding (LC-VNE) algorithms towards integrated node and link mapping," *IEEE/ACM Trans. Netw.*, vol. 24, pp. 3648–3661, Dec. 2016.

[11] R. Munoz *et al.*, "Integrated SDN/NFV management and orchestration architecture for dynamic deployment of virtual SDN control instances for virtual tenant networks," *J. Opt. Commun. Netw.*, vol. 7, pp. B62–B70, Nov. 2015.

[12] J. Yin *et al.*, "Experimental demonstration of building and operating QoS-aware survivable vSD-EONs with transparent resiliency," *Opt. Express*, vol. 25, pp. 15 468–15 480, 2017.

[13] Z. Zhu *et al.*, "Build to tenants' requirements: On-demand application-driven vSD-EON slicing," *J. Opt. Commun. Netw.*, vol. 10, pp. A206–A215, Feb. 2018.

[14] W. Fang *et al.*, "Joint defragmentation of optical spectrum and IT resources in elastic optical datacenter interconnections," *J. Opt. Commun. Netw.*, vol. 7, pp. 314–324, Mar. 2015.

[15] F. Tso, S. Jouet, and D. Pezaros, "Network and server resource management strategies for data centre infrastructures: A survey," *Comput. Netw.*, vol. 106, pp. 209–225, Sept. 2016.

[16] A. Mestres *et al.*, "Knowledge-defined networking," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 47, pp. 2–10, Jul. 2017.

[17] H. Dai, B. Dai, and L. Song, "Discriminative embeddings of latent variable models for structured data," in *Proc. of ICML 2016*, pp. 2702–2711, Jun. 2016.

Virtual-Network-Function Placement For Dynamic Service Chaining In Metro-Area Networks

Leila Askari, Ali Hmaity, Francesco Musumeci, Massimo Tornatore
Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy
E-mail: firstname.lastname@polimi.it

***Abstract*—The advent of new services with stringent requirements on bandwidth and latency has led to a downward curve in per-user revenues of telecom operators. This has stimulated a significant shift in the way operators provision their services, moving from the utilization of dedicated and static hardware to support network functions (as NATs, firewalls, etc.), to the deployment of Virtual Network Functions (VNFs) in the form of dynamically-reconfigurable virtual machines on low-cost servers and switches. These VNFs must be chained together and should be placed optimally to meet the Quality of Service requirements of the supported services. This problem consists in placing the VNFs and routing traffic sequentially among them and is known as Service Chaining (SC). Solving this problem dynamically based on how traffic evolves allows to achieve great flexibility in resource assignment in the existing infrastructure and to save operational expenditures. An effective algorithm for dynamic SC must promote consolidation in VNF placement (it is desirable to consolidate VNFs in the fewer possible number of network nodes), while maintaining low blocking probability and guaranteeing latency targets to the supported services. In this paper we propose an algorithm which performs dynamic SC in a metro-area network, while minimizing average number of nodes required to host VNF instances as well as the blocking probability. This algorithm can help telecom operators reduce their operational expenditures up to 50% by activating less nodes to host VNFs in the network, while maintaining an acceptable level of blocking probability.**

I. Introduction

Cost-effectively provisioning new services (as Augmented Reality) that might require high bandwidth, or be latency sensitive and have higher reliability requirements is a complex challenge for operators. In response to this pressure, the concept of Network Function Virtualization (NFV) has attracted the attention of operators, as it enables to reduce Capital Expenditures (CapEx) and Operational Expenditures (OpEx) by virtualizing network functions (as NATs, firewalls, etc.) using virtual machines (VMs) running on top of standard servers and switches, by promoting resource sharing, and by decreasing energy consumption thanks to the consolidation of many network functions within few shared facilities. By chaining these Virtual Network Functions (VNFs) together (i.e, by placing VNFs and route traffic sequentially among them), the operator can provide a specific service (e.g., Cloud Gaming, VoIP, etc.) referred to as Service Chain (SC) [1] [2].

As NFV decouples network functions from hardware-based network appliances, network operators, based on the current situation of the network, in terms, e.g., of the amount of traffic and type of SC requested by users, can activate VNFs (by assigning resources to them) and deactivate them (by releasing resources used by them) in different network nodes equipped with processing units (identified to as "NFV-nodes"). In this context, to provision a SC it is important to decide in which network node to locate VNFs and how to route the traffic among them, as a proper placement of the VNFs can lead to efficient resource utilization and better Quality of Service (QoS). Hence, network operators should use efficient dynamic service chaining algorithms which, on one hand, help them reduce the expenses by activating less VNFs on less nodes and, on the other hand, minimize the blocking probability. Note that, by activating more VNF instances in the network, the network operator can serve more traffic, but (since activating an instance of a VNF imposes additional cost on network operators in terms of hardware resources, required licenses for softwares and power consumption among others) at the same time will face an increase in OpEx. Hence an appropriate trade-off must be investigated.

Considering specific features of a metro network in terms of latency, type of services requested by users, changing traffic load, number of users and services requested by users, network operators need efficient algorithms for VNF placement able to provision SCs dynamically based on current network condition. Most of existing studies on service chaining deal with static provisioning of SCs while the dynamic service chaining problem has received little attention so far [3] [4]. In this paper we provide an algorithm for dynamic VNF placement for SC provisioning in a metro network where at each time instant a number of users, request a specific SC and based on the current condition of the network VNFs are placed on NFV-nodes in a way that, with minimum possible number of provisioned VNFs, the blocking probability is minimized. The algorithm performs VNF placement in such a way that the bandwidth requirements of links, computational requirements of the NFV-nodes and latency requirements of requested SC are satisfied and wavelength continuity at each node is enforced. Moreover, by provisioning different SCs in the same wavelength, grooming is done to exploit maximum capacity of the network. Our algorithm is able to balance the trade-off between minimizing latency violation, decreasing blocking probability and reducing OpEx.

The remainder of this paper is organized as follows. Section II provides a brief overview of the related works. Section III, describes the metro network architecture and topology considered in this paper. The proposed heuristic algorithm is

978-3-903176-07-2 © 2018 IFIP

presented in Section IV. Numerical results are presented in Section V. Finally, conclusion is discussed in Section VI.

II. Related Work

The problem of VNF placement and traffic routing for SC provisioning has been subject of intense investigation in the last years, especially in static settings. For static SC provisioning, most studies propose an integer linear programming (ILP) model to obtain the optimal solution. For example authors in [5] dealt for the first time with a formal modeling of service chaining problem, and defined it as a Mixed Integer Quadratically Constrained Program (MIQCP) which finds the placement of VNFs and chains them together considering resource limitation of the network. Authors in [6] design a dynamic programming algorithm to jointly place VNFs and route traffic between them. They divide the problem in smaller subproblems and solve them sequentially [7]. Also some heuristic algorithms for VNF placement have been already proposed, as in [8]. However, in all the above-mentioned works, provisioning of SCs is performed under a static traffic assumption.

Dynamic placement of VNFs is addressed in a very limited set of works. In [1] the dynamicity is accounted by considering that type, number and location of VNFs traversed by a given user's data flow may change in time. So a traffic-conditioning function is proposed which, based on Service Level Agreement (SLA) of each user, decides the type of traffic conditioning function (shaping or priority scheduling) suitable for a given user's data flow. Authors of [9] provide an online algorithm for VNF scaling to dynamically provision network services in a datacenter network. In [10] a Mixed Integer Programming formulation and a heuristic algorithm are provided to dynamically provision SC, again in a datacenter network, where an appropriate resource management is done based on number of users requesting SCs. The authors of [3] consider dynamic SC provisioning for two types of users in the network, new users and existing users that can change location in the network and change their requested SC. So, they propose at first an ILP model for service provisioning with the objective to maximize the profit of the service provider; then, to reduce time complexity, they provide a more scalable model based on column generation [4]. However, no existing work provides an algorithm for dynamic SC provisioning in a metro network capable of addressing the trade-off between QoS requirements of services, blocking probability and CapEx and OpEx of telecom operators, as the one provided in this paper.

III. Metro network architecture and Background on service chain provisioning

The topology that we considered in this paper is a 4-level hierarchical metro aggregation network that connects cell sites (to which users are connected) to Central Offices (COs). Access COs provides the connectivity between cell sites connected to them and Core CO through Main COs. In this network optical transparent switches are used which impose wavelength continuity. As it is shown in Fig. 1 a SC can be considered as a chain of VNFs that are virtual nodes, chained together using virtual links (i.e. connections between nodes along the SC) forming SC's path [11].

To perform service chaining the VNFs need to be placed on NFV-nodes and virtual links need to be mapped to (a set of) physical links. Each SC is also characterized by source of the SC request, which in our topology is a cell site, and destination of SC request, which for the SC shown in Fig 1 is Core CO.

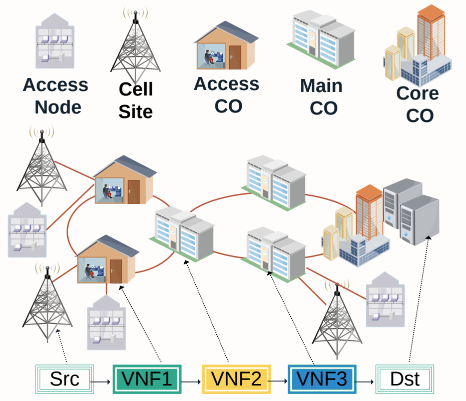


Fig. 1: Network topology.

To verify if latency requirements of SCs are satisfied, we model latency contributions as follows: i) context switching delay [12], that arises when a CPU is shared among multiple VMs and a delay is incurred for loading or saving the state of VMs. For each node n in NFV-nodes of SC ($n \in NFV_{SC}$), we consider a fixed value for context switching delay (DV_n) [13]; ii) propagation delay for each link l (DP_l) in SC's path ($Path_{SC}$); iii) Forward Error Correction (FEC) delay for (de)encoding optical signal in intermediate NFV-nodes of a SC; iv) optical-electrical-optical (OEO) conversion delay in intermediate NFV-nodes of a SC [14]. Overall, the end-to-end latency for a SC can be calculated as follows:

$$D_{SC} = \sum\nolimits_{n \in NFV_{SC}} (DV_n + OEO_n + FEC_n) + \sum\nolimits_{l \in Path_{SC}} DP_l$$

IV. Dynamic VNF placement Algorithm

In this section we describe the proposed dynamic algorithm, named Dynamic VNF Placement (DVNFP), that finds the placement for VNFs of a SC and route traffic between them at each time instant considering current state of the network.

A. Problem Statement

The problem can be stated, in a summarized form, as follows. We are given a hierarchical optical aggregation network which is composed of COs connected together using Wavelength Division Multiplexing (WDM) links. In this network, SC requests are dynamically generated by users (source of request) and they are generated in cell sites and terminated in a CO (destination of request) depending on the type of SC. Upon the arrival of a SC request, we need to decide the placement of its VNFs on NFV-nodes with the objective of minimizing the number of activated VNF instances as well as blocking probability, constrained by maximum link capacity, maximum node computational capacity and maximum tolerable SC latency.

A SC request is characterized by a given holding time and number of users requesting that specific SC. In addition, each SC requires a specific amount of bandwidth and has a maximum tolerated latency. Furthermore, for any SC, the corresponding VNFs require a specific amount of computational capacity in terms of fraction of CPU core usage per user.

B. Algorithm

In order to provision a SC we need to place all of its VNFs on NFV-nodes and route traffic between them. DVNFP first builds an auxiliary graph of network with all the nodes and links with their wavelengths. It takes as input a SC request which is specified by these properties:

- *src*: source of the SC request
- *dst*: destination of the SC request
- N_{vnf}: number of VNFs composing the SC
- F_{sc}: type of VNFs composing the SC
- N_u: number of users requesting the SC
- L_{sc}: latency requirements of the SC
- H_{time}: holding time of the SC

The pseudocode related to the placement of VNFs is shown in *Algorithm*1. The main steps of DVNFP for VNF placement can be defined as follows:

- *Reusing active VNFs:* Since activating an instance of a VNF imposes additional cost on network operators, when a SC request arrives, DVNFP tries to reuse already activated VNF instances in the network as much as possible. Therefore, as it is shown in *Algorithm* 1 line 4, for each VNF, DVNFP first checks if there is an already activated VNF instance of the same type in the network or not.
- *Selecting among active VNFs:* As it is shown in *Algorithm* 1 lines 5-21, if more than one VNF instance with enough capacity is already activated in the network, DVNFP uses a metric called "locality-awareness". This metric is obtained by summing up the length of the shortest path between source of SC request and the selected NFV-node, and the shortest path between that node and destination of SC request. It is worth mentioning that we use an adaptive Dijkstra algorithm to calculate the shortest path, in which the congested links are not included in the graph. DVNFP chooses the NFV-nodes with locality-awareness metric lower than a predefined threshold δ whose value can be decided based on the topology of the network. Among these NFV-nodes, our algorithm based on the requested SC decides which node to choose. So, if a SC request requires large computational resources e.g., Cloud Gaming [15], (requirements of these services will be quantified in Section VI) then the NFV-nodes closer to the Core CO, which are more likely to have large computational capacity are chosen. However, if the SC has stringent latency requirements (e.g. as happens for Smart Factory), DVNFP tries to serve that SC locally, using as NFV-nodes access COs or at least CO in lower level of the metro hierarchy. When the best NFV-node is found the VNF is placed on that node by allocating the required computational capacity.
- *Activating new VNF instance:* If no VNF instances of a certain VNF are already activated in the network, DVNFP tries to instantiate a new one. As it is shown in *Algorithm* 1 lines 22-30 at first it calculates the shortest path between source and destination of the SC request. Then it tries to place the VNF on the closest NFV-node to the source along the shortest path with enough computational capacity. If the VNF cannot be placed on any of NFV-nodes along the shortest path, the algorithm checks the capacity of all other NFV-nodes on the network and tries to place the VNF on the node with better locality-awareness and higher betweenness centrality (defined as number of shortest path passing through this node).

Note that, at each step, source of SC request is replaced by the NFV-node chosen to host a VNF at previous step and the above-mentioned procedures are repeated until all the VNFs of a SC are placed.

The pseudocode of QoS improvement is shown in *Algorithm* 2. When all the VNFs are placed, as it is shown in *Algorithm* 2 line 2, algorithm checks if latency requirement of the SC is satisfied. If it is the case, the SC is provisioned and when its holding time expires the resources used by this SC (link capacity and computational capacity of NFV-nodes) are released. These steps are shown in *Algorithm* 2 lines 3-6. If the VNF placement does not allow to meet latency requirements, algorithm calculates the latency of all virtual links and finds the one with the highest value of latency. After that, the resources on the endpoints of this virtual link are released and their VNFs are placed on their adjacent virtual nodes (if they have enough computational capacity). Then the shortest path between these new endpoints is calculated and is replaced with that virtual link with the highest latency. This procedure is referred to as VNFs grouping and is shown in *Algorithm* 2 lines 8-17. In Fig. 2 we demonstrate how grouping of VNFs is done. In this example the virtual link between VNF2 and VNF3 is the one which has the highest value of latency. Therefore, we need to release computational resources allocated to VNF2 on node 3 and to VNF 3 on node 6 and place VNF2 and VNF3 on node 2 and node 7 respectively. When the VNFs are placed on the new NFV-nodes the required computational resources on those nodes are allocated to these VNFs and the shortest path between these two NFV-nodes is calculated. Then the calculated shortest path is calculated and considered as the new virtual link connecting VNF2 and VNF3 instances. DVNFP repeats the same procedure until either latency requirement of the SC is satisfied, or all the VNFs are consolidated.

Algorithm 1 Placement of Virtual Network Functions

```
Given: Service Chain request $Req(src, dst, N_{vnf}, N_u, F_{sc}, L_{sc}, H_{time})$, actual network state $N_{state}$
\* Phase I *\
repeat
    if ∃ instance of VNF already placed then
        Select all the VNF instances.
        Sort NFV-nodes $f \in F_{li}$ where VNF instances
        are hosted by increasing value of $loc_f$ and
        select only the VNF instances which satisfies:
            $loc_f - length(sp) < \delta$.
        if ∃ more than one such VNF instance then
            Choose the node with less activated VNF
            instances.
            if ∃ more than one such NFV-node then
                if SC requires computational capacity then
                    Choose the NFV-node closer to
                    Core CO.
                else
                    Choose the NFV-node closer to src.
                end if
            end if
        end if
        Try to scale up the VNF instance in the selected
        NFV-nodes until success or all NFV-nodes have
        been tried.
        if success then
            update $N_{state}$
            continue
        end if
    else \* Find an NFV-nodes with enough capacity and place VNF *\
        Select in order the NFV-nodes on the shortest path
        between src and dst of the SCs.
        Sort the NFV-nodes by increasing number of active
        VNFs on nodes.
        Try placing the VNF instance on an NFV-node
        until all the NFV-nodes on the shortest path have
        been tried.
        if failed then
            Select the NFV-node on the network with better
            $loc_f$ and higher betweenness centrality.
            Try placing the VNF instance on an NFV-node
            until all the NFV-nodes have been tried.
            if failed then
                return SC request blocked due to capacity
            else
                Update $N_{state}$
            end if
        else
            Update $N_{state}$
        end if
    end if
until All the VNFs of the SC request are chained
```

Algorithm 2 QoS Improvement

```
\* Phase II *\
Check end-to-end latency of the embedded SC against
requirement.
if success then
    Provision SC request.
    Release the resources when $H_{time}$ expires.
    return SC request provisioned
else
    repeat
        Select the virtual link with highest latency.
        Release the resources of the VNFs on its
        end-points.
        Find the two closest nodes to two end-points on
        SC virtual path with enough capacity.
        if Such node not found then
            return blocked SC request due to latency.
        else
            Enable those VNFs on these two nodes
            Add virtual link between those two nodes to
            SC virtual path.
        end if
        if End-to-end latency is satisfied then
            Provision SC request.
            Release the resources when $H_{time}$ expires.
            return SC request provisioned.
        else if Consolidated all virtual links and latency not satisfied then
            return blocked SC request due to latency.
            Release all the resources provisioned earlier.
            Update $N_{state}$.
        end if
    until Latency satisfied or all VNFs have been
    consolidated
end if
```

C. Benchmark Algorithms

We considered two benchmark algorithms to evaluate the performance of DVNFP which are as follows:

- **Centralized service chaining:** In this approach we place all the VNFs in the node with highest computational capacity (Core CO in our topology) and we serve all the SCs using the VNF instances on that node.
- **Distributed service chaining:** In this approach we enable VNF instances on all the NFV-nodes whenever they are needed. In other words, even if there is already an activated instance of a VNF in the network, the algorithm enables a new instance on NFV-nodes along the shortest path between source and destination of the SC request. Algorithm repeats the same steps till all the VNFs are placed. If length of shortest path is less than number of VNFs that are needed to be placed to provision the SC, algorithm tries to put rest of VNFs on destination.

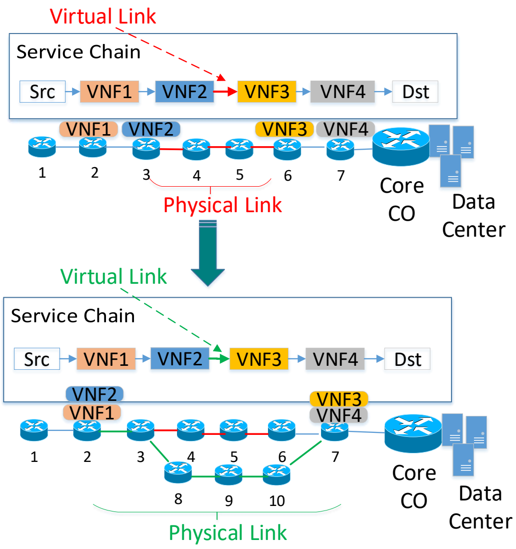


Fig. 2: VNF Grouping

V. Illustrative Numerical Results

In this section we compare the performance of DVNFP vs. the two benchmark cases, centralized service chaining and distributed service chaining. We use three performance metrics for comparison; blocking probability, which is calculated considering number of SC requests served out of total number of SC requests during the simulation, average number of active NFV-nodes, which is calculated considering number of NFV-nodes that have at least one running VNF instance at each time instant, and latency violation ratio which shows number of SC requests that violated latency requirements out of total SC requests.

Network modeling : We considered a topology similar to that shown in Fig.1 in which we have 80 nodes, 15 of which are NFV-nodes while the remaining nodes are forwarding nodes. The topology has 170 WDM links each supporting 16 wavelengths with 40 Gbit/s capacity. At each node wavelength continuity is enforced (we consider an optical network substrate), unless the node hosts a VNF, in which case, the intermediate virtual link is terminated and wavelength conversion is admissible. We assumed that for each SC request source is chosen randomly among cell sites while destination can be either Core CO or one of NFV-nodes based on the requested service type. Each NFV-node is equipped with 512 CPU cores, whereas the Core CO is assumed to have unlimited computational capacity.

Traffic/SC modeling: We conducted our simulative experiments using a C++ discrete-event driven simulator, that generates SC-requests as input traffic according to a Poisson-distribution of inter-arrival rates and negative-exponential distribution of the holding times (with average duration equal to one). All the plotted results have been obtained guaranteeing 95% statistical confidence and at most 5% confidence interval. We considered 6 different SC types as illustrated in Table I with different bandwidth and latency requirements [11] [15]–[18]. The VNFs are Network Address Translation (NAT), Firewall (FW), Traffic Monitor (TM), WAN Optimizer (WO), Video Optimization Controller (VOC), Intrusion Detection and Prevention System (IDPS). Each VNF requires specific amount of CPU resources per user. Table II illustrates the required amount of CPU (in terms of percentage of CPU per user) for each VNF [19]–[21].

TABLE I: Service Chains With Corresponding VNFs, Bandwidth and Latency Characteristics

Service Chain	Service Chain VNFs	Bandwidth	Latency
Cloud Gaming	NAT-FW-VOC-WO-IDPS	4 Mbps	80 ms
Augmented Reality	NAT-FW-TM-VOC-IDPS	100 Mbps	1 ms
VoIP	NAT-FW-TM-FW-NAT	64 Kbps	100 ms
Video Streaming	NAT-FW-TM-VOC-IDPS	4 Mbps	100 ms
MIoT	NAT-FW-IDPS	100 Mbps	5 ms
Smart Factory	NAT-FW	100 Mbps	1 ms

TABLE II: CPU Core Usage for VNFs

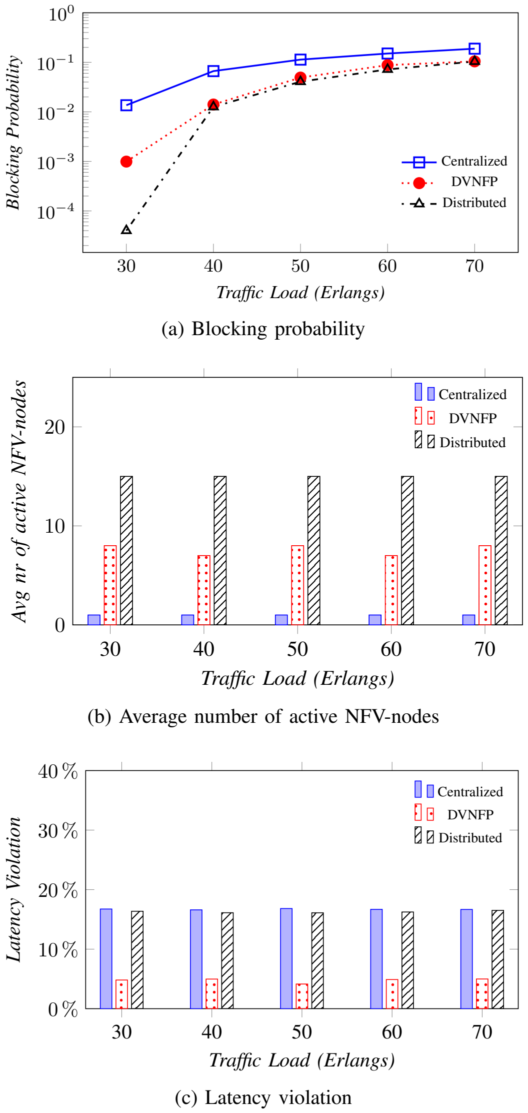
VNF Name	CPU Core Per User
NAT	0.00092
FW	0.0009
VOC	0.0054
TM	0.0133
WO	0.0054
IDPS	0.0107

A. *Comparison Between Algorithms*

We compare the three algorithms for increasing traffic load values. Fig 3(a) shows the blocking probability increase for increasing load in the network. We notice how blocking probability for DVNFP always lies in between blocking probability of two benchmark algorithms i.e. centralized and distributed, returning, for most cases, and especially for higher loads, results very similar to the distributed case. This observation is very promising, as it confirms that our algorithm guarantees a blocking close to the blocking lower bound (i.e, the one returned by a completely distributed service chaining approach). Fig 3 (b) plots the average number of active NFV-nodes. In this case we can see that DVNFP uses up to 50% less NFV-nodes in comparison with distributed for provisioning SC requests (even though the blocking probability is almost the same). In other words, as activating NFV-nodes imposes additional costs, using DVNFP telecom operators are able to almost halve the SC provisioning costs. Finally in Fig 3 (c), it is interesting to note that, although DVNFP requires to activate less NFV nodes, it still provides lower violation of QoS (i.e., latency) requirements in comparison to the distributed case. This is due to the fact that DVNFP performs VNF grouping whenever latency requirements of a provisioned SC is not satisfied. Moreover, less latency violations can be observed with respect to the centralized scenario, as DVNFP is able to choose NFV-nodes based on the requirements of SC (i.e. nodes closer to the source for latency sensitive SCs are chosen).

VI. Conclusion

We proposed an algorithm for dynamic placement of VNFs in the network which reduces operation costs in metro by consolidating VNFs as much as possible in network nodes, while



(a) Blocking probability

(b) Average number of active NFV-nodes

(c) Latency violation

Fig. 3: Simulation results

maintaining a low blocking probability. Simulation results show that DVNFP algorithm can balance the trade-off among three different metrics (blocking probability, average number of active NFV-nodes and latency violation) outperforming completely centralized or distributed solutions. In this paper we considered a fixed value for the context switching delay. However, in future work we will provide additional analysis considering that context switching delay may vary according to the number of VNFs activated in a node, and to the amount of traffic processed by each VNF. As resiliency is one of the challenges that network operators face while deploying virtualized services (i.e, service chains) future steps will focus in extending this work to provide protection against link and or node failures for dynamic VNF placement based on reliability targets defined for each service.

VII. Acknowledgement

The work leading to these results has been supported by the European Community under grant agreement no. 761727 *Metro-Haul* project funding.

References

[1] F. Callegati, W. Cerroni, C. Contoli, and G. Santandrea, “Dynamic chaining of virtual network functions in cloud-based edge networks,” in *Network Softwarization (NetSoft), 2015 1st IEEE Conference on*. IEEE, 2015, pp. 1–5.

[2] L. Qu, C. Assi, K. Shaban, and M. Khabbaz, “Reliability-aware service provisioning in NFV-enabled enterprise datacenter networks,” in *Network and Service Management (CNSM), 2016 12th International Conference on*. IEEE, 2016, pp. 153–159.

[3] J. Liu, W. Lu, F. Zhou, P. Lu, and Z. Zhu, “On dynamic service function chain deployment and readjustment,” *IEEE Transactions on Network and Service Management*, 2017.

[4] J.-J. Pedreno-Manresa, P. S. Khodashenas, M. S. Siddiqui, and P. Pavon-Marino, “On the need of joint bandwidth and NFV resource orchestration: A realistic 5G access network use case,” *IEEE Communications Letters*, vol. 22, no. 1, pp. 145–148, 2018.

[5] S. Mehraghdam, M. Keller, and H. Karl, “Specifying and placing chains of virtual network functions,” in *Cloud Networking (CloudNet), 2014 IEEE 3rd International Conference on*. IEEE, 2014, pp. 7–13.

[6] C. Ghribi, M. Mechtri, and D. Zeghlache, “A dynamic programming algorithm for joint VNF placement and chaining,” *Proceedings of the 2016 ACM Workshop on Cloud-Assisted Networking*, pp. 19–24, 2016.

[7] A. Gupta, B. Jaumard, M. Tornatore, and B. Mukherjee, “Service chain (SC) mapping with multiple SC instances in a Wide Area Network,” *arXiv preprint arXiv:1704.06716*, 2017.

[8] R. Cohen, L. Lewin-Eytan, J. S. Naor, and D. Raz, “Near optimal placement of virtual network functions,” in *Computer Communications (INFOCOM), 2015 IEEE Conference on*. IEEE, 2015, pp. 1346–1354.

[9] X. Wang, C. Wu, F. Le, A. Liu, Z. Li, and F. Lau, “Online VNF scaling in datacenters,” in *Cloud Computing (CLOUD), 2016 IEEE 9th International Conference on*. IEEE, 2016, pp. 140–147.

[10] A. Leivadeas, M. Falkner, I. Lambadaris, and G. Kesidis, “Resource management and orchestration for a dynamic service chain steering model,” in *Global Communications Conference (GLOBECOM), 2016 IEEE*. IEEE, 2016, pp. 1–6.

[11] A. Hmaity, M. Savi, F. Musumeci, M. Tornatore, and A. Pattavina, “Protection strategies for virtual network functions placement and service chains provisioning,” *IEEE International Workshop on Resilient Networks Design and Modeling (RNDM)*, 2016.

[12] M. Savi, M. Tornatore, and G. Verticale, “Impact of processing costs on service chain placement in network functions virtualization,” in *Network Function Virtualization and Software Defined Network (NFV-SDN), 2015 IEEE Conference on*. IEEE, 2015, pp. 191–197.

[13] F. M. David, J. C. Carlyle, and R. H. Campbell, “Context switch overheads for linux on ARM platforms,” in *Proceedings of the 2007 workshop on Experimental computer science*. ACM, 2007, p. 3.

[14] V. Bobrovs, S. Spolitis, and G. Ivanovs, “Latency causes and reduction in optical metro networks,” in *Optical Metro Networks and Short-Haul Systems VI*, vol. 9008. International Society for Optics and Photonics, 2014, p. 90080C.

[15] S. Choy, B. Wong, G. Simon, and C. Rosenberg, “The brewing storm in cloud gaming: A measurement study on cloud to end-user latency,” in *Proceedings of the 11th annual workshop on network and systems support for games*. IEEE Press, 2012, p. 2.

[16] G. Xiong, P. Sun, Y. Hu, J. Lan, and K. Li, “An optimized deployment mechanism for virtual middleboxes in NFV-and SDN-Enabling Network.” *TIIS*, vol. 10, no. 8, pp. 3474–3497, 2016.

[17] C. Westphal, “Challenges in networking to support augmented reality and virtual reality.” ICNC, 2017.

[18] The Metro-Haul project deliverables. [Online]. Available: https://metro-haul.eu/deliverables/

[19] M. F. Bari, S. R. Chowdhury, R. Ahmed, and R. Boutaba, “On orchestrating virtual network functions,” in *Network and Service Management (CNSM), 2015 11th International Conference on*. IEEE, 2015, pp. 50–56.

[20] M. Savi, M. Tornatore, and G. Verticale, “Impact of processing costs on service chain placement in network functions virtualization,” in *Network Function Virtualization and Software Defined Network (NFV-SDN), 2015 IEEE Conference on*. IEEE, 2015, pp. 191–197.

[21] A. Gupta. (2016) On service chaining using virtual network functions in operator networks. [Online]. Available: http://networks.cs.ucdavis.edu/presentation2016/Gupta-07-29-2016.pdf

Performance Analysis of QoT Estimator in SDN-Controlled ROADM Networks

Alan A. Díaz-Montiel*, Jiakai Yu†, Weiyang Mo†, Yao Li†, Daniel C. Kilper† and Marco Ruffini*
*CONNECT Centre, Trinity College Dublin, Ireland
†University of Arizona, Arizona, United States
Email: adiazmon@tcd.ie, jiakaiyu@email.arizona.edu, wmo/yaoli/dkilper@optics.arizona.edu and marco.ruffini@tcd.ie

***Abstract*—While new SDN control planes promise to provide faster and more dynamic provisioning of optical paths, mis-estimation of optical signal to noise ratio (OSNR) is still an issue that reduces the amount of capacity available in practice to allocate data paths. Typically, additional margins are applied to the estimation of OSNR for given paths, however, in the absence of detailed knowledge on the gain function of EDFAs, these margins are applied equally to all paths, leading to network under-utilisation. In this paper we show this effect over a simulated optical network based on the Telefonica Spanish national topology, emphasising the reduction in capacity due to the incomplete knowledge the network controller has on the exact wavelength and time based variation of EDFA amplifiers gain. We consider this to be of high relevance, as it opens up the road for further experimentation on the use of sparse optical performance monitoring to provide data that can be used to improve the QoT estimation from the control plane of an SDN-based system. The simulation used is based on the Mininet framework. This provides the advantage of testing SDN control planes that can be then utilised on experimental ROADM networks. In order to link SDN controller and the Mininet Emulation environment, we have developed an optical agent capable of simulating the behaviour of ROADMs systems.**

I. Introduction

Requirements driven by 5G services and the functional convergence of access, metro and cloud networking, are creating new challenges for the access-metro transport network [1], requiring fast dynamic reconfigurability of the optical transport layer above the current state of the art [2].

Dynamic add, drop and routing of wavelength channels can generate optical power dynamics which may result in signal quality degradation. This becomes even more complex in mesh networks, hence, the research community has been recently working on solutions for dynamic optical switching, both in proprietary [3] and open systems [4]. One of the key elements for enabling dynamic switching in ROADM networks is the presence of optical performance monitoring (OPM) functions [5], so that the control plane can operate on a feedback loop that takes into account the state of the active optical channels. OPM techniques are however still in their infancy and require highly specialised tools that are often limited in their capabilities. Recent studies have shown the beneficial impacts of OPM at intermediate nodes, enabling dynamic management decisions to reconfigure and optimise network channels [6]. While Software Defined Networking (SDN) approaches introduce a high level of flexibility for managing network resources, there is a lack of standardised interfaces for optical networking devices (i.e., optical switches), not to mention the absence of open interfaces in the OPM equipment, which leads to high-cost, complex solutions for monitoring in real-time the state of a network. Real-time analysis is crucial for dynamic light-path provisioning and network adaptation, overcoming the suboptimal solution of over-provisioning network resources for system-specific purposes [7]. Furthermore, the additional information provided by the OPM mechanism could not only assist reconfiguration and optimisation of the network performance, but also enable a better use of resources upon service setup (i.e., OSNR estimations, based on distance vs. modulation formats) [5]. However, on-site signal monitoring is still difficult to achieve, mainly because of the high CapEx and OpEx it generates.

To overcome the limitations imposed by OPM techniques, as it is the application of monitoring processes after the installation of network resources (i.e., wavelength allocation), multiple studies have proposed the inclusion of estimation functions for predicting the communication performance of an optical network [13] - [22]. While these approaches have given an insight into the physical impairments, this remains an area that could highly benefit from the use of modern technologies (i.e., machine learning techniques) to optimise the reliability of these functions for resource allocation and/or switching operations. Since it is possible that an estimation is not accurate, margins can be set in order to reduce potential failures. In this study, we analyse the performance achieved by applied fixed margins to a Quality of Transmission (QoT) estimator, which considers OSNR signal degradation, in order to achieve pre-allocation of light-paths, and dynamic switching.

While SDN is playing a major role in the control and management of electronic switching resources, its operation in the optical layer is still left to proprietary and disaggregated implementations. In this study, we also present an SDN control plane with OPM management capabilities based on the OpenFlow v1.5 recommendations, as an extension to the flow-rule capabilities of this protocol. Additionally, we have built an SDN-compatible ROADM network simulator to estimate the loss of performance due to the lack of information about the network. The latter was developed using open-source resources such as the Mininet framework [10], and software-switches [11].

The remainder of the paper is organised as follows. In section II, we provide a review of related work on the usage of QoT estimators that consider various physical impairments of an optical network. Then, in section III, we describe the

978-3-903176-07-2

architecture model of our system, as well as the physical impairments considered in the all-optical metro-network. In section IV, we present the experiments and the related results of the consideration of multiple margins to the QoT estimator. In the last section, V, we give our conclusions and present the direction we are taking for future work.

II. Related work

In [13], the authors studied the impact of OSNR levels of a signal towards near channels for a given set of connections. While they did not consider the addition of optical noise in the in-line optical amplifiers, the interest of a Quality of Transmission (QoT) estimator was first raised, considering some of the physical impairments in an optical network system. The idea of a QoT function was later introduced in [14], which, in combination with a customised routing algorithm, provided simulated performance studies on the feasibility of including these type of functions into the control plane. By considering a non-heterogeneous model of network elements, they evaluated transmission performance for different wavelengths, based mostly on the Chromatic Dispersion (CD) optical impairment. Since high computational resources are inappropriate for Routing and Wavelength Assignment (RWA) functions in the control plane, they determined it was necessary to quantify the estimation error when using the routing tool as a function of the network. Also in [14], the authors proposed the possibility to combine the QoT estimation with monitoring functions, by retrieving information from the optical nodes at fixed periods of time.

In [15] the authors proposed to use QoT estimations as a function of the CD map, to help derive appropriate margins on the dimensioning of an optical network. Under these considerations, they used the estimated results to determine the number of regenerators needed for a given network, as a function of the applied margins. In this study knowing the CD of the system helped reduce the errors of the QoT estimator.

Following a more statistical approach, the authors in [16] considered the introduction of confidence levels for adding margins in both fixed and adaptive manners. They also used the QoT function to determine the number of regenerators needed at a given transmission. However, it was concluded that comparing the required regenerators is not enough to assess the advantages related to a QoT estimation.

In [17], Leplingard et. al. analysed the application of adaptive margins to a QoT estimator, based on the amount of residual CD and non-linear phase experienced by a signal. In this study it was found that the utilisation of adaptive margins decreases the number of mis-estimations. Nonetheless, according to the authors, while the application of margins guarantees safer dimensioning, it is at the expense of including additional equipments.

Today, the evolution of coherent optical transmission has made it possible to easily recover from CD and Polarisation Mode Dispersion (PMD) using digital signal processing at the receiver, so that accurate QoT estimation for these impairments has become redundant [18]. However, in [18], Zami proposed that it is still relevant to consider the OSNR levels and crosstalk attenuation of signals, as input parameters for QoT functions. In addition, it is mentioned that analysing the performance of a transmission channel from a bandwidth perspective is important, especially when considering multiple physical impairments of a system. Zami proposed that the cumulated uncertainties along the light-paths must be also considered, e.g., as the aggregated noise caused by amplification systems.

Software Defined Optical Networks and Elastic Optical Networks are research areas that have been under development in the last decade. Overall, they propose the idea of having programmable optical elements that can dynamically adapt properties such as wavelength bandwidth or modulation formats [20], and the implementation of optical flexi-grid networking devices [19]. In [21] - [22], the authors studied the benefits of applying elastic modulation gains in the Microsoft's optical backbone in the US. They found that a capacity gain of at least 70% is achievable via elastic modulation. Also, they demonstrated how different wavelengths performed differently across the network, looking at multiple segments of it. From the latter, the authors concluded that different wavelengths might benefit from different modulation formats even while sharing paths.

In [23], the authors proposed an analytical framework for a QoT estimator considering spectrum dependent parameters. While assuming OPM monitoring functions capable of reporting the state of the network, the MATLAB simulations presented in this paper demonstrated that they were able to approximate the prediction of the network behaviour with high-accuracy. Although the analysis presented here lacks consideration of physical layer models, it provides an insight on how novel statistical techniques could improve the precision of a QoT function for signal performance in an optical network.

Bouda et. al. [24], proposed a prediction tool that is dynamically configurable considering optical physical impairments as these changed through the network. They included both linear and non-linear effects, such as Q-factor and non-linear fibre coefficients, for predicting accurate QoT. The authors were able to reduced the Q-estimation error to 0.6 dB. However, they believe the accuracy of the parameters in their model could be improved by considering more data variability, e.g., by changing the launch powers or considering measured OSNR levels.

In [25], a data-driven QoT estimator is analysed from a theoretical perspective. The authors commented on the advantages of approaches based on data analysis, in-gather than based on Q-factor estimation, overcoming the dependency of the consideration of physical layer impairments, eliminating the requirement of specific measurement equipment, as well as extensive processing and storage capabilities. While this approach presented high accuracy (between 92% and 95%) the neural network approach taken in this study presented high computational complexity, which is typically unsuitable for the management of all-optical networks.

Summarising, the literature review carried out in this section suggests that there are a number of areas that still need to be properly addressed. We consider there is a need for a QoT estimator that can accurately estimate OSNR signal degradation across nodes with unpredictable fluctuation in loss and gain. The latter could enable elastic configurations. Another missing component in this field is an SDN-based optical monitoring system, that can provide real-time data to help minimise the error of the QoT estimator discussed above.

III. Architecture and Model

The network topology we considered in this study is the Telefonica national Spanish telecommunication network model proposed in [31]. It consists of 21 nodes and 34 (inter-city) links, with varied distances. In order to analyse large-scale ROADM networks, we incremented the distances in the given Spanish network shown in Fig. 1 in order to operate on point-to-point connections ranging from 500 km to 4000 km. We have reproduced the topology on the Mininet emulator, developing an abstracted representation of an optical node through the use of OpenFlow software virtual switches (CPqD/ofsoftswitch v1.3 user-space software switch [11]). An example of an optical node architecture is reported in Figure 2: each of the ROADM components (WSS and EDFA) were emulated using separate virtual switches.

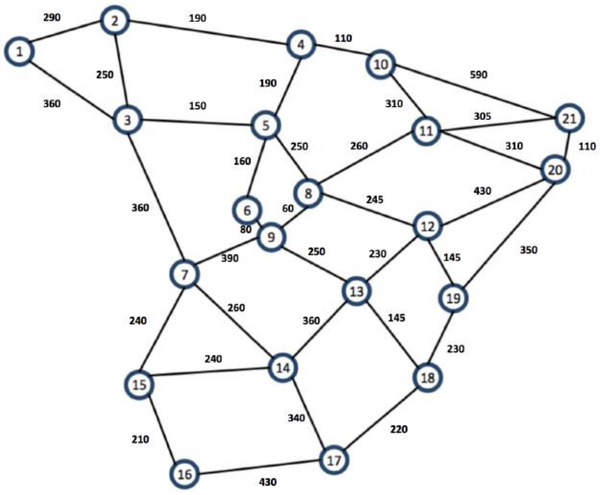


Fig. 1. Telefonica, national Spanish telecommunications network (link distances are reported in km) [31].

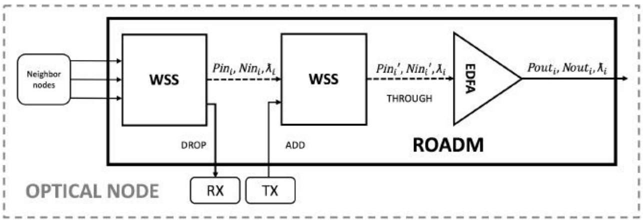

Neighbor nodes
WSS
Pin_i, Nin_i, λ_i
WSS
$Pin_i', Nin_i', \lambda_i$
THROUGH
EDFA
$Pout_i, Nout_i, \lambda_i$
DROP
ADD
ROADM
OPTICAL NODE
RX
TX

Fig. 2. Logical representation of the ROADM node.

Each output port has a post-amplification Erbium Doped Fibre Amplifier (EDFA), which compensates for the losses of the WSS in the system. In our abstracted optical network, at each link we installed an additional EDFA for each fibre span of 100 km, and an additional one at the end of a link to operate pre-amplification. In our model a colourless implementation was achieved by adopting WSS elements for both add and drop ports. The physical transmission and impairments parameters used in our simulations are given in Table I.

TABLE I
Parameters used for the physical transmission and impairments

Physical component	Physical impairment
Launch Power	-2 dBm
Fibre Attenuation	0.2 dB/km
WSS Loss	9 dB
EDFA Noise Figure	6
EDFA Gain	20 dB

In order to simulate the optical performance of a signal traversing the nodes, we encapsulate optical transmission parameters (e.g., signal power and noise) in customised Ethernet packets to allow the exchange of this information across the virtual switches. Following an SDN paradigm, we are able to monitor the exchanged packets via the OpenFlow Protocol calls to the devices. Figure 3 depicts the different layers of communication between the entities considered in our model. At the bottom, there is the data plane (DP), which is represented by the virtual network. In between the data and control plane we implemented an optical agent, which is the entity that simulates the optical behaviour of each network element. The optical agent has bidirectional TCP connections to both the Controller and the DP. The SDN controller is in charge of the network control and management operations (e.g., path computation, connection control) that could operate over a real ROADM network with support for OpenFlow Protocol version 1.5. The agent implementation in our model is used to handle the customised Ethernet packets traversing the network, in order to generate data structures for representing the optical performance, as it uses the values stored in the packet header to keep track of the signal power and noise at each port.

The optical agent utilises equations (1) and (2) for computing the signal power and noise values at a given port of a path:

$$P_o(P_i, G_t, \lambda) = P_i * G_t * f(\lambda) \tag{1}$$

$$N_o(N_i, G_t, \lambda) = N_i*(G_t*f(\lambda))+h(c/\lambda)*(G_t*f(\lambda))*NF*B \tag{2}$$

$$OSNR = P_o/N_o \tag{3}$$

In equation (1), P_o is the output power out of a given node, P_i the input power, G_t the target gain, and λ the wavelength. We determine the input power as the launched power of the system; target gain is a computed gain per EDFA to maintain the signal power, and $f(\lambda)$ is the ripple function that represents the detailed gain transfer function of the EDFA. In

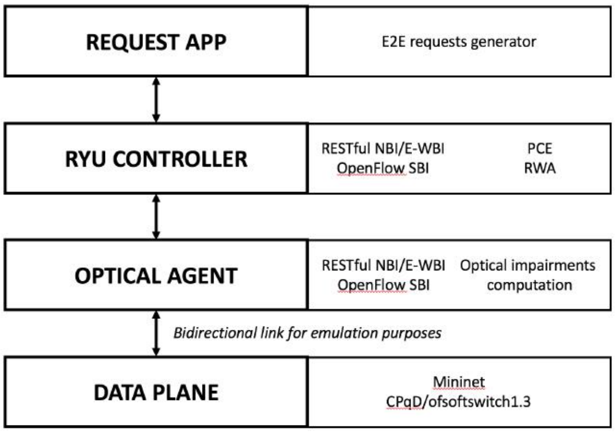


Fig. 3. Layered representation of the complete emulated system.

typical systems, a passive gain flattening filter is manufactured and applied to the amplifiers to compensate for the EDFA gain wavelength dependence. However, this is optimised for a specific operating point and is not tuned to each individual device, which brings in a certain degree of variability. In addition, the gain characteristic of the device will also vary over time. In order to reproduce this effect, since the optical power control stability problem regards the performance of each amplifier [26]-[28], we randomly allocated a different gain function to the different EDFAs of the system. This becomes indeed the main unknown variable in the system that affects the performance of the QoT estimator. In Figure 4, we present the gain functions that are considered for this study. Due to hands-on monitoring procedures and simulations at the CIAN testbed at the University of Arizona, we determined that the signal fluctuation imposed by amplification systems resembles a slowly varying sine function. Then, we shifted these monitored functions to left and right in order to add variability to the signal performance, and maintain the EDFA gain constant.

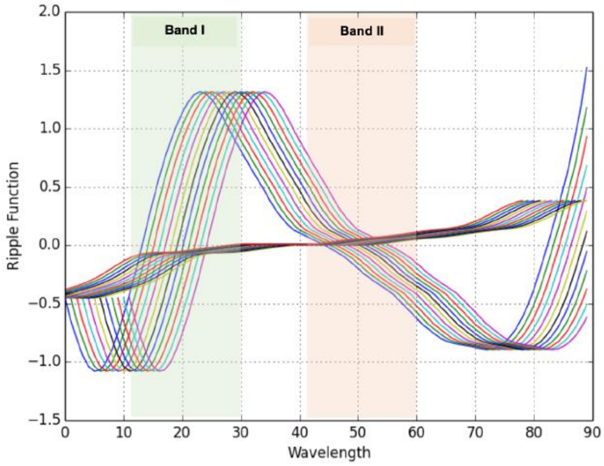


Fig. 4. Wavelength-dependent gain functions utilised for the simulation of detailed EDFA behaviour.

In equation 2, N_o is the output noise, N_i the input noise and G_t the target gain. We determined the input noise to be the generated noise after each phase, being 0 - or none - at the beginning of a single transmission. The system gain is the aggregated gain of the EDFAs that compose a link. We represent this by adding the ripple variation $f(\lambda)$ to the individual target gains *Gt*. Then, *h* is the Planck constant, *c* is the speed-of-light in an optical fibre, *NF* is the noise figure of the amplifiers, and *B* is the bandwidth of a channel in Hertz.

One of the novelties of our study is that we operate the simulations by computing the signal degradation at each node as the packets are traversing the network. Because of the model described in equations 1 & 2, the computation of both output power and noise at each node is dependent of the randomisation of ripple behaviour constraint at each optical amplifier.

In the control plane, we implemented a customised controller based on the Ryu framework proposed by NTT labs [29]. Apart from extending the OpenFlow Protocol descriptions to handle optical parameters, we included networking functions such as path computation, routing and re-routing, as well as OPM capabilities [30]. These are triggered by external applications to enable point-to-point connectivity, or by monitored data retrieved from the nodes. The Northbound Interface of our control plane is developed with RESTful API solutions, allowing for high-level requests from external applications. In addition to the generic networking functions at the control plane, we included an estimation module which performs a prediction of signal degradation given a point-to-point connection. The estimation function implemented in our controller uses the same tools presented in equations (1) & (2). However, assuming it has no detailed knowledge of the exact gain transfer function of each amplifier, it does not assume any such variation (e.g., it assumes a flat unitary ripple function). This calculation is triggered whenever there is a point-to-point request, and determines the feasibility of a light-path to be installed according to the OSNR levels, which are computed using equation (3).

IV. Experiments and Results

The traffic generated for this study consisted of 2,000 end-to-end paths of length between 500 and 4,000 km, across the network topology considered in Figure 1. The experiments were carried out over two different segments of the C-band, shown in Fig. 4, in order to take into account the effect of different gain transfer functions.

Similar to [21], we analysed the feasibility of all the paths in the monitored traffic to be transmitted at different modulation formats, considering the OSNR signal levels of each transmission channel. For the OSNR thresholds, we have assumed those values above BER pre-FEC reported in literature, specifically from [21], which are based on a symbol rate of 32 Gbaud. The modulation formats are QPSK, 8QAM, and 16QAM, with OSNR thresholds, respectively, of: 10 dB, 14 dB, and 17 dB. Carrying out an OSNR analysis of each path, we determined in our model that 36.9% of the traffic could be modulated using 16QAM, 50% at 8QAM, and the remaining 13.1% at QPSK for the first band (1534.8 to 1542 nm). For the second band (1546.8 to 1554 nm), 26.6% of the traffic could be modulated at 16QAM, 70% at 8QAM, and 3.4% at QPSK, when we apply no margins. This constitutes the maximum capacity that the selected paths could carry in the network, if the SDN controller had perfect knowledge on the QoT (in this case the OSNR levels) associated with all paths.

The QoT estimator implemented in our controller predicts the OSNR levels of a given signal traversing a path, as described in Section III. Because the estimation does not consider the optical power fluctuation caused by the amplifiers (only the noise figure is considered), the only option available to improve the likelihood of succeeding in creating a new path is to apply a margin to all the paths. Intuitively, adopting a more conservative margin also reduces the network capacity, as it reduces the number of paths generated.

We have thus analysed the performance achieved when applying different margins to the prediction of the OSNR levels, in order to verify the maximum capacity achievable. The margins are applied to the following formula, which is used to determine whether the results of estimation plus margin is above the required OSNR threshold:

$$OSNR_{est} + M > OSNR_{th} \tag{4}$$

In (4), $OSNR_{est}$ is the estimate OSNR from the controller (which does not know the specific amplifiers gain wavelength dependence), M is the margin applied to the path, and $OSNR_{th}$ the actual required OSNR threshold for setting up the working path. We considered margins from -6 dB (i.e., a conservative approach) to 6dB (i.e., with an aggressive approach). The results are reported in Figure 5. The maximum capacity of the system is the maximum capacity calculated by our simulation using all the possible paths. This would also be the capacity achieved by the SDN controller if it had exact knowledge of the OSNR levels for every path. The curves show that, on one hand, when implementing a conservative margin, i.e., under-estimating the OSNR levels (in the negative region), the QoT estimator progressively rejects the allocation of paths, and the overall capacity decreases accordingly. On the other hand, when the QoT estimator adopts a more aggressive strategy, i.e., over-estimating the OSNR levels (in the positive region), the capacity also decreases progressively, as a higher number of paths does not meet the minimum OSNR threshold for the selected modulation and thus cannot transport data. For the higher values of margin, the QoT estimation will assume that all paths can operate at 16QAM, and the achieved capacity settles at the value of 46% and 32%, respectively for the first and second bands of operation, which are related to the percentage of paths that can support the 16QAM modulation (as already mentioned at the beginning of this section). It is imperative to notice that both under-estimation and over-estimation at the control plane cause discrepancies in the performance of the network because of the misuse of network resources. While constant under-estimation would lead to the non-installation of paths, an over-estimation can lead to the installation of non-feasible paths, hence, installing non-usable resources.

According to our study, the optimal point of the OSNR margin adopted by the controller seem to be around the 0dB value, for both bands examined. However, even at the optimum, since such margins are adopted equally across all paths, the loss of capacity with respect to the maximum capacity is still of the order of 5%-17%.

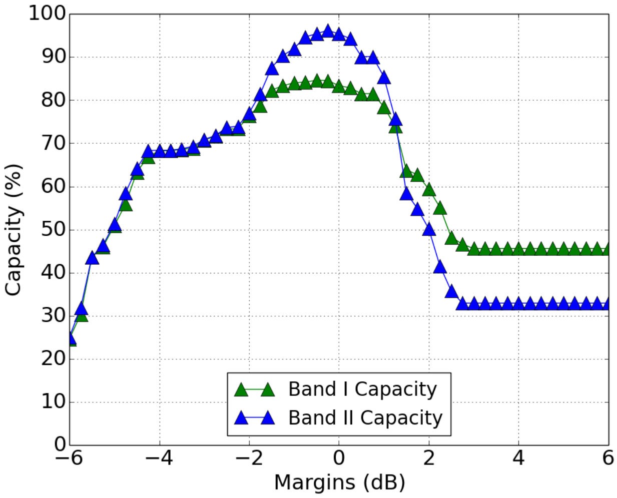


Fig. 5. Network capacity vs. OSNR margins applied by the control plane.

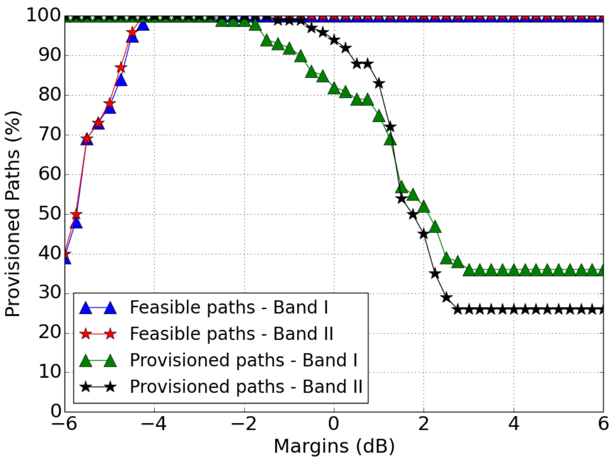


Fig. 6. Comparison of two analysis: Percentage of feasible paths that are provisioned for both bands vs. Margins (dB) (red and blue curves), and provisioned paths above required OSNR threshold for both bands vs. Margins (dB) (green and black curves).

In Figure 6, we show the success rate of two analysis: *(i)* the ratio of number of paths attempted to be installed, with respect to the total number of paths with OSNR levels above BER pre-FEC threshold. A more conservative approach would restrict significantly the attempts of installation, whereas a more aggressive prediction would fall into highly optimistic computations, attempting to install 100% of the paths; *(ii)* we analysed the success rate of the established paths. In other words, how the application of different margins to the QoT estimator affects the accuracy of feasible light-paths. Our results suggest that a conservative approach would increase the accuracy of the prediction of feasible paths, but at the expense of restricting network capacity. Contrasted with Figure 5, it is noted that when over-estimating the feasibility of light-paths, the maximum percentage of feasible paths is 36%, allowing for 46% of the network capacity for the first band. Similarly, this analysis determines that for the second band, only 26% of the paths can be successfully allocated, enabling for 32% of the network capacity.

V. CONCLUSIONS

In this study, we presented a study on the network under-utilisation issue brought by the absence of detailed knowledge on the behaviour of EDFA gain functions across wavelengths and over time. Our simulations, based on the use of a Ryu SDN controller operating over a Mininet emulation of the Spanish national network, showed the difference in network capacity between an optimal situation, where the gain function of the cascaded EDFAs is known in advance, and the real situation, where the controller only uses a flat gain function. By comparing the outcome of the simulator with the prediction of the controller, we could calculate how the use of different margin values for the OSNR affect network capacity. We showed that an aggressive margin strategy, that would tend to over-estimate the available OSNR, reduces the performance, as it favors the adoption of higher modulation rates, leading to situations where the paths created could not operate below the BER threshold. On the other hand, a too conservative strategy that tends to under-estimate the OSNR, would lead to a situation where the controller operates over lower modulation rates, and in some cases declines the creation of wavelength paths (which would instead have worked correctly, according to the Mininet simulation).

As future work, we plan to emulate the use of sparse OPM in the network to gain knowledge on the wavelength and time-varying behaviour of other EDFAs, and feeding the data to machine-learning based techniques to improve the per-path estimation of the OSNR, thus increase the network utilisation.

VI. ACKNOWLEDGMENTS

Financial support from Science Foundation Ireland (SFI) 15/US-C2C/I3132 and 13/RC/2077 (CONNECT) and NSF Center for Integrated Access Networks (CIAN) EEC-0812072 and NSF grant CNS-1650685 is gratefully acknowledged.

REFERENCES

[1] M. Ruffini, Multi-Dimensional Convergence in Future 5G Networks. IEEE/OSA Journal of Lightwave technology, Vol. 35, No. 3, March 2017

[2] Young-Jin Kim et al, Cross-layer orchestration for elastic and resilient packet service in a reconfigurable optical transport network.

[3] R. Muñoz et al, Dynamic distributed spectrum allocation in GMPLS-controlled elastic optical networks.

[4] Open ROADM. http://www.openroadm.org/home.html.

[5] D. C. Kilper and Y. Li, Optical physical layer SDN: Enabling physical layer programmability through open control systems.

[6] F. Meng et al, Field trial of a novel SDN enabled network restoration utilizing in-depth optical performance monitoring assisted network re-planning.

[7] K. Christodoulopoulos et al, Cross-layer and dynamic network orchestration based on optical performance monitoring.

[8] F. Pederzolli et al, YAMATO: The First SDN Control Plane for Independent, Joint, and Fractional-Joint Switched SDM Optical Networks.

[9] L. Velasco et al, Elastic Spectrum Allocation for Variable Traffic in Flexible-Grid Optical Networks.

[10] Mininet. https://github.com/mininet.

[11] CPqD Software Switch. http://cpqd.github.io/ofsoftswitch13/

[12] Collings, Brandon. "New devices enabling software-defined optical networks." IEEE communications magazine 51.3 (2013): 66-71.

[13] T. Zami, A. Morea and N. Brogard, "Impact of routing on the transmission performance in a partially transparent optical network," OFC/NFOEC 2008 - 2008 Conference on Optical Fiber Communication/National Fiber Optic Engineers Conference, San Diego, CA, 2008, pp. 1-3.

[14] Morea, Annalisa, et al. "QoT function and A* routing: an optimized combination for connection search in translucent networks." Journal of Optical Networking 7.1 (2008): 42-61.

[15] F. Leplingard, T. Zami, A. Morea, N. Brogard and D. Bayart, "Determination of the impact of a quality of transmission estimator margin on the dimensioning of an optical network," OFC/NFOEC 2008 - 2008 Conference on Optical Fiber Communication/National Fiber Optic Engineers Conference, San Diego, CA, 2008, pp. 1-3.

[16] A. Morea, T. Zami and F. Leplingard, "Introduction of confidence levels for transparent network planning," 2009 35th European Conference on Optical Communication, Vienna, 2009, pp. 1-2.

[17] F. Leplingard, A. Morea, T. Zami and N. Brogard, "Interest of an adaptive margin for the Quality of Transmission estimation for lightpath establishment," 2009 Conference on Optical Fiber Communication - incudes post deadline papers, San Diego, CA, 2009, pp. 1-3.

[18] T. Zami, "Physical impairment aware planning of next generation WDM backbone networks," OFC/NFOEC, Los Angeles, CA, 2012, pp. 1-3.

[19] Gerstel, Ori, et al. "Elastic optical networking: A new dawn for the optical layer?." IEEE Communications Magazine 50.2 (2012).

[20] Yin, Yawei, et al. "Software Defined Elastic Optical Networks for Cloud Computing." IEEE Network 31.1 (2017): 4-10.

[21] Ghobadi, Monia, et al. "Evaluation of elastic modulation gains in Microsoft's optical backbone in North America." Optical Fiber Communication Conference. Optical Society of America, 2016.

[22] Filer, Mark, et al. "Elastic optical networking in the Microsoft cloud." Journal of Optical Communications and Networking 8.7 (2016): A45-A54.

[23] Sartzetakis, I., et al. "Quality of transmission estimation in WDM and elastic optical networks accounting for space–spectrum dependencies." Journal of Optical Communications and Networking 8.9 (2016): 676-688.

[24] Bouda, Martin, et al. "Accurate prediction of quality of transmission with dynamically configurable optical impairment model." Optical Fiber Communications Conference and Exhibition (OFC), 2017. IEEE, 2017.

[25] Panayiotou, Tania, Sotirios P. Chatzis, and Georgios Ellinas. "Performance analysis of a data-driven quality-of-transmission decision approach on a dynamic multicast-capable metro optical network." Journal of Optical Communications and Networking 9.1 (2017): 98-108.

[26] D. Gorinevsky, and G. Farber, "System analysis of power transients in advanced WDM networks," Journal of lightwave technology, vol. 22, no. 10, pp. 2245, 2004.

[27] L. Pavel, "Dynamics and stability in optical communication networks: a system theory framework," Automatica, vol. 40, no. 8, pp. 1361-1370, 2004.

[28] D.C. Kilper, C. Chandrasekhar, and C.A. White, "Transient gain dynamics of cascaded erbium doped fiber amplifiers with re-configured channel loading," In Proc. Optical Fiber Communication Conference, 2006.

[29] Ryu SDN Controller. https://osrg.github.io/ryu/.

[30] Li, Yao, et al. "tSDX: Enabling Impairment-Aware All-Optical Inter-Domain Exchange." Journal of Lightwave Technology (2017).

[31] M. Ruiz, O. Pedrola, L. Velasco, D. Careglio, J. Fernández-Palacios, and G. Junyent, "Survivable IP/MPLS-Over-WSON Multilayer Network Optimization," J. Opt. Commun. Netw. 3, 629-640 (2011)

Optimization of Spectrally and Spatially Flexible Optical Networks with Spatial Mode Conversion

Mirosław Klinkowski*, Grzegorz Zalewski*, Krzysztof Walkowiak†,
*National Institute of Telecommunications, 1 Szachowa Street, 04-894 Warsaw, Poland; m.klinkowski@itl.waw.pl
†Department of Systems and Computer Networks, Wrocław University of Science and Technology, Poland

***Abstract*—Spectrally and spatially flexible optical networks (SS-FONs), which combine space division multiplexing (SDM) with flexible-grid elastic optical network (EON) technologies, bring additional complexity to network control due to the handling of a larger number of spatial modes in SDM than in conventional EONs. To effectively optimize such networks, in particular, to generate good-quality solutions of low optimality gaps in reasonable computation times, efficient algorithms are required. In this work, we focus on the routing, spatial mode, and spectrum allocation (RSSA) problem in the SS-FONs in which conversion of spatial modes in switching nodes is allowed. We propose and evaluate two enhancements in RSSA processing, namely, algorithm parallelization and application of dedicated data structures, which are built into a hybrid simulated annealing with greedy RSSA heuristic algorithm. To assess the quality of generated RSSA solutions, we develop suitable procedures for estimating solution lower bounds. The results of numerical experiments show the effectiveness of proposed techniques in speeding up the search for good-quality RSSA solutions.**

***Index Terms*—optical networks, space division multiplexing, routing, space and spectrum allocation, network optimization, offline planning, parallel algorithm, column generation**

I. Introduction

Space division multiplexing (SDM) is a forthcoming optical network technology going beyond the capabilities of fixed-grid wavelength division multiplexing (WDM) and flexible-grid elastic optical network (EON) systems by enabling parallel transmission of several co-propagating spatial modes in suitably designed optical fibers [1], [2]. SDM, when combined with multi-carrier (i.e., super-channel, abbreviated as SCh) and distance-adaptive transmission enabled by EONs, brings many benefits including enormous increase in transmission capacity, extended flexibility in resource management due to the introduction of the spatial domain, as well as potential cost savings thanks to the sharing of resources and the use of integrated devices [3]. The feasibility of the spectrally-spatially flexible optical network (SS-FON) concept has been demonstrated in [4].

The main concern in optical networking is provisioning of lightpath connections for transmitted signals. A lightpath is an optical path established between a pair of source-destination nodes. In SS-FONs, the lightpaths carrying SChs are routed through the network over the spatial modes of SDM suitable links within an appropriately assigned spectrum segment. Having a set of traffic demands, the routing of lightpaths requires a contention-free allocation of spectrum resources on spatial modes of each link belonging to the routing path of each connection realizing a demand. It translates into the problem of routing, spatial mode, and spectrum allocation (RSSA), which consists of finding lightpath connections, tailored to the actual width of transmitted signals (i.e., SChs), for end-to end demands that compete for spectrum and spatial resources [5]. The RSSA problem is present both in the phase of offline network planning and during its operation. The former case is more challenging since it involves the establishment of lightpath connections for a set of traffic demands, and such optimization problem is know to be $\mathcal{NP}$-hard [5]. The latter case concerns mainly the provisioning of lightpaths for connection requests that arrive and disappear stochastically (i.e., one-by-one). Even here, the complexity of RSSA may be high if in-operation planning (i.e., network re-optimization during its operation) is performed [6].

The handling of a much larger number of spatial modes than in single-mode EONs dramatically increases the complexity of both hardware and control functions in SDM networks. It results in a large set of decision variables in network optimization, which makes RSSA more complex than the routing and spectrum allocation (RSA) problem in EONs [7]. Consequently, efficient algorithms for solving the RSSA problem are required in such networks. In the following, we discuss related works and present our contributions.

A. Related Works

In network/connection planning, the decision how to allocate the traffic demands is made in an off-line manner, with relaxed processing time constraints (when compared to dynamic connection provisioning). Therefore, more complex and time consuming optimization methods can be applied for solving such decision problems. Among them, the most usual one considered for RSSA-related problems is the mixed-integer programming (MIP) modeling approach (see e.g., [8]). Also simple greedy algorithms are quite frequently used (e.g., [9]). The least popular are meteheuristics, which in most cases employ a simulated annealing (SA) approach (see e.g., [10]).

The advantage of MIP is that it yields exact (i.e., globally optimal) solutions. However, its key shortcoming is low scalability, i.e., it cannot provide optimal or even feasible results in a reasonable time for larger instances of complex problems, which is the case of most of optimization problems in optical networks. Contrarily, (meta-)heuristics can effectively generate feasible solutions to large-scale problems relatively quickly; however, they do not guarantee their optimality. Indeed, in

978-3-903176-07-2 © 2018 IFIP

many flexible-grid EON scenarios such methods may face scalability problems and have difficulties with providing solutions close to optimal ones [11].

One way to speed up the generation of good-quality solutions is to employ algorithm parallelization. The processing power of CPUs and GPUs in terms of the number of processing cores is continually increasing. It enables the possibility to run a number of threads, where each thread performs its own search for optimal solutions in the solution space. Effective application of GPUs in optimization of EONs has been demonstrated in [12]. To the best of our knowledge, algorithm parallelization in SS-FONs has not been studied yet.

Another way to increase the efficiency of optimization procedures is to use appropriate data structures. A basic representation of the allocation status of spectral-spatial resources is by means of a matrix structure, for instance, a matrix of binary variables, where each matrix element denotes whether given frequency slice on the corresponding spatial mode is used or not. In [13], another matrix representation that indicates the size of available contiguous spectrum blocks is proposed. Using this matrix, it is enough the check the value of a single matrix element so that to know if a demand can be accommodated in a given frequency slot, while a binary representation involves the iteration over a number of consecutive slices. As we show in this paper, further improvements can be achieved.

Since (meta-)heuristics do not have in-built optimality guarantees, additional procedures aiming at estimation of solution lower bounds (LBs) can be used for evaluation of solution quality. In particular, the lower the relative difference (referred to as *optimality gap*) between the values of generated feasible solution and LB the higher quality of the solution. Note that whenever these values equal (i.e., the optimality gap is 0%), then it is assured that the solution is (globally) optimal. In [14] and [15], the LBs are estimated in EON scenarios by solving an MIP model of the RSA problem in which the spectrum continuity constraint is relaxed. We have not found any similar works in the context of SS-FONs.

For more details on resource allocation schemes in SS-FONs as well as on optimization models and algorithms considered for the RSSA problem refer to our recent comprehensive literature survey presented in [5].

B. Assumptions and Contributions

We focus on an offline network planning problem, which concerns establishing lightpath connections for a set of traffic demands competing for spectral-spatial resources with a goal to optimize their utilization. The considered problem translates into an RSSA optimization problem in which the width of spectrum required in the network to allocate the demands is subject to minimization. As a case study scenario, we assume an SS-FON in which spectral SChs are carried by lightpaths over the spatial resources of optical links consisting of single-mode fiber bundles (SMFB). The lightpaths have assigned frequency slots that do not change on their routing paths (i.e., the spectrum continuity constraint is imposed). However, the network nodes allow for conversion of spatial modes, i.e., for lane changes and switching of modes between any input and output ports. Accordingly, different modes can be assigned to a lightpath in the links belonging to its routing path (i.e., the so-called spatial continuity constraint is relaxed). As discussed in [5], this is one of the most frequently considered scenarios in the literature in the context of SS-FONs.

Our main contribution is development of an efficient RSSA algorithm, based on a standard simulated annealing algorithm combined with a greedy RSSA heuristic, that introduces two enhancements in SS-FON optimization: parallel processing and improved data structures. We also develop suitable procedures for estimation of solution lower bounds, with the aim to evaluate the quality of generated RSSA solutions. As the obtained numerical results show, the proposed techniques significantly speed up the search for optimal RSSA solutions.

The rest of the paper is organized as follows. In Section II, we formulate the considered RSSA problem. In Section III, we describe the optimization algorithm. In Section IV, we present the procedures for estimation of lower bounds. In Section V, we present and discuss the results of numerical experiments. Eventually, we conclude the work in Section VI.

II. Problem Formulation

We formulate the RSSA problem as an MIP problem using the link-lightpath (LL) modelling approach [16].

The considered SS-FON is represented by graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V}$ is the set of optical nodes and $\mathcal{E}$ is the set of fiber links. The set of spatial modes available on each link is denoted as $\mathcal{M}$. On each mode $m \in \mathcal{M}$, for each network link $e \in \mathcal{E}$, the same bandwidth (i.e., optical frequency spectrum) is available and it is divided into set $\mathcal{S} = \{s_1, s_2, \ldots, s_{|\mathcal{S}|}\}$ of frequency slices of a fixed width. The set of (traffic) demands to be realized in the network is denoted by $\mathcal{D}$.

In the LL model, a notion of a lightpath is used. A lightpath is understood as tuple (p, c), where p is a route and c is a frequency slot. The route is a path through the network from the source node to the termination node of a demand ($p \subseteq \mathcal{E}$), while the frequency slot is a set of contiguous slices (the property called the spectrum contiguity constraint) assigned to the lightpath ($c \subseteq \mathcal{S}$). Frequency slot c should be wide enough to carry the bit-rate of demand d on path p, if it is supposed to satisfy this demand. Note that the width of c (i.e., $|c|$) may differ in the function of the length of path p. This fact allows us to model the distance-adaptive transmission, where the best possible modulation format is selected for each candidate path [11]. Frequency slot c is the same for each link belonging to the routing path (according to the spectrum continuity constraint). The set of allowable lightpaths for demand $d \in \mathcal{D}$ is denoted as $\mathcal{L}(d)$. Finally, let $\mathcal{L}$ be the set of all allowable lightpaths.

The RSSA problem in the considered SS-FON scenario with the spatial mode conversion simplifies to selecting one of the allowable lightpaths for each demand in such a way that the sum of lightpaths utilizing the same slice on the same link does not exceed the number of available spatial modes (as the spatial continuity constraint is relaxed). As a consequence, each lightpath is assigned a binary variable $x_{dl}, d \in \mathcal{D}, l \in \mathcal{L}(d)$, where $x_{dl} = 1$ indicates that lightpath l is actually set-up and it carries the traffic of demand d. Besides, each binary variable $y_{es}, e \in \mathcal{E}, s \in \mathcal{S}$, indicates if there is a used lightpath allocated on slice s of link e. Eventually, the use of slice s in the network is indicated by a binary variable $y_s, s \in \mathcal{S}$.

The corresponding MIP formulation of RSSA is as follows:

$$\text{minimize } z = \sum\nolimits_{s \in \mathcal{S}} y_s \tag{1a}$$

$$[\lambda_d] \quad \sum\nolimits_{l \in \mathcal{L}(d)} x_{dl} = 1 \qquad d \in \mathcal{D} \tag{1b}$$

$$[\pi_{es} \geq 0] \quad \sum\nolimits_{l \in \mathcal{L}(e,s)} x_{d(l)l} \leq |\mathcal{M}| y_{es} \quad e \in \mathcal{E}, s \in \mathcal{S} \tag{1c}$$

$$\sum\nolimits_{e \in \mathcal{E}} y_{es} \leq |\mathcal{E}| y_s \qquad s \in \mathcal{S}, \tag{1d}$$

where $\mathcal{L}(e, s)$ is the set of lightpaths routed through link e and slice s, and $d(l)$ is the demand realized by lightpath l. Optimization objective (1a) minimizes the number of the slices actually used (equal to the sum of variables y_s). Constraint (1b) assures that each demand will use exactly one lightpath from the set of allowable lightpaths. Constraint (1c) assures that there are no collisions of the assigned resources, i.e., the sum of lightpaths utilizing the same slice on the same link does not exceed the number of spatial modes. Finally, constraint (1d) defines variables y_s that indicate whether slice s is used on at least one link.

Note that the solution of (1) does not provide explicit information about the spatial modes utilized on consecutive links of the selected lightpaths. Therefore, to obtain a feasible assignment of modes to the lightpaths, a simple post-processing procedure presented in [8] can be applied.

Solving model (1) is difficult even in the case of EONs (i.e., for $|\mathcal{M}| = 1$) [15]. However, the linear relaxation of (1) (referred to as LP) can be useful in estimation of the solution lower bounds, for which we develop a suitable column generation-based procedure in Section IV-B. The procedure makes use of the dual variables associated with the respective constraints of LP, which are denoted in (1) by symbols λ_d and π_{es}.

III. Generating RSSA Solutions

In the search of optimal solutions for the RSSA problem formulated in Section II, we apply a similar optimization approach as in [17], which is a combination of greedy lightpath allocation (GLA) and simulated annealing (SA) – in this paper, we denote it as an SA-GLA algorithm. In particular, GLA processes demands one-by-one, according to a given demand order, and assigns to them lightpaths in such a way that each assignment minimizes cost function (1a) (i.e., spectrum usage). Here, the best routing path from the set of allowable paths $\mathcal{P}$ and vector of spatial modes is selected for each demand, whereas spectrum is allocated using a first-fit (FF) policy. The width of allocated frequency slot is calculated assuming the most spectrally efficient modulation format, but such that its transmission reach exceeds the path length. The demand order is being optimized iteratively by applying SA, in a similar way as in [18]. In particular, at each iteration, SA swaps the order of just two randomly selected demands and, for such new order, it calls the GLA procedure. If the new order leads to the improvement in the objective function, it is considered as the best one and accepted as the current order in next iterations. Otherwise, it is accepted as the current one with certain probability that decreases during SA processing. In the SA-GLA algorithm, GLA is capable of producing feasible RSSA solutions quickly, while SA explores the feasible solution space in the search for (locally) optimal solutions.

The time required to generate optimized RSSA solutions increases with the number of spatial modes and it can be considerable. Indeed, as reported in [7], algorithm processing times might be of the order of tens or hundreds of seconds even when using a simple greedy heuristic approach in SS-FONs supporting spectral-spatial SChs. Therefore, to speed-up the search for optimal RSSA solutions, we propose two enhancements in the SA-GLA algorithm processing, namely, parallelization of SA and application of suitable data structures for efficient search of free spectrum resources in GLA.

A. Parallel Simulated Annealing

In this work, we implement a basic approach for parallelization of our optimization algorithm, in which a number of parallel threads is run on a multi-core CPU, where each thread is associated with a logical CPU core, and each thread calls its own, independent instance of SA-GLA. The instances of SA-GLA are initialized with different orders of demands that, after calling the GLA procedure, result in different initial RSSA solutions. They perform random and uncorrelated swaps of pairs of demands and explore the solution space in the search for optimal RSSA solutions without any exchange of information about their current best solutions. After meeting the termination condition, which in our implementation happens either when the objective value of found solution equals the solution lower bound (estimated in a pre-processing phase) or given computation time limit is exceeded, the best solution found among all the threads is returned.

B. Efficient Spectrum Search

The GLA procedure aims at selecting the best lightpath (i.e., such that minimizes cost function (1a)) for each consecutively processed demand $d \in \mathcal{D}$. It checks iteratively all allowable paths from set $\mathcal{P}(d)$ and looks for a free frequency slot, the same on all links belonging to the path (due to the spectrum continuity constraint), on any spatial mode in each link of the path (since the space continuity constraint is relaxed). GLA applies the FF spectrum allocation policy. Namely, the allocation status of frequency slices in set $\mathcal{S}$ is checked starting from the lowest indexed slice (i.e., slice s_1) until the required number of free consecutive slices (denoted as N) that form the frequency slot is found on any spatial mode in each link of the path. If found, the frequency slot on the the lowest-indexed spatial mode (among possible ones) is selected.

The allocation status of spectral-spatial resources in a network link can be represented in a form of matrix $\mathbf{A}^{|\mathcal{M}| \times |\mathcal{S}|}$. Below, we present four alternative ways in which this data structure can be defined and processed.

1) **Slice Allocation Status (SAS)** – a basic approach used, e.g., in [15], in which $\mathbf{A} = (a_{ms}) \in \{0, 1\}^{|\mathcal{M}| \times |\mathcal{S}|}$, where element $a_{ms} = 0$ indicates that slice s on mode m is free, and $a_{ms} = 1$ otherwise; see Fig. 1(a). In SAS, the

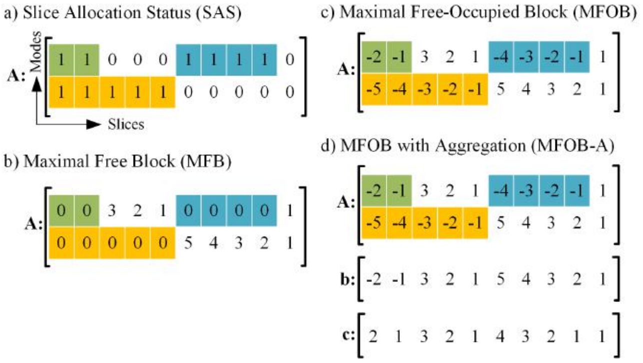
frequency slots in the order: $(s_1, ..., s_N)$, $(s_2, ..., s_{N+1})$, $(s_3, ..., s_{N+2})$, etc., are checked until the first one having all the slices unoccupied for some $m \in \mathcal{M}$ is found. Note that SAS involves the iteration over a number of slices to check the availability of a frequency slot.

2) **Maximal Free Block (MFB)** – a data structure proposed in [13], in which $\mathbf{A} = (a_{ms}) \in \mathbb{N}^{|\mathcal{M}| \times |\mathcal{S}|}$, where the (non-negative integer) value of element a_{ms} indicates the size of available contiguous spectrum block beginning from slice s on mode m; see Fig. 1(b). Similarly as in SAS, the scan of spectrum is performed for consecutive beginning frequency slices: $s_1, s_2, s_3, ...$, however, it is enough to check and satisfy the condition: $a_{ms} \geq N$ for certain $m \in \mathcal{M}$, so that to find the required frequency slot.
3) **Maximal Free-Occupied Block (MFOB)** – the approach that we propose in this paper, in which $\mathbf{A} = (a_{ms}) \in \mathbb{Z}^{|\mathcal{M}| \times |\mathcal{S}|}$, where the positive (integer) value of element a_{ms} indicates the size of available contiguous spectrum block beginning from slice s on mode m (similarly to MFB), and its negative value indicates the size of occupied contiguous spectrum block; see Fig. 1(c). As above, the search for a free frequency slot starts from beginning slice s_1 but, on the contrary to MFB, if $a_{ms_1} < N$, then the next one to be checked is at position $s_1 + |a_{ms_1}|$, and so on. In this way, the intermediate slices are skipped from processing since they obviously does not provide enough free resources to establish the required frequency slot.
4) **MFOB with Aggregation (MFOB-A)** – in this extended version of MFOB, which is suitable for SS-FONs with spatial mode conversion, auxiliary vectors $\mathbf{b}$ and $\mathbf{c}$ are used. Vector $\mathbf{b}$ is defined as $\mathbf{b} = (b_1, ..., b_{|\mathcal{S}|}) \in \mathbb{N}^{|\mathcal{S}|}$, where $b_s = max\{a_{ms} : m \in \mathcal{M}\}$ and it indicates the size of the largest free spectrum block beginning from slice s among all spatial modes. Vector $\mathbf{c}$ is defined as $\mathbf{c} = (c_1, ..., c_{|\mathcal{S}|}) \in \mathbb{N}^{|\mathcal{S}|}$, where $c_s = min\{|a_{ms}| : m \in \mathcal{M}\}$ and it indicates the size of the smallest spectrum block (either free or occupied) among all spatial modes that begins from slice s. The condition that terminates the search for a free frequency slot is: $b_s \geq N$, since it is assured that on at least one spatial mode there is such slot available that begins from slice s. If this condition is not met, the next check for free spectrum resources is performed at slice position $s + c_s$.

Note that both MFOB and MFOB-A have some processing overhead due to necessary updates of data structures after each lightpath allocation. Still, the overall benefits from accelerated spectrum search considerably outweigh this drawback.

IV. Estimating Lower Bounds

In this section, we develop two alternative methods for estimating the LBs of the RSSA problem formulated in Section II, one making use of the relaxed MIP formulation (referred to as LB-MIP) and the other employing linear problem relaxation supported with column generation and cut generation techniques (referred to as LB-CC). The quality of LBs estimated using these methods is evaluated in Section V.



a) Slice Allocation Status (SAS)

$$\mathbf{A}: \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

b) Maximal Free Block (MFB)

$$\mathbf{A}: \begin{bmatrix} 0 & 0 & 3 & 2 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 5 & 4 & 3 & 2 & 1 \end{bmatrix}$$

c) Maximal Free-Occupied Block (MFOB)

$$\mathbf{A}: \begin{bmatrix} -2 & -1 & 3 & 2 & 1 & -4 & -3 & -2 & -1 & 1 \\ -5 & -4 & -3 & -2 & -1 & 5 & 4 & 3 & 2 & 1 \end{bmatrix}$$

d) MFOB with Aggregation (MFOB-A)

$$\mathbf{A}: \begin{bmatrix} -2 & -1 & 3 & 2 & 1 & -4 & -3 & -2 & -1 & 1 \\ -5 & -4 & -3 & -2 & -1 & 5 & 4 & 3 & 2 & 1 \end{bmatrix}$$

$$\mathbf{b}: \begin{bmatrix} -2 & -1 & 3 & 2 & 1 & 5 & 4 & 3 & 2 & 1 \end{bmatrix}$$

$$\mathbf{c}: \begin{bmatrix} 2 & 1 & 3 & 2 & 1 & 4 & 3 & 2 & 1 & 1 \end{bmatrix}$$

Fig. 1. Different options for representation of availability of spectral-spatial resources in a network link.

A. Relaxing MIP Problem (LB-MIP)

One way to obtain an LB is to solve a simplified MIP problem that does not take the spectrum continuity constraints into account. Such approach has been shown to be effective in EONs [14], [15]. The corresponding problem for an SS-FON with spatial mode conversion can be formulated as follows:

$$\text{minimize } z^{lb} \tag{2a}$$

$$\sum\nolimits_{p \in \mathcal{P}(d)} x_{dp} = 1 \qquad d \in \mathcal{D} \tag{2b}$$

$$\sum\nolimits_{m \in \mathcal{M}} x_{dpem} = x_{dp} \qquad d \in \mathcal{D}, p \in \mathcal{P}(d), e \in p \tag{2c}$$

$$\sum_{d \in \mathcal{D}, p \in \mathcal{P}(d) : e \in p} n(d,p) \cdot x_{dpem} \leq z^{lb} \quad e \in \mathcal{E}, m \in \mathcal{M}, \tag{2d}$$

where x_{dp} is a binary variable that indicates if path p is used to realize demand d, x_{dpem} is a binary variable that indicates if spatial mode m is used to realize demand d in link e belonging to path p, z^{lb} expresses the (integer) number of slices required in the most utilized mode in a network link, and $n(d,p)$ is the number of slices requested by demand d on path p.

B. Solving LP with Column Generation and Cuts (LB-CC)

Solving MIP problem (2) may be difficult as it contains integer variables, which involves the use of a branch-and-bound algorithm. As an alternative way for estimating LBs, we can consider solving a linear relaxation of problem (1) (referred to as LP). Note that even the LP problem may be challenging if the number of problem variables and constraints is large. Therefore, to solve LP, we employ a column generation (CG) approach, which was shown to be effective in EONs [15], [19].

In CG, the LP problem is initiated with a limited set of problem variables (columns) and additional variables are iteratively generated and included into LP. Since in large problems most columns are irrelevant for the problem (their corresponding variables equal zero in any optimal solution), the processing complexity can be decreased by excluding these columns from the formulation. Note that an unalterable (possibly complete) set of columns is included into the problem when it is solved by an LP solver using a standard method.

The considered LP problem is initiated with a set of allowable lightpaths $\mathcal{L}$ that represents a feasible RSSA solution (found using the greedy algorithm described in Section III). This set is iteratively extended with new lightpaths that are provided by CG. A key element of CG is to formulate and solve a pricing problem (PP). For the RSSA problem in Section II, PP is defined as a problem of finding, for each demand $d \in \mathcal{D}$, a new lightpath l for which its so-called *reduced cost* is positive (and the largest). The reduced cost of primal variable x_{dl} is calculated as $\lambda_d - \sum_{e \in \mathcal{E}(l)} (\sum_{s \in \mathcal{S}(l)} \pi_{es})$, where $\mathcal{E}(l)$ and $\mathcal{S}(l)$ denote, respectively, the set of links and the set of slices used by lightpath l, and λ and π are the vectors representing an optimal dual solution obtained for the current LP. When found, variables x_{dl} representing new lightpaths are included into LP and the resulting LP problem is solved again (in next iteration). Note that after solving LP, the optimal values of dual variables λ_d and π_{es} are obtained directly from the LP solver. Finally, if no such a lightpath exists for all demands, the CG procedure terminates. For more details on CG the reader is referred to [19].

Eventually, similarly as in [15], it is worth to strengthen the LP with the following valid equalities (cuts): $y_s = 1$, $s \in \{1, 2, \ldots, \lceil z^{lb} \rceil\}$, where z^{lb} is the optimal objective value of LP after solving it with CG. Indeed, z is integer in (1) and, therefore, $z \geq \lceil z^{lb} \rceil$ holds. After adding cuts, the resulting LP problem is solved again using CG. This loop (adding cuts and solving LP with CG) is repeated until z^{lb} is an integer value. In this case, rounding it up and setting $y_s = 1$ for $s \in \{1, 2, \ldots, \lceil z^{lb} \rceil\}$ is worthless since it does not have impact on the solution of LP, and the estimation of LB is terminated.

V. Numerical Results

In this section, we evaluate the proposed RSSA optimization procedures in a European network of 28 nodes and 82 links (EURO28) (shown in the bottom-right corner of Fig. 2). We assume the flexgrid of 12.5 GHz granularity and the number of spatial modes $\mathcal{M} \in \{7, 12\}$. The transmission is realized using spectral SChs and polarization division multiplexing. A spectral SCh consists of a number of optical carriers (OCs), each OC occupying 37.5 GHz, and a guard-band of 12.5 GHz. For OCs, we consider four modulation formats: BPSK, QPSK, 8QAM, and 16QAM, of the transmission reach 6300, 3500, 1200, and 600 km [20], and the carried bit-rate 50, 100, 150, and 200 Gbit/s per OC, respectively. We consider that the OCs forming an SCh use the same modulation format. To generate routing paths, we apply a k-shortest path algorithm (assuming physical path lengths) with $k = 10$ paths per demand, and we exclude the paths of length exceeding the maximum transmission reach. Traffic demands have randomly generated end nodes and bit-rates between 50 Gbit/s and 1 Tbit/s, with 50 Gbit/s granularity. All the results are obtained and averaged over 10 randomly generated demand sets.

Numerical experiments are performed on a dual-processor 2.2 GHz 10-core Xeon-class machine (40 logical cores in total) with 128 GB RAM. The performance of SA depends on its parameters, which are: cooling rate and initial temperature coefficient. We apply the same values of these parameters for the studied EON28 network as in [17], namely, the cooling rate is 0.99 and the initial temperature coefficient is 0.05.

We begin with evaluating the LB estimation procedures presented in Section IV, i.e., LB-MIP (with 1-hour run-time limit) and LB-CC. In addition, we include the results of LP relaxation (solved using CG without cuts). To solve the MIP and LP models, we use CPLEX v.12.6.3 [21] (run in a parallel mode and with default settings). In Table I, we show the obtained LB values (z^{lb}) and processing times (T) for $|\mathcal{M}| \in \{7, 12\}$ and $|\mathcal{D}| \in \{200, 400\}$. We can see that LB-CC offers the best performance in terms of z^{lb} (the highest values) and is much faster than LB-MIP. Solving LP (using CG) can be very fast, however, the obtained LBs are of very low quality and it holds also for other, not reported here values of $|\mathcal{M}|$. Therefore, for the following analysis of optimality gaps of the SA-GLA heuristic, we select LB-CC.

TABLE I
Comparison of estimated lower bounds; T in seconds.

		LB-MIP		LB-CC		LP	
$\lvert\mathcal{M}\rvert$	$\lvert\mathcal{D}\rvert$	z^{lb}	T	z^{lb}	T	z^{lb}	T
7	200	58.2	1137	58.4	64	23.9	1
	400	104.2	3600	104.2	482	49.0	3.6
12	200	57.4	40.6	57.4	7.2	14.0	0.6
	400	61.5	2911	62.5	141	28.6	1.3

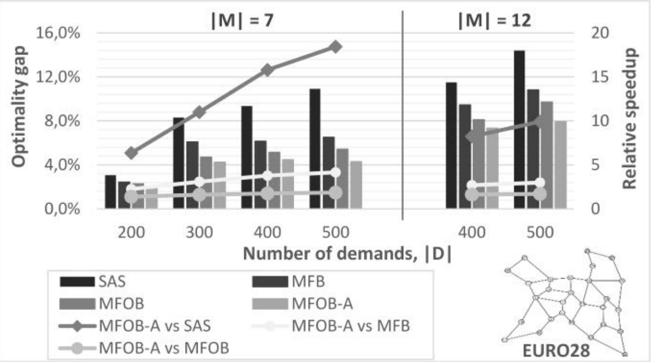


Fig. 2. Comparison of spectrum search procedures in serial SA-GLA with 300 sec. run-time limit, and relative speedup of MFOB-A (vs other options).

Next, we compare the spectrum search procedures presented in Section III run with a serial version of SA-GLA (i.e., 1 thread) with a 300 sec. run-time limit (the short time is due to the large number of executed experiments). In Fig. 2, we can see that the lowest optimality gap, defined as a relative difference between the objective values of LB and heuristic solutions, is obtained with MFOB-A, and the difference between MFOB-A and SAS is of some percents. The optimality gaps increase with the number of demands ($|\mathcal{D}|$) and spatial modes ($|\mathcal{M}|$), which is a result of the complexity of solving larger problem instances. In Fig. 2, we show also a relative speedup of MFOB-A with respect to other spectrum-search options, defined as a ratio of the run-times of a single iteration of the SA-GLA using MFOB-A and the SA-GLA using other option. We can see the the highest speedup of MFOB-A is with respect to SAS (up to 18 times for $|\mathcal{M}| = 7$ and $|\mathcal{D}| = 500$), and it increases with $|\mathcal{D}|$, but decreases with $|\mathcal{M}|$. This gain results from the much faster scanning of spectrum in MFOB-A than in other approaches since MFOB-A skips from processing the entire blocks of

TABLE II
PERFORMANCE OF THE SA-GLA HEURISTIC UNDER DIFFERENT ALGORITHM SETTINGS FOR $|\mathcal{M}| = 7$ AND $|\mathcal{D}| \in \{200, 300, 400\}$.

				$\|\mathcal{D}\| = 200$		$\|\mathcal{D}\| = 300$		$\|\mathcal{D}\| = 400$	
#	Search	Threads	Run-time	z	Δ	z	Δ	z	Δ
(0)	(initial RSSA solution)			89.9	34.4%	114.7	31.5%	140.9	25.6%
(1)	MFOB-A	1	1000 sec.	59.1	1.2%	80.7	3.3%	108	3.5%
(2)	MFOB-A	40	1000 sec.	58.8	0.7%	79.7	2.0%	107.2	2.8%
(3)	SAS	1	1 hour	59.4	1.8%	81.5	4.2%	109.2	4.6%
(4)	MFOB-A	40	1 hour	58.8	0.7%	79.2	1.4%	106.4	2.1%

spectrum that are not suitable for allocation of a demand (as explained in Section III in more details). Consequently, much more iterations of SA-GLA can be performed within the given 300 sec. run-time limit and, thus, better performance results are achieved with MFOB-A than with other approaches.

In Table II, we present performance results of SA-GLA, namely, the objective value of best solution (z) and its optimality gap (Δ), obtained for $|\mathcal{M}| = 7$ and $|\mathcal{D}| \in \{200, 300, 400\}$. Here, our main goal is to compare different aspects of the SA-GLA algorithm and, therefore, the presented results have been obtained for different appropriately selected algorithm settings (indexed from (1) to (4)). For reference, we also include adequate values of initial solutions that initialize the SA-GLA algorithm (under index (0)). First, when comparing (0) with (1)-(4), we can see that SA-GLA is capable of significantly improving the initial solutions. Next, a direct comparison of settings (1) and (2), in which the same spectrum search procedure is applied (i.e, MFOB-A) but different number of threads is assumed, shows that application of parallel processing improves the SA-GLA performance. For instance, the difference in Δ is 1.3% for $|\mathcal{M}| = 300$ if SA-GLA is run with 40 threads instead of 1. Note that the efficiency of parallel processing can be further increased if more sophisticated techniques are applied (e.g., such as exchanging information between threads). Settings (3) and (4) represent, respectively, a baseline version of SA-GLA (run with the SAS approach and 1 thread) and the fully enhanced version of the algorithm. The comparison of (3) and (4) shows that the overall improvement in Δ that comes from both the use of MFOB-A and algorithm parallelization is between about $1\% - 3\%$ (depending on $|\mathcal{D}|$) after 1 hour of SA-GLA performance. Eventually, a comparison of settings (2) and (4), which differ in the considered run-time limit (i.e., 1000 seconds vs 1 hour), shows that the increase of the algorithm processing time may allow SA-GLA to generate better quality solutions.

VI. CONCLUSIONS

We have focused on optimization of RSSA in SS-FONs with spatial mode conversion. We have proposed an efficient optimization algorithm as well as effective procedures for estimating solution lower bounds. We have shown that application of parallel processing and use of dedicated data structures can significantly speed up the search for optimal RSSA solutions. The quality of obtained solutions is high (the optimality gaps reach about $1\%-2\%$ and below), which is a good result taking into account the size of optimized network instances (in terms of both the number of spatial modes, demands, and network dimension). In future works, we will aim at improving the efficiency of parallel algorithm and will address other SS-FON scenarios including the network without lane changes.

ACKNOWLEDGMENTS

The work of M. Klinkowski was supported by National Science Centre, Poland under Grant 2016/21/B/ST7/02212. The work of Krzysztof Walkowiak was supported by statutory funds of the Department of Systems and Computer Networks, Wroclaw University of Science and Technology.

REFERENCES

[1] G. M. Saridis *et al.*, “Survey and evaluation of space division multiplexing: From technologies to optical networks,” *IEEE Commun. Surv. & Tutorials*, vol. 17, no. 4, pp. 2136–2156, 2015.

[2] E. Agrell *et al.*, “Roadmap of optical communications,” *J. of Optics*, vol. 18, pp. 1–40, 2016.

[3] D. M. Marom and M. Blau, “Switching solutions for WDM-SDM optical networks,” *IEEE Comm. Mag.*, vol. 53, no. 2, pp. 60–68, 2015.

[4] B. Shariati *et al.*, “Realizing spectrally-spatially flexible optical networks,” *IEEE Photon, Society Newsletter*, pp. 4–9, Dec. 2017.

[5] M. Klinkowski *et al.*, “Survey of resource allocation schemes and algorithms in spectrally-spatially flexible optical networking,” *Opt. Switch. and Netw.*, vol. 27, pp. 58–78, January 2018.

[6] L. Velasco and M. Ruiz, *Provisioning, Recovery, and In-operation Planning in Elastic Optical Network*. John Wiley & Sons, 2017.

[7] P. Lechowicz *et al.*, “Selection of spectral-spatial channels in SDM flexgrid optical networks,” in *Proc. of ONDM*, Budapest, Hungary, May 2017, pp. –.

[8] K. Walkowiak *et al.*, “ILP modeling of flexgrid SDM optical networks,” in *Proc. of NETWORKS*, Montreal, Canada, Sep. 2016, pp. 1–3.

[9] A. Muhammad *et al.*, “Resource allocation for space division multiplexing: Optical white box vs. optical black box networking,” *IEEE J. Lightw. Technol.*, vol. 33, no. 23, pp. 4928–4941, 2015.

[10] J. Perello *et al.*, “Flex-grid/SDM backbone network design with intercore XT-limited transmission reach,” *IEEE/OSA J. of Opt. Commun. and Netw.*, vol. 8, no. 8, pp. 540–552, 2016.

[11] K. Walkowiak, *Modeling and Optimization of Cloud-Ready and Content-Oriented Networks*, ser. Studies in Systems, Decision and Control. Springer, 2016, vol. 56.

[12] L. Gifre *et al.*, “Experimental assessment of a high performance backend PCE for flexgrid optical network re-optimization,” in *Proc. of OFC*, San Francisco, USA, Mar. 2014.

[13] Z. Shi *et al.*, “Contaminated area-based RSCA algorithm for superchannel in flex-grid enabled SDM networks,” in *Proc. of ACP*, Wuhan, China, Nov. 2016.

[14] K. Christodoulopoulos *et al.*, “Elastic bandwidth allocation in flexible OFDM based optical networks,” *IEEE J. Lightw. Technol.*, vol. 29, no. 9, pp. 1354–1366, 2011.

[15] M. Klinkowski *et al.*, “Solving large instances of the RSA problem in flexgrid elastic optical networks,” *IEEE/OSA J. of Opt. Commun. and Netw.*, vol. 8, no. 5, pp. 320–330, 2016.

[16] L. Velasco *et al.*, “Modeling the routing and spectrum allocation problem for flexgrid optical networks,” *Phot. Netw. Commun.*, vol. 24, no. 3, pp. 177–186, 2012.

[17] M. Klinkowski and K. Walkowiak, “On performance gains of flexible regeneration and modulation conversion in translucent elastic optical networks with superchannel transmission,” *IEEE J. Lightw. Technol.*, vol. 34, no. 23, pp. 5485–5495, 2016.

[18] ——, “A simulated annealing heuristic for a branch and price-based routing and spectrum allocation algorithm in elastic optical networks,” in *Proc. of IDEAL*, Wroclaw, Poland, 2015.

[19] M. Ruiz *et al.*, “Column generation algorithm for RSA problems in flexgrid optical networks,” *Phot. Netw. Commun.*, vol. 26, no. 2-3, pp. 53–64, 2013.

[20] P. S. Khodashenas *et al.*, “Comparison of spectral and spatial superchannel allocation schemes for SDM networks,” *IEEE J. Lightw. Technol.*, vol. 34, no. 11, pp. 2710–2716, 2016.

[21] IBM, “ILOG CPLEX optimizer,” 2017, http://www.ibm.com.

A Cost-effective and Energy-efficient All-Optical Access Metro-Ring Integrated Network Architecture

Dibbendu Roy, Sourav Dutta, Brando Kumam and Goutam Das

***Abstract*—All-optical access-metro networks avoid costly OEO conversions which results in subsequent reduction of infrastructure costs and improvement in energy-efficiency of the network. However, avoiding OEO conversions imply that OLTs are unable to route packets to the ONUs due to unavailability of processing provisions, which necessitates setting up of lightpaths between ONUs. Setting up lightpaths, require a control mechanism, which considers requests from all ONUs in the metro-ring and informs them about the lightpath to be set up. Owing to high data rate of optical networks, the lifetime of a lightpath may be minuscule (few microseconds), which enforces the control mechanism to perform these operations frequently, which incurs large control overhead and processing complexity. This turns out to be a major bottleneck in all existing proposals, which can be alleviated if lightpaths are set up between OLTs instead of ONUs. However, in this case, facilitating data transmission between source and destination ONUs becomes a challenge, due to inability of OLT to process and buffer data packets. In this paper, we resolve this issue by proposing a novel architecture which supports a plausible MAC protocol. The proposed architecture demonstrates significant reduction in cost and power-consumption figures with a slight improvement in reach when compared to traditional architectures.**

***Index Terms*—all-optical, access-metro integration, metro-ring network**

I. Introduction

Passive Optical Network (PON) has emerged to be a widely accepted mature access technology that promises support for wide range of emerging bandwidth-hungry Internet services and applications. A PON typically comprises of a centrally located Optical Line Terminal (OLT), connected to several Optical Network Units (ONUs) located at customer's premises and one or more Remote Nodes (RNs) realized using passive power splitters (PS) or arrayed waveguides (AWGs) [1]. While access networks employ packet switching due to bursty nature of traffic from users, they are often connected to backbone networks through circuit-switched metro network. This results in inefficient handling of bursty data leading to a bandwidth bottleneck termed as metro gap. Moreover, circuit-switched metro, and packet-switched access network necessitates expensive Optical-Electrical-Optical (OEO) conversions at the OLT which account for major infrastructure costs [2]. Several packet-switched metro-network architectures and protocols have been studied to alleviate metro gap [3]–[6]. Packet-switched optical access and metro networks provide an opportunity of avoiding OEO conversions which develops into the notion of all-optical access-metro networks.

Several all-optical access-metro network architectures have been proposed in the literature [2]–[5], [7]. The authors of [3], [5] employ Optical Burst Switching (OBS) to facilitate an all-optical network, which adds significantly to cost and power consumption [2]. Another architecture, STARGATE, proposes infusion of a star network along with the existing metro-ring topology [4]. The OLTs supported by a metro-ring are connected both by the metro-ring network and the star network. Introduction of a star network requires laying of additional cables within the ring, thereby increasing infrastructure costs. These proposals avoid OEO conversions between a source-destination pair, which indicates unavailability of routing provisions at any intermediate node. Thus, they set up optical lightpaths (a wavelength channel without any OEO conversion) between a source ONU and a destination ONU. Since ONUs are equipped with a single transceiver, only one lightpath can be set up for a source-destination pair at any given time to avoid collisions. This calls for a control mechanism which maintains essential network information and performs related processing for setting up new lightpaths. Then, the information about the new lightpath has to be passed over to source, destination and the intermediate nodes, which is achieved through control messages. Therefore, architectures proposed in [3]–[5] separates data and control planes, wherein the control packets undergo OEO conversions. The data packets are sent over the established lightpaths without any OEO conversion.

As discussed above, the aforementioned architectures set up lightpaths by sending control messages among ONUs. We illustrate that such a scheme suffers from large overhead due to control messages and processing complexity. Consider a metro-ring network that supports N OLTs. Suppose a lightpath exists between ONUs of OLT_i and OLT_j. This lightpath may pass through some intermediate nodes in the ring. The same lightpath (wavelength channel) cannot be used for these intermediate nodes unless the lightpath expires. Once the lightpath expires, the corresponding wavelength is released and a new lightpath can be set up. Hence, the control mechanism has to consider all ONUs connected to OLT_i, OLT_j and those connected to the intermediate nodes as well for setting up the new lightpath. In the worst case, the control mechanism considers all ONUs supported by the ring. Thereafter, in order to avoid collisions, the information regarding the new lightpath has to be passed to all ONUs in the metro-ring, which incurs large control overhead. Since the lifetime of a lightpath may be minuscule (few microseconds), the control mechanism has to perform these operations (processing and message passing) very frequently.

Another approach in this regard (all-optical access-metro integration) is to replace a metro-ring network by few metro-

978-3-903176-07-2 © 2018 IFIP

core (MC) nodes which serve both access and core network [7]. This eradicates the overhead involved in message passing as discussed above. However, this scheme suffers from the following drawbacks. Since access fibers are directly linked to MC nodes, the reach of PON is extended (termed as long reach PON or LR-PON). Thus, metro-core nodes serve several PONs (large number of ONUs) employing more power splitters which impairs the power-budget. Further, packets undergo OEO conversions only at ONUs and MC nodes. Since the MC nodes replace the metro-ring, they support as many ONUs as the ring used to support which leads to a manyfold increase in the scheduling complexity. The power-budget and complexity issues would restrict the scalability of such networks.

The above discussions demonstrate that, while replacing metro-ring by metro-core nodes suffers from scalability and processing issues, the other approaches incur large control overhead and processing complexity. Thus, in order to realize a cost-effective all-optical access-metro network, it is essential to reduce the number of control messages involved in setting up of lightpaths without replacing the metro-ring by few centralized nodes. This can be achieved if lightpaths are set up between OLTs instead of ONUs, which reduces the control complexity and overhead drastically (as many times as the number of ONUs connected to an OLT). Further, all-optical network avoids OEO conversions at OLT which indicates that the OLT is devoid of electrical buffering and processing provisions. This leads to the following challenges:

- During the lifetime of a lightpath between two OLTs (say OLT_i and OLT_j), OLT_i can only transmit data intended for OLT_j. Thus, OLT_i must be able to segregate data packets for OLT_j from its ONUs in the upstream (US). Since an OLT is unable to process packets, this segregation poses a challenge.
- The OLT, being unable to process packets, is unaware of its destination address (ONU). Thererfore, routing packets to a destination ONU from the OLT appears to be a challenging task.

In this paper, in order to address these issues, first, we propose a plausible MAC protocol for a metro-ring network. Then we present a cost-effective, energy-efficient All-optical Access Metro Ring Integrated Network (AMRIN) architecture which supports the proposed MAC protocol. AMRIN avoids OEO conversions for the data packets while two control channels are maintained for scheduling in upstream and downstream, wherein the control packets undergo OEO conversions. AMRIN exhibits significant reduction in cost and power-consumption with a slight improvement in reach when compared to broadcast and select (BS) and wavelength split (WS) architectures [1].

The rest of the paper is organized as follows. Section II describes the mentioned issues in detail followed by solution proposals which result in the proposed AMRIN architecture. We evaluate the performance of AMRIN with respect to traditional ones in Section III. Section IV concludes the paper with remarks on the advantages and applicability of AMRIN.

II. Proposed Solutions

As discussed above, we aspire to set up lightpaths between OLTs which reduces the control complexity. We clarify that the lightpaths are set up between OLTs if they belong to the same metro-ring network. Otherwise, lightpath is set up between OLT of a metro ring and the gateway edge router of the same metro-ring. In this section, we develop the intuition behind our proposed architecture which aims at mitigating the issues presented in Section I. First, we look into the challenges in the MAC layer which serves as a motivation for the proposed architecture. Thereafter, we propose architectures for supporting the downstream (DS) and upstream (US) data transmissions respectively.

A. MAC Design

Here, we discuss the issues in the Medium Access Control (MAC) layer and propose solutions for realizing all-optical access-metro networks. First, we describe a process of setting up a lightpath by an OLT. In order to do so, we consider that each metro node is equipped with a reconfigurable optical add-drop multiplexer (ROADM) [8]. The ROADM of metro nodes associated to the source and destination OLTs of a lightpath, adds (for source) or drops (for destination) the wavelength corresponding to the lightpath during its lifetime. All intermediate nodes bypass this wavelength during the lifetime of the lightpath. Thus, ROADMs need to be configured, which requires knowledge of the lightpath to be set up. One of the simplest possible approach to set up lightpaths is by dividing all wavelengths into fixed time slots where each slot is statically assigned to a certain source-destination pair (metro-nodes) similar to slotted rings without channel inspection protocol [8]. Since metro-nodes are aware of the static allocation, the ROADM can be easily configured to add or drop wavelengths during their slots (lightpaths). Variable-sized time slots and their dynamic allocation can also be managed by sending a control message (token). This requires designing an efficient metro-ring MAC protocol which is beyond the scope of this paper.

The protocol discussed above facilitates setting up of lightpaths in the metro-ring. We now describe how the US data from ONUs reach an intended OLT through the set up lightpaths. Traditionally, an OLT schedules the US data without considering the metro-ring MAC protocol. This US data is then processed and buffered at the OLT which allows US data to be sent to an intended OLT in its corresponding slot. The OLT then processes and sends the received data to the destination ONU. However, all-optical networks are devoid of buffering and processing provisions at the OLT. A reasonable approach would then be to segregate packets according to the intended OLTs (or a gateway edge-router), at an ONU. This can be achieved if each ONU maintains separate buffers for all OLTs in the metro-ring. When a packet arrives at an ONU, it discerns the address of the intended OLT, and stores the packet in the corresponding buffer. During the slot between two OLTs (say OLT_j and OLT_k), OLT_j polls its ONUs, to upstream data from the buffers corresponding to OLT_k with help of control messages sent through separate control channels (for US and DS). The OLT schedules multiple ONUs such that the upstream (US) data from these ONUs occupy the respective lightpath (slot) by avoiding collisions. The US data requires a propagation delay to reach an OLT which has to be considered while polling.

The above process enables US data transmission from an ONU to an intended OLT in an all-optical network. However, the inability to process packets at OLT, implies that the OLT is unaware of the destination address (ONUs) of any packet. This suggests that a packet may not be able to reach the destination ONU directly. Thus, in our proposed approach, an OLT bypasses a DS packet, which reaches an arbitrary ONU. This ONU then routes the packet to the destination ONU without sending the packet again to the OLT, which solves the routing issue. Next, we propose an architecture which facilitates the US and DS solutions.

B. Proposed Architecture

Traditionally, access architectures employ two stages of remote nodes for the optical distribution network [9]. As discussed above, the DS routing issue is addressed if all ONUs served by an OLT can share data among themselves (local sharing). Fortunately, we have already proposed an architecture which supports content sharing (CS-OAN) among ONUs [9] by modifying the remote nodes and the ONUs as shown in Fig. 1 and Fig. 2. CS-OAN maintains the passive nature of the remote nodes which is a desirable feature for any PON architecture. The content sharing feature and its operation is briefly discussed as follows:

The downstream (DS) and upstream (US) of CS-OAN operate on non-overlapping sets of wavelengths denoted by $\lambda_1^{'}, \lambda_2^{'}, \ldots, \lambda_N^{'}$ and $\lambda_1, \lambda_2, \ldots, \lambda_N$ respectively. In Fig. 1, the US and DS AWGs (AWG_{US} and AWG_{DS}) ensure that each ONU under a second stage remote node (say $ONU_{x_l,i}$ under RN_{2,x_l}) operates on a unique US and DS wavelength (say λ_i and $\lambda_i^{'}$ respectively). $ONU_{x_l,i}$ upstreams data to OLT only on λ_i and thus the other free US wavelengths $(\lambda_1, \ldots, \lambda_{i-1}, \lambda_{i+1}, \ldots, \lambda_N)$ can be used to reach the rest of the ONUs under same remote node $(ONU_{x_l,1}, \ldots, ONU_{x_l,i-1}, ONU_{x_l,i+1}, \ldots, ONU_{x_l,N})$. This is achieved by employing a $N \times N$ AWG (AWG_{CS} in Fig. 1) which acts as a routing device. If λ_i is incident on j^{th} input port of AWG_{CS}, λ_i appears at $((N - j + i)\ mod\ N)^{th}$ output port of the AWG_{CS} (symmetric routing property of AWG [9]). Thus, a wavelength and input port pair uniquely maps an output port of AWG_{CS}. Connecting each output port to an ONU of the same remote node allows an ONU to access all other ONUs by selecting suitable input port and wavelength pair. For example, in Fig. 1, if $ONU_{x_l,k}$ connected to the k^{th} input port of AWG_{CS}, has to share data to $ONU_{x_l,j}$, $ONU_{x_l,k}$ can transmit the data on wavelength $\lambda_{((N-j+k)\ mod\ N)}$ which appears at the j^{th} output port of AWG_{CS}. The j^{th} output port then is fed to a 3:1 combiner at the DS of $ONU_{x_l,j}$ which now caries both DS and content shared (CS) data. Since the CS data is carried by an US wavelength, there is no collision in the fiber with the DS data of $ONU_{x_l,j}$. Fiber Bragg Grating, filters the US data from the CS data where the US data is forwarded to the OLT through AWG_{US} while CS data is forwarded to the destination ONU through AWG_{CS} and a 3:1 combiner as shown in Fig. 1.

At the ONU, the tunable transmitter transmits the US data as well as data for content sharing. The DS data (now combined with CS data) is passed through a band splitter which segregates the DS and US wavelengths. The DS data is then received by a fixed receiver tuned at corresponding DS wavelength while the CS data is received by a broadband photo detector which tunes into US wavelengths as shown in Fig 2. However, ONUs are usually equipped with a single line card which cannot process both CS and DS data at the same time (receiver collision). This calls for proper scheduling of CS data which may be performed by a control message. Since the OLT can detect DS data only on its arrival, control message cannot be sent before the arrival of the data. Thus, DS data has to be delayed such that the control message informs the ONUs on the same DS wavelength to suspend their reception of CS data. This can be achieved by employing delay lines (DL) at the OLT for each wavelength as shown in Fig. 1.

It is important to note that content sharing is limited within the ONUs of a second stage remote node. Thus, our problem is only partially solved. It is not guaranteed that the packets will reach an ONU which is under the same remote node as the destination ONU (as explained in the example above with RN_{2,x_l}). This problem can be dealt with, if we ensure that packets reach at least one of the ONUs under all second stage remote nodes. It is then easy to conceive that one possible way to achieve this is by realizing the first remote node (RN_1) as a power splitter (PS) which broadcasts a packet to all remote nodes. Since the OLT bypasses any DS packet, the DS architecture operates on all wavelengths supported by the metro-ring network. Thus, each packet reaches one of the ONUs of each second-stage remote node. If the packet has already reached the destination ONU, content sharing will not be necessary. Otherwise, the packet is locally shared (under same remote node) to the destination ONU.

In our proposed AMRIN architecture, we consider that the DS utilizes all wavelengths used for setting up lightpaths in the metro-ring. The ONUs of AMRIN operate on single wavelengths for upstreaming their data to the OLT. However, the lightpaths between two OLTs can be of any DS wavelengths $(\lambda_1^{'}, \lambda_2^{'}, \ldots, \lambda_N^{'})$. This necessitates an OLT to convert US wavelengths to the lightpath wavelength. Since wavelength conversions are not required in the DS, the US data is segregated from the DS at OLT by using a circulator as shown in Fig. 1. First, the US data passes through an AWG which de-multiplexes the wavelengths. Separate wavelength converters (WCs) are employed for each wavelength which ensures full flexibility. Since wavelengths can be converted to any arbitrary DS wavelength, connecting the WC outputs to ROADM ports at the metro ring will require proper switching. However, this switching can be avoided by using a power combiner and AWG combination as shown in Fig. 1 which

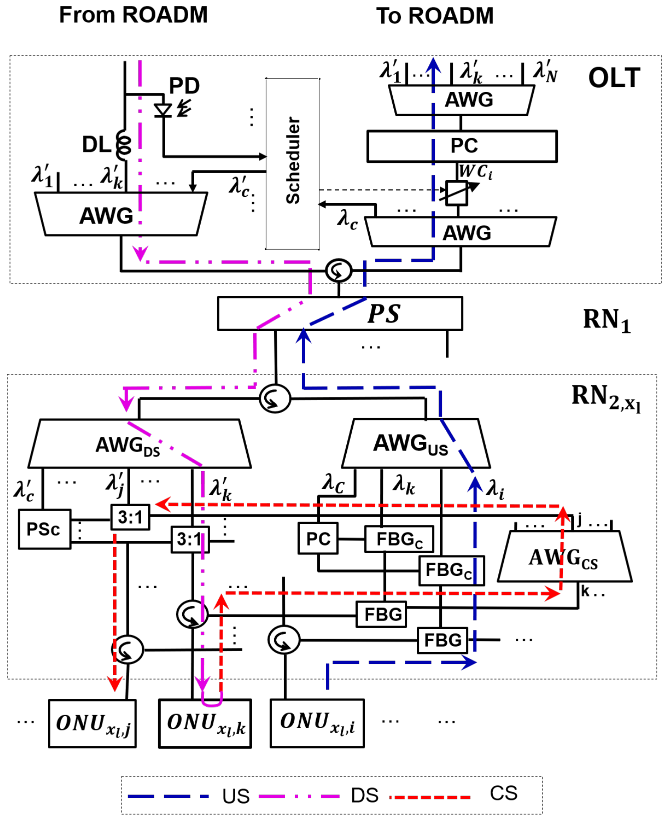


Fig. 1. Proposed Architecture for OLT and Remote Nodes. PC- Power Combiner, 3:1- 3:1 PC, DL- Delay Line, PD- Photo Detector

outputs fixed DS wavelengths for the metro node ROADM.

The above discussions demonstrate the operation of AMRIN for US, DS and CS data. As discussed before, scheduling of CS and US data requires sending control message in the DS direction as shown in Fig 1. In AMRIN, this control message is sent through the control channel λ'_c, which is shared among the ONUs. Following the DS path as described above, the control message reaches the power splitter, PS_C (refer Fig. 1). Each output of PS_C is connected to a 3:1 combiner for each ONU and thus the control packets reach ONU through the DS path. In order to schedule the ONUs, the OLT requires buffer reports from ONUs, which is sent after US data through the control channel λ_c. In order to facilitate this, the US data is passed through a fiber bragg grating (FBG_C) which separates the control wavelength from the US wavelengths. This control wavelength (λ_c) of all ONUs is combined by a power combiner, and is connected to AWG_{US}. The US control packets then reach OLT by following the US path (refer Fig. 1). Since same control channel (both λ'_c and λ_c) is shared by all ONUs, the OLT has to avoid collisions in control messages.

C. Data transmission in AMRIN

We illustrate the working of AMRIN with help of an example. Consider a scenario where a lightpath has been set up between OLT_1 and OLT_2 at wavelength λ'_k. We describe the process by which an ONU (say $ONU_{x_l,i}$) under OLT_1 can facilitate data transmission to any ONU (say $ONU_{x_m,j}$) under OLT_2. Let λ_i, λ_j and λ'_i, λ'_j be the corresponding US and DS wavelengths respectively for ONUs $ONU_{x_l,i}$ and $ONU_{x_m,j}$. For sake of illustration, in Fig. 1, we have shown the path for US from $ONU_{x_l,i}$ to its corresponding metro-node. OLT_1 being aware of the start and end times of the lightpath, sends a control message through the DS path to poll $ONU_{x_l,i}$. In addition, it configures the wavelength converter associated to λ_i (WC_i) such that it converts λ_i to the DS wavelength of the lightpath (λ'_k). The control message informs the address of the OLT associated with the lightpath (in this case OLT_2) to $ONU_{x_l,i}$. This enables $ONU_{x_l,i}$ to upstream (at λ_i) its data from the buffer maintained for OLT_2 at the ONU after receiving the control message. The US data passes through the FBG at RN_{2,x_l} which forwards λ_i to AWG_{US} and reaches OLT_1 through RN_1. The pre-configured WC_i then converts the US wavelength to λ'_k which reaches to the metro node.

The metro node adds λ'_k with help of a pre-configured ROADM to the metro-ring as discussed in Section II-A. The US data from $ONU_{x_l,i}$ is now carried from the metro-node by the lightpath on λ'_k as DS data for OLT_2. Since OLT_2 also implements similar architecture as OLT_1, the path for DS of OLT_2 is shown in Fig. 1 as well. Further, in Fig. 1, we illustrate the data transmission from $ONU_{x_l,i}$ of OLT_1 to $ONU_{x_m,j}$ of OLT_2 for $m = l$. The DS data, received by OLT_2 is bypassed via RN_1 and reaches the k^{th} ONU of all second stage remote nodes. Then $ONU_{x_m,k}$ has to locally share the data to $ONU_{x_m,j}$ shown in Fig. 1. OLT_2 sends a control message to $ONU_{x_m,k}$ for scheduling the local sharing to $ONU_{x_m,j}$. On arrival of this control message, $ONU_{x_m,k}$ upstreams the data on $\lambda_{(N-j+k)modN}$. The data comes out of the j^{th} port of the AWG and finally reaches $ONU_{x_m,j}$ where it will be detected by the broadband photo detector (BPD) as shown in Fig 2. In the case when source and destination ONUs are under same OLT but different remote nodes, the data is routed through ROADM.

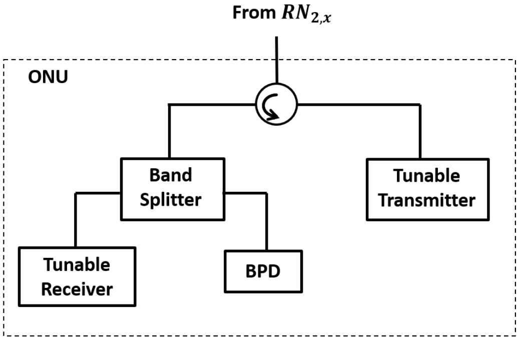


Fig. 2. Architecture for ONUs

III. Results and Discussion

In this section, we compare the proposed AMRIN architecture with two popular TWDM access network architectures:

TABLE I
COST, POWER-BUDGET AND POWER-CONSUMPTION FIGURES OF COMPONENTS

Component	Cost (€)	Power Budget (dB)	Power Consumption (W)	Component	Cost (€)	Power Budget (dB)	Power Consumption (W)
OLT Transceiver	140	NA	6	PS (2X2)	40	-3	0
OLT Line Card	4000	NA	7	Fiber Braggg Grating	30	-1	0
OLT Rack	82800	NA	100	EDFA	8000	30	8
Wavelength Converter	100	NA	0.5	Street Cabinet	150	NA	0
Circulator	30	-1	0	ONU Transceiver	140	NA	2.3
AWG (32X32)	500	-3	0	ONU Line Card	260	NA	4
AWG (4X4)	60	-3	0	Tunable Filter	60	0	0
PS (32X32)	132	-15	0	Band Splitter	60	-2	0

TABLE II
COMPARISON OF PERFORMANCE MEASURES FOR DIFFERENT ARCHITECTURES

	Cost Per User (€)				Power Consumption (W)				Reach (km)				Flexibility	Supports Local Sharing
BW/ONU (Mbps) / Architecture	320	160	80	40	320	160	80	40	320	160	80	40		
AMRIN	839	776	744	728	6.58	6.44	6.37	6.34	150	135	120	105	HIGH	YES
BS	2368	1434	967	732	10.8	8.55	7.43	6.86	130	115	100	85	HIGH	NO
WS	2308	1374	907	673	10.8	8.55	7.43	6.86	145	130	115	100	MODERATE	NO
WS-LS	2556	1623	1155	922	10.8	8.55	7.43	6.86	145	130	115	100	MODERATE	YES

Broadcast and Select (BS) and Wavelength Split (WS) [1]. We consider overall cost per user, power-budget (reach) and power-consumption as metrics for the comparison. The effects of increasing the number of OLTs supported by the metro-ring and technology up-gradation from existing Coarse Wavelength Division Multiplexing (CWDM) to Dense Wavelength Division Multiplexing (DWDM) have also been demonstrated.

Table I enlists the metric values for various components used in the architectures [1]. We consider a metro-ring network supporting 8 OLTs and 32 wavelengths (10 Gbps each) with a total data-rate support of 320 Gbps in the ring, unless mentioned otherwise. For conventional architectures (BS and WS) 32 transceivers and wavelengths will be required at the OLT to support the 32 wavelengths in the metro-ring. In addition, we assume that the total data rate is uniformly distributed between the OLTs which implies that access network would require four 10 Gbps transceivers and line cards (at OLT) since each PON supports an average data rate of 40 Gbps. In AMRIN, as the OLT does not perform any OEO conversions for data packets, only one line card is sufficient for processing the control packets. Since the metro-network is not modified by AMRIN, its components would add to a fixed cost and power consumption figure for all architectures which is ignored in the calculations. The reach is calculated for the access network considering 0 dBm power at the OLT. Further, we consider a receiver sensitivity, system margin and fiber attenuation loss of -25 dBm, 6dBm and 0.2 dB/Km respectively.

A. *Comparison of AMRIN with various access technologies*

As discussed above, we calculate the performance metrics (cost per user, reach and power-consumption) for AMRIN, BS and WS using Table I. We consider the metro-ring network to support 8 OLTs and 32 wavelengths (10 Gbps each) as mentioned above. In Table II, the performance figures have been tabulated for the architectures by varying the average bandwidth per ONU (BW/ONU) which is equivalent to varying the number of ONUs connected to each OLT. For example, a BW/ONU of 320 Mbps $\Longrightarrow$ (40 Gbps)/320 Mbp = 128 ONUs are being served by an OLT. We observe (from Table II) that AMRIN costs almost a third as much compared to BS and WS for BW/ONU = 320 Mbps (128 users). Also, the power consumption figure for AMRIN is lower than BS and WS architectures by 4.22 W. Increasing the number of users, leads to a reduction in gain for both cost and power consumption figures. This is due to the following reason. AMRIN requires a single transceiver and line card at the OLT whereas a WS or BS OLT require 36 transceivers and line cards. The cost of these components are shared among the ONUs and thus the gain reduces by increasing the number of ONUs. For 1024 ONUs (BW/ONU = 40 Mbps), AMRIN incurs a higher cost compared to WS architecture. However, AMRIN supports a technically advanced feature of content sharing which serves the purpose of enhancing user bandwidth [9]. Thus, for comparison, we consider a modified WS architecture that supports local sharing (WS-LS) which emerges to be costlier than AMRIN. The PON reach has been evaluated for varying number of users as well. In addition, AMRIN provides better reach compared WS, WS-LS and BS.

B. *Effect of Number of OLTs in Metro-ring*

Increasing the number of OLTs in the metro-ring manifests scaling of metro-network to support larger number of users. Table III compares the effect on performance measures (cost per user, power-consumption and reach) on increasing the number of OLTs (from 8 to 16) for the architectures discussed above. Table III shows that an increment in the number of

TABLE III
EFFECT OF THE NUMBER OF OLTS IN METRO RING

	Cost Per User (€)				Power Consumption (W)				Reach (km)			
BW/ONU (Mbps)	80		40		80		40		80		40	
No. of OLTs Architecture	8	16	8	16	8	16	8	16	8	16	8	16
AMRIN	744	776	728	744	6.37	6.44	6.34	6.37	120	135	105	120
BS	967	1402	732	951	7.43	8.45	6.86	7.37	100	115	85	100
WS	907	1342	673	891	7.43	8.45	6.86	7.37	115	115	100	100
WS-LS	1155	1591	922	1139	7.43	8.45	6.86	7.37	115	115	100	100

TABLE IV
EFFECT OF NUMBER OF WAVELENGTHS IN METRO RING

	Cost Per User (€)				Power Consumption (W)				Reach (km)			
No. of ONUs/OLT	512		1024		512		1024		512		1024	
No. of Wavelengths Architecture	32	64	32	64	32	64	32	64	32	64	32	64
AMRIN	744	751	728	731	6.37	6.40	6.34	6.35	120	135	105	120
BS	967	1256	732	877	7.43	8.34	6.86	7.32	100	100	85	85
WS	907	1196	673	818	7.43	8.34	6.86	7.32	115	130	100	115
WS-LS	1155	1438	922	1060	7.43	8.34	6.86	7.32	115	130	100	115

OLTs, increases the cost for all architectures. However, the rate of increment for AMRIN is much lower compared to others. For example, in case of BW/ONU = 80 or 512 ONUs/OLT, AMRIN reduces this rate by three times compared to others. This benefit improves further for larger number of users (1024 ONUs/OLT or BW/ONU = 40). Further, power consumption in AMRIN increases by 0.07 W compared to 1.02 W for others, while reach improves by 20 Km.

C. *Effect of Technology Up-gradation*

Technology Up-gradation refers to increase in the overall average data rate supported by the metro-ring. This may be achieved by increasing the number of wavelengths supported by the ring from 32 wavelengths to 64 wavelengths using Dense Wavelength Division Multiplexing (DWDM). Table IV shows that the rate at which cost increases in AMRIN is almost 40 and 50 times lower than other architectures (BS and WS) for 512 and 1024 ONUs respectively. While AMRIN consumes an additional power of 0.03 W and 0.01 W, BS and WS consumes an additional power of 0.91 W and 0.46 W for 512 ONUs and 1024 ONUs respectively. Thus, tables III and IV demonstrate that AMRIN is quite insensitive to scaling and up-gradation of the metro-ring network compared to other architectures which is highly desirable.

IV. CONCLUSION

In this paper, we proposed an access-metro integrated architecture, AMRIN, which facilitates data transmission among ONUs just by setting up lightpaths between OLTs, instead of ONUs, without undergoing OEO conversions at OLT. In the process, the computational complexity and costs, involved in setting up lightpaths, is reduced drastically when compared to existing all-optical access-metro architectures. AMRIN reaps a significant benefit in infrastructure cost (CAP-EX) and overall access-network energy consumption (300% and 64% respectively for 8 OLTs, 32 wavelengths, and 128 ONUs) over traditional TWDM architectures even though it supports a technically advanced feature of content sharing. AMRIN exhibits low sensitivity towards scaling and technology up-gradation compared to the conventional architectures, which is a desirable feature for any future optical network architecture. Also, an AMRIN OLT does not require back plane switch which suggests significant cost benefit. Further extensions of the present work would involve design of efficient protocols both for orchestrating local sharing operation and setting up dynamic lightpaths. In AMRIN, since DS data is routed through ONUs, these ONUs may not forward the data. An efficient mechanism design is required which enforces the local sharing of DS data.

REFERENCES

[1] C. Bhar, G. Das, A. Dixit, B. Lannoo, D. Colle, M. Pickavet, and P. Demeester, “A novel hybrid wdm/tdm pon architecture using cascaded awgs and tunable components,” *Journal of Lightwave Technology*, vol. 32, no. 9, pp. 1708–1716, 2014.

[2] A. Shami, M. Maier, and C. Assi, “Broadband access networks.” Springer, 2009.

[3] J. Segarra, V. Sales, and J. Prat, “An all-optical access-metro interface for hybrid wdm/tdm pon based on obs,” *Journal of lightwave technology*, vol. 25, no. 4, pp. 1002–1016, 2007.

[4] L. Meng, C. M. Assi, M. Maier, and A. R. Dhaini, “Resource management in stargate-based ethernet passive optical networks (sg-epons),” *IEEE/OSA Journal of Optical Communications and Networking*, vol. 1, no. 4, pp. 279–293, 2009.

[5] T. Muciaccia and V. M. Passaro, “A novel sdn-like dwdm all-optical metro-access network architecture,” 2016.

[6] M. Ruffini, “Multidimensional convergence in future 5g networks,” *Journal of Lightwave Technology*, vol. 35, no. 3, pp. 535–549, 2017.

[7] ——, “Metro-access network convergence,” in *Optical Fiber Communication Conference*. Optical Society of America, 2016, pp. Th4B–1.

[8] M. Maier, *Optical switching networks*. Cambridge University Press, 2008.

[9] C. Bhar, A. Mitra, G. Das, and D. Datta, “Enhancing end-user bandwidth using content sharing over optical access networks,” *IEEE/OSA Journal of Optical Communications and Networking*, vol. 9, no. 9, pp. 756–772, 2017.

Analysis of mean packet delay in DR-MPCP limited service using queueing theory

Tomoko Kamimura, Sumiko Miyata
Shibaura Institute of Technology, 3-7-5 Toyosu, Koto-ku, Tokyo 135-8548, Japan
ma17030@shibaura-it.ac.jp, sumiko@shibaura-it.ac.jp

***Abstract*—Ethernet passive optical access network (EPONs) have become widespread optical access network. In EPON uplink communication, communication is reserved by polling from an optical network unit (ONU) to an optical line terminal (OLT) using multi-point control protocol (MPCP). A method of delaying the transmission timing of REPORT messages in MPCP, called delayed REPORT messages-MPCP (DR-MPCP) has been proposed as a way of reducing the mean packet delay time of EPON uplink communication [1][2]. However, this method cannot limit the transmission window size of each ONU (making it gated service). In real networks, the transmission window size should be limited to be fair to each ONU (i.e. it should be limited service). This paper derives the upper limit and theoretical expression of the mean packet delay time in DR-MPCP limited service using the queueing theory M/G/1 model. It also analyzes the characteristics of mean packet delay.**

***Keywords*—Ethernet Passive Optical Network (EPON), MPCP, DR-MPCP, M/G/1-model, Mean packet delay.**

I. Introduction

Access networks connect subscribers such as business offices or house with a service provider's station, which connects to a metropolitan area network or wide area network. Recently, fiber to the home (FTTH) is widely used among access networks. A passive optical networks (PON) is one technology that support FTTH. PONs connect an optical splitter to an optical fiber to which an optical line terminal (OLT) is connected, branches the optical signal, and broadcasts to optical network unit (ONU). Among PON technologies, Ethernet passive optical networks (EPONs) are one of the most popular methods at present [3]. However, EPON has a problem: the mean waiting time of arriving packets (mean packet delay) is long. Thus, the mean packet delay must be decreased to create more efficient access networks. In downlink communication, i.e. packet transmission from the OLT to the ONU, a packet is broadcast through the optical splitter, so the delay of each packet does not significantly affect communication quality. However, in an EPON uplink, i.e. the communication from the ONU to the OLT, time division multiple access (TDMA) is used. This requires dynamic bandwidth allocation (DBA) to appropriately scheduling data to avoid packet collisions. With DBA, each ONU sends its transmission request to an OLT, and then each ONU can reserve the network resources through a reply from the OLT.

The exchange of messages for this reservation is defined by the multipoint control protocol (MPCP) media access method [4]. In MPCP, each ONU sends its set of the transmission requests to the OLT as a 64-byte REPORT message. The OLT calculates both the transmission window and transmission starting time for each ONU. The transmission window is the data size, which means that an ONU can transmit only a reserved amount of packets in each cycle. The OLT sends the calculation result as a GATE message. Then, the ONU sends the packets. Thus, the communication time in the uplink is divided into a reservation interval for the packet transmission control and a data interval for the packet transmission itself.

In EPONs with MPCP, interleaved polling with adaptive cycle time (IPACT) is a common polling method [5]. However, there is a problem with the IPACT method. For this reason, Miyata et al. proposed advanced MPCP called delayed REPORT messages-MPCP (DR-MPCP), in which the transmission timing of the REPORT message is delayed [1][2]. They also modeled DR-MPCP and derived its mean packet delay using the M/G/1 queueing model. They found that the mean packet delay of DR-MPCP is shorter than that of IPACT. However, in [1][2], they analyzed only gated service despite limited service being used in real networks.

Therefore, we analyzed the theoretical expression and characteristics of mean packet delay in limited service using M/G/1 queueing model. To analyze limited service, we first extended the analysis that converts gated service to limited service [1][2]. Using the analysis, we analyzed the upper bound of the mean packet delay of DR-MPCP limited service. After that, we derived an exact solution for the mean packet delay of DR-MPCP limited service while considering round trip time (RTT) for reservation and showed the effectiveness of DR-MPCP using numerical calculation and a simulation.

II. Classification of system model

The basis of the queueing theory analysis used in this research is the polling system [6]. In the polling system, illustrated in Fig. 1, N users send packets in order. The number in this figure means the number of the data interval and reservation interval of the ONU. The total interval obtained by combining the reservation and data intervals of N users is called a cycle. The packets that arrive within a cycle are reserved to be transmitted simultaneously in one reservation interval. There are three systems for determining which packets

978-3-903176-07-2 © 2018 IFIP

are transmitted during the data interval of each cycle: a gated system, an exhaustive system, and a partially gated system [6]. In a gated system, packets that arrive before the reservation interval are transmitted in the data interval. In an exhaustive system, packets that arrive in the reservation and data intervals are all transmitted in the data interval. In a partially gated system, packets that arrive before the data interval are transmitted in the data interval, even if they arrive after the reservation interval. In gated service, all requested packets in a cycle are transmitted within their own data interval. In contrast, in limited service, the amount of transmittable packets is limited (pre-determined) for each data interval.

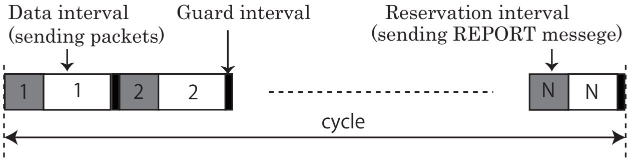


Fig. 1. Polling system.

Table I and II show the differences between IPACT [5] and DR-MPCP [1][2], both of which analyze on the basis of this polling system. In IPACT, the order of the data and the reservation interval are reversed using the polling system. IPACT gated service can be modeled using the gated service of the polling system. Because limited service is a method that restricts the data interval, it can be modeled using the limited service of the polling system. Note that the gated system and gated service are different things.

TABLE I. GATED SERVICE

IPACT [5]	DR-MPCP [1][2]
Only those packets that arrived prior to the ONU's preceding reservation interval are transmitted.	In IPACT gated service, the transmission timing of the REPORT message is shifted.

TABLE II. LIMITED SERVICE

IPACT [5]	DR-MPCP [1][2]
Each ONU's data interval is limited by the maximum transmission window.	In IPACT limited service, the transmission timing of the REPORT message is shifted.

III. DR-MPCP

A. System model

As shown in Fig. 2, it is assumed that in EPONs, N ONUs are connected to an OLT through an optical splitter, and the distance between the OLT and each ONU is the same. Moreover, the arrival rate and service time of each packet are assumed to be independent. In the case of EPONs, uplink communication must consider the DBA in order to avoid the collision of packets. For these reason, we only focused on uplink communication. In this system, the OLT is controlled by cycle polling using DR-MPCP limited service. Let T_{cycle} be one cycle time of the data and reservation intervals of all the ONUs. In limited service, the OLT allocates a data interval for each ONU. The maximum value of this data interval is the upper limit value T_{max}. Because the length of the data interval varies depending on the traffic demand, Tcycle also varies.

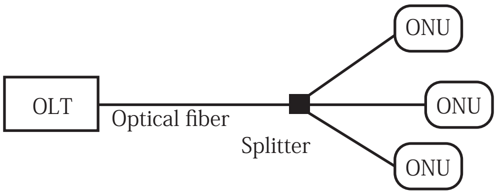


Fig. 2. System overview.

It is also assumed that a packet arriving at any of the ONUs waits in a queue until it receives the GATE message from the OLT and starts to be transmitted. The waiting packets are transmitted in first-in first-out (FIFO) order in accordance with the assigned data interval. It is assumed that the buffer size of each ONU is sufficiently larger than the amount of arriving packets and that there is no packet loss due to queue overflow. A guard time is set between the reservation interval of the ONU and the data interval of the next ONU. Further, the probability of a packet arriving at the queue of each ONU follows an independent Poisson distribution λ/N, and the primary and secondary moments of the packet service time are $\overline{X}$ and $\overline{X^2}$, respectively. Also, the primary and secondary moments of the reservation interval are $\overline{V}$ and $\overline{V^2}$, respectively, and the variance is σ_v^2. The traffic intensity of all ONU packets is assumed to be $\rho = \lambda\overline{X}$.

In the DR-MPCP method, the timing at which ONUs receive the GATE message (as well as the timing at which REPORT messages are transmitted) is delayed. The amount by which it is delayed is the sum of the reservation and data intervals of m ONUs $(0 \leq m < N)$. That is, immediately after the data interval of ONU n, the reservation interval of the $(N-m+n)$th ONU comes.

B. Mean packet delay

Assuming that the mean packet delay of a packet is $\overline{W}$, it can be given by the expression $\overline{W} = \overline{W_F} + \overline{W_Q} + \overline{W_R}$. Here, $\overline{W_F}$ is the residual service time, i.e. the mean remaining time until an arrived packet's service time is complete. $\overline{W_Q}$ is the mean time for the transmissions of packets ahead of the arrived packet and $\overline{W_R}$ is the mean time for the reservations of packets ahead of an arrived packet.

Parameters $\overline{W_R}$ and $\overline{W_Q}$ are common among the polling system, IPACT, and DR-MPCP for gated service. However, $\overline{W_R}$ is a different expression [1][2]. Even if DR-MPCP is expanded for limited service, the only difference is that there is an upper limit on the length of the data interval. Thus, parameters $\overline{W_R}$ and $\overline{W_Q}$ use only the following Eq.s (1) and (2), and we derive only $\overline{W_R}^{dr,lim}$, meaning W_R of DR-MPCP limited service. Here, the expressions of $\overline{W_R}$, $\overline{W_Q}$, and $\overline{W_R}^{dr,gt}$ in DR-MPCP gated service are shown below.

$$\overline{W_F} = \frac{\lambda\overline{X^2}}{2} + \frac{(1-\rho)\overline{V^2}}{2\overline{V}} \tag{1}$$

$$\overline{W_Q} = \rho\overline{W} \tag{2}$$

$$\overline{W_R}^{dr,gt} = \frac{1}{2}(3N-2m-1)\overline{V} \tag{3}$$

The mean packet delay in DR-MPCP gated service can be expressed as follows:

$$\overline{W}^{dr,gt} = \frac{\lambda\overline{X^2}}{2(1-\rho)} + \frac{(3N-\rho-2m)\overline{V}}{2(1-\rho)} + \frac{\sigma_v^2}{2\overline{V}}. \tag{4}$$

C. Differences between gated and limited services

In EPONs, a guard interval is provided between the reservation and data intervals. The sum of the data and reservation intervals, including the guard interval, is called cycle time T_{cycle}. When this cycle time increases, the packet delay also increases. When this cycle time decreases, the proportion of the cycle time to the guard interval increases.

Let T_{cycle_max} be the maximum value of cycle time. It can be expressed as follows:

$$T_{cycle\_max} = N(T_{max} + \overline{V}). \tag{5}$$

Note that we assume that the guard interval is included in $\overline{V}$. In gated service, T_{max} is set according to the buffer size of each ONU. However, T_{cycle} for limited service is restricted by T_{max}.

IV. Analysis of DR-MPCP limited service

In DR-MPCP gated service, all packets that arrive before the REPORT message is transmitted can be transmitted in the data interval of the next cycle. However, gated service is not realistic. This is because it creates unfairness of transmission window arises among users. In limited service, if the arrived packets cannot all be transmitted in the data interval, the remaining packets are reserved again with the REPORT message of the next cycle.

A. Upper bound of the mean packet delay

First, we analyzed the upper limit of the mean packet delay in DR-MPCP limited service. We assumed a traffic situation with a high load at which the data interval becomes T_{max}. The behavior of limited service is similar to that of gated service. In this paper, the upper bound of the limited service was analyzed using the gated service model [1][2].

Traffic situations can be roughly divided into two types: a low traffic load and a high traffic load. As shown in Fig. 3, N users send packets, and the traffic in ONUs from 1th to $(m+1)$th is low. This traffic intensity $\tilde{\rho}$ is $\lambda\tilde{\overline{X}}$. In contrast, the traffic in ONUs from Nth to $(m+2)$th is high. This traffic intensity ρ' is $\lambda'\overline{X} > 1$. In this traffic situation, not all packets can be transmitted in a data interval because many packets arrive. This means that the data interval is the upper limit value. That is, the sum of each data and reservation interval is a fixed value $\tilde{V}$. Here, $\overline{V}$ used in the analysis of the gated service is also a fixed value. Thus, we can replace $\tilde{V}$ with $\overline{V}$. This $\tilde{V}$ can be written as $\tilde{V} = (N-m)\overline{V} + (N-m-1)T_{max}$ from Fig. 3. From this, the limited service model with an upper-bound traffic situation can be approximated using gated service with only one ONU.

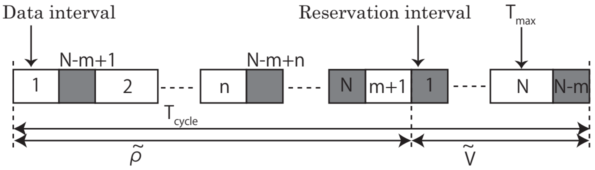


Fig. 3. DR-MPCP upper limit

This can be applied to Eq. (4) but with N, λ, ρ, $\overline{V}$, σ_v^2, and m replaced with $\tilde{\lambda}$, $\tilde{\rho}$, $\tilde{V}$, $N\sigma_v^2$, and 0, respectively.

$$\overline{W}_{simp}^{dr,lim} = \frac{\tilde{\lambda}\overline{X^2}}{2(1-\tilde{\rho})} + \frac{(3-\tilde{\rho})\tilde{V}}{2(1-\tilde{\rho})} + \frac{N\sigma_v^2}{2\tilde{V}} \tag{6}$$

As shown in Fig. 4, we compared the theoretical results with simulation results. The number of ONUs is set to 16 and 32, and m is set to 12 and 24, respectively. The bandwidth of the uplink communication C_{up} is $1Gbps$. The guard time t_g is $1\mu s$, and the REPORT message size L_R is 64 bytes, as is the MPCP standard [4]. The mean reservation $\overline{V} = t_g + 8\frac{L_R}{C_{up}}$ is set to $1.512\mu s$ with $\sigma_v^2 = 0$. The packet payload size is distributed as 64 bytes (47%), 300 bytes (5%), 594 bytes (15%), 1300 bytes (5%), 1518 bytes (28%) between 64 bytes and 1518 bytes [8], with $\overline{X} = 5.090\mu s$ and $\overline{X^2} = 51.468(\mu s)^2$. The simulation code is written using MATLAB.

As shown in Fig. 4, the simulation and theoretical values are almost identical. This indicates our theory is valid. In addition, the mean packet delay increases sharply as the traffic density increases. However, this model cannot be applied data with an interval time less than T_{max}. Therefore, the next section performs theoretical analysis of mean packet delay when the data interval time is equal to or less than T_{max}.

V. Mean packet delay analysis of DR-MPCP limited service

With DR-MPCP limited service, the mean packet delay can be given by the expression $\overline{W}^{dr,lim} = \overline{W_F} +$

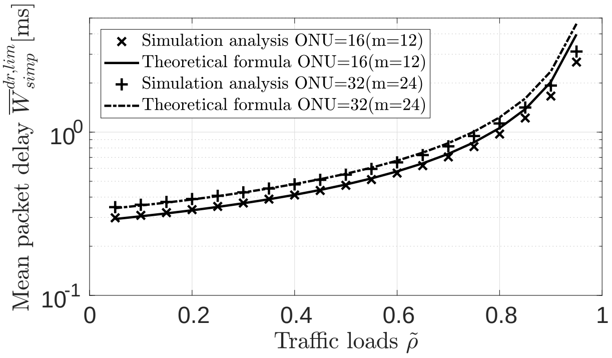


Fig. 4. Mean packet delay in the DR-MPCP upper-limit model.

$\overline{W_Q} + \overline{W}_R^{dr,lim}$. The mean packet delay of the limited service can be analyzed using the same $\overline{W_F}$ of Eq. (1) and $\overline{W_Q}$ of Eq. (2). This is because limited service is the same as gated service except for the upper value of the time for the data interval. Thus, we derive $\overline{W}_R^{dr,lim}$ by considering the difference from $\overline{W}_R^{dr,gt}$.

As shown in Fig. 5, the waiting time of an arrived packet falls into four categories $C_d = \{D_b; R_b; D_a; R_a\}$ on the basis of the time the packet arrive. Without loss of generality, we assume that an arriving packet is for the ONU 1 [7]. In the case of D_b, a packet arrives in the data interval before its ONU's REPORT message. In the case of R_b, a packet arrives in the reservation interval before its ONU's REPORT message. In the case of D_a, a packet arrives in the data interval after its ONU's REPORT message. In the case of R_a, a packet arrives in the reservation interval after its ONU's REPORT message. As shown in Table III, the probability of a packet arriving in a data interval is $\frac{\rho}{N}$, and the probability of a packet arriving in a reservation interval is $\frac{1-\rho}{N}$.

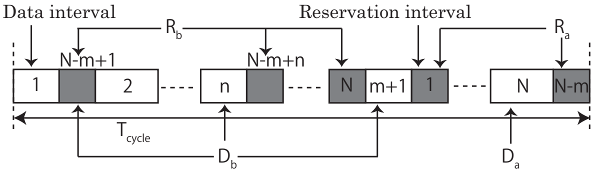


Fig. 5. Cases of packet arrival in first ONU in DR-MPCP

TABLE III. CLASSIFICATION OF PACKET ARRIVAL PROBABILITY OF $\overline{W}_R^{dr,gt}$ [1][2]

C_d	$\overline{W}_R^{dr}$ for DR-MPCP	Probability	Range of n
D_b	$(N-n+1)\overline{V}$	$\frac{\rho}{N}$	$n = 1, ..., m+1$
R_b	$(N-n)\overline{V}$	$\frac{1-\rho}{N}$	$n = 1, ..., m$
D_a	$(2N-n+1)\overline{V}$	$\frac{\rho}{N}$	$n = m+2, ..., N$
R_a	$(2N-n)\overline{V}$	$\frac{1-\rho}{N}$	$n = m+1, ..., N$

In limited service, the size of the transmission window is limited. If many packets arrive at an ONU, some cannot be transmitted within the transmission window and are instead transmitted in the the ONU's data interval in the next cycle. In this situation, additional cycles occur. Therefore, in limited services, it is necessary to newly analyze the additional cycle time.

Let N_Q be the number of packets already queued in the queue of all the ONUs when a packet arrives. At this time, the average number of packets already waiting in each ONU can be written as $\frac{N_Q}{N}$. According to Little's law, it is equal to $\lambda\overline{W}$. That is, the average service time of packets already waiting in the queues of each ONU can be expressed by the following formula.

$$\frac{N_Q}{N} \times \overline{X} = \frac{\lambda \overline{WX}}{N} = \frac{\rho\overline{W}}{N} \tag{7}$$

In these words, the average of the number of cycles required to process the packets waiting in the queues of all ONUs can be expressed as $(\rho\overline{W}/N)/T_{max}$ by using the maximum value T_{max} of the data interval.

However, in the case of D_a or R_a in Table III, the sum of the service time for arrive packets exceeds T_{max}. The packets that cannot be transmitted to the OLT in the cycle are handled in the next cycle. Therefore, the average number of cycles in D_a and R_a is $(\rho\overline{W}/N)/T_{max} - (N - m)/N$. Because the time for arrive packets in D_a and R_a is outside the data interval of ONU 1, the probability that the packet arrives in ONU 1 in these cases is shown as $(1 - \rho/N)q$. Here, q is the probability that the sum of the services time for the arrive packets requested by the REPORT message exceeds T_{max}. In this situation, excess packets are transmitted in the next cycle. In the case of D_b or R_b, the average of the number of cycles is $(\rho\overline{W}/N)/T_{max}$ because the data interval of ONU 1 is less than T_{max}. The probability of packets arriving in these cases is $1 - (1 - \rho/N)q$.

Table IV shows the division packet arrival in these cases. The additional mean waiting time in DR-MPCP with cycle time added by the limitation of transmission window for the limited service can be expressed by the following equation:

$$\begin{aligned}\Delta\overline{W}_R^{dr,lim} &= \left\{\frac{\rho\overline{W}}{NT_{max}} \times (1 - ((1 - \frac{\rho}{N})q))\right.\\ &\left.+ (\frac{\rho\overline{W}}{NT_{max}} - (\frac{N-m}{N})) \times ((1 - \frac{\rho}{N})q)\right\} N\overline{V}\\ &= \left\{\frac{\rho\overline{W}}{NT_{max}} - (\frac{N-m}{N}) \times (1 - \frac{\rho}{N})q\right\} N\overline{V}.\end{aligned}$$

where, $\Delta W_R^{di,lim} = W_R^{dr,lim} - W_R^{dr,gt}$. Thus, we can derive $W_R^{lim,lim}$,

$$\begin{aligned}\overline{W}_R^{dr,lim} &= \frac{(3N-2m-1)\overline{V}}{2} + \frac{\rho\overline{WV}}{T_{max}}\\ &\quad -q(\frac{N-m}{N})(N-\rho)\overline{V}.\end{aligned} \tag{8}$$

TABLE IV. DR-MPCP CASES FOR LIMITED SERVICE.

Cases	Average number of cycles	Probability
D_b or R_b	$(\rho\overline{W}/N)/T_{max}$	$1-(1-\rho/N)q$
D_a or R_a	$(\rho\overline{W}/N)/T_{max}-\frac{N-m}{N}$	$(1-\rho/N)q$

Next, we derive the probability q. Let T' be the average of the data interval that is satisfied by a value less than T_{max}, and the following equation holds.

$$\frac{\rho}{1-\rho}=q\frac{T_{max}}{\overline{V}}+(1-q)\frac{T'}{\overline{V}} \tag{9}$$

This formula can be summarized as follows:

$$\begin{aligned}\frac{\rho}{1-\rho}\overline{V}&=qT_{max}+T'-qT'=q(T_{max}-T')+T'\\ q=\frac{\frac{\rho}{1-\rho}\overline{V}-T'}{T_{max}-T'}&=\frac{\frac{\rho}{1-\rho}\overline{V}-T'+T_{max}-T_{max}}{T_{max}-T'}\\ &=1-\frac{T_{max}-\frac{\rho}{1-\rho}\overline{V}}{T_{max}-T'}.\end{aligned} \tag{10}$$

The parameter T' is an unknown parameter. In this work, we use the approximation used in [7].

$$T_{max}-T'\approx T_{max} \tag{11}$$

By using Eq. 11, q becomes as follows:

$$q\approx 1-\frac{\rho\overline{V}}{T_{max}(1-\rho)}. \tag{12}$$

Therefore, by substituting Eq. (12), the mean packet delay in DR-MPCP limited service is as follows:

$$\begin{aligned}\overline{W}^{dr,lim}&=\frac{\lambda\overline{X^2}+(3N-2m-\rho)\overline{V}}{2(1-\rho-\frac{\rho\overline{V}}{T_{max}})}\\ &-\frac{(N-m)(N-\rho)q\overline{V}}{(1-\rho-\frac{\rho\overline{V}}{T_{max}})}+\frac{(1-\rho)\sigma_v^2}{2\overline{V}(1-\rho-\frac{\rho\overline{V}}{T_{max}})}.\end{aligned} \tag{13}$$

VI. NUMERICAL ANALYSIS

In this section, we compare the mean packet delay, Eq. (13) derived in the previous chapter with that obtained by simulation. Parameters for numerical analysis were set in the same way as in Section IV-A.

In general, if the distance from the OLT to the ONU increases, the RTT (i.e. the time from sending the REPORT message to returning the GATE message) for the reservation of the transmission also increases. The time form transmitting the REPORT message (reserving a packet) to transmitting the reserved packet with DR-MPCP is shorter than that with IPACT. If the RTT exceeds than the time from packet reservation to packet transmission, an idle interval occurs because packets cannot be sent until the GATE message arrives. Therefore, in the DR-MPCP system, it is necessary to take this into consideration when setting the shifting amount m. In this numerical analysis, m with RTT taken into consideration is as follows:

$$m^*=\max\left(\left\lfloor\frac{(\overline{T_{cycle}}-RTT)\times N}{\overline{T_{cycle}}}\right\rfloor,0\right), \tag{14}$$

where, $\overline{T_{cycle}}$ is the mean value of T_{cycle} for each ONU. In this study, it is assumed that propagation delay occurs only in RTT. RTT was calculated using the group refractive index $n_g=1.46$ of quartz optical fiber [9]. Further, T_{max} was set after setting the maximum value of the cycle time as $T_{cycle\_max}=2[ms]$ [5]. From Eq. (5), T_{max} is set as follows:

$$T_{max}=\frac{T_{cycle\_max}}{N}-\overline{V}[s]. \tag{15}$$

Table V gives the value of m^* using Eq. (14). However, when ρ is 0.55 or less, m^* is always 0, so it is omitted from the table. The results in this table reveal that a range of $m^*=0$ exists. When $m*=0$, DR-MPCP is the same as IPACT, which does not shift the REPORT message.

TABLE V. m^* (DISTANCE FROM OLT TO ONU IS $10\ km$)

N / ρ	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95
16	0	0	0	0	0	3	6	9	12
32	3	6	9	12	15	19	22	25	28

The theoretical formula (13) of $\overline{W}^{dr,lim}$ and the simulation values were compared on the basis on the above traffic parameters. The results are shown in Fig. 6.

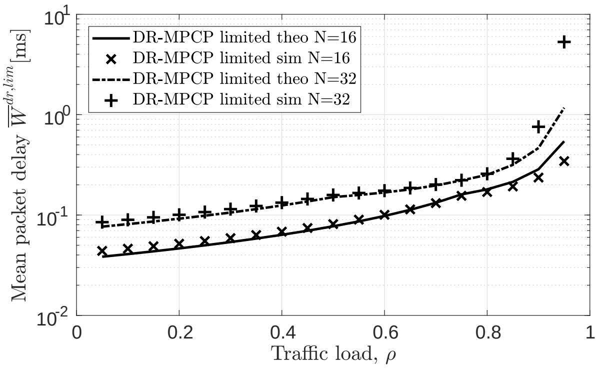


Fig. 6. Mean packet delay of DR-MPCP limited service. Distance from OLT to ONU is $10\ km$.

As shown in Fig. 6, the simulation and theoretical results match closely. However, when the traffic intensity is high, the value of the theoretical expression deviates from the simulation value. This is because the approximation formula (11) assumes that the processing time for packets arriving at the ONU is equivalent to the set T_{max}. However, in a real network, if the traffic intensity is high, the value of T' also increases. In that case, Eq. (11) does not hold. That is, it can be considered that the value of q in Eq. (12) is estimated to be lower than its actual value. Therefore, in order to approximate the theoretical formula to the simulation value, it is necessary to appropriately set the value of T'(that is, q).

Next, we analyze the case where the distance between the OLT and the ONU is changed, shown in Fig. 7. In this analysis, from the values in Table VI, the ONU selects traffic $\rho = 0.8$ in which m^* is not 0 in both $N = 16$ and 32. The value of m in this figure is shown in Table VI. As shown in Fig. 7, the value of the mean packet delay increases as the distance increases. This is because the RTT increases as the distance increases. In this traffic situation, the allowable shifting amount m must be set to 0, as is the case for IPACT. In these words, as the distance from the ONU to the OLT increases, the effectiveness of DR-MPCP's decrease in the delay time in comparison with IPACT decreases.

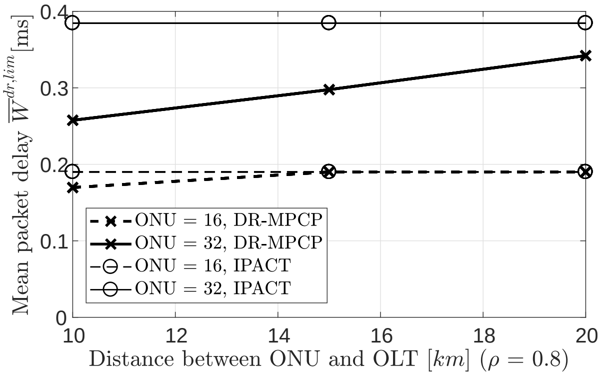


Fig. 7. Mean packet delay of DR-MPCP limited service

TABLE VI. MAXIMUM VALUE OF m IN RANGE WHERE IDLE INTERVAL CANNOT BE FORMED WHEN CONSIDERING RTT($\rho = 0.8$)

N / Distance from OLT to ONU $[km]$	10	15	20
16	3	0	0
32	19	12	6

VII. CONCLUSION

In this research, we derived the mean packet delay of DR-MPCP limited service and showed the validity of our theoretical analysis by comparing it with a simulation. However, we found that the value of our theoretical analysis became less accurate when the traffic intensity was high. In the future, we will improve this part of our theoretical analysis. In addition, as a result of comparing the mean packet delay while consider the RTT according to the distance from the OLT to the ONU, we found that the effectiveness of the DR-MPCP increases as the distance decreases. In the future, we will propose an extended DR-MPCP to decrease mean packet delay even for long distance. Moreover, the result of this research can be applicable not only to EPON but also to NG-PON2 [10] and others. For this reason, we would like to adapt to other systems in the future. Finally, since the traffic pattern assumed in this work is single, we would like to extend it in the case of multi-dimensional traffic [11] in the future.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number JP16K16050.

REFERENCES

[1] S. Miyata, K. Baba, K. Yamaoka, H. Kinoshita, "DR-MPCP: Delayed REPORT message for MultiPoint Control Protocol in EPON," Proc. of IEEE RNDM 2015, pp. 237-242, Oct. 2015.

[2] S. Miyata, K. Baba, K. Yamaoka, "Exact mean packet delay for Delayed REPORT message MultiPoint Control Protocol in EPON," Journal of Optical Communications and Networking, IEEE/OSA Journal of Optical Communications and Networking, vol. 10, issue. 3, pp. 209 – 219, March. 2018.

[3] Derek Nesset, "PON Roadmap [Invited]," Journal of Optical Communications and Networking Vol. 9, Issue 1, pp. A71-A76, Jan. 2017.

[4] G. Kramer, "Ethernet Passive Optical Networks," New York: Mc-Graw Hill, 2005.

[5] G. Kramer and B. Mukherjee, "IPACT: A Dynamic Protocol for an Ethernet PON (EPON)," IEEE Communications Magazine, Feb. 2002.

[6] D. P. Bertsekas and R. G. Gallager, *Data Networks*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1992.

[7] S. Bharati, P. Saengudomlert, "Analysis of Mean Packet Delay for Dynamic Bandwidth Allocation Algiorithms in EPONs," IEEE, Jornal Of Lightware Technology, Vol. 28, No.23, pp. 3454-3462, Dec. 1, 2010.

[8] C. G. Park et al., "Packet delay analysis of symmetric gated polling system for DBA scheme in an EPON," Telecommun. Syst., vol.30, no. 1-3, pp. 13-34, 2005.

[9] J. M. Senior, *Optical Fiber Communications Principles and Practice,* Prentice Hall: PEARSON, 2009.

[10] ITU-T Recommendation G.989 :" 40–Gigabit–capable passive optical networks (NG-PON2): Definitions, abbreviations and acronyms, " 2015.

[11] M. D. Logothetis, I. D. Moscholios, A. C. Boucouvalas and J. S. Vardakas, " Delay Performance of WDM-EPON for Multi-dimensional Traffic Under the IPACT Fixed Service and the Multi-Point Control Protocol", Proc. of 2nd European Teletraffic Seminar, Blekinge Institute of Technology, Karlskrona, Sweden, 30 September - 2 October, 2013.

Photonic Sub-Lambda Transport: An Energy-Efficient and Reliable Solution for Metro Networks

Masahiro Nakagawa, Kana Masumoto, Hidetoshi Onda, Kazuyuki Matsumura
NTT Network Service Systems Laboratories
NTT Corporation
3-9-11, Midori-cho, Musashino-shi, Tokyo, 180-8585 Japan
nakagawa.masahiro@lab.ntt.co.jp

***Abstract*—Telecom carriers need to reduce operational expenditures (OPEX) to reduce total network cost. Such OPEX include power consumption, maintenance, and repair related costs, all of which must be considered, especially when providing various network services nationwide. This paper thus presents the Photonic Sub-Lambda transport network (PSL network), an energy-efficient and reliable optical network architecture for metro networks. Numerical results reveal that the PSL network can simultaneously reduce power consumption by 30%+, failure-recovery operations by 40%+, and repair costs by 80%+ compared with reconfigurable optical add/drop multiplexer (ROADM)-based networks.**

***Keywords*—optical network architecture; metro network; operational expenditures; numerical analysis**

I. INTRODUCTION

The transport networks of telecom carriers generally consist of a large amount of transport equipment deployed in thousands of office buildings [1, 2]. Such buildings are located everywhere from densely populated urban areas to sparsely populated rural ones for providing nationwide network service coverage. Moreover, such transport networks consist of many metro and core networks, and metro networks are more numerous and have much more equipment deployed in them than core ones (e.g., by two orders of magnitude each). Furthermore, telecom carriers need to deal with ever-diversifying user requirements while keeping both capital expenditures (CAPEX) and operational expenditures (OPEX) under control [3], especially in metro networks.

So far, telecom carriers have applied evolving optical transmission technologies and packet switching technologies, which reduce CAPEX per transported bit, to their networks. However, on-going OPEX are currently becoming more and more important than initial CAPEX for telecom carriers. For instance, it is indicated that yearly OPEX are now typically 2–5 times higher than CAPEX [4]. Note that OPEX in transport networks can be divided into several categories, such as continuous cost resulting from power consumption and space, maintenance and repair, service provisioning, and service management [5]. Business process optimization and automation of operations with SDN/NFV technologies have been widely investigated for OPEX savings [6, 7], and these methods can effectively save on service provisioning and management related OPEX. However, continuous power consumption, maintenance, and repair related OPEX (i.e., the major contributors to network OPEX in many cases) remain to be tackled [7]. Naturally, such OPEX in metro networks can be considerable when operating 1k-building scale network infrastructure. Therefore, in addition to SDN/NFV efforts, a promising metro network architecture is needed that not only further reduces CAPEX but also lowers power consumption and suppresses maintenance/repair frequency in order to reduce total network cost.

To meet this need, this paper presents the Photonic Sub-Lambda transport network (PSL network), an optical network architecture that not only requires low CAPEX but also offers energy-efficiency and reliability. The basic concept of the proposed architecture has already been presented [8], but our previous study [8] focused only on CAPEX reduction. In contrast, this paper describes how the PSL network can reduce power consumption, maintenance, and repair related OPEX and extensively analyzes such OPEX. Specifically, the PSL network consumes less power and has lower failure frequency than traditional metro networks since it minimizes O/E/O conversions and leverages optical passive devices. Moreover, numerical results quantitatively clarify the OPEX benefits of the PSL network.

This paper is organized as follows. In Sec. II, we describe the conventional metro networks and summarize some related work. Section III presents our PSL network in detail. Then we show and discuss the results of numerical analysis in terms of power consumption, maintenance, and repair related OPEX in Sec. IV. Finally, we provide conclusions in Sec. V.

II. METRO NETWORK ARCHITECTURES AND RELATED WORK

Metro networks are basically aggregation networks between several access and core networks. Whereas limited numbers of buildings (i.e., nodes) in urban cities are interconnected by optical links in core networks, metro networks aggregate/distribute various traffic demands between access and core networks. Naturally, metro networks are more numerous and have more equipment deployed in them than core networks. Note that traffic volume to be accommodated in metro networks strongly depends on the area: there are large differences in traffic volume between rural and urban areas.

978-3-903176-07-2 © 2018 IFIP

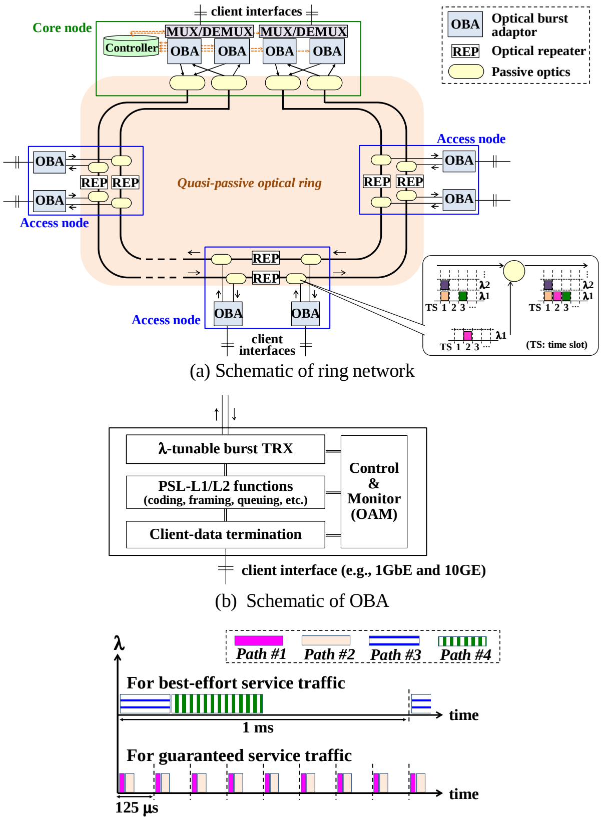
Today, two main architectures have been widely deployed in metro networks: the reconfigurable optical add/drop multiplexer (ROADM)-based wavelength-routed network and the electronic-switch based opaque network. The former can often waste lambda capacity since traffic volume can be smaller than rigid and coarse-grained path bandwidth, which can result in high CAPEX. On the other hand, the latter enables flexible resource utilization, but O/E/O conversion and electronic processing are required in every node, which can lead to high power consumption. Thus, neither can optimize both CAPEX and power consumption.

From an operational point of view, redundant configurations and failure-recovery operations must be executed to maintain service quality. Note that component failures are inevitable, and telecom carriers need to conduct numerous failure-recovery operations, especially in metro networks, when operating 1k-building scale networks. In general, such failure-recovery operations in transport networks require human intervention and transportation to/from buildings in which the failed equipment is deployed. This can be a significant cause of OPEX, hence a metro network architecture that has lower network-failure frequency would be useful. However, in the abovementioned conventional network architectures, end-to-end paths traverse multiple optical or electronic switches (active components) at every node. Thus, the failure rate of such switches affects end-to-end reliability and network-failure frequency. Their failure rates are not negligible, so both component failure rates and number of active components used must also be considered to minimize total network costs.

Several optical metro networks have recently been proposed [9–12]. Although these solutions can flexibly utilize optical fiber capacity while reducing electronic processing, the component cost of high-end devices such as high-speed optical switches needs to be considered, especially when traffic volume to be accommodated is small. This is mainly because expensive solutions cannot suppress CAPEX per transported bit in small-traffic areas. Moreover, the failure rate of high-end devices and the power consumption of corresponding drivers must be considered to suppress OPEX. Also, an “open” optical transport solution is now actively being discussed [13], which has the potential to prevent vendor lock-in scenarios and reduce CAPEX. However, a new OPEX factor of integrating various equipment of various vendors needs to be considered. Furthermore, the architectural change from the current networks is marginal and cannot lead to significant OPEX savings. Therefore, to drastically reduce not only CAPEX but also OPEX even in small-traffic areas, a new network architecture is needed, which is presented in the next section.

III. Photonic Sub-Lambda Transport Network (PSL Network)

The PSL network is intended to aggregate various services’ traffic from geographically separated nodes in a low-CAPEX, energy-efficient, and reliable way. For achieving this, time division multiplexing in the optical domain with optical passive devices is utilized, which enables resources to be flexibly utilized without electronic switching or sophisticated components. An outline of the PSL network is shown in Fig. 1,



(a) Schematic of ring network

(b) Schematic of OBA

(c) Example of transmission schedule

Fig. 1. Outline of PSL network.

where a particular node (core node) is connected to the core network, while the other nodes (access nodes) are connected to access networks. As shown in Fig. 1, this network mainly consists of a quasi-passive optical ring, optical transceiver (TRX) modules, and electronic functions. Note that quasi-passive means that some access nodes require optical repeaters (REPs) for supporting transmission distance in metro networks. Every node has optical passive devices (e.g., couplers and arrayed waveguide gratings (AWGs)) and optical burst adaptors (OBAs; see Fig. 1(b)) that encapsulate the traffic from client interfaces in an optical burst and execute an optical burst transmission. To avoid optical burst collisions, the controller at a core node manages a burst transmission schedule. An example of such a schedule is illustrated in Fig. 1(c), in which best-effort service traffic and guaranteed service traffic is simultaneously accommodated while multiple wavelength resources are utilized. This collision avoidance mechanism enables multiple bursts/paths to be multiplexed in the optical domain with optical passive devices that consume no power and have quite a long lifetime. As a result, resources can be shared across many paths while minimizing O/E/O conversions and electronic functions such as header processing, buffering, and electronic switching. Moreover, the number of OBAs required at each node can be flexibly determined to meet traffic conditions, which allows right-sized solutions and pay-as-you-grow designs. Thus, the PSL network can use fewer TRXs than

ROADM-based networks, which can lead to lower CAPEX, power consumption, and failure frequency.

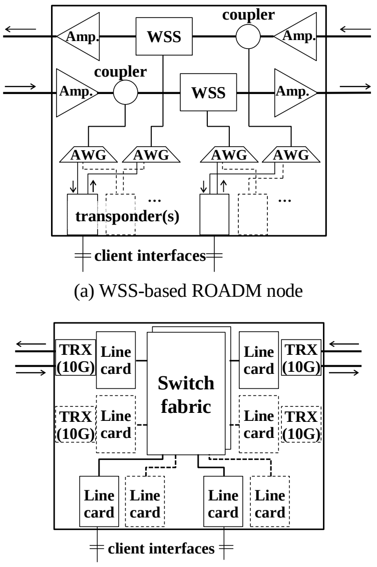
It is important to note that optical burst transmission is already a mature technology in passive optical network (PON) systems in access networks. In addition, the capacity of PON systems is continuously increasing, and emerging PON technologies such as next-generation PON stage 2 (NG-PON2) [14] are making WDM burst transmission feasible for practical use. NG-PON2 systems and related devices are now commercially available. Therefore, in the PSL network, commodity low-power PON devices such as TRXs and LSIs can be used instead of proprietary components. Also note that conventional optical burst amplification technologies (e.g., [15]) can be utilized for longer-reach optical burst transmission while using standard EDFAs. Hence, the PSL network can also be a highly practical solution for flexible metro networks.

IV. Power Consumption and Failure-Recovery Related Cost Evaluation

This section evaluates power consumption and failure-recovery related costs to quantify the OPEX benefits of the PSL network through numerical analysis. In the following, we first describe the assumed network model including detailed node architectures of the PSL network and comparative networks. Second, we compare network power consumption of the PSL network to those of comparative networks to evaluate the energy efficiency of the PSL network. We then estimate failure-recovery related costs in large-scale network infrastructure where a number of metro networks are in operation to verify how effectively the PSL network reduces OPEX.

A. Network Model

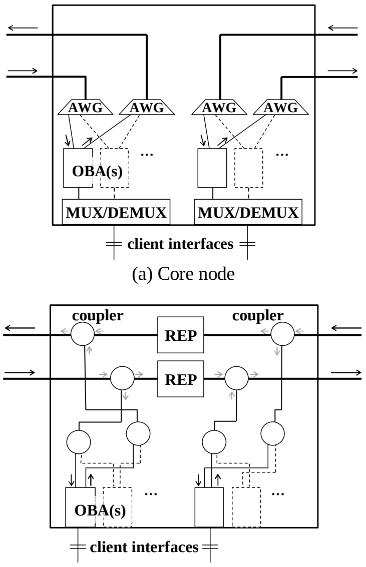
In this paper, we basically assume a 9-node bi-directional ring network with 1 core node and 8 access nodes. Traffic is assumed to flow between each core and access node pair, where the volume of each flow is static and uniformly distributed. Note that a protection switching function is assumed to be implemented in external routers/switches connected to transport equipment to simplify the transport layer, just as in previous work [16]. Specifically, data duplication and select are executed at such routers/switches, and transport networks simply provide two disjoint paths to each traffic demand. Moreover, we select a ROADM-based network and a packet transport network (PTN) using multiprotocol label switching - transport profile (MPLS-TP) as comparative architectures, both of which are already used in metro networks. Simplified ROADM and PTN nodes are illustrated in Fig. 2. As shown in Fig. 2(a), a ROADM is assumed to be a wavelength selective switch (WSS)-based architecture and does not have colorless, directionless, and contentionless (CDC) functionality. In PTN nodes, the number of line cards and required switching capacity strongly depend on traffic volume to be accommodated. Additionally, node architecture in the PSL network is depicted in Fig. 3, which shows how to use optical passive devices. As shown in Fig. 3, AWGs are utilized in core nodes since traffic is aggregated to core nodes from access nodes and more wavelengths than access nodes need to be handled. On the other hand, optical couplers are utilized for multiplexing and distribution in access nodes. Note that optical filters equipped at receivers of OBAs select and extract the desired data signals, just as in PON systems. Each access node in the PSL network is assumed to be equipped with REPs for simplicity. In addition, we used the component cost, power consumption, and mean time between failure (MTBF) values in Table I, on the basis of previous work [17–21]. Note that OBAs in the PSL network are assumed to be implemented with PON devices, and the effect of forward error correction (FEC) (e.g., RS (248, 216)) is considered.



(a) WSS-based ROADM node

(b) PTN node

Fig. 2. Comparative architectures.



(a) Core node

(b) Access node

Fig. 3. Node architecture in PSL network.

TABLE I. COST, POWER CONSUMPTION, AND MTBF VALUES OF EACH COMPONENT

Components				Relative cost	Power consumption	MTBF
Common	Optical coupler			0.6	0	12,000,000 h
	Amplifier			15	12 W	1,000,000 h
	AWG (1 : N)			0.3×N	0	4,000,000 h
ROADM	10G transponder			18.75	50 W	350,000 h
	WSS			37.5	30 W	250,000 h
PTN	Switch fabric[a]			1.45 /10G	10 W /10G	400,000 h
	10G line card			9.84	50 W	350,000 h
	1G×10 line card			1.87	40 W	350,000 h
PSL network	Core node	OBA	10G burst TRX	2.5	2.5 W	500,000 h
			L1/L2 LSI	5	6 W	450,000 h
		MUX/DEMUX[a]		1 /10G	10 W /10G	400,000 h
	Access node	OBA	10G burst TRX	2.5	2.5 W	500,000 h
			L1/L2 LSI	0.6	3 W	450,000 h
		REP		40	100 W	1,000,000 h

[a] Component cost and power consumption depend on switching or MUX/DEMUX capacity.

B. Power Consumption Evaluation

We evaluate network power consumption by multiplying power consumption of each component by the required number of components under the given condition. Note that the ratio of power consumption to total transmission capacity is a widely used metric suitable for core networks but not metro networks since traffic volume in metro networks covering rural areas may be much smaller than the overall transmission capacity of high-capacity systems. The calculated power consumptions in Fig. 4 show that the PSL network can achieve the lowest power consumption of the three architectures. The results show that the PSL network can reduce power consumption by more than 30% compared with ROADM-based networks when traffic volume per access node is smaller than 2 Gbps or larger than 10 Gbps, even when power-hungry REPs are used in all access nodes. This is due to leveraging optical passive devices, sharing TRXs, and avoiding the use of proprietary transponders. Also note that PTNs can be more energy-efficient than ROADM-based networks when traffic volume is quite small, though power consumption of PTNs sharply increases as the traffic volume increases. However, the PSL network can share resources as flexibly as a PTN while reducing the amount of electronic processing and consumes 80% less power than a PTN when traffic volume per access node is larger than 4 Gbps.

To clarify the power consumption structure and discuss the characteristics of the three architectures, Fig. 5 shows a breakdown of network power consumption when traffic volume per access node is set to 4, 8, and 12 Gbps. In a ROADM network, the main contributor to total network power consumption is the transponder, since the amplifier and WSS are optical devices that consume less power. In addition, in a PTN, the main contributor is the 10G line card, and power consumption of all components increases as traffic increases because of an opaque solution. On the other hand, in the PSL network, the main contributor is REP when traffic volume is small. Although the numbers of optical burst TRXs and L1/L2 LSIs increase as traffic increases, such components (PON devices) consume much less power than proprietary components, and resource sharing can suppress the required number of such components. As a result, total power consumption does not sharply increase when traffic increases. Note that reducing power consumption of REPs can naturally achieve more energy-efficient networks.

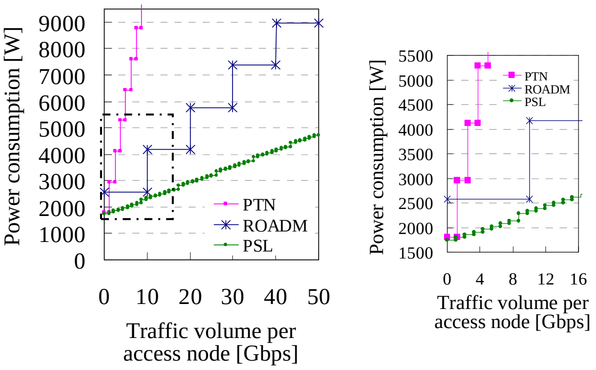


Fig. 4. Network power consumption of three architectures. (Right graph is a closeup of the dash-dotted square in the main graph.)

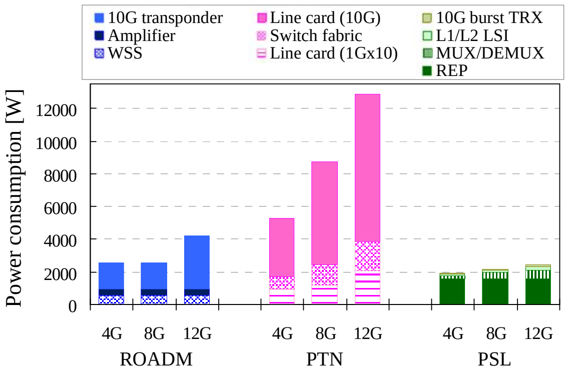


Fig. 5. Power consumption breakdown of three architectures for various traffic scenarios.

C. *Failure-Recovery Related Cost Evaluation*

We quantify the reduction of failure recovery operations and repair costs achieved with the PSL network and assess the impact of different network architectures on reliability. In this paper, individual component failures are assumed to occur randomly in accordance with MTBF values, and failed components need to be replaced. For simplicity, the repair costs are calculated by multiplying each component cost by the number of failures per component. A number of metro networks are assumed to be operated, where each network is a 9-node ring network as previously described and traffic volume per access node is set to 4 Gbps. Annual required numbers of failure recovery operations with various network numbers are shown in Fig. 6. The results verify that the PSL network using a quasi-passive optical ring can reduce operations by 40% compared with ROADM-based solutions. Thus, the PSL network can be very effective, especially when the number of operating networks is large, since the absolute number of the required recovery operations is naturally large. For instance, 350 operations can be eliminated per year when 500 networks are operated. This can directly save OPEX, though maintenance strategies and cost structures may vary among network operators. Moreover, such a reduction would be very beneficial for rural areas that generally occupy a large share of land [22, 23]. This is because smaller traffic volume tends to make cost per transported bit higher and long-distance transportation from network operation centers is required in many cases. Note that the PSL network requires 60% fewer failure recovery operations than a PTN, which has a larger number of active components. In addition, the calculated repair costs in Fig. 7 reveal that the PSL network can reduce repair costs by more than 80% compared with a ROADM network. This is due to not only reducing failure-frequency as shown in Fig. 6 but also leveraging mass-produced low-cost components instead of proprietary and/or high-end components. As a result, both failure recovery operations and repair costs can be reduced, which will lead to significant OPEX savings in transport networks of telecom carriers.

To demonstrate the impact of active/passive components on failure frequency (i.e., failure recovery operation), the contribution of each component is shown in Fig. 8 for traffic volumes per access node of 4, 8, and 12 Gbps when operating 500 networks. In a ROADM network, the contributions of the WSS and transponder are comparable when traffic volume is small, and the transponder becomes the main contributor as traffic increases. On the other hand, in a PTN, the main contributor is the 10G line card, the number of which to deploy strongly depends on traffic volume, just as in Fig. 5. Moreover, in the PSL network, the major contributors are naturally the 10G burst TRX and L1/L2 LSI, the sum of which is comparable to that of the transponder in a ROADM network. Thus, differences in the number of failure recovery operations between a ROADM network and the PSL network result from optical devices used (active WSS or passive coupler). In addition, the contribution of each component to repair costs in 500 networks is shown in Fig. 9. In a ROADM network, the main contributor is not the transponder but the WSS when traffic volume is small, which results from the difference in component cost. A PTN has smaller total repair costs than a ROADM network even though it has higher failure frequency since its well-matured components are low cost. Furthermore, in the PSL network, the main contributor to cost is REPs since other active components (i.e., TRXs and LSIs) are mass-produced, inexpensive PON devices. Therefore, we can conclude that leveraging a quasi-passive optical ring and PON devices effectively suppress both the failure frequency and repair costs.

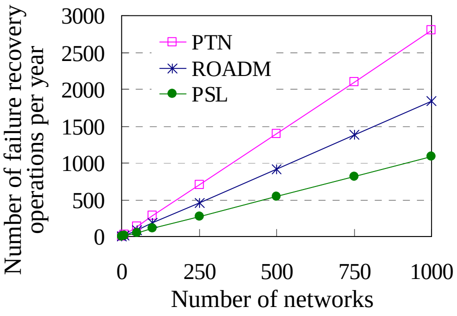


Fig. 6. Number of failure recovery operations in three architectures.

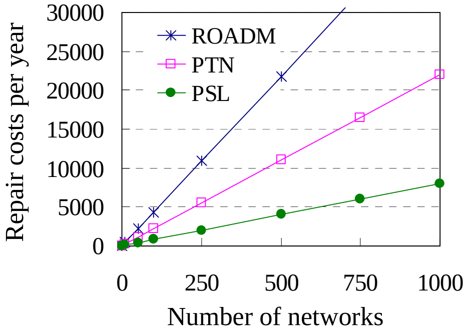


Fig. 7. Repair costs of three architectures.

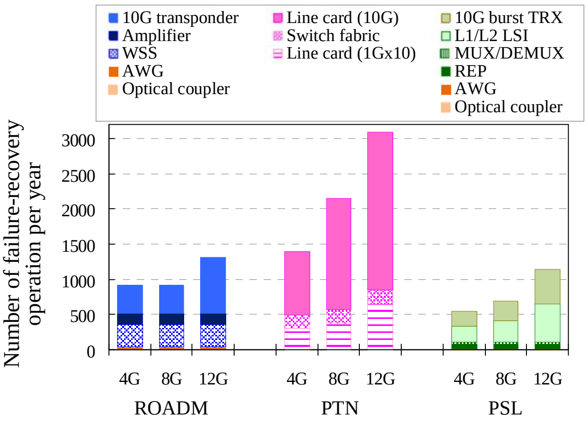


Fig. 8. Comparison of component failure frequency per year for various traffic scenarios.

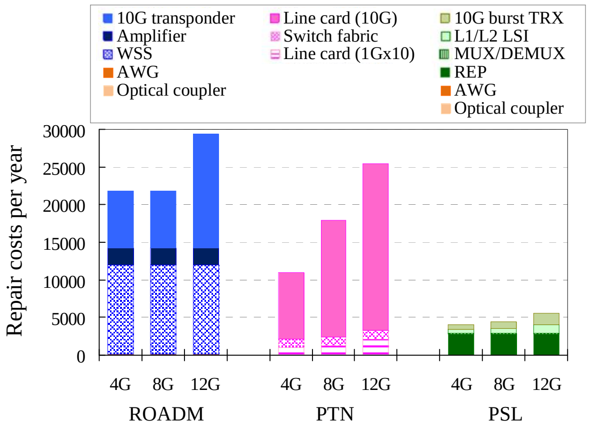


Fig. 9. Comparison of component repair costs per year for various traffic scenarios.

Remark: Most of conventional flexible metro networks leverage not only proprietary transponders (e.g., optical OFDM transponders in elastic optical networks (EONs) [12], optical OFDM burst transponders in a TISA network [11]) but also many active components (e.g., high-speed optical switches in a POADM based network [9] or a TSON [10], bandwidth-variable WSSs in EONs) at every node, leading to more failure-recovery operations. Such proprietary components generally have higher costs than mass-produced PON devices, leading to higher repair cost. Therefore, we can conclude that the PSL network offers lower OPEX than conventional flexible metro networks.

V. Conclusion

Telecom carriers need to optimize operational expenditures (OPEX), especially in metro networks, to cost-effectively provide various network services nationwide. In this paper, we presented a promising optical network architecture, the Photonic Sub-Lambda transport network (PSL network), as an energy-efficient and reliable solution for metro networks. The PSL network utilizes a quasi-passive optical ring, enabling lower power consumption and higher reliability than conventional optically or electronically switched architectures such as a reconfigurable optical add/drop multiplexer (ROADM)-based network and a packet transport network (PTN). We also provided numerical evaluations, which revealed that the PSL network can effectively reduce power consumption and failure recovery-related OPEX. Such results indicate that our PSL network can be effective even in 1k-building scale scenarios where accommodated traffic volume per node is smaller than wavelength capacity.

References

[1] L. Peterson, A. Al-Shabibi, T. Anshutz, S. Baker, A. Bavier, S. Das, J. Hart, G. Palukar, and W. Snow, “Central office re-architected as a data center,” IEEE Commun. Mag., vol. 54, no. 10, pp. 96–101, Oct. 2016.

[2] Y. Uematsu, S. Kamamura, H. Date, H. Yamamoto, A. Fukuda, R. Hayashi, and K. Koda, “Future nation-wide optical network architecture for higher availability and operability using transport SDN technologies,” IEICE Trans. on Commun., vol. E101-B, no. 2, pp. 462–475, Feb. 2018.

[3] C.G. Gruber, “Capex and opex in aggregation and core networks,” Proc. OFC 2009, OThQ1, Mar. 2009.

[4] M. Walker, “A growth opportunity for vendors: telco opex,” OVUM, Oct. 2012.

[5] S. Verbrugge, D. Colle, M. Jager, R. Huelsermann, F.-J. Westphal, M. Pickavet, and P. Demeester, “Impact of resilience strategies on capital and operational expenditures,” Proc. ITG Tagung Photonical Networks, pp. 109–116, May 2005.

[6] E. Hernandez-Valencia, S. Izzo, and B. Polonsky, “How will NFV/SDN transform service provider opex?” IEEE Netw., vol. 29, no. 3, pp. 60–67, May/June 2015.

[7] B. Naudts, M. Kind, S. Verbrugge, D. Colle, and M. Pickavet, “How can a mobile service provider reduce costs with software-defined networking?” Int. J. Netw. Manag., vol. 26, no. 1, pp. 56–72, Jan./Feb. 2016.

[8] M. Nakagawa, K. Masumoto, K. Hattori, T. Matsuda, M. Katayama, and K. Koda, “Flexible and cost-effective optical metro network with photonic-sub-lambda aggregation capability,” Proc. OECC/PS 2016, ThA2-2, July 2016.

[9] D. Chiaroni, G. Buforn, C. Simonneau, S. Etienne, and J.-C. Antona, “Optical packet add/drop systems,” Proc. OFC 2010, OThN3, Mar. 2010.

[10] G.S. Zervas, J. Triay, N. Amaya, Y. Qin, C.C. Pastor, and D. Simeonidou, “Time shared optical network (TSON): a novel metro architecture for flexible multi-granular services,” Opt. Express, vol. 19, no. 26, pp. B509–B514, Dec. 2011.

[11] P. Gavignet, E. Le Rouzic, E. Pincemin, B. Han, M. Song, and L. Sadeghioon, “Time and spectral optical aggregation for seamless flexible networks,” Proc. PS 2015, pp. 43–45, Sept. 2015.

[12] P. Layec, A. Dupas, D. Verchère, K. Sparks, and S. Bigo, “Will metro networks be the playground for (true) elastic optical networks?,” J. Lightwave Technol., vol. 35, no. 6, pp. 1260–1266, Mar. 2017.

[13] M.D. Leenheer, T. Tofigh, and G. Parulkar, “Open and programmable metro networks,” Proc. OFC 2016, Th1A.7, Mar. 2016.

[14] D. Nesset, “NG-PON2 technology and standards,” J. Lightwave Technol., vol. 33, no. 5, pp. 1136–1143, Mar. 2015.

[15] H.H. Lee, J.H. Lee, and S.S. Lee, “All-optical gain-clamped EDFA using external saturation signal for burst-mode upstream in TWDM-PONs,” Opt. Express, vol. 22, no. 15, pp. 18186–18190, July 2014.

[16] T. Hofmeister, V. Vusirikala, and B. Koley, “How can flexibility on the line side best be exploited on the client side?” Proc. OFC 2016, W4G.4, Mar. 2016.

[17] F. Rambach, B. Konrad, L. Dembeck, U. Gebhard, M. Gunkel, M. Quagliotti, L. Serra, and V. Lopez, “A multilayer cost model for metro/core networks,” J. Opt. Commun. Netw., vol. 5, no. 3, pp. 210–225, Mar. 2013.

[18] FP7 OASE project deliverable D4.2.2, “Technical assessment and comparison of next-generation optical access system concepts,” June 2013.

[19] W.V. Heddeghem, F. Idzikowski, W. Vereecken, D. Colle, M. Pickavet, and P. Demeester, “Power consumption modeling in optical multilayer networks,” J. Photon. Netw. Commun., vol. 24, no. 2, pp. 86–102, Oct. 2012.

[20] S. Verbrugge, D. Colle, M. Pickavet, P. Demeester, S. Pasqualini, A. Iselt, A. Kirstädter, R. Hülsermann, F.-J. Westphal, and M. Jäger, “Methodology and input availability parameters for calculating OpEx and CapEx costs for realistic network scenarios,” Journal of Optical Networking, vol. 5, no. 6, pp. 509–520, June 2006.

[21] FP7 OASE project deliverable D4.3.2, “Operational impact on system concepts,” Apr. 2012.

[22] Ericsson whitepaper: Full Service Broadband Metro Architecture, Nov. 2007.

[23] Ericsson whitepaper: Microwave Towards 2020, Sept. 2014.

SINR-Oriented Flexible Quantization Bits for Optical-Wireless Deep Converged eCPRI

Longsheng Li*, Meihua Bi*†, Wei Wang*, Yan Fu*, Xin Miao* and Weisheng Hu*
*State Key Laboratory of Advanced Optical Communication Systems and Networks, Shanghai Jiao Tong University, Shanghai, China, email: wshu@sjtu.edu.cn
† School of Communication Engineering, Hangzhou Dianzi University, Hangzhou, China, email: bmhua@hdu.edu.cn

***Abstract*—The split-PHY and Ethernet-based eCPRI is mostly advanced for reducing fronthaul line rate and leveraging Ethernet protocol. In this paper, a fundamental method is proposed to compress the eCPRI data by dynamically adjusting the quantization bits of IQ sample according to the corresponding wireless signal quality of user equipment. With the aid of this mechanism, the fronthaul traffic is correlated to the eventual end-user traffic instead of the air bandwidth, and hence the flexibility and the bandwidth efficiency of the interface can be further improved. To verify this proposal, a simulation model achieving the complete low-MAC layer and PHY layer processing for LTE uplink and strictly following the 3GPP specifications is built. The results indicate that for the typical mobile environment, the proposed scheme can statistically save ~20% interface bandwidth.**

***Keywords*—eCPRI, split-PHY, SINR, quantization bit, flexibility**

I. Introduction

Fueled by the low-cost broadband services and the high-level coordination technologies, the centralization of baseband units (BBU) becomes the mainstream [1,2] for the 5G access network. To support this architecture, the exploration of novel interface for mobile fronthaul (MFH) is critical to efficiently deliver the radio signal data among BBUs and remote radio heads (RRHs). As one of the major standardized interfaces, the common public radio interface (CPRI) [3] transmits the continuous time-domain in-phase and quadrature signal (IQ) samples with fixed 30 quantization bits, therefore it would consume large bandwidth and limits the flexibility and feasibility of MFH. To solve this problem, the split-PHY architecture combined with Ethernet networking has been widely investigated [4],[5]. And the standardized eCPRI [6] has also been proposed to reduce the MFH traffic and leverage the mature packet transport standard. In eCPRI, the function split point for uplink is implemented after the resource element demapping. Base on this split, the interface bandwidth is significantly reduced and the MFH traffic volume can scale flexibly according to the number of physical resource blocks (RBs), or the used air bandwidth. This feature highly promotes the bandwidth efficiency and benefits the statistical multiplexing, whereas, the eventual MFH efficiency, or the fronthaul to backhaul bandwidth ratio [7], is still significantly limited by the quality of wireless channel. Since in a practical mobile communication system, a certain number of RBs is assigned to each user equipment (UE) for its transmission. And according to the number of RBs and the channel quality of the UE, a specific modulation and coding scheme (MCS) is chosen to modulate the data of end-users, namely the transport block (TB), onto the RBs. In this situation, for the uplink transmission, owing to the transmission of wireless Rayleigh fading channel, the signal qualities of UEs' RBs vary dramatically, and the end-user data rate supported by different UEs' RBs also strongly fluctuate [8]. Therefore, if all RBs is uniformly packaged in MFH interface regardless of its real data capacity, the eventual MFH efficiency for a specific UE is highly determined by its wireless signal quality, or correspondingly its MCS. As shown in Fig. 1, with 6 physical RBs and 10-bit quantization for IQ sample, the ratio of MFH line rate contributed by IQ sample to the end-user data rate decreases with the MCS index [9] (the high-order MCS is applied to high-quality signals). This result indicates that the eventual MFH efficiency for the RBs filled with low-quality signal is inefficient. Fortunately, in eCPRI specification, the quantization bit width and format for RB samples, or the frequency-domain IQ samples, are vendor specific [6], which offers a desirable interface to intelligently package the data of IQ sample. Moreover, as presented in [14], the frequency-domain IQ signals with diverse signal quality are not equally sensitive to the quantization resolution, among which the signal with a low signal to interference and noise ratio (SINR) has a large margin for quantization resolution and hence demands fewer quantization bits. Consequently, based on the SINR monitoring, the redundant bits for the quantization of low-quality signal should be located and saved in eCPRI.

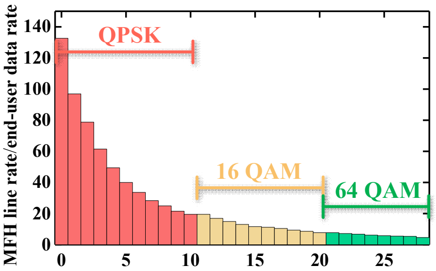


Fig. 1. Bandwidth efficiency of MCS. MCS: modulation and coding scheme.

In this paper, for the first time, we propose and demonstrate a flexible quantization bits (FQB) scheme for uplink eCPRI, which leverages the monitoring of signal quality in wireless system to lower the MFH traffic. The potential redundant quantization bits are predicted in our scheme, where the RRH is informed of the signal quality by the BBU through real-time control information on user plane, based on which the optimal

978-3-903176-07-2 © 2018 IFIP

quantization bits are configured for every uplink LTE subframe. With this mechanism, the redundant bits for low-quality RBs are reduced, and hence the eventual MFH traffic. The results of the simulation system with the configuration specified in [9,13] convince that about 17~23% traffic can be saved by the FQB.

The rest of the paper is organized as follows. Section II introduces the relevant background of the mobile communication system. Section III describes the architecture of the proposed MFH link and explains its feasibility. Section IV shows the simulation results comparing the FQB scheme with the ordinary uniform quantization bits (UQB) scheme, and Section V concludes this paper.

II. Background and Motivation

A. Influence of split-PHY architecture on MFH interface

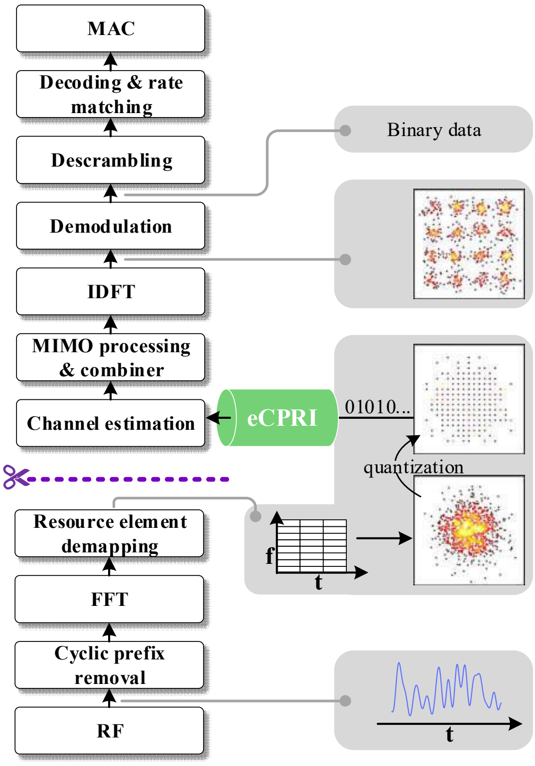


Fig. 2. LTE function layers and signal format at each interface.

Figure 2 is the schematic of LTE function layers with signal formats attached. As illustrated in this figure, the user plane data of eCPRI is the frequency-domain IQ sample from the resource element demapping, which should be quantized before packaged into the eCPRI frame. Obviously, more quantization bits bring less quantization noise at the cost of losing bandwidth efficiency. Different from the time-domain quantization [15] in CPRI, the quantization noise in eCPRI does not spread over the entire physical uplink frequency band, instead, it only takes effect among the corresponding frequency-domain IQs which are then together used for the demodulation of a specific UE. In other words, the frequency division multiplexed UEs can be served by independent quantization resolution.

B. Fidelity of frequency-domain quantization

Due to the transmission of frequency selective Rayleigh fading channel, the qualities of received frequency-domain signals are diverse. In the model given in 3GPP specification [9] section A.2.2, for typical uplink scenarios, the variance among the SINR of UEs is more than 10 dB. These signals with different SINR also have different sensitivities to the quantization resolution. Figure 3 shows the relationship between the SINR of quantized signal and that of the unquantized signal. To give this result, the simulation model following 3GPP standard as described in section III is used. It can be seen that the deterioration caused by quantization is significant for the high-quality signal, in comparison, the low-quality signal is much less sensitive to quantization noise, hence its fidelity can be satisfied by relatively fewer quantization bits. Based on this feature, the major idea of our proposal is to detect the SINR of UE's RBs, and according to which the optimal quantization bits are configured to prevent redundancy bits.

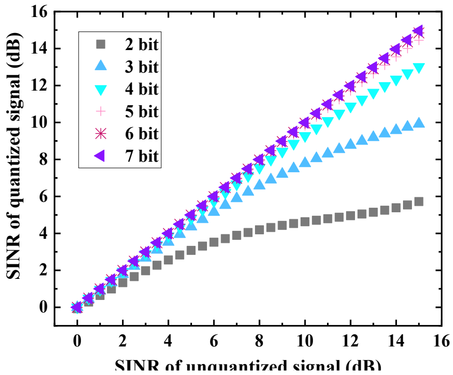


Fig. 3. SINR of quantized and unquantized signal.

III. Architecture and Simulation Model of eCPRI with THE FQD

A. System architecture based on eCPRI

Figure 4 illustrates how the SINR monitoring and quantization bit control fit into to the eCPRI-based MFH interface to achieve the SINR-oriented FQB, note that the referential eCPRI protocol stack can be found in [6] section 3 and the interfaces for the C&M and the synchronization plane are irrelevant to THE FQD, thus not presented here. The eCPRI user data plane contains three types of information: user data (the IQ sample), real-time control data and other eCPRI services. The FQB for uplink works as follows. At the RRH, the IQ samples are firstly quantized by the initial quantization bits. Then, together with other information, the user data are transmitted through MFH link with eCPRI. At the BBU, the received IQ samples are used for low-PHY demodulation, where the SINR is also measured. This SINR is passed to a quantization bit manager which adjusts the number of quantization bits by looking up a preset table. The quantization bit control command is added to the real-time control information and passed back to the RRH. And afterward, the RRH read this command and adjust the quantization bits for the forthcoming samples. Note that the overhead for quantization bit control is negligible since one SINR measuring presents the signal quality of one subframe consisting of a large number of IQ samples from one UE. Besides, this FQB control has a delay approximate to the one-way delay of eCPRI. Fortunately, this shall be low enough to follow the change of wireless channel. For comparison, the delay of adaptive modulation and coding (AMC) in LTE is no less than 7 subframe cycles [13], namely ~7 ms.

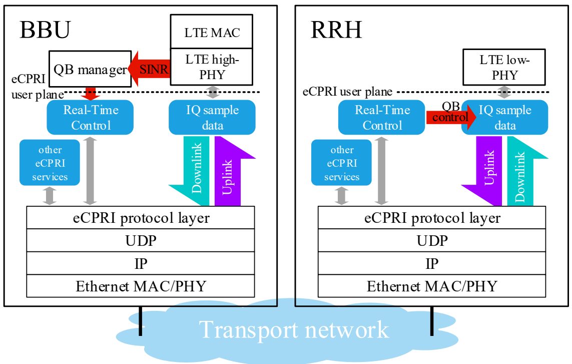


Fig. 4. System architecture of eCPRI with the SINR-oriented FQB. QB: quantization bit.

B. Simulation model

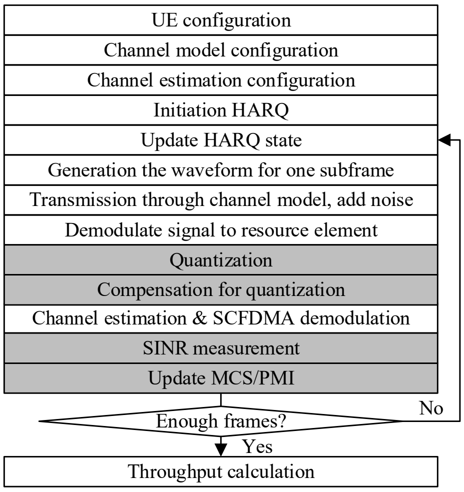

UE configuration
Channel model configuration
Channel estimation configuration
Initiation HARQ
Update HARQ state
Generation the waveform for one subframe
Transmission through channel model, add noise
Demodulate signal to resource element
Quantization
Compensation for quantization
Channel estimation & SCFDMA demodulation
SINR measurement
Update MCS/PMI
Enough frames?
No
Yes
Throughput calculation

Fig. 5. Working flow for uplink simulation with adaptive modulation and coding.

To verify the feasibility of the proposal, we build a simulation system achieving the entire PHY-layer processing and Low-MAC layer processing ending at HARQ. The simulation platform is Matlab, and all functions, unless otherwise specified, are based on Matlab standardized LTE System Toolbox [16]. The major objects of the simulation are to a) determine the optimal FQB strategy, b) compare the bandwidth efficiency between the FQB scheme and the UQB scheme, and c) find the FQB's influence on final end-user throughput. Note that although the simulation is based on LTE air interface, this proposal also works for the forthcoming 5G situation, since the SINR diversity will always exist.

Figure 5 presents the processing flow of the simulation, in which the grey function modules are additionally added or customized compared to the standard system. For the quantization processing, all IQ samples within one subframe are quantized by the same number of bits. The resolution of quantization bit in this system is 0.1, and non-integer bit is achieved by quantizing a portion of the samples with the bit number rounded toward negative infinity, and the rest with the bit number rounded toward positive infinity. The compensation for quantization is done by filling the lost lowest bits with random binary bits. In realistic LTE link, the AMC is employed, which lets UE with high-quality adopt high-order MCS to realize better spectral efficiency. Nevertheless, in 3GPP specification, the strategy of selecting appropriate MCS according to channel quality is not given and should be vendor specific. Based on this situation, we employ a maximum throughput strategy for AMC. For a given SINR, the concept of this strategy is simply to choose the MCS that can achieve the maximum throughput with cyclic redundancy check (CRC) success ratio considered. Hence, in section 0, the throughput versus SINR for all 29 MCSs [12] is measured to produce an SINR-MCS table for AMC. In AMC processing, the updated quantization bits together with the updated MCS take effect after 7-subframe delay as defined in [13]. The major parameters of the simulation are listed in TABLE I.

TABLE I
PRIMARY PARAMETERS OF THE SIMULATION

Parameter	Quantity
SRS[a] periodicity	2 ms
Delay of MCS update	7 ms
Transmitter antenna number	2[b]
Receiver antenna number	1[b]
Channel Delay profile model	EPA-5[c]
Rayleigh fading model type	GMEDS[d]
Transmission bandwidth	90 KHz (6 RBs)
Frame number	500
Average SNR	5 dB

[a] Sounding reference signal.
[b] [13].
[c] Extended Pedestrian A model with 5-Hz Doppler frequency.
[d] Generalized Method of Exact Doppler Spread.

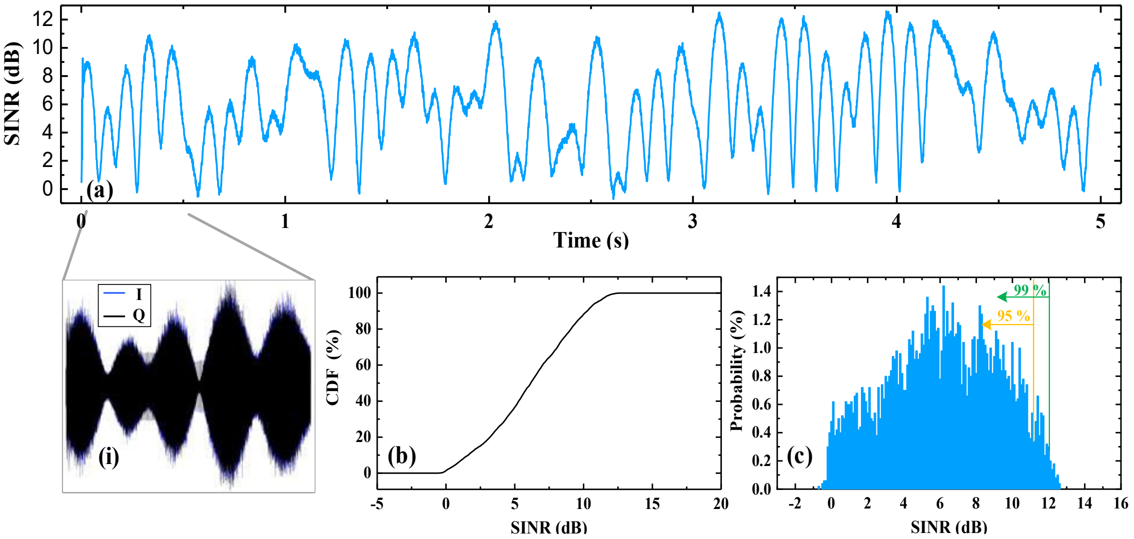


Fig. 6. (a) SINR variation during 5 ms, (b) the cumulative distribution and (c) probability distribution of SINR.

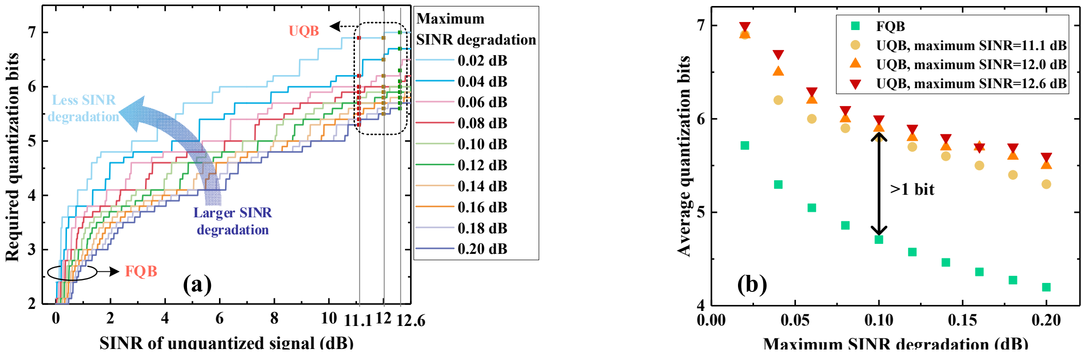


Fig. 7. (a) minimum quantization bits satisfying SINR degradation threshold versus SINR of the unquantized signal in the FQB and UQB scheme, (b) statistical average quantization bits versus SINR degradation threshold. The threshold here is set for maximum SINR deterioration.

IV. SIMULATION RESULT

A. *Reduction of quantization bits*

To make the result convincing and significant enough, the SINR distribution of the wireless signal in our model is firstly simulated which is utilized for further statistical measurement. Figure 6(a) presents the SINR variation over 5 seconds, namely 500 frames, and the inset (i) gives the corresponding time-domain IQ waveform during the first 0.5 second. Based on this result, the probability distribution and the cumulative distribution of SINR are further shown in Fig. 6(b) and (c). The CDF of SINR obtained here is approximate to the reference model given in 3GPP specification [13] Section A.2.2.

Then, the relationship between the SINR of unquantized signal and that of quantized signal is tested, which is achieved by bypassing the transmission channel module and only attaching the additive noise. Therefore, the noise power can be easily controlled and any possible unquantized SINR can be covered. The fitting results for the integer quantization bits situations are presented in Fig. 3. In this paper, the signal fidelity of MFH interface is depicted by the maximum tolerable SINR degradation induced by quantization process. With the SINR relationship between unquantized and quantized signal, the minimum quantization bits satisfying the threshold of SINR deterioration can be determined at each unquantized signal SINR, and the SINR to bit number mapping table is built for the FQB. As shown in Fig. 7(a), the curves of the FQB scheme are non-decreasing functions, where the low-SINR signal demands relatively fewer bits. In general, stricter tolerance of SINR degradation is supported by more quantization bits. As for the UQB scheme, the tolerable SINR degradation should be satisfied at the maximum unquantized SINR. Based on the result of SINR distribution, three typical values, 12.6, 12.0 and 11.1 dB are treated as the maximum unquantized SINR in turn, which can cover 100%, 99% and 95% of the SINR distribution respectively. The corresponding required bits for the three values in the UQB are circled in Fig. 7(a).

With the SINR distribution obtained and the quantization strategies determined, the statistical average quantization bits for MFH interface can be measured. The results are given in Fig. 7(b), which exemplify that the FQB scheme can noticeably reduce the quantization bit by ~1 bit for any SINR degradation requirement compared to the UQB scheme. Typically, with Maximum SINR degradation being 0.1 dB, the FQB scheme can save 20.2% quantization bits compared to the UQB scheme with the maximum SINR set as 12.0 dB.

B. End-user throughput

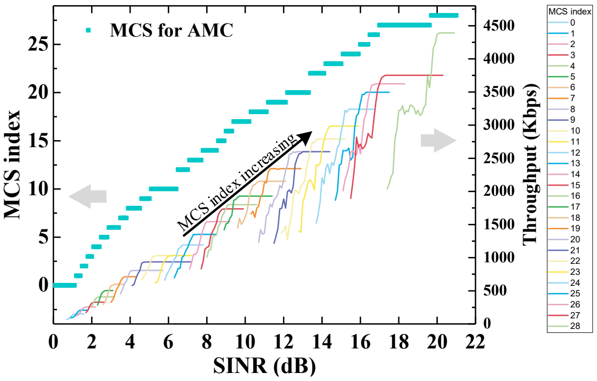


Fig. 8. Throughput versus SINR with fixed MCS index and chose MCS index versus SINR in AMC.

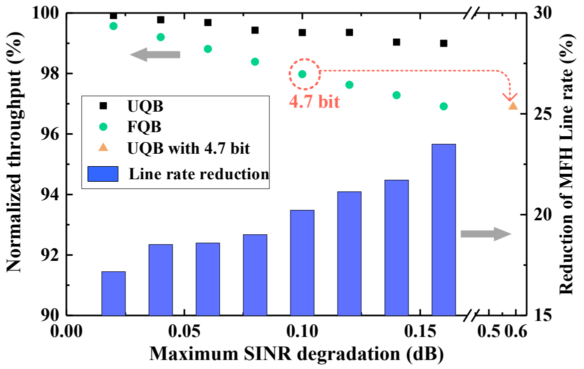


Fig. 9. Working flow for uplink simulation with adaptive modulation and coding.

The proposed FQB scheme and the UQB scheme are applied to the uplink system including the channel transmission and the AMC to benchmark their influence on the final end-user throughput. Firstly, to determine the working range of each MCS for AMC, the throughput versus SINR for each MCS is individually tested. The measured throughput and the chosen MCS index at corresponding SINR are exhibited in Fig. 8. Based on this uplink system, the throughput for the FQB scheme and the UQB scheme with maximum unquantized SINR set as 12.0 dB are compared, and the result is depicted in Fig. 9. Here, the throughput is normalized by dividing the throughput of the unquantized system. It is revealed that the throughput for the FQB is decreased by only 1% while the MFH line rate is saved by 17~23% compared to the UQB counterpart. Therefore, it is convinced that the FQB is able to compress the eCPRI interface traffic without significantly sacrificing the end-user throughput. For reference, the UQB with 4.7 quantization bits, which are the statistical average bits of FQB scheme at 0.1 dB SINR degradation, is tested. Compared to its FQB counterpart, the UQB scheme with 4.7 bit has ~1% lower throughput and ~0.58 dB worse maximum SINR degradation. It can be seen from this result that compared to directly decreasing quantization bits, the FQB can cause less injury to the wireless signal.

V. CONCLUSION

In this paper, we propose a novel MFH quantization scheme which exploits the existing function in the wireless system to assist the transmission of split-PHY-based MFH. The number of quantization bits for the MFH interface is flexible according to the signal quality of the frequency-domain samples, thus the redundant bits are significantly decreased. Following the 3GPP specification, the simulation results reveal that ~20% uplink MFH can be compressed, and compared to traditional uniform quantization scheme, the end-user throughput is barely sacrificed in the FQB scheme. Therefore, the proposed scheme could be an effective solution to improve the MFH efficiency.

ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China under Grant 61431009, Grant 61521062, and Grant 61575123.

REFERENCES

[1] Mobile, China, “C-RAN: the road towards green RAN.” White Paper, version 2.5 (2013), http://labs.chinamobile.com/cran.

[2] NGMN, “NGMN 5G White Paper”, White Paper, version 1.0 (2015), https://www.ngmn.org/fileadmin/ngmn/content/downloads/Technical/2015/NGMN_5G_White_Paper_V1_0.pdf.

[3] CPRI Specification V7.0, “Common Public Radio Interface (CPRI); Interface Specification,” (2015).

[4] S. Zhou, X. Liu, F. Effenberger, and J. Chao, “Low-Latency High-Efficiency Mobile Fronthaul With TDM-PON (Mobile-PON),” Journal of Optical Communications and Networking 10(1), A20-A26 (2018).

[5] K. Miyamoto, S. Kuwano, T. Shimizu, J. Terada, and A. Otaka, “Performance Evaluation of EthernetBased Mobile Fronthaul and Wireless CoMP in Split-PHY Processing,” Journal of Optical Communications and Networking 9(1), A46-A54 (2017).

[6] eCPRI Specification V1.0, “Common Public Radio Interface: eCPRI Interface Specification,” August 2017.

[7] China Mobile Research Institute, “White Paper of Next Generation Fronthaul Interface.” White paper v1.0, 2015.

[8] S. Catreux, V. Erceg, D. Gesbert, and R. W. Heath, “Adaptive Modulation and MIMO Coding for Broadband Wireless Data Networks,” IEEE Communications Magazine 40(6),109-115,2002.

[9] 3GPP, TS 36.101, v15.0.0, “User Equipment (UE) radio transmission and reception (Release 15),” (2017).

[10] 3GPP, TS 36.104, v15.0.0, “Base Station (BS) radio transmission and reception (Release 15),” (2017).

[11] 3GPP, TS 36.211, v13.0.0, “Physical channels and modulation (Release 13),” (2016).

[12] 3GPP, TS 36.213, v12.7.0, “Physical layer procedures (Release 12),” (2015).

[13] 3GPP, TS 36.814, v9.2.0, “Further advancements for E-UTRA physical layer aspects (Release 12),” (2017).

[14] U. Dötsch, M. Doll, H.P. Mayer, F. Schaich, J. Segel, and P. Sehier, “Quantitative analysis of split base station processing and determination of advantageous architectures for LTE,” Bell Labs Technical Journal 18(1), 105–128, 2013.

[15] D. Samardzija, J. Pastalan, M. MacDonald, S. Walker, and Reinaldo Valenzuela, “Compressed transport of baseband signals in radio access networks,” IEEE Transactions on Wireless Communications, 11(9), 3216-3225, 2012.

[16] H. Zarrinkoub, “Understanding LTE with MATLAB: from mathematical modeling to simulation and prototyping,” John Wiley & Sons, 2014.

Midhaul Transmission Using Edge Data Centers with Split PHY Processing and Wavelength Reassignment for 5G Wireless Networks

Jiakai Yu[1], Yao Li[2], Mariya Bhopalwala[1], Sandip Das[3], Marco Ruffini[3], Daniel C. Kilper[2]
[1]Department of Electrical and Computer Engineering, University of Arizona
[2]Optical Science College, University of Arizona
[3]CONNECT Research Centre, University of Dublin, Trinity College, Dublin, Ireland
{jiakaiyu, mariyab}@email.arizona.edu, {yaoli, dkilper}@optics.arizona.edu, dassa@tcd.ie, marco.ruffini@scss.tcd.ie

***Abstract*—Distributed processing of edge data centers in a metropolitan area is considered to reduce the large data traffic load due to Cloud Radio Access Network (C-RAN) fronthaul digitized radio-over-fiber protocols. A dynamic PHY split strategy is examined for high-capacity optical Dense Wavelength Division Multiplexing (DWDM) based C-RANs with limited edge data center resources. A network performance simulation model is developed based on a regional optical network in the New York metropolitan area to evaluate the dynamic midhaul approach. The use of a midhaul network improves the network performance by reducing traffic congestion and enhancing wavelength channel utilization. Simulation results show a 45% reduction in the required optical capacity in our proposed adaptive midhaul network compared to a traditional CPRI fronthaul network.**

Keywords—radio access networks, optical fiber networks; Functional PHY Split; Data Center; routing and wavelength assignment

I. Introduction

Wireless networks continue to face rapidly growing traffic demands while supporting an increasingly wide range of services and applications. Cellular radio access networks with baseband processing at every access point may not scale well for the high capacity and large numbers of small cells expected in 5G networks. Cloud radio access networks (C-RAN) have been proposed as a scalable solution by separating the radio components from the baseband unit (BBU), in order to gain the efficiencies of cloud computing for radio networks [1, 2]. Shared processing resources and commodity hardware used in the C-RAN architecture provide various benefits, such as low energy consumption, statistical multiplexing gain, and coordinated multi-point (CoMP) transmission/reception [3].

Optical networks can provide high capacity to satisfy the growing traffic needs in 5G networks. An effective method to facilitate 5G C-RAN architectures is the use of optical DWDM [4, 5]. However, the resulting fronthaul (FH) network in a C-RAN between the Remote Radio Heads (RRHs) and BBU pools requires high capacity in order to handle digitized radio signals. The raw I/Q waveform samples are bi-directionally transmitted over optical fiber by using bandwidth inefficient transmission protocols such as the Common Public Radio Interface (CPRI), which requires 2.5 Gbps optical bandwidth for a 150Mbps wireless transport rate with 2x2 MIMO and 20MHz carrier spectra in a small cell for downlink transmission [1]. In order to reduce FH optical transmission capacity requirements, functional split points in the baseband processing chain have been investigated with partial functionalities of BBUs placed into RRHs [6, 7]. This dual-site processing runs into trouble because it violates several main goals of C-RAN. Increasing use of distributed processing can increase cost and reduce the effectiveness of techniques such as CoMP [8]. Furthermore, optimal functional split points in FH might vary depending on different base station configurations, access network topology, network traffic load, and signal transmission routing, when Total Cost of Ownership (TCO) is considered.

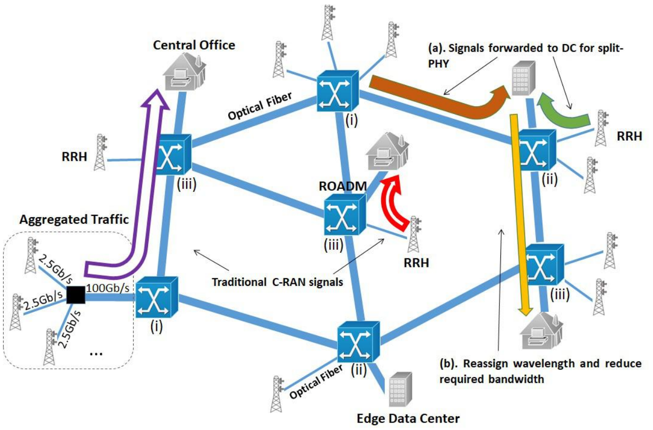


Fig. 1: Adaptive midhaul network architecture and strategy. Hollow arrow shows fronthaul-only C-RAN while solid arrow shows data center midhaul method. Different colors mean different wavelength channels. (i) pure access node with only RRHs; (ii) Data Center node with local data centers; (iii) BBU node with local BBU pool or Central Office.

Flexible centralization in C-RANs should not be limited to functional split processing between BBUs and RRHs. Adaptive PHY splits and processing job placement at multiple sites can also be considered, because a pre-designed architecture might be far from optimal, considering variable 5G application requirements over time and location [9]. Edge Data Centers (DCs) have been used in metropolitan areas for Ethernet switching services and are gaining attentions for telecom networks [10 - 12]. For example, Central Office Re-architected as a Datacenter (CORD) is a platform to bring data center economies to telecom networks using SDN, NFV, and other technologies [13]. These DCs can be utilized or more widely deployed to support midhaul networks, which provide partial PHY processing for signals enroute to their destination

978-3-903176-07-2c © 2018 IFIP

baseband processing location. Fig. 1 shows the use of midhaul links, taking advantage of edge data centers. Once certain data centers in DWDM-based C-RAN are implemented with PHY split processing and optical signal processing technologies, wireless signals are forwarded to these data centers, and split-PHY signals are transported over fiber between data centers and BBUs or Central Office. The midhaul connection shown in brown, green, and yellow arrows will not only be more efficient in transporting user data, but also allows for wavelength re-assignment and grooming of the optical signals. It is important to understand the impact of this approach on the optical network resources. In this work, we examine how midhaul links impact the optical network capacity requirement and wavelength blocking. We further consider adaptive PHY split processing in which different split points can be used for individual digitized radio signals.

The rest of the study is organized as follows. Section II introduces the DWDM based C-RAN architecture, PHY split technology, and edge datacenter development for 5G networks. In section III, we present the adaptive midhaul C-RAN approach. Section IV reports the simulation results of our framework compared with fronthaul-only C-RAN to validate the advantages of midhaul links in terms of overal optical network capacity. Section V concludes the paper.

II. BBU Functional Splits in C-RANs

A. DWDM based C-RAN Architecture

Fig. 1 illustrates an adaptive midhaul network in an optical DWDM based C-RAN architecture which we intend to examine. Each optical node is a reconfigurable optical add-drop multiplexer (ROADM) that serves as a hub for connecting various systems, such as RRHs, BBUs, and edge DCs via optical fibers. Depending on if there is a BBU pool (Central Office) or DCs connected with the ROADM, we classify these optical nodes into 3 categories: (i) wireless access point node, (ii) DC node, and (iii) BBU node. Wireless traffic requests can be sourced from any RRH in the optical nodes listed above, and traffic from the same source ROADM node and destination ROADM node are aggregated at the source ROADM node into a 100 Gb/s DWDM channel for transmission through the network. We assume that each optical link between ROADMs can support up to 40 high-capacity 100 Gb/s wavelength channels for upstream and downstream transmission.

In our reference C-RAN architecture, the light-path connection is set up directly between two ROADMs connecting the source RRH, and destination BBU pool respectively. In source ROADM, the fronthaul rates from all antennas of multiple sectors are aggregated into a single wavelength channel. Therefore, the final CPRI bit rate at source ROADM can be obtained from the equation [1]:$B_{CPRI} = 2 \times (16/15) \times S \times A \times f_S \times b_S \times LC$, where the 2 and 16/15 are IQ processing and overhead factors, respectively, and remaining factors are S number of sectors, A number of antennas per sector, f_s sample rate, b_s number of bits per sample, and LC line rate. Therefore, following equation [1], we obtain the aggregated CPRI rate of a cell with one mobile network operator, each coming with 3 sectors, 2x2 MIMO, and 20+20 MHz as 15 Gb/s. Functional Split in L1 Layer

Due to the large optical transmission capacity requirements of CPRI based fronthaul, functional splits in BBU-RRH digitized IQ data processing has been investigated such that some processing functions of the BBUs are moved to the RRHs. In order to support distributed MIMO and CoMP techniques, and also given the data rate reduction is not significant for split points higher than the L1 layer, current functional split processing architecture designs are focused on the MAC-PHY or PHY splits [6, 14]. Fig.2 (a) illustrates the general PHY split points of fronthaul. Although various methods are used to implement PHY splits, this analysis uses the capacity reduction factors corresponding to different splits shown in TABLE I, which were derived considering central small cell function virtualization with LTE HARQ approach [15].

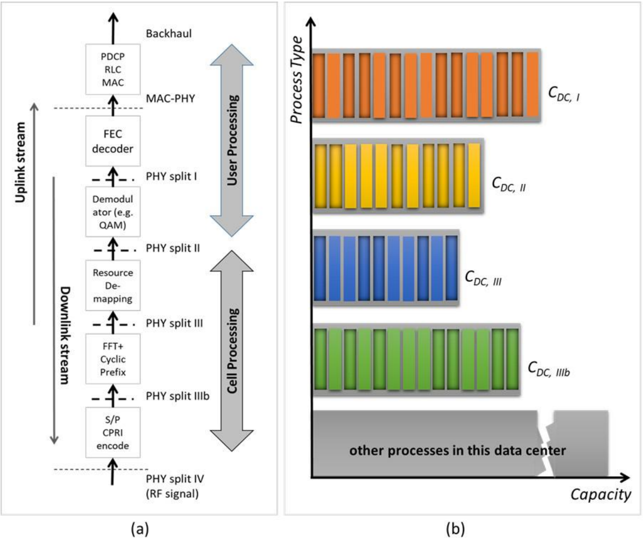


Fig. 2: Edge data center PHY split and processing capacity model. (a) Four PHY split steps. (b) DCs are randomly assigned processing capacity for each PHY split. Ocuppied resources are marked with shadow.

In a typical LTE processing chain, the radio frequency (RF) signals are received, and transformed to baseband. Then serial-to-parallel (S/P) conversion and CPRI encoding are applied. The cyclic prefix (CP) is removed and symbols are transformed to the frequency domain. Next, resource demapping is processed to disassemble sub-frames. Signal demodulator processing, such as Quadrature amplitude modulation (QAM) demodulation, is performed for each user. Finally, forwarded error correction (FEC) en-/de-coding is performed before sending the data to the higher layer functions and protocol processing, such as medium access control (MAC), radio link control protocol (RLC), and packet data convergence protocol (PDCP). The RRH-BBU processing split can be adopted after any of these processing components as mentioned above. Depending on the position of the split, the splits can be labelled accordingly as shown in the Fig. 2(a).

Various functional split options can be selected in a dual-site processing RAN framework to relieve the traffic load in fronthaul edge networks. However, decreasing the cost of fronthaul increases the cost of RRHs. This trade-off motivates the use of adaptive split PHY processing in edge data centers.

TABLE I. EFFECT OF SPLIT OPTIONS IN L1 LAYER [15]

Possible Split Levels	*Downlink bandwidth*	*Uplink bandwidth*
MAC - PHY	152 Mbps	49 Mbps
PHY split I: Soft Bit	173 Mbps	452 Mbps
PHY split II: Subframe data	933 Mbps	903 Mbps
PHY split III: Subframe	1075 Mbps	922 Mbps
PHY split IIIb: Subframe	1966 Mbps	1966 Mbps
PHY split IV: CPRI encoding	2457.6 Mbps	2457.6 Mbps

III. ADAPTIVE MIDHAUL C-RANS

A. Midhaul Networking

Generally, digitized baseband base station processing can be implemented in DCs. In a metropolitan area network, edge DCs with limited capacity or resources may conserve processing by partitioning split PHY processes. This strategy has the following benefits. First, these DCs are already widely deployed and can be easily implemented with PHY split processing functionality in a cost-efficient way. Secondly, they may have better processing performance than cost and power constrained RRHs. Thirdly, the PHY split point can be reconfigured and the resources can be tuned based on network or application requirements. Lastly, the wavelength of optical signals received by edge DCs can be dynamically reassigned. Those DCs implemented with PHY split processing act as temporary reconfigurable remote baseband processing or digital units, and work in coordination with the Central Office or BBU pools (which themselves may be implemented in a big data center). This midhaul strategy is a potential solution to the current trade-off problem between RRH placement expenditure and FH optical capacity requirements.

By deploying PHY split and wavelength reassignment in edge DCs, the DWDM based C-RAN architecture can be very flexible. The main features of this adaptive midhaul approach are as follows:

(1) All intra PHY split processing in RRHs is removed, only CPRI encoder remains
(2) Light-paths for RRH-BBU services can be multiple hops via DC nodes.
(3) Edge DCs have limited capacity to process various PHY split processing, and different functional split points require different capabilities and resource capacities.
(4) When a DC is the intermediate node along an RRH-BBU path, the wavelength channel can be reassigned.
(5) In a metropolitan area, the total length of fiber between source RRH and destination BBU pool should be less than 40km in order to meet ultra-low latency requirements [1].
(6) DCs can adapt processing resources for each PHY split function when other service resources are spare. The PHY split point is flexible for each signal, and it is dependent on current available resources in DCs.
(7) PHY split processing can still benefit from multiplexing gain when traffic is heavy.

We illustrate how edge data centers in this architecture work as follows. A RRH requests to set up a lightpath connection with a nearby Central Office or BBU pool to transmit the RF signal via CPRI. If there is a lightpath already set up from the aimed source to destination node, the signals are transmitted via an available channel or an occupied channel by grooming. Along the established light-path, any DC node can process PHY split with its available capacity. The preference of PHY split options for each DC node is from split I to split IIIb to best reduce the traffic data rate and save optical bandwidth. Every time the signal data rate is reduced in a DC node, signal is re-assigned and re-groomed into a new wavelength channel for transmission. If there is no available DC along the connection path, the original CPRI data rate is transmitted, and it acts like a traditional FH network connection in the C-RAN architecture. Fig.1 illustrates the edge datacenter midhaul strategy in DWDM based C-RANs.

B. Midhaul C-RAN based Routing and Wavelength Assignment

To evaluate this midhaul approach, a simulation model is needed. The key factor in this simulation model is to design an algorithm for routing and wavelength assignment (RWA) supporting split PHY processing and wavelength channel reallocation in DC nodes for RRH-BBU light-path connections.

We assume the final CPRI line rate per connection request from a RRH is aggregated by multiples of the 2.5 Gb/s basic CPRI rate signals. Besides, RF signals from different sectors can be groomed into a same channel. For example, a 3-sector cell with 2x2 MIMO, 20 MHz will occupy 3 channels, and each channel transmits a 7.5 Gb/s CPRI signal by using CPRI aggregated bit rate equation. In our work, we consider aggregated radio signals groomed into 100 Gb/s capacity optical channels. The simplified algorithm we designed is presented as below.

Algorithm: adaptive DC-PHY Split RWA Algorithm

```
Parameters:
    network Graph $G$,
    connection request $R_{s,d}$,
    source node $s$,
    destination node $d$,
    allocated channel $chnl$,
    occupied bandwidth $bw$,
    routing path $rp_{s,d}$,
    segmented routing path $srp_{s,d}$ (segmented points are DC nodes),
    final connection path $CP$,
    $i$-$th$ DC node $DC_i$ $(i = 1, 2, 3\dots)$,
    available DC capacity for split $C_{i,j}(j = I, II, III, IIb)$,
    resource exhausted by split $C_{PHY\text{-}k}$ $(k = I, II, III, IIIb)$

Input: network Graph $G$,
       connection request $R(s, d)$

Output: connection path $CP$

for each connection request $R_{s,d}$ do
If $s==d$ source equals destination do
    BBU node handles this local request
    return path results $CP$ with NONE
```

```
if $s!=d$ not local request do
    find $K=5$ candidate routing paths $rp_{s,d}$ in $G$
    for each $rp_{s,d}$ do
        for there is any $DC_i$ in $rp_{s,d}$ do
            routing path is segmented into $srp_{s,d}$
        if $srp_{s,d}$ ==NONE, no DC node along $rp_{s,d}$, do
            assign an available $chnl$, $bw$ (2.5 Gbps)
            if $chnl$ or $bw$ == NONE, no available resources
                continue to next routing path $rp_{s,d}$
            else return $CP$ with $rp_{s,d}$, $bw$ and $chnl$
        else for each segment in $srp_{s,d}$ do
            assign an available $chnl$ per segment
            if $chnl$ == NONE break to next path $rp_{s,d}$
            if the source node of the segment is $DC_i$ do
                find the traffic required split point
                for $k$ = from $I$ to required split point do
                    if $C_{i,k} > C_{PHY\text{-}k}$ do
                        $C_{i,k} = C_{i,k} - C_{PHY\text{-}k}$
                        $bw$ is allocated based on processed split
                        break
                    else required $bw$ does not change
            return $CP$ with $srp_{s,d}$, a list of $(bw, chnl)$
if no successful routing path $rp_{s,d}$ is found do
    return $CP$ with blocking, service fails
```

To better understand the performance of adaptive midhaul networks, we introduce two routing selection policies implemented in adaptive DC-PHY Split RWA: (1) Direct Link First (DLF) which searches the shortest candidate routing paths from source to destination, and (2) long multi-hop routing paths via DCs First (DCF) method that is greedy to find nearby DC nodes. If the direct path is chosen first, it means there is the least number of datacenters with PHY split processing along the traffic path, so that the C-RAN will consume the most capacity in the optical network. Otherwise, more DC resources are used to reduce the required capacity for the overall network.

C. Edge Data Center Model

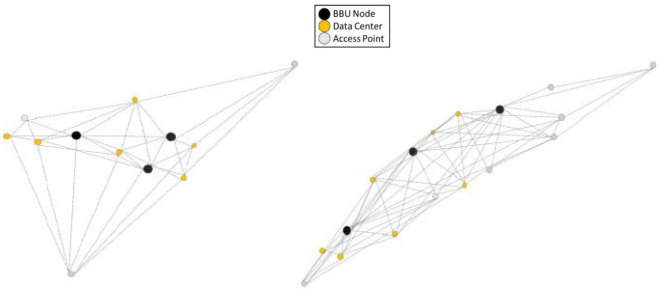
Our edge data center capacity model accounts for unique hardware processing capabilities and PHY split processing capacities. The model splits the PHY processing into four separate processing steps. The processing resources required for each step can be uniquely specified as well as the capacity within each data center for processing the corresponding steps. In this way, the model can account for pre-assigned resources for different processing steps and unique accelerated hardware for the steps. The model also allows for the steps to be grouped in different combinations or altogether for a uniform computing model. In practice, the specific resources for a given step in the PHY processing can be a complex function of the various hardware components or server configurations. By parameterizing the different processing steps, the impact of different processing constraints can be studied. This also enables the analysis of CoMP strategies utilizing different split points [16].

Our edge data center model is explained in Fig. 2. Fig. 2(a) illustrates PHY split processing steps deployed in data centers. And Fig. 2(b) shows the specific resources assigned to wireless signal PHY split processing in data centers. Different PHY split points need different processing equipment and capacity. This may relate to CPU performance and memory size. For example, PHY split point III and IIIb may use FPGA to finish processing, while computing split point I and II is using VMs deployed in performance servers [17], in order to meet the 5G latency requirement. In our simulation, each PHY split step is randomly assigned certain processing capacity for each PHY split step in a DC to mimic various data center conditions. When traffic is handled in this DC, it first determines if this traffic needs PHY split processing and what the split point is. Then accordingly, the appropriate split processing is applied when there is available capacity. And the handled traffic consumes its according PHY split capacity. For example, traffic processed only at PHY split point IIIb in a DC is forwarded into another DC for split point I. Then this new DC will only consume capacity $C_{PHY\text{-}I}$, $C_{PHY\text{-}II}$, and $C_{PHY\text{-}III}$ to complete the task, since the signal already consumes the capacity $C_{PHY\text{-}IIIb}$ in the previous DC.

IV. Simulation Results and Evaluations

A. Simulation Setup

The metropolitan network topologies were developed from commercial fiber networks deployed in New Jersey and Manhattan [www.zayo.com], shown in Fig. 3. The total 12 nodes are divided into 3 BBU nodes, 6 DC nodes and 3 wireless access nodes in the New Jersey topology, while 17 nodes are divided into 3 BBU nodes, 7 DC nodes, and 7 pure access points in the Manhattan topology. A discrete-event simulator is developed and 5000 Poisson traffic requests are generated with their source node and destination BBU node uniformly distributed in the topologies.



(a) New Jersey topology. (b) Manhattan topology.

Fig. 3: New Jersey and Manhattan radio access networks.

The selected DC nodes are randomly assigned with limited processing capacity for functional PHY split processing to simulate various sizes of data centers. The resulting CPRI data rate after the processing to a give split point are shown in TABLE I. For simplification, we only consider the downstream direction in our simulation.

Besides, the assigned limited capacity in edge DCs is scaled by a factor k (k = 0.5, 1.0 and 2.0). For example, certain capacity is randomly assigned to DCs for four PHY split processing in the beginning of the simulation. When scaling factor is 1.0, DC capability keeps same. When the factor becomes 0.5, PHY split points in this data center is assigned only for IIIb and III. And when the factor is 2.0, the data center can still process all PHY split point, and also have twice capacity for each split-PHY step than the factor set as 1.0.

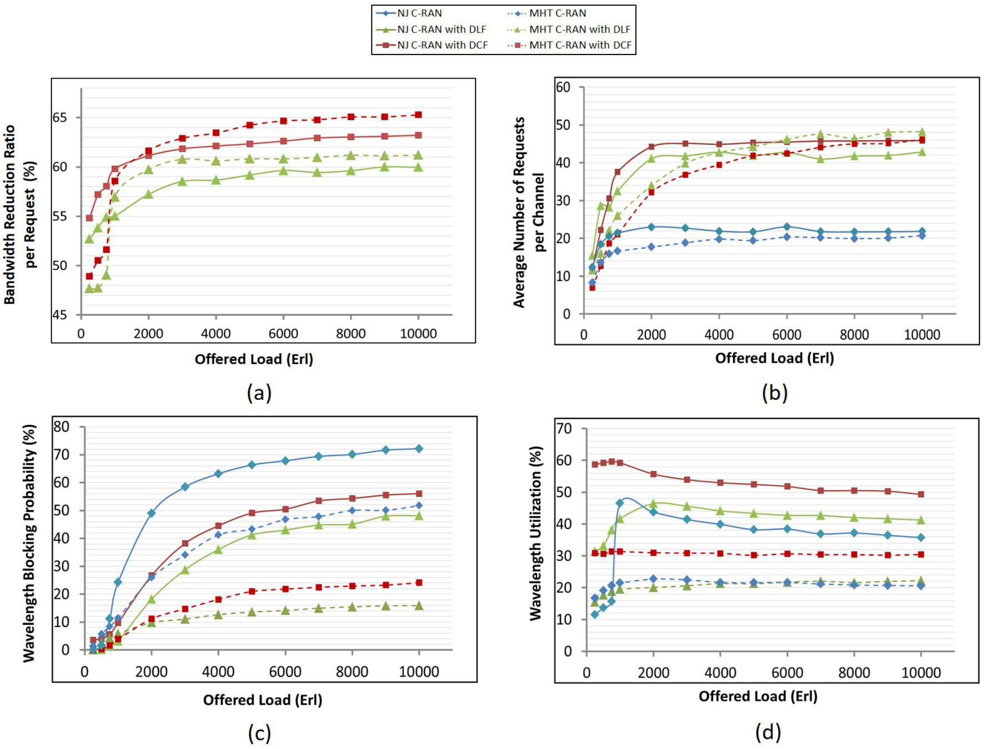


Fig. 4: New Jersey (NJ) topology simulation results of midhaul C-RAN with adaptive PHY split processing and wavelength reassignment in edge data centers compared with traditional C-RAN. Two routing policies implemented in adaptive DC-PHY split RWA are used: direct link first (DLF) for shortest path from source to destination, and long link via nearby data center first (DCF) algorithms.

B. Simulation Results

Fig. 4 shows our adaptive DC-PHY split midhaul networks simulation results compared with a pure fronthaul C-RAN. In Fig. 4(a), the average bandwidth reduction ratio per request is determined by calculating the average data rate per connection taken over all links in the network compared with the pure FH C-RAN case for the same set of demands. With assistance from PHY split processing in edge DCs, the average optical bandwidth used per request can be reduced more than 45% using midhaul networking. Also, the ratio keeps increasing as the traffic load increases. This result can be understood considering that for both networks there are more DC nodes than BBU nodes. This means a large portion of traffic is transported via the light-paths with DC nodes when the traffic load is heavy. Our midhaul network approach will use DC nodes to process traffic dynamically and flexibly as much as possible, by taking different PHY split points and discovering all available wavelengths for each request. When comparing DLF and DCF methods, we find the DCF method can have more effect on reducing bandwidth, due to its priority for using split PHY processing in DC nodes, while DLF provides a lower blocking rate by using shorter light-path.

Since the average data rate used per request becomes lower, the average number of processed connections per channel becomes higher, as illustrated in Fig. 4(b). Therefore, the utilization per wavelength channel is improved when compared with pure C-RAN. In midhaul networking, the average number of handled connections per channel can be at least twice that of a fronthaul C-RAN architecture in the both topologies.

As Fig. 4(c) shows, blocked connection requests in midhaul C-RAN are much less than traditional C-RAN. In midhaul networking, any DC node along the path can reduce the required data rate for each request, re-allocate the wavelength channel, and re-groom the signal into a working but bandwidth-spare light-path. This reduces the blocking rate due to the smaller wavelength management granularity of the networks. In this way, small path segments in midhaul C-RAN can be fully used for CPRI signal transmission.

Lastly, the results of wavelength utilization (total number of occupied wavelengths/number of all wavelengths) in our simulation model are shown in Fig. 4(d). The DCF method in C-RAN can provide significant improvement in channel utilization. Comparing C-RAN with midhaul C-RAN using the DLF method, there is a period between offered traffic load 1000 and 2000 Erlang in which C-RAN has higher utilization. This results from handling fewer requests per channel on average, as explained in Fig. 4(b). As the traffic load becomes high, more and more channels are used to set up light-path connections in midhaul networking, while fewer channels are available in C-RAN due to the high blocking rate. The link-level wavelength management on the segmented light-paths can keep midhaul C-RAN performing with better utilization at high traffic load.

To evaluate the impact of the processing resources of the edge datacenters, we linearly change the PHY split processing capacity of the DCs with a multiplicative factor, as Fig. 5 shows. With DC capacity scaling factor 2.0, DC nodes have twice the resources for PHY split processing. With scaling factor 0.5, the processing capacity in most edge DCs is too low to process higher PHY split points since we assume only large DCs have qualified equipment to process lower PHY split points. As a result, the transmitted bandwidth of the CPRI signal is close to the unprocessed data rate. So the blocking ratio remains as high as pure FH C-RAN.

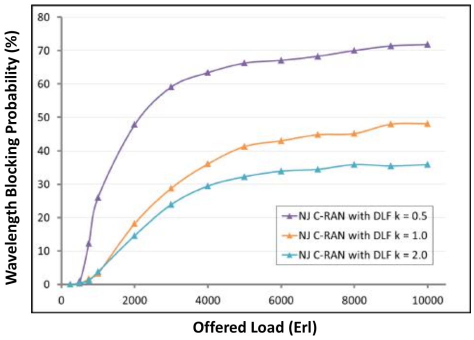


Fig. 5: Evaluating the effect of DC processing capacity by scaling the total capacity with a factor *k*. The result is based on adaptive DC-PHY split RWA with DLF policy in New Jersey topology.

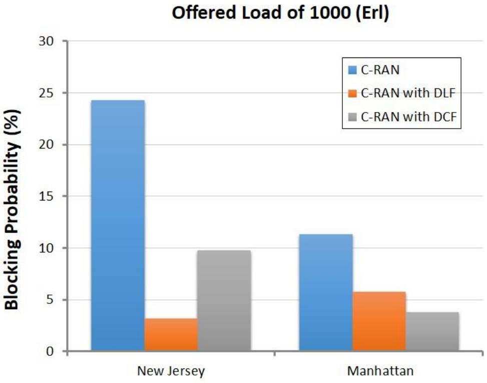


Fig. 6: Blocking probability of two policies in two network topologies when offered load is 1000 Erlang.

Simulations using the Manhattan topology show similar results except that the DLF method has a higher blocking rate than the DCF method when offered traffic is low, while this result is opposite in the New Jersey topology case, shown in Fig. 6. In the New Jersey networks, access points are widely spread, while access points with low connectivity degree are centralized in the northeast in Manhattan areas. When traffic is sourced from these access points in the Manhattan network, it is difficult to offer sufficient resources to handle these signals in edge DCs, when using the DCF method. Additionally, long multi-hop paths would cause severe traffic congestion, since there are insufficient resources for channel re-allocation in nearby DC nodes.

V. Conclusion

This paper examines the network efficiency gains through using midhaul networks in DWDM based C-RANs. PHY processing is adaptively split based on available edge data center processing resources in metropolitan areas. Midhaul networks are shown to reduce the CPRI optical bandwidth by deploying functional PHY split processing in edge datacenters. At the same time, functionality of dynamic wavelength reassignment deployed in edge data centers can improve optical network performance, such as blocking ratio and channel utilization, due to segment-scale granularity management of the C-RAN architecture. Besides, midhaul networks provide reconfigurable and cost-efficient performance, since functionalities and resources in edge data centers can be tuned to fit the network topology and traffic load pattern.

Acknowledgements

This work was supported by Science Foundation Ireland (SFI) 15/US-C2C/I3132 and 13/RC/2077, and the NSF Center for Integrated Access Networks (CIAN) EEC-0812072.

References

[1] Pfeiffer, Thomas. "Next generation mobile fronthaul and midhaul architectures." *Journal of Optical Communications and Networking*, vol. 7, no. 11, pp. B38-B45, 2015.

[2] Jun, Wu, et al. "Cloud radio access network (C-RAN): a primer." *IEEE Network*, vol. 29, no. 1, pp. 35-41, 2015.

[3] Rost, Peter, et al. "Cloud technologies for flexible 5G radio access networks." *IEEE Communications Magazine*, vol. 52, no. 5, pp. 68-76, 2014.

[4] Fiorani, Matteo, et al. "Challenges for 5G transport networks." *IEEE International Conference on Advanced Networks and Telecommunication Systems (ANTS)*, pp. 1-6, 2014.

[5] Skubic, Björn, et al. "The role of DWDM for 5G transport." *European Conference on Optical Communication (ECOC)*, pp. 1-3, 2014.

[6] Dötsch, Uwe, et al. "Quantitative analysis of split base station processing and determination of advantageous architectures for LTE." *Bell Labs Technical Journal*, vol. 18, no.1, pp. 105-128, 2013.

[7] Checko, Aleksandra, et al. "Evaluating C-RAN fronthaul functional splits in terms of network level energy and cost savings." *Journal of Communications and Networks*, vol. 18, no. 2, pp. 162-172, 2016.

[8] Miyamoto, Kenji, et al. "Performance evaluation of Ethernet-based mobile fronthaul and wireless CoMP in split-PHY processing." *Journal of Optical Communications and Networking*, vol. 9, no. 1, pp. A46-A54, 2017.

[9] Xinbo, Wang, et al. "Centralize or distribute? A techno-economic study to design a low-cost cloud radio access network." *IEEE International Conference on Communications (ICC)*, pp. 1-7, 2017.

[10] Bhaumik, Sourjya, et al. "CloudIQ: A framework for processing base stations in a data center." *Proceedings of the 18th annual international conference on Mobile computing and networking*. pp. 125-136, 2012.

[11] Wubben, Dirk, et al. "Benefits and impact of cloud computing on 5G signal processing: Flexible centralization through cloud-RAN." *IEEE signal processing magazine*, vol. 31, no. 6, pp. 35-44, 2014.

[12] Al-Fares, Mohammad, Alexander Loukissas, and Amin Vahdat. "A scalable, commodity data center network architecture." *ACM SIGCOMM Computer Communication Review*. vol. 38, no. 4, pp. 63-74, 2008.

[13] Peterson, Larry, et al. "Central office re-architected as a data center." *IEEE Communications Magazine*, vol. 54, no. 10, pp. 96-101, 2016.

[14] Miyamoto, Kenji, et al. "Split-PHY processing architecture to realize base station coordination and transmission bandwidth reduction in mobile fronthaul." *Optical Fiber Communications Conference and Exhibition (OFC)*, pp. 1-3, 2015.

[15] "Small Cell Virtualization: Functional Splits and Use Cases", Small Cell Forum whitepaper, rel. 7, Jan. 2016.

[16] Shibata, Naotaka, et al. "System level performance of uplink transmission in split-PHY processing architecture with joint reception for future radio access." *IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pp. 1375-1379, 2015.

[17] Amos, Doug, Austin Lesea, and René Richter. FPGA-based prototyping methodology manual: Best practices in design-for-prototyping, 2011.

On the Benefits of Elastic Spectrum Management in Multi-Hour Filterless Metro Networks

Jose-Juan Pedreno-Manresa*, Jose-Luis Izquierdo-Zaragoza†, Filippo Cugini‡ and Pablo Pavon-Marino*
*Universidad Politécnica de Cartagena, Plaza del Hospital, 1, 30202 Cartagena, Spain
†Universitat Politècnica de Catalunya, Jordi Girona, 31, 08034 Barcelona, Spain
‡Consorzio Nazionale Interuniversitario per le Telecomunicazioni, Via Moruzzi 1, 56124 Pisa, Italy
E-mail: josej.pedreno@upct.es

***Abstract*—The dawn of 5G is pushing operators to deploy high-capacity, agile networks capable of adapting to time-varying traffic patterns, especially into metro sections. ROADMs are key enablers for agility in the optical layer, however the benefits of this agility do not always compensate for increased costs. As such, filterless optical networks are emerging as a cost-effective and reliable solution compared to active photonics, thanks to a winning combination of coherent transponders and passive splitters/couplers. However, spectrum allocation management policies are of paramount importance to maximize the overall network throughput. In this paper, we focus on a filterless metro network where the hourly variation of the demands traffic is known, coming from historic data estimations. Then, we observe how the knowledge of the traffic profiles can be exploited. To assess this, we evaluate the performance, in terms of throughput, of three different spectrum management approaches: (i) fixed, where lightpaths remain static along time once allocated; (ii) semi-elastic, where lightpath-bandwidth vary according to current traffic requirements, but central frequency remains fixed; and (iii) hitless full-elastic, where any lightpath parameter may be reconfigured without disrupting the traffic. Besides, we consider two transponder types equipped with (i) shared or (ii) independent tunable lasers for transmission and reception, which affects to spectrum allocation of bidirectional connections. According to our results, the semi-elastic approach clearly outperforms the fixed approach (23-33% more throughput) with a reduced gap to the hitless full-elastic case (10-24% less throughput), especially considering that the latter is not commercially available yet. Interestingly, using dual-laser transponders only yields a 10% gain with respect to single-laser transponders for the semi-elastic scenario, and thus may not justify the extra hardware.**

***Index Terms*—Filterless optical networks, time-varying traffic, elastic spectrum allocation.**

I. Introduction

Filterless optical networks (FONs) have been introduced to target a significant cost reduction compared to existing optical networks. Optical nodes based on reconfigurable optical add/drop multiplexers (ROADMs) and wavelength selective switches (WSS) provide agility in optical transport but incurring in additional capital expenditures (CAPEX), which does not always compensate the benefits of agility. Leveraging coherent transponders (TXPs) and passive splitters/couplers, filterless nodes are deemed to be a cost-effective and reliable alternative to active optical nodes [1]. In FONs several optical trees are built over the same physical topology. In each of these trees, optical connections are based on *light-trees* and optical signals reach all nodes. Thanks to the coherent technology, each signal can be properly selected at the receiver (RX) of the corresponding destination node, following a drop and waste (D&W) strategy.

Despite of cost savings and simple control and maintenance operations, FONs present some drawbacks: (i) there is a significant waste of spectrum resources since each signal occupies the entire associated fiber tree; (ii) simple topologies such as horseshoes or trees with no physical loops must be used to avoid power recirculation and lasering effects of amplified spontaneous emission (ASE) noise; and (iii) careful control of the overall power entering each TXP is required, as all light channels enter each of them, provided that power levels must be above sensitivity.

As of today, filterless solutions have not been considered for large scale deployments [2]. The main motivation is that coherent technologies have been mainly adopted in backbone networks, where large ROADM-based mesh topologies are employed and the aforementioned filterless drawbacks have practically prevented the deployment of such D&W solutions.

However, the traffic growth occurring in metro networks is driving the replacement of traditional direct-detection 10G cards with more advanced solutions at 100G. In this context, where network operators usually adopt simple topologies (i.e., horseshoes, rings…) over relatively short distances (up to 150 km) such transmission constraints become less critical, and coherent TXP over filterless transport may represent a suitable and cost-effective option. Nonetheless, the potential application of these solutions in the metro opens the way to additional and yet undiscussed design and technological aspects.

In this paper, we focus on filterless solutions in the context of metro networks, specifically addressing the following two aspects: (i) the impact of traffic dynamicity, which is significantly more intense than the one experienced in backbone networks, in the design of spectrum assignment (SA) solutions for filterless networks; and (ii) the availability of either a single or two tunable lasers within the coherent TXP, which may limit flexibility on bidirectionality.

The contribution of this work is two-fold. First, an SA algorithm to (re-)allocate demands adapting to the new traffic conditions is presented, considering three different variants, namely *fixed*, *semi-elastic* and *hitless full-elastic* [3], depending on the tunable parameters on the transponder, that is,

978-3-903176-07-2 © 2018 IFIP

central frequency (CF) and bandwidth (BW). Second, we evaluate whether the investment in TXP with dual-laser is beneficial compared to single-laser variants. As a benchmark, we use a horseshoe metro FON subject to a *multi-hour traffic profile*, being throughput the metric under consideration.

The rest of the paper is organized as follows. In Section 2, we review some previous works in filterless networks. In Section 3, we provide with a background on the role of TXPs in filterless metro networks. In Section 4, we describe our SA algorithm. In Section 5, we report and discuss the results of our case study. Finally, Section 6 concludes the paper.

II. Related Work

The concept of FONs was first introduced in [4]. Aside of considerations from the photonics (i.e., power control, coherent detection) [5] or control plane [6] perspectives, in this section we review the efforts in the last decade in topics related to network planning and resource allocation.

The first problem arising in FONs is connectivity. In fact, these networks present several differences in terms of planning and operation with respect to ROADM-based networks. Actually, the routing and spectrum assignment (RSA) problem is augmented to a topology, routing and spectrum assignment (TRSA), where several physical sub-topology (or fiber trees, onwards) are built on top of a shared fiber topology (topology subproblem) and demands are alternatively assigned to some of them (routing subproblem). In addition, the broadcast nature within a tree means that a more careful SA is needed [7], aiming to take advantage of time-dependent spectrum sharing to optimize its utilization.

To summarize, as a result of the TR subproblems, we must ensure connectivity between all node pairs while avoiding laser loop and fulfilling reachability constraints. Finally, the SA provides the spectrum allocation across the corresponding fiber tree.

Authors in [8] propose a TRSA algorithm (with static traffic and focus on fixed-grid) to perform a techno-economic analysis comparing ROADM-based networks and FONs. Their results demonstrate cost savings up to two orders of magnitude, but lack of a throughput analysis. In [7], they extend their analysis to a dynamic scenario, where lightpaths are setup upon request over a previously deployed filterless network. Here, they provide a comparison of ROADM-based, static filterless and dynamic filterless in terms of wavelength usage over a given demand set but, still, there is nothing about a throughput analysis varying the overall network load. By contrast, authors in [9] present a multi-goal optimization solution to this problem in a pilot network, and they observed that up to 2.5 times more traffic can be supported by their reference network considering active photonics instead of filterless solutions. Finally, authors in [10] provide a comparison in terms of spectrum utilization similar to [7], but focused on flex-grid. This latter work also considers a multi-period scenario, where traffic grows year-over-year.

To the best of knowledge, our work is the first considering the potential implications the SA problem in the context of the particular constraints imposed by FONs for multi-hour traffic, where rate and spectrum occupation vary along time. The benefits of having such information about variations in traffic patterns, and its exploitation for improved network efficiency, have been studied in the past for fixed-grid [11] and flex-grid [3]. Here, we approach this problem in order to analyze the implications of different SA schemes and TXP technologies in terms of throughput and identify potential techno-economic aspects.

III. Filterless Technologies in Dynamic Metro Networks

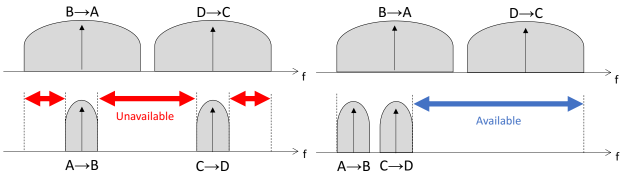
Next generation metro networks are expected to be driven by technological solutions where all cost contributions (e.g., hardware components within the coherent TXP) need to be carefully considered. Moreover, compared to backbone networks, high traffic dynamicity will be experienced, leading to the potential adoption of cost-effective highly reconfigurable solutions. These two aspects are discussed in the following subsections.

A. Transponder technologies

Coherent TXPs are built in hardware as *transceivers*, that handle within the same card both directions, i.e. transmitter (TX) and RX. According to the adopted technology and equipped hardware capabilities, transmission parameters may or not be configured in the same way both directions. For example, a connection between nodes A and B may or not be capable of supporting different CF from A to B (i.e., TX at A) and from B to A (i.e., RX at A). Such capability depends on the availability of either one or two tunable lasers and related electronic circuits within the card. If just one tunable laser is present, both TX and RX have to be operated over the same CF, i.e., the TX laser is also used as local oscillator at the RX side. Instead, if two tunable lasers are available in the same card, TX and RX can be operated over independent frequencies.

A single-laser provides cost savings, but introduces assignment constraints in spectrum allocations. Such constraints may be particularly relevant in next-generation metro networks, where significant traffic asymmetry is expected. For example, huge amount of traffic flows from a content delivery network (CDN) network to the access (i.e., downstream traffic to end-users) while the reverse direction (i.e., upstream traffic from users) is significantly less utilized.

Fig. 1 shows two asymmetrical bidirectional connections. From nodes A to B and C to D a large amount of spectrum is occupied to guarantee the required bandwidth capacity. Fig. 1a shows the case where a single-laser is shared between the TX and local oscillator at RX. In Fig. 1b two independent laser sources are utilized, enabling a free allocation of each central frequency. As it can be noticed, in the first scenario, the constraint on the same CF for both TX and RX leads to high fragmentation and, in turn, to relevant wasting of spectrum resources, which is avoided when each card is equipped with two independent laser sources.



(a) Single-laser TXPs (b) Dual-laser TXPs

Figure 1. Resource allocation with different TXPs technologies.

B. *Traffic dynamicity*

Regardless the bidirectional tunability issue described in the previous subsection, of special interest in this work is taking advantage of bandwidth-variable transponders (BV-Ts) to follow traffic variations adapting rate and spectrum occupation.

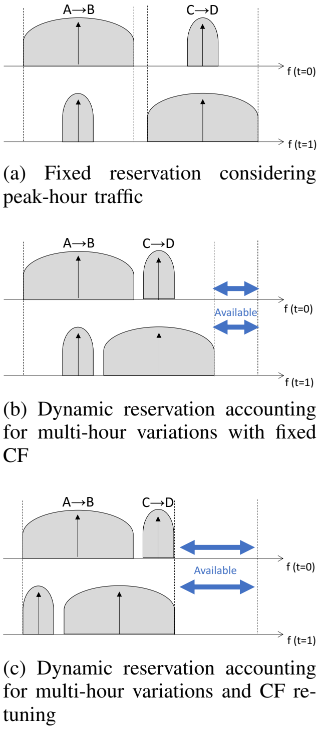
Fig. 2 shows two neighboring connections from nodes A to B and C to D (for simplicity, here just one direction is shown). Both connections experience time dependent bandwidth utilization. For example, the first connection serves a business district where most of the bandwidth is requested in working hours (time t=0) with scarce use, e.g., in the evening (t=1). Conversely, the second connection serves a residential area where limited bandwidth is needed during working hours but larger service requests are experienced in the evening. To this respect, different SA policies can be applied potentially considering time-dependent spectrum sharing.

In Fig. 2a, the resource allocation is performed in a fixed manner, considering peak-hour traffic volumes. In Fig. 2b, the allocation is performed accounting for multi-hour variations and not on the peak-hour values. This way, spectrum resources used by connection A to B at t=0 can be reused by connection C to D at t=1. In this case, fixed CF are assumed. Compared to Fig. 2b, in Fig. 2c we consider re-tuning of CF as an additional degree of flexibility. Such re-tuning can be implemented by adopting the *push-pull defragmentation* technique presented in [12], where the automatic frequency control of the coherent RX is properly exploited to track the gradual shift deliberately applied to the TX, without affecting high-layer traffic. Such adaptation is further simplified in FONs networks given the absence of filters.

It is worth mentioning that, as discussed in the subsequent sections, the combination of common/independent CF assignment in bidirectional connections (as discussed in Section III-A and such fixed/multi-hour assignment with/without central frequency adaptation may lead to remarkably different network utilization performances, with potential high impact on deployment costs and use of resources before fiber exhaustion, even for multi-period optimization where traffic only “grows”.

IV. Proposed Algorithm

As described in Section II, the TRSA problem in filterless networks can be decomposed into three subproblems: (i) generate the set of fiber trees (T), (ii) associate end-to-end demands to one of the trees (R), and (iii) allocate spectrum for each of the demands into the trees (SA).



(a) Fixed reservation considering peak-hour traffic

(b) Dynamic reservation accounting for multi-hour variations with fixed CF

(c) Dynamic reservation accounting for multi-hour variations and CF re-tuning

Figure 2. Resource allocation at different time of the day/week.

In this paper, we focus on horseshoe topologies, containing one bidirectional fiber tree, targeting the SA problem for all the demands within a tree. Due to the nature of FONs, topology and routing subproblems are implicitly solved for horseshoes.

The aim of our algorithm is to guarantee the correct provisioning of all demands while minimizing spectrum utilization. The input data of the algorithm are the following:

- FON topology containing a single fiber tree ($G(N, E)$), composed of a set of nodes with BV-Ts and bidirectional fibers interconnecting them, each supporting up to S slices at a line rate of R Gbps/slice.
- A list containing 24 different traffic matrices ($M^t_{(i,j)}$ where (i, j) is the source-destination node pair and $t \in [0, 23]$ the time period), containing per-hour traffic volume.

Our multi-hour spectrum allocation (MH-SA) initially considers a *semi-elastic* approach [3]: each lightpath uses a fixed CF whereas the number of occupied slices is adapted according to the offered traffic on each period. Next subsection is devoted to describe the implementation of our MH-SA algorithm, whereas further subsections will describe variations of the algorithm according to the technologies and techniques explained in Section III.

A. *Implementation*

Because of the $\mathcal{NP}$-complexity of the SA problem, we resort to a heuristic algorithm based on biased random-key genetic algorithm (BRKGA) [13], a well-known variant of the genetic algorithm (GA) meta-heuristic. The algorithm (pseudocode shown in Fig. 3) consists on several phases that are executed sequentially within a loop until a feasible solution is found or the maximum number of iterations (given as an input parameter) is reached.

- *GeneratePopulation*. Each candidate solution (also called *chromosome*) is encoded as an array, where each index (key) represents the demand identifier and its value the CF. To generate a chromosome, all demands are randomly sorted, and one-by-one we try to find a valid CF (using a first-fit approach) that ensures enough slices available for each hour of the day. To guarantee all demands can be initially serviced, we assume unlimited slice capacity (that is, an arbitrarily large *virtual spectrum*), whose rationale is further explained. The total number of chromosomes generated in this phase is given as an input parameter of the algorithm.
- *Crossover*. New chromosomes (*offspring*) are generated in this phase. Each new chromosome inherits its values from two 'parent' chromosomes: one of them is always selected among the *elite* population, whereas the other can be *elite* or not. Without loss of generality, in the first iteration, there is no *elite* population and both parents are selected at random. We iterate over each key (i.e., demand) selecting randomly one value (i.e., CF) among the two parents. The offspring chromosomes will then become part of the general population and the parents are discarded. The number of offspring generated in this phase is given as an input parameter.
- *CostComputation*. In order to test the feasibility of each candidate solution, we iterate over the 24 sets of demands. For each demand we select the appropriate CF from the chromosome and try to allocate the required number of slices. In case of overlapping, we consider the solution as infeasible. Once we have iterated over all sets of the demands, we select those slices that have never been used and we delete them from the virtual spectrum (shifting the remaining ones to the left). If the new shrinked spectrum size is lower or equal than the *real spectrum* size (input parameter), we consider the solution feasible and the algorithm ends. Each unfeasible solution is assigned a cost, equal to the percentage of blocked traffic.
- *Mutation* To avoid stagnation, in each iteration a portion of the general population is renovated, introducing newly created chromosomes (as explained in the first phase).

B. Transponder technologies

To reflect the effect of using different coherent TXP equipped with single or dual-lasers, we added an input parameter to the algorithm to control the bidirectional policy of the CF: (i) *different CF* with no restriction applied, and (ii) *same CF* which forces lightpaths to share the same CF as its bidirectional counterpart.

The former is implemented executing two independent instances of the MH-SA algorithm, one per direction of the horseshoe. In case a solution can be found for both, the scenario is feasible. Conversely, the latter is implemented in a tricky way: we concatenate traffic demands from each (i,j) in both directions in a virtual day of 48 hours. As such, we only execute an instance of the MH-SA algorithm.

```
Algorithm 1 Multi-Hour Spectrum Allocation
Require: $G(V,E)$, $M^t_{(i,j)}$, $S, R$
  GeneratePopulation
  $iteration = 0$
  while $iteration \leq maxIterations$ do
    Crossover
    CostComputation
    if feasible solution is found then
      end algorithm
    end if
    Select 'Elite' among the offspring
    Mutation
    iteration++
  end while
```

Figure 3. Pseudocode for the MH-SA algorithm.

C. Spectrum allocation approaches

To assess the quality of our algorithm, we compare the results with two boundary solutions in terms of achievable throughput. The lower bound is given by a *non-multihour, non-elastic* approach, where spectrum allocation is made for the worst case, that is, the daily maximum (peak) across all time intervals. The upper bound is given by a multi-hour aware *full-elastic with defragmentation* approach, where CF may also change, and therefore all connections are packed each time period using *push-pull defragmentation* techniques [12].

In order to model non-elastic/full-elastic, some minor modifications were added to our MH-SA algorithm. To implement the *non-elastic* option, we removed the dynamicity of the traffic information, considering only a single demand per node-pair where the offered traffic was equal to the maximum among the 24-hour time. For the *hitless full-elastic* approach, we consider all allocated demands can be packed, removing all spectrum between connections, for each single slot, but maintaining the same left-to-right order (in terms of slice indexes) of demands. Note that keeping such left-to-right order means that the variations in the CF can be accomplished using non-disruptive slow spectrum-shifting techniques.

V. Case Study

In this section, we report the results collected from testing our algorithm in a real-life scenario. We aim to analyze the total throughput achieved by combining different modulations and spectrum allocation approaches to find possible trade-offs.

The results were obtained using the offline network design tool Net2Plan [14]. With this tool, users can design and dimension networks assuming some static information (e.g. physical topology and traffic matrix). The algorithm was develop in Java, implementing public and well-documented interfaces. For the purpose of inspection and validation, both the source code of the algorithm and Net2Plan are available on the website [15].

A. Testing scenario

In order to test the algorithm we use the horseshoe topology in Fig. 4 as reference scenario. It is composed of 2 metro-core edge nodes (MCENs) providing connectivity toward core networks, 5 access-metro edge nodes (AMENs) as gateways for end-users (actual producers and consumers of traffic), and 3 nodes acting as CDN. For simplicity, we consider asymmetric, bidirectional traffic demands for node pairs MCEN-MCEN, MCEN-CDN, CDN-CDN, CDN-AMEN and AMEN-AMEN. MCEN-CDN and CDN-AMEN traffic is only considered for the closest node pairs. For example, there is only MCEN-1-CDN-1 traffic but not MCEN-2-CDN-1 traffic, as well as CDN-1-AMEN-1 traffic but not CDN-1-AMEN-5 traffic.

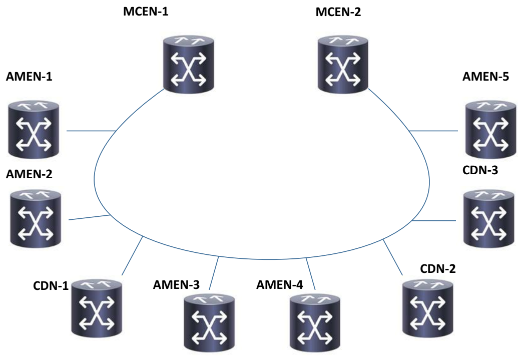


Figure 4. Reference horseshoe topology.

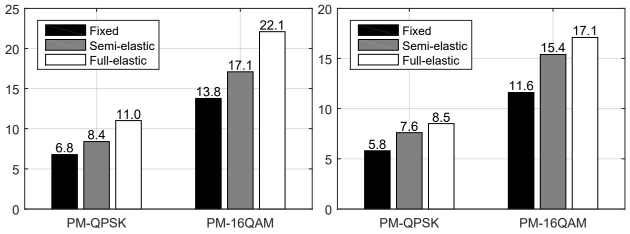
A combination of several different methods were used to generate a realistic multi-hour traffic matrix. First, a reference traffic matrix was built using a population-distance model described in [16], using scaling factors coming from analysis, trends and forecast for the incoming years [17]. Then, to recreate a multi-hour scheme, an activity factor was applied to the aforementioned traffic to simulate variance on each hour of the day [18] using a bimodal distribution to model peak and idle periods. Parameters of these distribution like width of the peak period and peak-to-idle ratio where selected at random. More details can be obtained directly from the source code.

B. Results

Different tests were performed using two different modulations: (i) 32 GBaud PM-QPSK over 37.5 GHz, with a total bitrate of 100 Gbps (R=16.6 Gbps per 6.25 GHz slice) and (ii) PM-16QAM, leading to 200 Gbps over 37.5 GHz (R=33.3 Gbps per 6.25 GHz slice). We establish a spectrum of 2.5 THz (S=400 slices). Using the scheme explained in Section V-A, we scaled the total offered traffic increasingly until blocking occurs. Then, we determine spectrum requirements SU for traffic (i, j) in time period t according to Eq. (1):

$$SU^t_{(i,j)} = \left\lceil M^t_{(i,j)}/R \right\rceil \tag{1}$$

We tested the algorithm MH-SA for the three SA approaches, single/dual-laser TXPs and the two modulation formats. Throughput results, indicating the peak-hour carried traffic before resource exhaustion, are presented in Fig. 5.



(a) Dual-laser TXPs (b) Single-laser TXPs

Figure 5. Maximum throughput (in Tbps).

We observed that the semi-elastic approach outperforms the fixed approach by a 23-33% in overall throughput. As expected, the algorithm performs poorly compared to the *full-elastic* scheme (with losses of 10-24% of throughput). These numbers are presented in Table I. Interestingly, even though it is clear that a full-elastic approach would be preferred in all instances, we would like remark that despite the fact that push-pull techniques [12] have been fully tested and demonstrated, unfortunately are not yet commercially available.

Table I
THROUGHPUT VARIATION (PERCENTAGE) WHEN USING SEMI-ELASTIC ALLOCATION COMPARED TO OTHER STRATEGIES

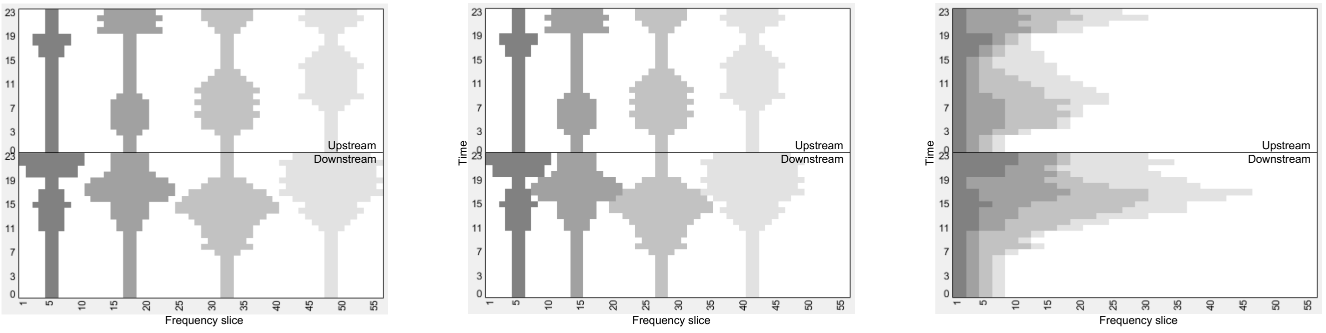
SA approach	dual-laser	PM-QPSK	PM-16QAM
Fixed	Yes	23.53%	23.91%
	No	31.03%	32.76%
Full-elastic	Yes	-23.64%	-22.62%
	No	-10.59%	-9.94%

From the results, it is worth discussing about the gain in terms of throughput using dual-laser TXPs instead of single-layer TXPs. Surprisingly, the use of the former only yields to an improvement of 10-11% of throughput and, as such, may not justify investment on these devices.

Table II
THROUGHPUT GAIN (PERCENTAGE) WHEN USING DUAL RESONATOR TRANSPONDERS INSTEAD OF SINGLE RESONATOR

SA approach	PM-QPSK	PM-16QAM
Fixed	17.2%	18.9%
Semi-elastic	10.5%	11%
Full-elastic	29.4%	29.2%

Before concluding, we illustrate the efficiency of our MH-SA algorithm in Fig. 6. This picture presents a partial snapshot of the spectrum utilization during a day in the case of using same CF for TX/RX. In Fig. 6a we can see the poor spectrum utilization of the fixed approach, since there is no change for spectrum sharing. Conversely, for the semi-elastic approach (see Fig. 6b) it is clear that because of bandwidth adaptation, slices can be shared by different demands in different time intervals, thus reducing the overall spectrum requirements. However, because of using same CF for both directions in this scenario, spectrum sharing cannot be totally exploited, thus reducing the overall throughput with respect to dual-laser TXPs. Finally, the full-elastic approach yields to the best spectrum utilization.



(a) Fixed SA (b) Semi-elastic SA (c) Full-elastic SA

Figure 6. Spectrum occupation for four example demands.

VI. Conclusions and Further Work

In this paper, we have presented an algorithm to determine the maximum throughput which can be achieved in filterless optical networks according to different spectrum management policies. In our results, a market-ready semi-elastic approach (considering state-of-the-art BV-Ts) represents a reasonable solution compared to a non-commercial hitless full-elastic scenario. Besides, we have illustrate the fact that single-layer TXPs may be a cost-effective alternative to dual-laser ones with limited throughput losses.

Extensions to this work may include mesh topologies, where fiber trees are also a decision variable, and/or techno-economic studies (e.g., independent tunability of TX/RX, comparison with ROADM-based solutions, and so on).

Acknowledgment

The research leading to these results has received funding from the EC through the METRO-HAUL project (G.A. n^{o} 761727), from the Spanish MINECO TWINS (TEC2017-90097-R) and ONOFRE-2 (TEC2017-84423-C3-1-P), the FPU Spanish fellowship program (ref. FPU14/04227), and the “Programa de Formación del PDI a través de la Movilidad de la UPCT” (PMPDI-2017).

References

[1] C. Tremblay, É. Archambault, M. P. Bélanger, J.-P. Savoie, F. Gagnon, and D. V. Plant, “Passive filterless core networks based on advanced modulation and electrical compensation technologies,” *Telecommunication Systems*, vol. 54, no. 2, pp. 167–181, Oct. 2013.

[2] C. Tremblay, P. Littlewood, M. P. Bélanger, L. Wosinska, and J. Chen, “Agile Filterless Optical Networking,” in *Proceedings of the 21st Conference on Optical Network Design and Modeling (ONDM 2017)*, Budapest (Hungary), May 2017.

[3] M. Klinkowski, M. Ruiz, L. Velasco, D. Careglio, V. Lopez, and J. Comellas, “Elastic Spectrum Allocation for Time-Varying Traffic in FlexGrid Optical Networks,” *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 1, pp. 26–38, Jan. 2013.

[4] C. Tremblay, F. Gagnon, B. Châtelain, É. Bernier, and M. P. Bélanger, “Filterless Optical Networks: A Unique And Novel Passive WAN Network Solution,” in *Proceedings of the 12th OptoElectronic and Communications Conference and 16th International Conference on Integrated Optics and Optical Fiber Communication (OECC/IOOC’07)*, Yokohama (Japan), Jul. 2007.

[5] F. Cugini, C. Porzi, N. Sambo, A. Bogoni, and P. Castoldi, “Receiver Architecture with Filter for Power-Efficient Drop&Waste Networks,” in *Proceedings of the Optical Fiber Communications Conference and Exhibition 2016 (OFC 2016)*, Anaheim, CA (United States), Mar. 2016.

[6] G. Mantelet, C. Tremblay, D. V. Plant, P. Littlewood, and M. P. Bélanger, “PCE-Based Centralized Control Plane for Filterless Networks,” *IEEE Communications Magazine*, vol. 51, no. 5, pp. 128–135, May 2013.

[7] G. Mantelet, A. Cassidy, C. Tremblay, D. V. Plant, P. Littlewood, and M. P. Bélanger, “Establishment of Dynamic Lightpaths in Filterless Optical Networks,” *Journal of Optical Communications and Networking*, vol. 5, no. 9, pp. 1057–1065, Sep. 2013.

[8] É. Archambault, D. O’Brien, C. Tremblay, F. Gagnon, M. P. Bélanger, and É. Bernier, “Design and Simulation of Filterless Optical Networks: Problem Definition and Performance Evaluation,” *Journal of Optical Communications and Networking*, vol. 2, no. 8, pp. 496–501, Aug. 2010.

[9] S. Krannig *et al.*, “How to design an optimized set of fibre-trees for filterless optical networks – The elegance of a multi-goal evolutionary Pareto optimization versus a deterministic approach,” in *Proceedings of 17th ITG-Workshop on Photonic Networks*, Leipzig (Germany), May 2016.

[10] É. Archambault *et al.*, “Routing and Spectrum Assignment in Elastic Filterless Optical Networks,” *IEEE/ACM Transactions on Networking*, vol. 24, no. 6, pp. 3578–3592, Dec. 2016.

[11] R. Aparicio-Pardo, N. Skorin-Kapov, P. Pavon-Marino, and B. Garcia-Manrubia, “(Non-)Reconfigurable Virtual Topology Design Under Multihour Traffic in Optical Networks,” *IEEE/ACM Transactions on Networking*, vol. 20, no. 5, pp. 1567–1580, Oct. 2012.

[12] F. Cugini *et al.*, “Push-Pull Defragmentation Without Traffic Disruption in Flexible Grid Optical Networks,” *Journal of Lightwave Technology*, vol. 31, no. 1, pp. 125–133, Jan. 2013.

[13] J. F. Gonçalves and M. G. C. Resende, “Biased random-key genetic algorithms for combinatorial optimization,” *Journal of Heuristics*, vol. 17, no. 5, pp. 487–525, Oct. 2011.

[14] P. Pavon-Marino and J.-L. Izquierdo-Zaragoza, “Net2Plan: An Open Source Network Planning Tool for Bridging the Gap between Academia and Industry,” *IEEE Network*, vol. 29, no. 5, pp. 90–96, Sep.-Oct. 2015.

[15] “Net2Plan - The open-source network planner,” [Last accessed: January 2018]. [Online]. Available: http://www.net2plan.com

[16] R. S. Cahn, *Wide Area Network Design: Concepts and Tools for Optimization*, 1st ed., ser. The Morgan Kaufmann Series in Networking. Morgan Kaufmann, May 1998.

[17] “Cisco – The Zettabyte Era: Trends and Analysis,” [Last accessed: January 2018]. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/vni-hyperconnectivity-wp.htm

[18] J. Milbrandt, M. Menth, and S. Kopf, “Adaptive Bandwidth Allocation: Impact of Traffic Demand Models for Wide Area Networks,” Institute of Computer Science, University of Würzburg, Würzburg (Germany), Research Report 363, Jun. 2005.

Burst-Mode FEC Performance for PON Upstream Channels with EDFA Optical Transients

Nicola Brandonisio, Daniel Carey, Stefano Porto, Giuseppe Talli, and Paul D. Townsend
Tyndall National Institute,
University College Cork,
Cork, Ireland
nicola.brandonisio@tyndall.ie

Abstract—**The performance of forward error correction (FEC) based on Reed-Solomon coding is analyzed experimentally for a burst-mode upstream channel within a passive optical network (PON) testbed. During this analysis, the upstream FEC performance is impaired by inducing correlated and localized errors within the burst through the injection of optical transients. These transients emulate the optical signal variation associated with the add-and-drop events of wavelength channels within a long-reach optical link based on a chain of erbium doped fiber amplifiers (EDFAs). The robustness of the FEC has been analyzed by measuring the post-FEC bit error rate (BER) as a function of the amplitude of the emulated transients and their delay with respect to the transmitted bursts. A margin of approximately 4dB is demonstrated for the transient amplitude before the FEC degradation. Furthermore, while the post-FEC BER is strongly degraded by the emulated transients, the pre-FEC BER stays below the FEC threshold, demonstrating the importance of measuring the post-FEC BER in order to correctly characterize the FEC performance in PON upstream channels affected by optical transients.**

Keywords—forward error correction; passive optical networks; burst-mode transmission; field-programmable gate arrays; erbium doped fiber amplifiers; optical transients.

I. Introduction

Long-reach passive optical networks (PONs) have been recently demonstrated capable of providing cost-efficient and flexible solutions for service convergence within the access and metro layers of the Internet infrastructure [1–5]. Different PON architectures have been proposed according to the geographical distribution of the customers which are covered by the network. In particular, a PON architecture based on a single amplifier node (AN) has been demonstrated in [3,4] to provide efficient solutions for densely populated metropolitan areas. Moreover, a PON architecture based on chains of optical ANs, including erbium doped fiber amplifiers (EDFAs), has been demonstrated in [5] to deliver efficient solutions for sparsely populated rural areas.

In all long-reach PON architectures, burst-mode forward error correction (FEC) plays a crucial role for the implementation of the upstream channels within the stringent optical power budget prescribed by the current PON standards [6,7]. In order to achieve the desired performance of the burst-mode FEC each network component should be carefully designed. In fact, burst-mode FEC can be affected by strongly correlated and localized errors within the burst, which can be introduced by transient behaviour of the network components.

In this work Reed-Solomon (RS) coding is employed for implementing the burst-mode FEC in accordance with the recommendations of the current PON standards [6,7]. The theoretical performance of RS-based FEC is obtained when errors are uniformly distributed within the burst. However, the PON upstream links are characterized by several sources of impairments which can cause strongly correlated and localized errors within the burst. These sources include the electrical and optical turn-on transients of the transmitter or the distortion potentially introduced by the gain setting of the electrical amplifier in the burst-mode receiver. An example of this concept has been demonstrated in [8] for a state-of-the-art long-reach PON testbed based on a single AN. In this case, the burst-mode FEC performance has been strongly degraded by inducing correlated and localized errors within the burst through the transient behaviour of a clock-and-data recovery unit. Other examples of burst-mode FEC performance degraded by correlated errors can be found in [9] for a PON upstream channel and in [5] for a long-reach PON testbed composed of a chain of EDFA-based ANs. However, in the investigation presented in [5], the burst-mode FEC performance has been only marginally degraded by introducing optical transients within the burst through on-off switching of several wavelength channels.

This paper aims to extend the work presented in [5] by analysing the burst-mode FEC performance with respect to the amplitude of optical transients which can arise in long-reach PONs that employ chains of EDFA-based ANs, when wavelength channels are added and dropped. In this work, these optical transients have been emulated within a simplified testbed in order to adjust the amplitude of the optical transients and their delay with respect to the optical bursts. This analysis has led to determine a margin of approximately 4dB for the amplitude of these optical transients before the FEC degradation. In this work the importance of measuring the post-FEC bit error rate (BER) is also demonstrated for correctly evaluating the burst-mode FEC performance in PON upstream links affected by optical transients.

Financial support from Science Foundation Ireland (SFI) (grants 12/IA/1270, 12/RC/2276, and 10/CE/I1853), and EU FP7 project DISCUS (grant CNECT-ICT- 318137) is gratefully acknowledged.

978-3-903176-07-2 © 2018 IFIP

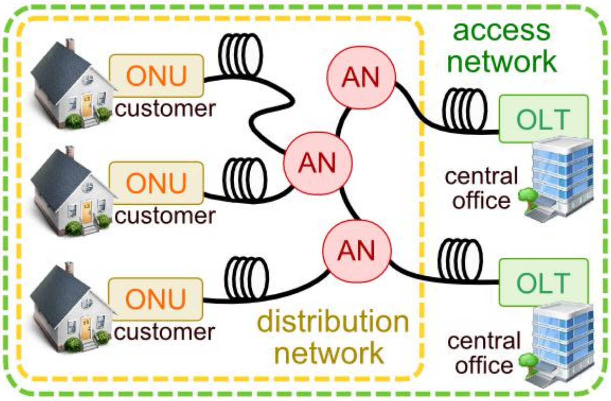


Fig. 1. Long-reach PON based on a chain of amplifier nodes.

This paper is organized as follows. In Section II, the PON testbed is presented with the description of the implemented network components. The technique used for emulating the optical transients is also described. In Section III, the FEC performance within this PON setup is experimentally investigated showing how the measured FEC performance can be degraded by the emulated optical transients. In this way, the robustness of the burst-mode FEC is evaluated with respect to these transients. Finally, in Section IV the conclusions of this work are presented.

II. Passive Optical Network Testbed

The long-reach PON access system considered in this work is based on the three main components schematically represented in Fig. 1: the optical network units (ONUs), the optical line terminals (OLTs), and the optical distribution network (ODN) [1-3]. The ONUs are used by the customers to access the PON infrastructure and are connected to a chain of EDFA-based ANs through the ODN, as described in [5]. The OLTs are located at the central offices and are connected to the ANs through the backhaul links, enabling the connection between the access network and the main Internet infrastructure.

This long-reach PON architecture has been emulated by the testbed used in this work. As shown in Fig. 2, two ONUs and one OLT have been fully developed in hardware using Xilinx Virtex-7 field-programmable gate arrays (FPGAs) on the VC709 development boards. Each ONU employs a semiconductor optical amplifier (SOA) for carving the optical bursts, which are transmitted and received using a standard small form-factor pluggable (SFP+) transceiver. These SFP+ transceivers are tunable across 100 wavelength channels with 50GHz spacing within the C-band and have a transmission rate of 10Gb/s. The upstream bursts generated by these two ONUs are time-multiplexed using a synchronization protocol implemented through the downstream data and managed by the OLT. The ONUs have been calibrated in order to transmit bursts with approximately the same optical power close to +1dBm. An SFP+ transmitter is also used by the OLT for generating the downstream data. The upstream burst traffic is received by the OLT using a linear burst-mode receiver (LBMRx), which is followed by a clock-and-data recovery (CDR) unit [10,11].

Real-time 10Gb/s burst-mode FEC has been implemented on the FPGAs using the RS(248,216) coding technique. This algorithm encodes a block of 216 data symbols into a block of 248 symbols by adding 32 check symbols, where each symbol is a group of 8 bits. When a block of 248 symbols is received, the RS(248,216) algorithm uses the 32 check symbols for correcting up to 16 corrupted symbols within the received block. This algorithm leads to a post-FEC BER lower than 1e-12 for a pre-FEC BER lower than 1.1e-3, assuming that the errors are uncorrelated and uniformly distributed within the burst. The pre-FEC BER value of 1.1e-3 is defined as the FEC threshold. Here, the optical power received by the OLT has been adjusted using a variable attenuator in order to obtain a pre-FEC BER for both ONUs close to the FEC threshold. This situation represents the worst case scenario for the FEC performance and leads to a conservative estimation of the impact of the emulated optical transients on the FEC.

As shown in the PON testbed depicted in Fig. 2, the power transients within the upstream burst traffic are emulated by driving an SOA with a waveform representative of the optical signal variation generated by chained EDFAs within a long-reach PON upstream link. In comparison with the approach of creating the power transients directly through a chain of EDFAs as implemented in [5], the approach used here has the advantage of varying the amplitude of the transients, which corresponds to a change in the number of chained EDFAs, by simply scaling the waveforms that drives the SOA. Moreover, the generation of the emulated optical transients can be synchronized with the optical bursts transmitted by the ONUs in order to analyse the effect of these transients as a function of their position within the upstream bursts. In particular, the OLT FPGA provides the electrical trigger for the waveform generator and controls the ONUs through the downstream protocol.

The waveform used for reproducing the profile of the residual optical transient has been measured at the output of a chain of 5 gain-stabilized EDFAs using a continuous-wave probe channel at 1550.12nm when the background channels (between 1546.12 – 1563.85nm) have been turned on ("Add Event") and off ("Drop Event") with a period of 10µs. These background channels are emulated by employing amplified spontaneous emission (ASE) which is generated by an EDFA before being filtered and flattened using a wavelength selective switch (WSS) and modulated using SOAs to emulate the typical burst mode power variations of upstream traffic as

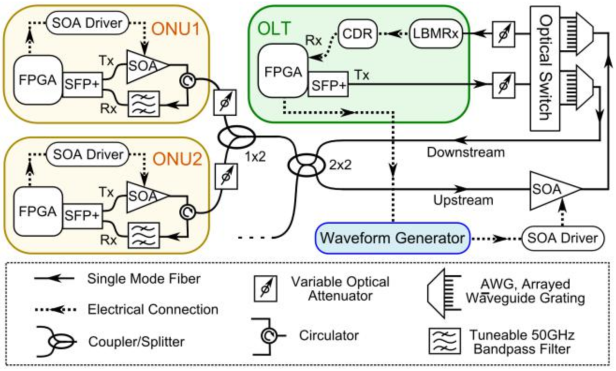


Fig. 2. PON testbed for EDFA transient emulation.

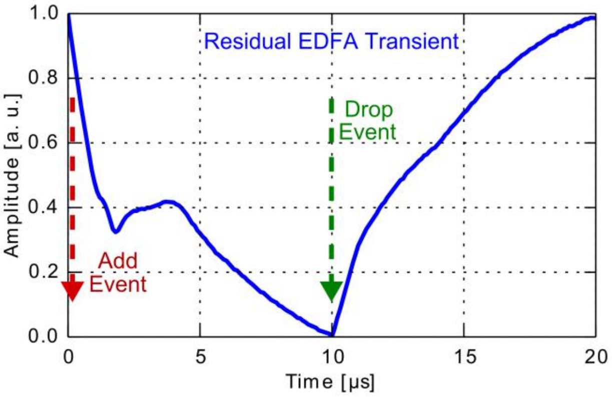


Fig. 3. Measured residual optical transient for an upstream optical link based on a chain of 5 gain-stabilized EDFAs with background traffic turned on ("Add Event") and off ("Drop Event") every 10µs.

described in [5]. The measured optical transient profile is plotted in Fig. 3. This profile is the result of the intrinsic gain dynamics of the EDFA coupled with the fast automatic gain stabilization inside the EDFA modules, which can lead to a rather complex shape as shown in Fig. 3. For a more detailed analysis of the dynamics involved we refer the interested reader to [12].

III. Burst-Mode Transient Analysis

The burst-mode FEC performance has been analysed as function of the amplitude of the emulated optical transients and their delay with respect to the upstream bursts. In order to equally affect the two ONUs, the duration of the bursts transmitted by the ONUs has been fixed to approximately 5µs, which is half of the period (10µs) of add-and-drop events. This scenario is schematically depicted in Fig. 4, where the duration of the sketched transients is compared with the duration of the time-multiplexed upstream bursts of the ONUs. From this figure one can see that the transient delay needs to be varied between 0 and 10µs in order to analyse all possible alignments between the optical transients and the optical bursts.

These considerations have led us to the characterization presented in Fig. 5, where the measured pre- and post-FEC BER for the two ONUs is plotted with varying transient delay and with a transient amplitude of 5.2dB. This particular value has been experimentally determined in order to be close to the minimum transient amplitude before the FEC degradation. In Fig. 5 the measured pre-FEC BER of both ONUs always stays below the FEC threshold, hence the post-FEC BER of both ONUs is expected to be lower than 1e-12. However, for a transient delay close to 0 the post-FEC BER of ONU1 clearly raises above 1e-12, while for a transient delay close to 5µs the post-FEC BER of ONU2 increases. This degradation of the FEC performance occurs because the optical transients introduce correlated and localized errors within the bursts. Hence, Fig. 5 demonstrates that, in upstream PON links affected by optical transients, the verification of a pre-FEC BER lower than the FEC threshold is not sufficient for assuring the correct FEC performance. In these cases, the post-FEC BER must also be characterized in order to verify the correct FEC behaviour.

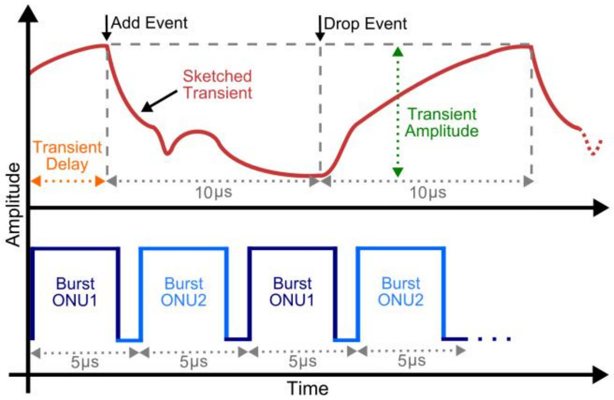


Fig. 4. Duration of sketched emulated transients compared with duration of upstream bursts transmitted by two time-multiplexed ONUs.

The measurements presented in Fig. 5 have been used to identify the following transient delays which represent the two worst cases for the FEC degradation: 0.1µs for ONU1 and 5.2µs for ONU2. These two cases, which correspond to an "Add Event" aligned with the start of the burst, have been analysed in Fig. 6 (a) and (b) respectively, where the measured pre- and post-FEC BER of both ONUs has been plotted as a function of the emulated transient amplitude. In Fig. 6 the pre-FEC BER of both ONUs increases with the increase of the transient amplitude. This behaviour is expected considering that the increase of the transient amplitude leads to the increase of the number of errors within the burst. In Fig. 6 a margin of approximately 5dB can be estimated by considering solely a pre-FEC BER below the FEC threshold. However, for transient amplitudes larger than 4dB the post-FEC BER increases above 1e-12 with a pre-FEC BER lower than the FEC threshold, indicating the degradation of the FEC performance. The difference of approximately 0.3dB between the margins measured for ONU1 and ONU2 is expected to be within the fabrication tolerances of the different SFP+ transmitters employed in the ONUs. The difference of approximately 1dB between the pre- and post-FEC margins could be particularly relevant when designing PON architectures within the strict optical power budget of the current PON standards. Finally, this work also demonstrates the relevance of characterizing the

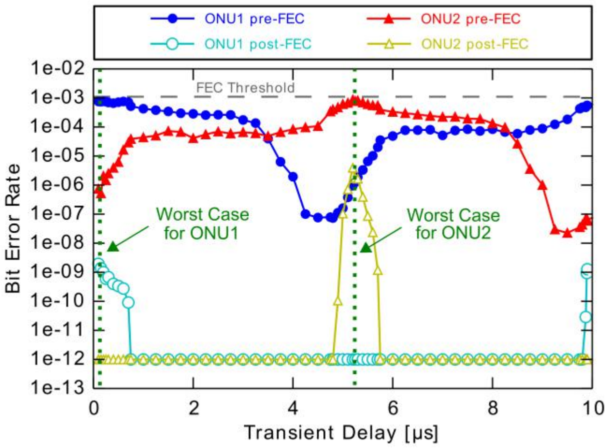


Fig. 5. Measured burst-Mode FEC performance with fixed transient amplitude of 5.2dB and with transient delay varied with respect to the upstream bursts.

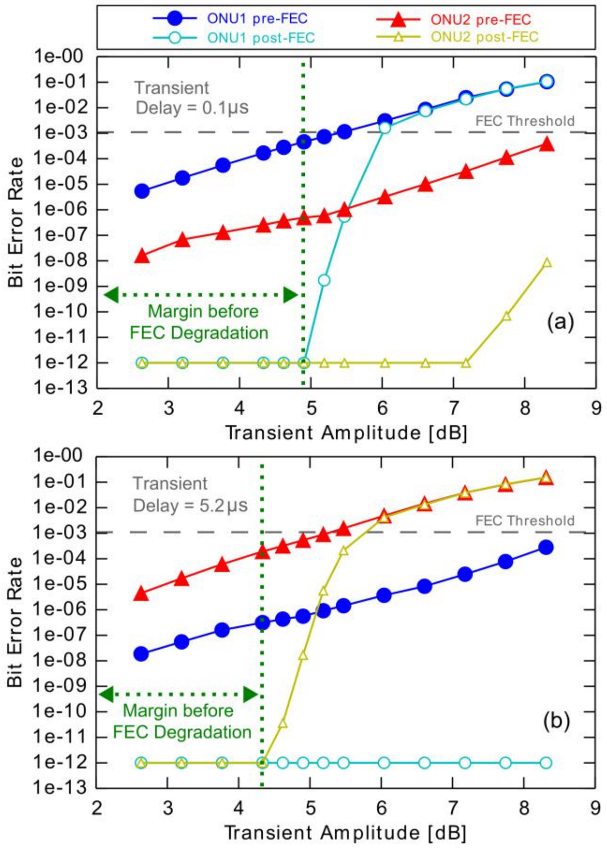


Fig. 6. Measured burst-mode FEC performance for two ONUs when the amplitude of the emulated EDFA transients is varied with transient delays of 0.1µs (a) and 5.2µs (b).

post-FEC BER in PON upstream channels as a function of the temporal distribution of the errors within the burst [8].

IV. Conclusions

The analysis of burst-mode FEC performance within a PON upstream channel with emulated optical transients has been presented. These transients have been emulated in order to mimic the optical signal variation caused by added and dropped wavelength channels within long-reach PON upstream links with chained EDFA-based ANs. The amplitude of the optical transients and their delay with respect to the bursts have been varied in order to investigate the robustness of the FEC. During this analysis, the emulated optical transients have degraded the FEC performance by introducing correlated and localized errors within the bursts. Two worst cases for the FEC performance have been analysed for the two ONUs, demonstrating a margin of approximately 4dB for the transient amplitude before the FEC degradation. This work has also demonstrated the importance of characterizing the post-FEC BER in PON upstream channels affected by optical transients, which can lead to the degradation of the FEC performance while leaving the pre-FEC BER below the FEC threshold.

References

[1] D. Nesset, “PON roadmap [invited],” IEEE/OSA Journal of Optical Communications and Networking, vol. 9, no. 1, pp. A71-A76, 2017.

[2] M. Ruffini, M. Achouche, A. Arbelaez, R. Bonk, A. Di Giglio, N. J. Doran, M. Furdek, R. Jensen, J. Montalvo, N. Parsons, T. Pfeiffer, L. Quesada, C. Raack, H. Rohde, M. Schiano, G. Talli, P. Townsend, R. Wessaly, L. Wosinska, X. Yin, and D. B. Payne, “Access and metro network convergence for flexible end-to-end network design [invited],” IEEE/OSA Journal of Optical Communications and Networking, vol. 9, no. 6, pp. 524-535, 2017.

[3] G. Talli, F. Slyne, S. Porto, D. Carey, N. Brandonisio, A. Naughton, P. Ossieur, S. McGettrick, C. Blümm, M. Ruffini, D. Payne, R. Bonk, T. Pfeiffer, N. Parsons, and P. Townsend, “SDN enabled dynamically reconfigurable high capacity optical access architecture for converged services,” Journal of Lightwave Technology, vol. 35, no. 3, 2017.

[4] G. Talli, S. Porto, D. Carey, N. Brandonisio, A. Naughton, P. Ossieur, F. Slyne, S. McGettrick, C. Blum, M. Ruffini, D. Payne, R. Bonk, T. Pfeiffer, N. Parsons, and P. Townsend, “Demonstration of SDN enabled dynamically reconfigurable high capacity optical access for converged services,” Optical Fiber Communications Conference (OFC), Postdeadline Paper Th5B.1, 2016.

[5] D. Carey, N. Brandonisio, S. Porto, A. Naughton, P. Ossieur, N. Parsons, G. Talli, P. Townsend, “Dynamically reconfigurable TDM-DWDM PON ring architecture for efficient rural deployment,” European Conference on Optical Communication (ECOC), pp. 692-694, 2016.

[6] ITU-T G.987.3, “10-gigabit-capable passive optical networks (XG-PON): transmission convergence (TC) layer specification,” 2014.

[7] ITU-T G.989.2, “40-gigabit-capable passive optical networks 2 (NG-PON2): physical media dependent (PMD) layer specification,” 2014.

[8] N. Brandonisio, S. Porto, D. Carey, P. Ossieur, G. Talli, N. Parsons, and Paul Townsend, “Forward error correction analysis for 10Gb/s burst-mode transmission in TDM-DWDM PONs,” Optical Fiber Communications Conference (OFC), Paper Th2A.28, 2017.

[9] E. I. de Betou, E. Mobilon, B. Angeli, P. Öhlen, A. Lindström, S. Dahlfort, E. Trojer, “Upstream FEC performance in combination with burst mode receivers for next generation 10 Gbit/s PON,” European Conference on Optical Communication (ECOC), Paper Mo.2.B.5, 2010.

[10] P. Ossieur, N. A. Quadir, S. Porto, C. Antony, W. Han, M. Rensing, P. O’Brien, and P. D. Townsend, “A 10Gb/s linear burst-mode receiver in 0.25 um SiGe:C BiCMOS,” IEEE Journal of Solid-State Circuits, vol. 48, no. 2, pp. 381-390, 2013.

[11] S. Porto, C. Antony, A. Jain, D. Kelly, D. Carey, G. Talli, P. Ossieur, and P. D. Townsend, “Demonstration of 10Gbit/s burst-mode transmission using a linear burst-mode receiver and burst-mode electronic equalization [invited],” IEEE/OSA Journal of Optical Communications and Networking, vol. 7, no. 1, pp. A118-A125, 2015.

[12] A. Kaszubowska-Anandarajah, R. Oberland, E. Bravi, A. Surpin, O. Aharoni, U. Ghera, R. Giller, E. Connolly, E. K. MacHale, M.Todd, G. Talli, and D. McDonald, “EDFA transient suppression in optical burst switching systems,” International Conference on Transparent Optical Networks (ICTON), Paper Mo.B2.4, pp. 1-4, 2012.

On Learning Bandwidth Allocation Models for Time-Varying Traffic in Flexible Optical Networks

Tania Panayiotou[1], Konstantinos Manousakis[1], Sotirios P. Chatzis[2], Georgios Ellinas[1]
[1] KIOS Research and Innovation Center of Excellence,
Department of Electrical and Computer Engineering, University of Cyprus
[2] Department of Electrical Engineering, Computer Engineering and Informatics,
Cyprus University of Technology,

Abstract—**We examine the problem of bandwidth allocation (BA) on flexible optical networks in the presence of traffic demand uncertainty. We assume that the daily traffic demand is given in the form of distributions describing the traffic demand fluctuations within given time intervals. We wish to find a predictive BA (PBA) model that infers from these distributions the bandwidth that best fits the future traffic demand fluctuations. The problem is formulated as a Partially Observable Markov Decision Process and is solved by means of Dynamic Programming. The PBA model is compared to a number of benchmark BA models that naturally arise after the assumption of traffic demand uncertainty. For comparing all the BA models developed, a conventional routing and spectrum allocation heuristic is used adhering each time to the BA model followed. We show that for a network operating at its capacity crunch, the PBA model significantly outperforms the rest on the number of blocked connections and unserved bandwidth. Most importantly, the PBA model can be autonomously adapted upon significant traffic demand variations by continuously training the model as real-time traffic information arrives into the network.**

I. Introduction

With the emergence of new types of applications and services, the Internet traffic is exponentially growing [1]. Next generation optical networks are expected to support both the ever increasing traffic demand and the increased uncertainty in predicting the sources of this traffic. Over the last few years, and as the currently deployed optical networks are nearing a capacity crunch, they have undergone significant changes.

Flexible optical networks are considered today as a promising solution for coping with the increasing demand, due to their capability of efficiently utilizing the available spectrum resources [2]. Flexible optical networks are based on bandwidth variable transceivers (BVTs), a flexible grid, and network nodes that can adapt to the actual traffic needs [2]. In this type of networks, for establishing a connection, the Routing and Spectrum Assignment (RSA) problem must be solved. The routing (R) problem deals with finding a route for a source and destination pair. The spectrum allocation (SA) problem deals with allocating spectral resources to the routing path (the spectrum slots are occupied symmetrically around the nominal central frequency of the channel). The allocated spectrum must meet the slot continuity and contiguity constraints [3], subject to the constraint of no frequency overlap. Once a connection is established the spectrum width can be dynamically adapted (if feasible) in response to bandwidth variations. The RSA problem for time-varying traffic has been studied in [4]-[7] with the aim of best fitting the bandwidth requirements upon demand variations. A survey regarding the methods developed for the R problem can be found in [8], whereas regarding the SA problem, a number of SA policies have been developed that are in general categorized into fixed, semi-elastic, and elastic [5], [8].

In the *fixed* SA policies [4], [5] the allocated spectrum and the central frequency remain static for the entire lifetime of a connection. These policies lead to a sub-optimal use of the available resources as much of the allocated spectrum is most of the time wasted. In the *semi-elastic* SA policies [4], [5] the central frequency remains static but the allocated spectrum width can be expanded/reduced according to the actual bandwidth demand. The main difference with the fixed SA policies is that the unutilized slots can now be used for subsequent connection requests providing higher flexibility and better resource utilization. In the *elastic* SA policies [4]-[7], [9] both the allocated central frequency and the spectrum width can change. The spectrum width can be expanded/reduced according to the actual bandwidth demand and the central frequency can be shifted [5], [6], [9], [10]. The elastic SA policies offer better resource utilization but require the highest computational complexity and complex algorithms in the Path Computation Element for minimizing traffic interruptions if a reallocation policy is followed [5], [8]. Further, control plane extensions are still required for allowing dynamically adjusting both the allocated spectrum and the central frequency.

Most SA policies are based on daily Internet traffic patterns that can be known a priori due to the periodic behavior of Internet traffic [5]-[7], [9]. The traffic patterns include information regarding the estimated peak rate of each connection request for each time interval (usually 24-hour patterns). The estimated peak rates are used by the SA policy followed in order to allocate just enough bandwidth for each connection. For handling a situation where more bandwidth is eventually requested than the estimated one, the estimated peak rate is multiplied by a certain oversubscription ratio [4].

Motivated by the fact that the Internet traffic demand has been shown to follow the log-normal distribution [11], in this work, instead of assuming that the daily traffic patterns are given in the form of estimated peak rates, we assume that they are given in the form of distributions describing the traffic

978-3-903176-07-2 © 2018 IFIP

demand uncertainty (the mean and variance of the log-normal distribution are given). In this work, the assumption throughout is that the distribution describes the aggregate traffic resulting from multiple users. We wish to infer from these distributions a predictive bandwidth allocation (PBA) model that best fits the future bandwidth demands. In particular, we wish to find a bandwidth allocation (BA) model that is capable of predicting the number of spectrum slots that will best fit the traffic demand fluctuations of the next time interval. We have formulated the problem as a Partially Observable Markov Decision Process (POMDP) as POMDPs have been proven to be very effective for addressing planning domain problems with uncertainty [12]-[14]. For finding the PBA model, the POMDP is solved by means of dynamic programming. Note that the approach used for training the PBA model does not need to know the underlying traffic demand distributions. It can utilize real-time information for continuously adjusting the model upon variations on the traffic demand. The training procedure can be performed continuously offline, given that enough traffic information is available. Large amounts of traffic information can be easily collected by monitoring the traffic demand fluctuations within short time intervals. Nevertheless, given the fact that we do not have available real traffic information, in this work, we made the assumption that the traffic demand distributions are known. These distributions are used as traffic demand data generators for training and evaluating the effectiveness of the proposed PBA model.

We assume a network that is elastically reconfigured at the beginning of each time interval (24 hourly intervals). For each network reconfiguration, an RSA heuristic is executed offline. The SA must adhere to the BA model followed. A connection is blocked if a feasible route and SA cannot be found. Between network reconfigurations, the bandwidth for the established connections is semi-elastically expanded/reduced according to the fluctuations of the actual traffic demand. If the allocated bandwidth is higher or equal to the requested one, then the connection bandwidth is semi-elastically expanded/reduced or it remains unchanged. If the allocated bandwidth is less than the requested one, then some of the requested bandwidth remains unserved.

The PBA model is evaluated and compared to a number of benchmark BA models that naturally arise from the assumption of traffic demand uncertainty. Specifically, the PBA is compared to the Highest BA (HBA), to the Maximum Probability BA (MPBA), and to the Expected BA (EBA) models on a network that is operating at its capacity crunch.We show that the PBA model significantly outperforms the rest regarding the unserved bandwidth.

II. Bandwidth Allocation Models

We assume that the traffic demand is log-normally distributed [11] and that traffic demand information is available for a 24-hour period and for N source-destination pairs (connections). In particular, we assume that each connection is described by a set of traffic demand distributions, with each distribution describing the traffic demand fluctuations within a single time interval. In general, the log-normal distribution is asymmetrically distributed around its mean value and is suitable for describing data with heavy-tails and skewness.

The traffic demand fluctuations for each time interval $\{t\}_{t=1}^{24}$ and for each connection $\{n\}_{n=1}^{N}$ are described by $Z_{tn} \sim LN(\mu_{tn}, \sigma_{tn}^2)$. We assume that $z_{tn} \in (0, B)$ and that $B < B'$, where $z_{tn} \in Z_{tn}$, B is equal to the feasible rate of the BVTs, and B' is equal to the total link capacity (all network links occupy B' spectrum slots). For making the learning procedure of the PBA model computationally tractable, we have discretized the distributions according to specific rate intervals. Specifically, we have divided B into a intervals in such a way that the a^{th} interval is given by $B_a = [(a-1)k, ak]$, where $(a-1)k$ is the minimum rate of B_a, ak is the maximum rate of B_a, and $a = 1, 2, .., \frac{B}{k}$. Then we evaluated for each time interval t, for each connection n, and for each B_a, the probabilities $p_{tn}^a = P[z_{tn} \in B_a]$, where p_{tn}^a is the probability of connection n requesting at t a number of spectrum slots between $(a-1)k$ and ak. Since the traffic demand distributions are in this work randomly generated and may not be perfectly fitted to the tunability capabilities of the BVTs assumed, we have also evaluated $p_{tn}^0 = P[z_{tn} > B]$ to handle the distributions that generate rates above the feasible rate of the BVTs. By doing so, we managed to generate a valid discrete probability distribution. Without loss of generality, we assume that $B_0 = 0$ with probability p_{tn}^0.

For a network that is already configured and operating at t', a BA model indicates for each connection n the bandwidth allocation action a that must be taken for reconfiguring the network at the next time interval t. If the BA model indicates an action a for the connection n, then the number of spectrum slots Δ_{tn} that must be allocated to connection n are given by $\Delta_{tn} = \max\{B_a\}$. Note that the actions are actually the indices to the B_a intervals, and thus, for simplicity, the same notation is used for both the actions and the indices of the rate intervals. We assume that a network reconfiguration takes place at the beginning of each time interval t and is computed offline during the previous time interval t'. We now proceed with the description of the BA models developed.

1) Highest BA (HBA) Model: Indicates for each connection n and each upcoming time interval t, the BA action a that corresponds to the highest possible bandwidth demand. Specifically, $\Delta_{tn} = \text{argmax}_{a|p_{tn}^a>0}\{\max\{B_a\}|a = 0, 1, .., k\}$.

2) Maximum Probability BA (MPBA) Model: Indicates for each connection n and each upcoming time interval t, the BA action a that corresponds to the bandwidth interval with the maximum probability. Specifically, $\Delta_{tn} = \text{argmax}_a\{p_{tn}^a|a = 0, 1, .., k\}$.

3) Expected BA (EBA) Model: Indicates for each connection n and each upcoming time interval t, the BA action a that corresponds to the bandwidth interval in which the expected bandwidth of the distribution of interest belongs. Specifically, given that the expected bandwidth is $E[\Delta_{tn}] = \sum_{i=0}^{k} \max\{B_i\}p_{tn}^i$, then $\Delta_{tn} = \max\{B_a\}$, where $E[\Delta_{tn}] \in B_a$.

4) Predictive BA (PBA) Model: Our stochastic BA problem is formulated as a Partially Observable Markov Decision Process

(POMDP). POMDPs generalize Markov Decision Processes (MDPs) that are usually used in heuristic search and planning for accommodating stochastic actions and full state observability [15]. POMDPs differ from MDPs in that the states are not observable but are estimated from observations.

Formally, a POMDP is defined as a tuple $\{S, A, T, O, \Omega, b_0, R, \gamma\}$, where S is the set of states. A is the set of actions, $T(s'|s,a)$ defines the distribution over next state s' to which the agent may transition after taking action a from state s, O is the set of observations, $\Omega(o|s,a)$ is a distribution over observations o that may occur as a result of taking action a and entering state s, $R(s,a)$ is the reward function that specifies the immediate reward for taking action a at state s, $\gamma \in [0,1)$ is the discount factor that weighs the importance of current and future rewards, and b_0 is the vector of initial state distribution such that $b_0(s)$ denotes the probability of starting at state s.

In general, at each time step, the environment is at some state $s \in S$. The agent takes an action $a \in A$, and the environment transitions to state s' with probability distribution $T(s'|s,a)$. At the same time, the agent receives an observation $o \in O$ which is associated with the latent (unobservable) state s' according to some conditional likelihood function $\Omega(o|s',a)$. Finally, the agent receives a reward equal to $R(s,a)$. Then the process repeats. The goal is for the agent to choose actions at each time step t that maximize its expected future discounted reward $E[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t)]$.

In our BA problem, let us consider that the correlation between optimal network configuration and traffic demand patterns is not static, but may fluctuate on the grounds of longer term temporal dynamics. In that case, we must be capable of inferring these changes and adapting our policies accordingly. The essence of POMDPs addresses this consideration; POMDPs effect this goal by postulating that, at each time point, the modeled system has some latent state, s. Depending on the latent state, s, the same traffic demand requires a different policy of network reconfiguration, due to the different longer-term trends/dynamics that this latent state information encapsulates.

On this basis, for formulating the POMDP according to our BA problem, S, A, T, O, Ω, b_0 and R are now defined, for each connection n in the network, as follows:

- $S = \{s|s = 0, 1, ..., k\}$ with each state s representing the number of spectrum slots assigned to connection n.
- $A = \{a|a = 0, 1, ..., k\}$ with each action a representing the interval B_a, and hence the number of spectrum slots Δ_* that must be allocated to n.
- $T(s'|s,a)$ defines the probability of transitioning to state s' if action a is taken at s. Note that for each connection n we assume that a spectrum size transition is always possible (for simplicity a network with infinite capacity is assumed - the network capacity limitations are considered during the RSA algorithm in which the trained BA models are incorporated).
- $O = \{o|o = 0, 1, ..., k\}$ with each observation o representing the interval B_o in which the requested (observed) rate belongs.
- $\Omega_n(o|s,a) = p_{tn}^o$ is the observation distribution of connection n. The observation distribution generates at each time step t the true bandwidth demand of n.
- $R(s,a)$ is the reward function that specifies the immediate reward for taking action a at state s, and cannot be known a priori. The immediate reward for each state-action pair depends on what the agent observes at s' after action a is taken at s. On this basis, it is evaluated on the fly during the learning and exploration procedure of the POMDP (see Algorithm 1). For evaluating $R(s,a)$, we define instead a reward function $r(s',a,o)$. Each element of $r(s',a,o)$ specifies the reward received when o is observed at s', after action a is taken at s. Specifically,

$$r(s',a,o) = \begin{cases} -C, & \text{if } a < o \\ \exp[M(k-a+o)], & \text{otherwise} \end{cases} \quad (1)$$

Equation 1 indicates that if the requested demand (o) is higher than the allocated bandwidth (a), then the reward function r returns the constant negative reward $-C$, penalizing the action taken at s. On the other hand, if the requested demand (o) is lower than the allocated bandwidth (a), then a positive reward is received. According to Eq. 1 the positive reward is calculated as $\exp[M(k-o+a)]$, where M is a constant number, and returns a greater reward when the requested bandwidth is closer to the allocated one. Note that the reward is increasing exponentially as the requested bandwidth becomes closer to the allocated one, in order to allow the PBA model to learn the importance of allocating a bandwidth that is near the requested one. Equivalently, the PBA model is guided to avoid allocating at each time interval the highest possible bandwidth in an attempt to ensure a positive reward. By doing so, we aim at reducing both the unserved bandwidth as well as the unutilized allocated bandwidth (PBA is guided to strike a balance between the unserved bandwidth and the allocated one). Note that b_0 is set to $b_0(s) = \frac{1}{k}$ $\forall s$ indicating that connection n can be initialized at any possible state s.

Commonly, POMDPs are solved by formulating them as completely observable MDPs over the *belief states* (posterior probability) of the agent [16]. Specifically, in POMDPs, as the true state is not observable, the agent must choose its actions based only on past actions and observations. Normally, the best action to take at time step t depends on the entire history of actions and observations that the agent has taken so far. However, the probability distribution over current states, known as the belief, is a sufficient statistic for a history of actions and observations [13]. In discrete state spaces, the belief state at step $t+1$ can be computed from the previous belief, b_t, the last action a, and observation o, by the following application of Bayes rule [13]

$$b_{t+1}^{a,o}(s) = \Omega(o|s,a) \sum_{s' \in S} T(s|s',a) b_t(s') / Pr(o|b,a), \quad (2)$$

where $Pr(o|b,a) = \sum_{s' \in S} \Omega(o|s',a) \sum_{s \in S} T(s'|s,a) b_t(s)$. The Bellman equation for the resulting belief MDP is [13]:

$$V_t^*(b) = \max_{a \in A} Q_t(b,a), \quad (3)$$

$$Q_t(b,a) = R(b,a) + \gamma \sum_{o \in O} Pr(o|b,a) V_t(b^{a,o}), \qquad (4)$$

where the value function $V(b)$ is the expected discounted reward that an agent will receive if its current belief is b, $Q(b,a)$ is the value of taking action a at belief b, and $R(b,a)$ is the expected reward given by $\sum_{s \in S} R(s,a)b(s)$. As the exact solution of the Bellman equation (Eq. (3)) is intractable for large spaces [17], in this work, the Real-Time Dynamic Programming-Bel (RTDP-Bel) [18] heuristic algorithm is used for finding an optimal policy. In RTDP-Bel a greedy policy π_V is used for finding an optimal policy, where $\pi_V(b) = \text{argmax}_{a \in A} Q_t(b,a)$.

The RTDP-Bel is an asynchronous value iteration algorithm that converges to the optimal value function and policy over the relevant belief states without having to consider all the belief states in the problem. For achieving this, the RTDP-Bel uses an admissible heuristic function or lower bound h as the initial value function. Provided with such a lower bound, RTDP-Bel selects for update the belief over the states that are reachable from the initial state b_0 through the greedy policy π_V in a way that interleaves simulation and updates. For the implementation of the RTDP-Bel, the estimates $V(b)$ are stored in a hash table that initially contains only the heuristic value of the initial state, b_0. Then, when the value of a belief $b^{a,o}$ that is not in the table is needed, a new entry for $b^{a,o}$ with value $V(b^{a,o}) = h(b^{a,o})$ is allocated. These entries are updated following Eq. (3) when a move from s is performed. The RTDP-Bel algorithm is described analytically in [18].

In this work, the state-of-the-art RTDP-Bel algorithm is slightly modified to fit our problem formulation, incorporating the reward function defined in Eq. 1. The modified RTDP-Bel algorithm is described in Algorithm 1. Algorithm 1 is independently executed for each connection n in the network, and hence for each connection a different PBA model is evaluated. In Algorithm 1, an *episode* is defined as the sequence of actions and observations received for all the time intervals $\{t\}_{t=0}^{24}$. According to Algorithm 1, in each time interval t a single observation is sampled from $\Omega_n(o|s,a)$. It is true, however, that within t a number of traffic demand fluctuations may occur. The algorithm will eventually obtain enough observations and will converge to an optimal PBA through the iteration over a large number of episodes. In Algorithm 1 the target belief is at $t = 24$.

III. Routing and Spectrum Allocation

The RSA heuristic is executed for each time interval $\{t\}_{t=1}^{24}$ and for each connection $\{n\}_{n=1}^{N}$, during the previous time interval t'. Network reconfiguration takes place at the beginning of each time interval t. For each t, the RSA is solved without considering the network configuration at t' (complete connection reallocation is allowed). Specifically, for each t, the RSA finds a route and a spectrum allocation for each connection n, starting with the connection, n', requesting the maximum number of slots $\Delta_{tn'}$. For the R problem, the k-shortest path algorithm is used [19], while for the SA problem the first-fit algorithm is used, subject to the spectrum continuity, spectrum contiguity, and no frequency overlap constraints [3]. An ILP formulation was also developed for BA model evaluation, demonstrating that the proposed PBA model outperforms the benchmark BA models (omitted due to space limitations).

Algorithm 1 Modified RTDP-Bel alg. for each connection n

```
**Start** with $b = b_0$.
**Sample** state $s$ from its probability distribution $b(s)$.
**Evaluate** each action $a$ at belief state $b$ as:
    $Q(b,a) = R(b,a) + \gamma \sum_{o \in O} Pr(o|b,a) V(b^{a,o})$,
    initializing $V(b^{a,o})$ to $h(b^{a,o})$ if $b^{a,o}$ is not in the hash.
**Select** action $a$ that maximizes $Q(b,a)$.
**Update** $V(b)$ to $Q(b,a)$.
**Sample** next state $s'$ from its probability distribution $T(s'|s,a)$.
**Sample** observation $o$ from its probability distribution $\Omega_n(o|s',a)$
**Sample** reward $r$ from the reward function $r(s',a,o)$
**Set** $R(s,a)$ equal to $r(s',a,o)$.
**Compute** $b^{a,o}$ using (2).
**Finish** if $b^{a,o}$ is target belief, else $b := b^{a,o}$, $s := s'$, and go to 3.
```

IV. Performance Evaluation

The performance of the BA models was evaluated and compared on the generic Deutsche Telekom (DT) network [4]. Each spectral slot in the network was set at 12.5GHz, with each fiber link utilizing $B' = 180$ slots. The feasible range of the BVTs was set to $B = 100$ slots. Note that this link capacity was chosen for reducing the computational time in our MATLAB machine with a CPU @2.60GHz and 8GB RAM. Bandwidth B was divided into $k = 10$ rate intervals $\{B_a\}_{a=0}^{k}$. Hence, each BA model can choose at each t and for each n amongst 11 spectrum allocation actions. Each action a indicates that $\Delta_{tn} = a \times k$ spectrum slots must be allocated at time interval t for connection n. Twenty-four time intervals were assumed.

In total 14 connection were considered, with seven of the connections following the log-normal distribution and the rest set to be static. The static connections were added as a simple approach for bringing the network at its capacity crunch and enabling the performance evaluation of the BA models on such a network. Regarding the stochastic connections, their traffic demand parameters, for each connection n and time interval t, are given by the $(\mu_{tn}, \sigma_{tn}^2)$ parameters of the log-normal distribution. The σ^2 parameters were uniformly generated in the range $[0,1]$ and the μ parameters were uniformly generated in the range $[0,5]$. Note that for simplicity, and without loss of generality, we did not consider that the mean rate value (μ) between sequential (in time) traffic distributions increases/decreases smoothly. Such a consideration would not affect the learning procedure or the efficiency of the PBA model. Regarding the static connections, their bandwidth demand Δ_* was set to be constant for all the time intervals. Δ_* values were randomly generated in the range $[20,60]$.

A. Training the PBA Model

For training the PBA model, the discount factor γ was set to 0.95 (typical value for POMDP training). Constants C and

M of the reward function (Eq. 1) were set to 10000 and 10, respectively. Note that a complete examination of how γ, C, and M values affect the trained PBA model could not be performed in this paper due to space limitations, and it is left for future work. A unique PBA model was trained for each one of the seven stochastic connections. For each PBA model, RTDB-Bel was iterated over 6000 episodes of learning, which interleaved simulation and model updates (the model was updated after every 20 simulated episodes). After each model update, 200 test episodes were generated with the model fixed, for evaluating the model's efficiency. For each test episode, the model returned the total reward, the total allocated bandwidth, and the total number of negative rewards received. These values were averaged over all 200 episodes.

Figures 1-3 illustrate how the average reward, the average allocated bandwidth, and the average number of negative rewards evolve over the training time of the PBA model. Training time is given in hours and corresponds to the time required for training and testing the model (for the 6000 episodes). A model update is indicated with a circle in Figs. 1-3 (250 total model updates). Figures 1-3 correspond to the PBA model of connection $n = 1$ (similar figures were obtained for all the other connections but are omitted due to space limitations). Figure 1 shows that the PBA model performs better as the training procedure evolves. The average reward increases with the number of model updates (training time) as the agent learns to take better bandwidth allocation decisions. Fewer negative rewards are received (Fig. 3) and the allocated bandwidth converges near the requested one (Fig. 2).

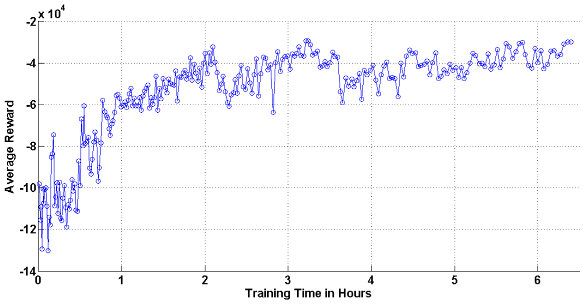


Fig. 1: Average reward over training time.

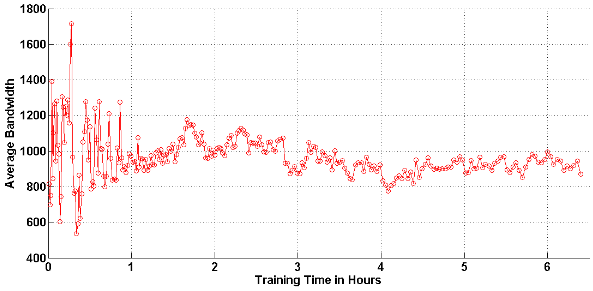


Fig. 2: Average allocated bandwidth over training time.

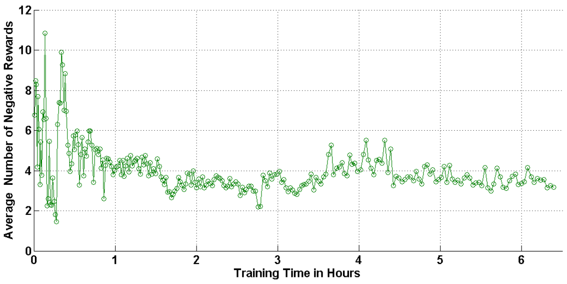


Fig. 3: Average negative rewards over training time.

In our simulations, each connection was trained for the same number of episodes and the last PBA model obtained was utilized for the network reconfigurations (during the RSA heuristic). Each model required up to 6 hours of training and testing. An action was generated within milliseconds from each model. Note that the models for each connection can be trained in parallel and independently from each other, and thus the number of time-varying connections does not affect the scalability of the PBA model. Further, the training procedure can be continuously performed for automatically adjusting the models upon significant variations on the traffic demand distributions; an important capability of the proposed method, given that the future traffic demand is expected to increase in uncertain ways (we cannot know the magnitude of a future traffic demand or the sources of this traffic).

Table I demonstrates how each trained PBA model performs against HBA, MPBA, and EBA. For each BA model we generated 200 episodes of actions and observations assuming a network with infinite capacity. The allocated bandwidth and the number of times an observation was greater than the action taken (negative reward) were averaged over these episodes. Note that a single observation was drawn for each action taken. Table I shows both the average allocated bandwidth and the average number of negative rewards.

TABLE I: BA Model Comparison

	Average Allocated Bandwidth				Average No. of Negative Rewards			
n	HBA	MPBA	EBA	PBA	HBA	MPBA	EBA	PBA
1	1490	560	690	868	0	4.77	3.3	3.1
2	1730	500	470	1171	0	7.2	4.2	2.1
3	1310	370	480	950	0	3.48	2.5	0.9
4	1400	370	580	750	0	6	3.3	4.5
5	1690	350	488	665	0	6.13	4	3.5
6	1420	320	520	830	0	6.7	4.3	4.1
7	1700	380	480	667	0	6.1	4.1	3.4

According to Table I, PBA tends to allocate fewer slots compared to HPBA and more slots compared to MPBA and EBA. Hence, PBA increases the negative rewards received compared to HBA that never receives a negative reward. MPBA and EBA receive on the average more negative rewards than PBA as they tend to allocate fewer slots than PBA. This is a consequence of the reward function (Eq. 1) defined for PBA training that aims at allocating at each time interval a bandwidth that is close to the requested one.

B. Network Performance Evaluation

The RSA algorithm was solved on the DT network for each BA model and each time interval t. For each t, an action was generated for each connection n and RSA was solved having as inputs the rates Δ_{tn} indicated by the model's actions. RSA required at most 15 seconds for finding a feasible solution for each time interval. Between network reconfigurations the traffic demand fluctuated according to the given set of traffic demand distributions. For the traffic demand fluctuations we have drawn from each Z_{tn} the samples $\{z_{tn}^i\}_{i=1}^{60}$ representing the traffic demand fluctuations every minute of the hour. Sample $\delta_{tn}^i = z_{tn}^i$ denotes that connection n requests δ_{tn}^i spectrum slots at the i^{th} minute of time interval t.

For each established connection, the allocated Δ_{tn} slots were compared to each δ_{tn}^i in order to calculate the unserved slots and the excess (unutilized) allocated slots. The unserved slots for each episode are given by $U = \frac{1}{60\times24}\sum_t\sum_n\sum_i|\Delta_{tn} - \delta_{tn}^i|$, if $\Delta_{tn} < \delta_{tn}^i$. The excess slots for each episode are given by $E = \frac{1}{60\times24}\sum_t\sum_n\sum_i(\Delta_{st} - \delta_{tn}^i)$, if $\Delta_{tn} > \delta_{tn}^i$. Two-hundred episodes were generated for each BA model and the unserved and excess slots were averaged over these episodes. Table II shows the average number of unserved ($\bar{U}$) and excess ($\bar{E}$) slots per time interval. It also shows the average number of blocked connections ($\bar{\Pi}$) per episode.

TABLE II: BA Model Comparison on DT Network

	HBA	MPBA	EBA	PBA
Av.# of Excess Slots ($\bar{E}$)	337	35.3	82	155
Av.# of Unserved Slots ($\bar{U}$)	23	49	48	20.3
Av.# of Blocked Connections ($\bar{\Pi}$)	16	0	0	0

According to Table II, as expected, HBA allocates on the average a higher number of excess slots (337) compared to the other models. The high number of excess slots led, on the average, to 16 blocked connections (these connections are entirely terminated, each for an hour during a day). HBA is clearly not a feasible solution for a network operating at its capacity crunch. If we assume that the end user behavior remains the same during the unavailability period, the 16 blocked connection lead to 23 unserved slots (greatly unbalanced between the connections). Under this consideration, PBA outperforms HBA by 11%.

Table II shows that MPBA, EBA, and PBA significantly reduce the average excess slots by 80%, 75%, and 54%, respectively, compared to HBA. Consequently, these models, unlike HBA, did not cause any blocking. However, the traffic demand fluctuations within each time interval resulted in some unserved slots. In particular, PBA results on the average in 20.3 unserved slots, while MPBA and EBA, result on the average in 49 and 48 unserved slots, respectively. Hence, PBA outperforms MPBA and EBA, in terms of unserved slots, by approximately 58%. Overall, PBA predicts a bandwidth that more efficiently handles traffic demand fluctuations.

V. Conclusion

We proposed an effective formulation of a state-of-the-art POMDP method that learns by means of DP an optimal predictive BA model from a given set of traffic demand distributions that is consequently used for bandwidth allocation decisions during network reconfigurations. PBA is compared to the naturally arising HBA, MPBA, and EBA techniques and it is shown that it outperform HBA on the number of blocked connections, as well as MPBA and EBA on the unserved bandwidth that may occur during traffic demand fluctuations.

Acknowledgment

This work has been supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No 739551 (KIOS CoE) and from the Government of the Republic of Cyprus through the Directorate General for European Programmes, Coordination and Development. This work has also received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under REA Grant Agreement no. 630853.

References

[1] Cisco white paper, "The Zettabyte Era: Trends and Analysis," 2017.

[2] O. Gerstel, et al., "Elastic Optical Networking: A New Dawn for the Optical Layer?," *IEEE Comm. Mag.*, 50(2):s12–s20, 2012.

[3] K. Christodoulopoulos, et al., "Elastic Bandwidth Allocation in Flexible OFDM-Based Optical Networks," *IEEE/OSA J. Lightwv. Techn.* 29(9):1354–1366, 2011.

[4] K. Christodoulopoulos, et al., "Time-Varying Spectrum Allocation Policies and Blocking Analysis in Flexible Optical Networks," *IEEE J. on Selected Areas in Comm.*, 31(1):13–25, 2013.

[5] M. Klinkowski, et al., "Elastic Spectrum Allocation for Time-Varying Traffic in FlexGrid Optical Networks," *IEEE J. on Selected Areas in Comm.*, 31(1):26–38, 2013.

[6] S. Shakya, et al., "Spectrum Allocation for Time-varying Traffic in Elastic Optical Networks using Traffic Pattern," *Proc. OFC*, 2014.

[7] G. Shen, et al., "Maximizing Time-dependent Spectrum Sharing between Neighbouring Channels in CO-OFDM Optical Networks," *Proc. ICTON*, 2011.

[8] B. C. Chatterjee, et al., "Routing and Spectrum Allocation in Elastic Optical Networks: A Tutorial," *IEEE Comm. Surveys & Tutorials*, 17(3):1776–1800, 2015.

[9] K. Christodoulopoulos, et al., "Dynamic Bandwidth Allocation in Flexible OFDM-based Networks," *Proc. OFC*, 2011.

[10] F. Cugini, et al., "Push-Pull Defragmentation Without Traffic Disruption in Flexible Grid Optical Networks," *IEEE/OSA J. Lightwv. Techn.*, 31(1):125–133, 2013.

[11] I. Antoniou, et al., "On the Log-normal Distribution of Network Traffic," *Physica D: Nonlinear Phenomena*, 167(1–2):72–85, 2002.

[12] L.P. Kaelbling, et al., "Planning and Acting in Partially Observable Stochastic Domains," *Art. Intell. Journal*, 101(1–2):99–134, 1998.

[13] F.D.-Velez, "The Infinite Partially Observable Markov Decision Process," *Advances in Neural Information Proc. Systems 22*, 2009.

[14] S.P. Chatzis, D. Kosmopoulos, "A Non-stationary Infinite Partially-Observable Markov Decision Process," *Proc. ICANN*, 2014.

[15] D. Bertsekas, "Dynamic Programming and Optimal Control, (2Vols)". *Athena Scient.*, 1995.

[16] E. Sondik, "The Optimal Control of Partially Observable Markov Decision Processes over the Infinite Horizon: Discounted Costs", *Oper. Res.*, 26(2):282–304, 1978.

[17] C. Papadimitriou and J. Tsisiklis, "The Complexity of Markov Decision Processes," *Math. of Operations Research*, 12(3):441–450, 1987.

[18] B. Bonet, H. Geffner, "Solving POMDPs: RTDP-Bel vs. Point-based Algorithms," *Proc. IJCAI*, 2009.

[19] J.Y. Yen, "Finding the k shortest loopless paths in a network," *Management Science*, 17(11):712-716, 1971.

Provisioning of 5G Services Employing Machine Learning Techniques

Antonia Pelekanou(1), Markos Anastasopoulos(2), Anna Tzanakaki(1)(2), Dimitra Simeonidou (2)

(1) National Kapodistrian University of Athens, GR, (2) HPN Group, University of Bristol, UK

***Abstract* - This study proposes a modeling framework for optimal online 5G service provisioning, based on low computational complexity machine learning techniques such as Neural Network (NNs). NNs are trained to take optimal decisions adopting an offline Integer Liner Programming (ILP) model. This framework is used to solve the generic joint Fronthaul (FH) and Backhaul (BH) service provisioning problem over a converged high capacity and flexibility optical transport aiming at minimizing the overall energy consumption of the 5G infrastructure. Our modeling results indicate that the proposed approach adopting NN based real time service provisioning can provide very similar performance to the one derived adopting the high complexity but accurate ILP approach.**

Index Terms - Machine Learning, Optimization, 5G, ILP, LSTM, MLP, optical transport.

I. INTRODUCTION

As the demand for high-speed mobile internet access connectivity increases at a rapid pace, Radio Access Networks (RANs) deployments need to be transformed into open, scalable and dynamic ecosystems able to support a large variety of demanding applications and services in a flexible and efficient manner. The fifth generation (5G) mobile networks address this need through a set of hardware and software technology innovations targeting both the data and the control plane. Suitable solutions include the adoption of new centralized control and management frameworks based on Network Functions Virtualization (NFV)/Software Defined Networking (SDN) principles. In addition, new architectural models allow to migrate from highly distributed and inefficient structures to more centralized approaches relying on concepts such as the Cloud-RAN (C-RAN) approach. C-RAN and its more recent variant including the notion of dynamic functional splits [1] introduce the need for fronthaul (FH) services interconnecting remote units (RUs) with processing units to allow centralization and ultimately softwarization of the RAN. Through pooling and coordination gains of softwarized/centralized RANs, significant cost reduction as well increased scalability and flexibility over current RAN solutions can be achieved.

To successfully deploy the concept of softwarized RAN together with the increased backhaul (BH) requirements imposed by the current and upcoming 5G services at the data plane, there is a need for a high capacity transport network interconnecting the remote antennas with the compute resources where softwarized versions of the RAN protocol stack are executed. This will be enabled by a control plane solution able to manage and optimize the operation of a large number of highly heterogeneous network and compute elements, taking decisions related to: *i) optimal embedding* of service requests and creation of service chains over the converged network resources [2], [3], *ii) optimal infrastructure slicing* across heterogeneous network domains [4], *iii) optimal sharing* of common resources in support of Information and Communication Technology (ICT) and vertical industry services [5], *iv) optimal fronthaul* deployment strategies including optimal placing of central units with respect to remote units, functional split selection etc. [6], [7].

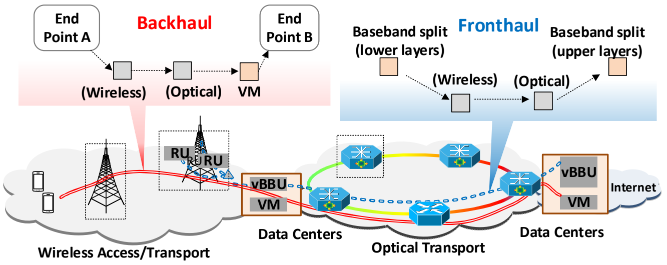


Figure 1: Provisioning of FH and BH services over a common 5G network infrastructure.

These problems are traditionally solved by a centralized controller considering in many cases multiple objectives and constraints (ranging from Capital and Operational Expenditure minimization, energy consumption, latency, resource availability etc.), adopting a variety of mathematical modeling frameworks based on integer linear [8] and non-linear [9] programming, stochastic linear and nonlinear programming formulations [10] etc. Although these schemes can be effectively used to identify the optimal operational points of the whole system, their increased computational complexity and slow convergence time makes them unsuitable for real time network deployments. To cope with the increasing computational complexity inherent in these models, alternative modular optimization schemes have been proposed. These aim at decomposing large optimization problems into smaller and easier subproblems that are able to handle a large number of variables.

Towards this direction, this study proposes a modular framework to enable optimal 5G service provisioning taking advantage of the optimal decisions taken through offline tools based on Integer Linear Programming (ILP) and less computationally intensive online tools based on Neural Networks (NNs). More specifically,

978-3-903176-07-2 © 2018 IFIP

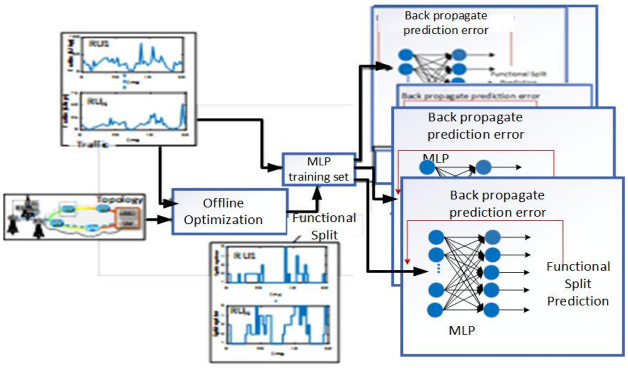


Figure 3: Construction of the training set that will be used for the design of the NN-based 5G optimization framework. The offline optimization block represents the ILP model, the solution of which gives the optimal baseband split for each RU. The optimal baseband splits as they have been obtained by the ILP and the traffic statistics as they collected by the RUs constitute the training set of the MLP-NN models. MLP-model learns to map each value of traffic statistics to the value of the optimal functional split through the backpropagation algorithm for each RU.

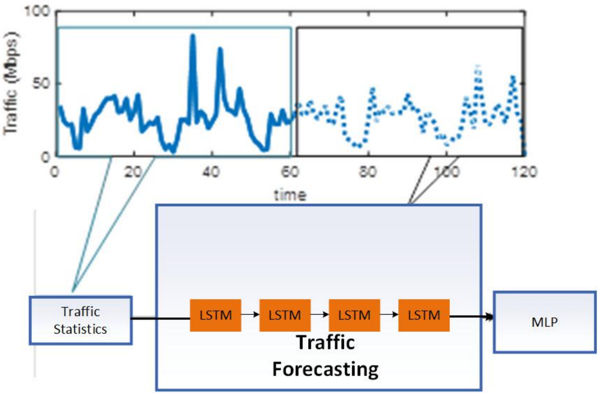


Figure 2: NN model-based, LSTM and MLP, for the optimization of the 5G network in the upcoming time instants. The input of the LSTM model is the traffic at time step t and the output is the traffic at next time step t+1. The value of traffic as derived from the LSTM is given as input to the MLP which can predict the optimal functional split in upcoming time instant.

the offline ILP model is first used to create a set containing the optimal design policies for converged 5G network environments. NNs then use the output of the ILP as a training set. Once NNs have been trained, they can be used by the centralized controller for real time optimal decision making. To demonstrate the efficiency of the proposed technique, in the present study we consider the generic joint FH and BH service provisioning problem over a converged high capacity and flexibility optical transport network environment [4]. In this converged network part of the optimization problem is associated with identifying the optimal fronthaul service. This is directly related with the identification of the optimal split option adopted [4]. To solve the problem of optimizing the fronthaul services, a NN model is proposed to identify, in real time, the optimal functional split for each RU. Although NNs have been widely adopted to model various problems with remarkable performance in telecommunication networks (see [13]-[18]), to the best of the authors knowledge this is the first time where ILP and NNs are appropriately combined for the design of converged 5G Networks.

The rest of the paper is organized as follows. Section II provides a brief description of the problem under investigation, emphasizing on the offline ILP scheme and the proposed two-stage NN model. Optimal design strategies of the proposed NN model are provided in Section III whereas performance evaluation under a realistic network configuration is carried out in Section IV. Finally, Section V concludes the paper.

II. PROBLEM DESCRIPTION

This paper focuses on the generic case of a converged 5G infrastructure interconnecting a set $\mathcal{R}$ of R remote units (RUs) with a set $\mathcal{S}$ of S Central Units (CUs). This infrastructure, integrates wireless access and optical transport networks together with compute elements and is used to support FH and BH services. As already discussed for the FH services, we consider the dynamic functional split option approach, where the baseband signal processing tasks of the antennas can be divided, allocating some functions at the RUs and the remaining ones at the CUs. As discussed in [4], the decision to execute these functions locally or remotely depends on various parameters including the network topology, the availability of resources, the service characteristics etc. For the BH services, we consider content delivery type of services, where mobile devices offload their compute intensive tasks to the cloud [4]. For this type of services, specific compute and network resources need to be also reserved across the converged 5G infrastructure. A graphical representation of this concept is shown in Figure 1.

To date, the problem of joint FH/BH service provisioning over a common infrastructure has been formulated and solved based on Integer Linear Programming (ILP) [19]. Specifically, assuming that $\mathcal{D}_F, \mathcal{D}_B$ is the set of FH and BH demands, respectively, $\mathcal{E}$ is the set of links, κ_e, the cost per link e in the infrastructure with $e \in \mathcal{E}$, κ_s the remote processing cost at $s \in \mathcal{S}$, κ_r the processing cost at the RU $r \in \mathcal{R}$ measured in Giga Operations per Second- GOPS, $u_{FH,e}, u_{BH,e}$ is the link e capacity allocated for FH and BH services, $\pi_{FH,s}$, $\pi_{BH,s}$ is the processing capacity for the FH/BH services and $\mathcal{U}_e$ is the capacity of link e, then, the joint FH/BH optimization problem can be formulated as follows:

$$\min \mathcal{F}\,(\boldsymbol{u},\boldsymbol{\pi}) = [\mathcal{FH}(\boldsymbol{u},\boldsymbol{\pi}),\quad \mathcal{BH}(\boldsymbol{u},\boldsymbol{\pi})] \quad (1)$$

where

$$\mathcal{FH}(\boldsymbol{u},\boldsymbol{\pi}) = \sum\nolimits_{e\epsilon\mathcal{E}} \kappa_e\, u_{FH,e} + \sum\nolimits_{s\epsilon\mathcal{S}} \kappa_s \pi_{FH,s} + \sum\nolimits_{d\epsilon\mathcal{D}_F} \kappa_d\, \pi_{FH,d} \qquad (2a)$$

$$\mathcal{BH}(\boldsymbol{u},\boldsymbol{\pi}) = \sum\nolimits_{e\epsilon\mathcal{E}} \left[\mathcal{U}_e - u_{FH,e} - u_{BH,e}\right]^{-1} + \sum\nolimits_{s\epsilon\mathcal{S}} \left[\mathcal{C}_s - \pi_{FH,s} - \pi_{BH,s}\right]^{-1} (2b)$$

subject to capacity, functional split and demand constraints. $(2a)$ minimizes the expected cost for the FH services while $(2b)$ the

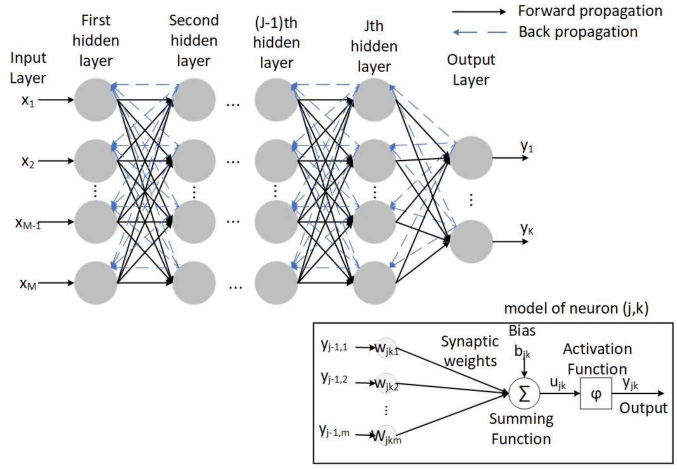


Figure 4: Graphic illustration of an MLP NN structure with backpropagation algorithm.

associated costs for the BH services. A description of the ILP-based modeling approach together with the relevant implementation details is given in [19] . Using as inputs network topology details and traffic statistics, the location where each function/task will be processed together with the required network and compute resources can be determined. Although the above-mentioned optimization framework can effectively identify the optimal operational point of the whole system, its increased computational complexity and its slow convergence time makes it impractical to optimize the operation for real time network deployments. To address this limitation, a two-step Neural Network-based optimization framework is proposed. This framework allows real time identification of the optimal operational strategies per RU. In the first step, using a specific set of training data, a novel Multilayer Perceptron (MLP) - based NN model is constructed that in-real time can identify the optimal operational policies for the whole 5G infrastructure. A high-level view of this process is shown in Figure 3 for a specific case where the MLP-NN is used to identify the optimal split per RU. To achieve this, a training set combining data from history traffic statistics as well as data extracted from the offline - optimization framework described above is considered. An algorithmic approach that allows the identification of optimal MLP-NN architecture is provided in the following section.

Once the model has been trained, the MLP-NN model is combined with a trained Long Short-Term Memory (LSTM) NN model used for traffic forecasting. This aims at identifying the optimal operating conditions for the 5G infrastructure in the upcoming time periods. The flowchart of this process is provided in Figure 2.

III. REAL TIME OPTIMIZATION FOR 5G

A. Artificial Neural Network Preliminaries

Artificial NNs are defined as systems of interconnected computational units, known as neurons, that interact with the environment. Each neuron has a non-linear, differentiable function, known as activation function, used to compute a weighted sum of the outputs of the previous-layer. In NNs, knowledge is stored in interneuron connection strengths, known as synaptic weights using a learning algorithm. The learning algorithm is a function that updates the value of synaptic weights during the learning operation. The Backpropagation algorithm is the most popular learning algorithm for training NNs and comprises two phases, the forward phase and the backward phase. Through the first phase, the signal is transmitted from the input to the output on a layer by layer basis, keeping the synaptic weights' unaltered. In the second phase, the comparison between the network's output and the desired response leads to an error signal. The error signal is propagated backwards through the network, starting from the output, and then the synaptic weights are re-evaluated to minimize the loss function. The loss function is a function that calculates the divergence between predicted and expected network's response values [12]. Figure 4 shows a typical MLP neural network with J hidden layers over which the backpropagation algorithm is applied.

Table 1: Overview of the Backpropagation Algorithm applied to the MLP-NN

Parameters:
M = dimensionality of the input space and number of neurons in hidden layers, *m = 1,2, …, M*
J = number of hidden layers, j = 1,2, …, J
K = number of neurons in output layer, k = 1,2, …, K
N = number of epochs, n = 1,2, …, N
$\boldsymbol{w_m}$=*synaptic weight vector of neuron m*
d = desired response vector
y = neuron output
$\boldsymbol{\delta_k}$ =*local gradient at neuron k*
η = learning rate
$\boldsymbol{e_k}$ = *error*

Initializations:
Set the synaptic weights of the algorithm to small values selected from uniform distribution.

Computations:

- **If neuron *k* is an output neuron then***:*

(3.1) $\boldsymbol{u_k(n) \ = \sum_{m=1}^{M} w_{km}(n) \ \cdot y_{Jm}(n) + b_k}$

(3.2) $\boldsymbol{y_k(n) \ = \varphi(u_k(n)\ )}$

(3.3) $\boldsymbol{e_k(n) = d_k(n) - y_k(n)}$

(3.4) $\boldsymbol{E(n) = \sum_{k\epsilon C} e_k{}^2(n)/2}$

(3.5) $\boldsymbol{\Delta w_{ki}(n) = -\eta \frac{\partial E(n)}{\partial w_{ki}(n)}}$

(3.6) $\boldsymbol{\frac{\partial E(n)}{\partial w_{ki}(n)} \ = \ -e_k(n)\varphi'{}_k\big(u_k(n)\big)\cdot y_{ji}(n)}$

(3.7) $\boldsymbol{\delta_k(n) = e_k(n)\varphi'{}_k\big(u_k(n)\big)}$

(3.8) $\boldsymbol{\Delta w_{ki}(n) = -\eta \ \delta_k(n)\cdot y_{ji}(n)}$

(3.9) $\boldsymbol{w'_{ki}(n) = w_{ki}(n) + \ \Delta w_{ki}(n)}$

else if it is a hidden neuron at layer j:

(3.10) $\boldsymbol{u_{jm}(n) \ = \sum_{m=1}^{M} w_{jm}(n) \ \cdot y_{j-1,m}(n) \ + b_{jm}}$

(3.11) $\boldsymbol{y_{jm}(n) \ = \varphi\big(u_{jm}(n)\ \big)}$

(3.12) $\boldsymbol{e_k(n) = d_k(n) - y_k(n)}$

(3.13) $\boldsymbol{E(n) = \sum_{k\epsilon C} e_k{}^2(n)\ /2}$

(3.14) $\boldsymbol{\Delta w_{ji}(n) = -\eta \frac{\partial E(n)}{\partial w_{jm}(n)}}$

(3.15) $\boldsymbol{\delta_{jm}(n) = \varphi'{}_{jm}\Big(u_{jm}(n)\Big)\cdot \sum_{k=1}^{K} \delta_{jk}(n)\cdot w_{jk}(n)}$

(3.16) $\boldsymbol{\Delta w_{jm}(n) = -\eta \ \delta_{jm}(n) y_{j-1,m}(n)}$

(3.17) $\boldsymbol{w'_{jm}(n) = w_{jm}(n) + \ \Delta w_{jm}(n)}$

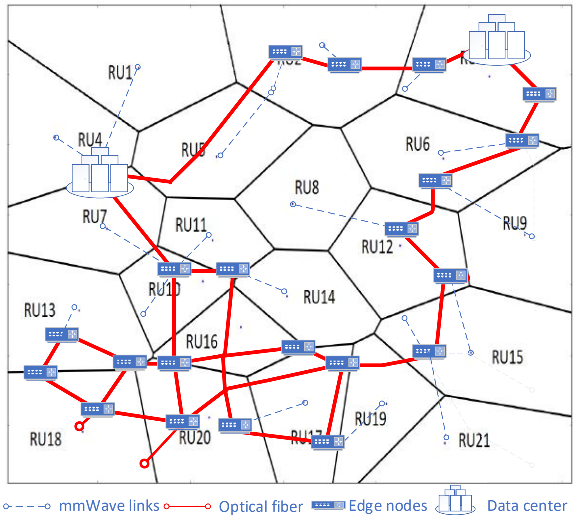


Figure 5: 5G network topology under investigation

The modeling details of the backpropagation algorithm for the MLP network are summarized in Table 1. Specifically, if neuron k is an output neuron, then the linear combiner output $u_k(n)$ is calculated by the weighted sum of the inputs· $y_{Jm}(n)$ with the respective synaptic weights $w_{km}(n)$ using equation (3.1) (see lower part of Figure 4). $u_k(n)$ is then applied to an activation function φ, which limits the amplitude of the output of neuron k (3.2) resulting to the final output of neuron k at the n iteration, namely $y_k(n)$**.** The estimation error at the output of neuron k is calculated through (3.3), while the total instantaneous error $E(n)$ of the whole network is calculated using (3.4). The error is propagated backward and the correction Δw_{ki} is applied to the synaptic weight w_{ki} (3.5) - (3.9). A similar set of equations is applied for the hidden neurons (3.10) - (3.17).

Our objective is to identify an MLP network that maps any input x to the corresponding output y. Output y is obtained from the solution of the corresponding ILP formulation, while x represents the set of history observations. As an example, consider the scenario for which we apply to the MLP a training set that comprises a set of pairs (x, y), where x represents the traffic statistics for a particular RU at a given point in time, while y represents the functional split. The optimal functional split per RU over time has been obtained through the solution of the ILP model described in Sec. II. This training set is given as input to the MLP neural network in order to learn how to map each input x to the corresponding output y. Once the system has been trained, the MLP can predict the functional split given any new data without solving the corresponding ILP. The parameters of the MLP model can be derived executing the algorithm of Table 1 for different parameters' values (batch size, number of hidden layers etc.). At the end of the experiments, the combination of parameter value is chosen according to their ability to maximize the prediction accuracy.

B. Traffic forecasting using Long Short-Term Memory Neural Networks

Long Short-Term Memory (LSTM) is a special case of Recurrent Neural Network (RNN) capable to learn long-term dependencies, since it can remember information that was acquired in previous steps of the learning process. LSTM contains a set of recurrent blocks, known as memory blocks, each of which has one or more memory cells. Each cell is composed of three basic units, the input, output and forget gate that are responsible to decide whether to forget, keep, update or output information that has been acquired previously. LSTM is the most successful model for predicting long-term time series [1].

In the present study, the LSTMs are optimally designed to forecast the traffic load of each RU based on history traffic data available. The LSTM input vector corresponds to the traffic at an arbitrary time step t while the LSTM output vector corresponds to the traffic at time step t+1. To train the LSTMs, the dataset containing history measurements of each RU is split into two parts, the training set and the test set. The training set is used during the training of the LSTM network, while the test set is used to validate the effectiveness of each LSTM designed. To identify the optimal LSTM architecture for each RU, an extensive set of experimentations is performed. Given that the LSTM architecture can be fully characterized by the number of hidden layers, neurons, epochs and the batch size, our objective is to the identify how these parameters can be optimally combined to minimize the forecasting error. This process is summarized as follows:

Step 1- Batch size. The batch size is the number of training instances used in each iteration. The weights are updated after each batch propagation. We choose the value for the batch size that minimizes forecasting error keeping all other parameters constant.

Step 2- Number of epochs: The number of epochs determine the maximum number of passes over the training dataset. Various values for the number of epochs are tested in order to identify the optimal one that minimizes the forecasting error.

Step 3- Number of neurons. In this step, our objective is to identify the optimal number of neurons that achieves optimal traffic forecasting accuracy.

Step 4 – Number of hidden layers. The last parameter that we study is the number of hidden layers. As before, after extensive experimentations we choose the number of hidden layers that

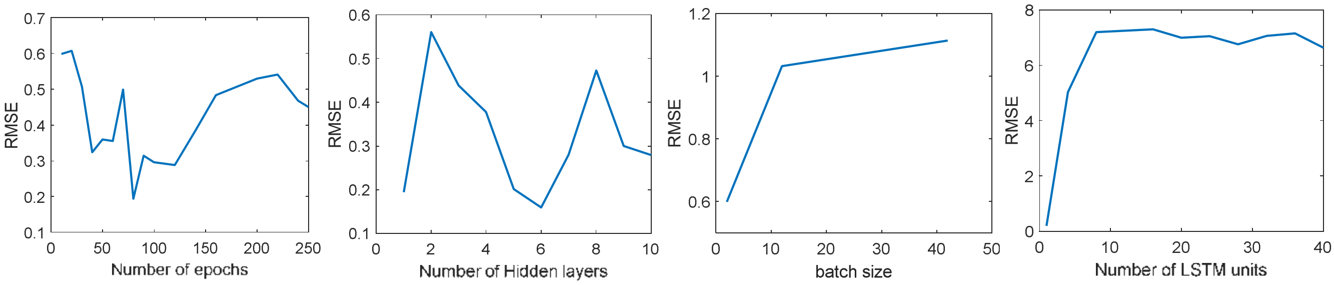


Figure 6: Learning curves of the LSTM for RU16

Table 2:Parameter Settings of Neural Networks for the RU16

	Batch Size	Number of Epochs	Hidden Layers of the Network	Number of Neuron for Output Layer	Activation Function for Hidden Layers	Activation Function for Output Layer
LSTM	2	80	6 hidden layers with 1 neuron each layer	1	Relu	-
MLP	2	600	1 hidden layer with 10 neurons	5	Relu	Softmax

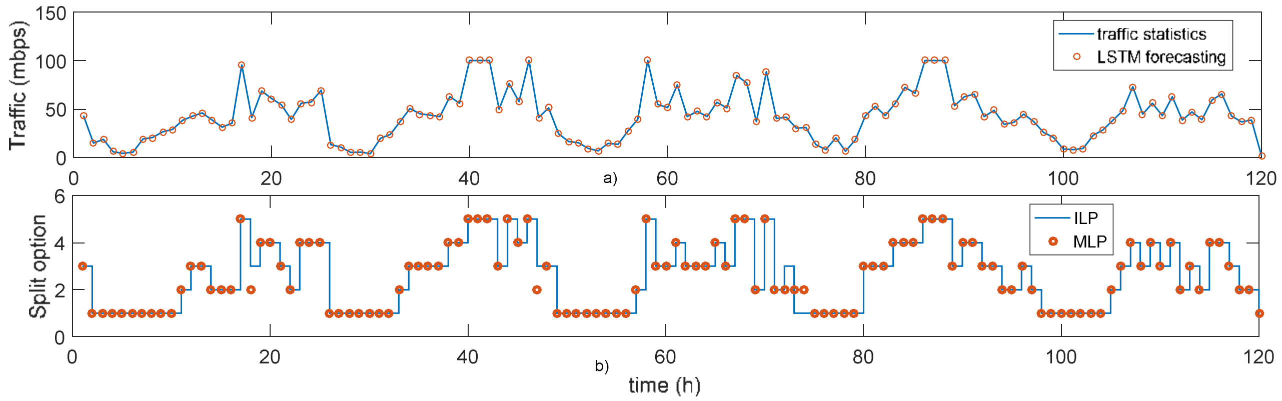


Figure 7: a) Traffic forecasting for RU16 using LSTM. b) Optimal functional split prediction for RU16 using the results obtained from the ILP and the MLP.

minimizes the forecasting error calculated through the root-mean-squared-error formula (RMSE).

IV. Numerical results

A. *Topology description and assumptions*

The validity of the proposed NN-based optimization framework is evaluated using the optical transport network topology presented in [4] over which 21 RUs are deployed. The coverage area of each RU is shown in Figure 5. For this topology, mobile devices served by the corresponding RUs generate demands according to real datasets reported in [11]. Each RU is connected to the optical transport through microwave point-to-point links with 2 Gb/s bandwidth, and 45W power consumption. The optical transport has a single fiber per link, 4 wavelengths of 10 Gb/s each per fiber, and minimum bandwidth granularity of 100 Mb/s [4]. The processing requirements of the mobiles devices and the RUs are supported through a set of DCs. For this network topology, our objective is to design a NN model that approximates the optimal ILP described in Sec. II and solved in [4]. To keep the analysis tractable, the results provided are correspond to the optimal functional split of RU16, however, similar studies have been conducted for all compute/network elements of Figure 5 focusing on other parameters of interest such as, network capacity for optical links, compute capacity for DCs, locations where demands are processed for demands etc.

B. *Neural Network/Learning topology optimization*

To design the two-stage NN model using the LSTM/MLP models, the methodology presented in Sec. III is applied to all network components. For each component, our objective is to design an NN that approximates with very high accuracy the optimal policies obtained through the corresponding ILP model.

To identify the optimal NN models, the learning curves showing the RMSE as a function of the number of epochs, hidden layers, neurons and batch size are first obtained. Based on these curves, the optimal values of the parameters that minimize the corresponding error can be readily determined. A typical set of learning curves for the LSTM model of RU16 is shown in Figure 6 and the corresponding optimal values are provided in Table 2.

C. *Traffic forecasting based on LSTM Neural Networks*

Once the optimal LSTM NN structure has been determined, the model is trained using the history dataset and the corresponding synaptic weights are determined. The test set is applied to the LSTM model to evaluate its forecasting performance. A snapshot showcasing the performance accuracy of the LSTM model for RU16 is illustrated in Figure 7 a), where an RMSE of 0.16 is obtained corresponding to a forecasting error in the order of 0.3%.

D. *Prediction of operational parameters: Optimal Functional Split based on MLP Neural Networks*

Following a similar approach to the LSTM problem design, once the MLP network has been defined, the derived model is trained and validated using the training set obtained from the ILP formulation. Figure 7 b) shows the performance of the proposed model where it is observed that the MLP is able to identify the optimal functional split with a 95% accuracy.

E. *Total power consumption*

Finally, the performance of the proposed NN scheme is compared to the ILP based optimization approach presented in [4] in terms of total network power consumption. It is observed in Figure 8 that the power consumption over time for both schemes takes very close values, indicating the effectiveness of the proposed NN scheme to identify the optimal operational

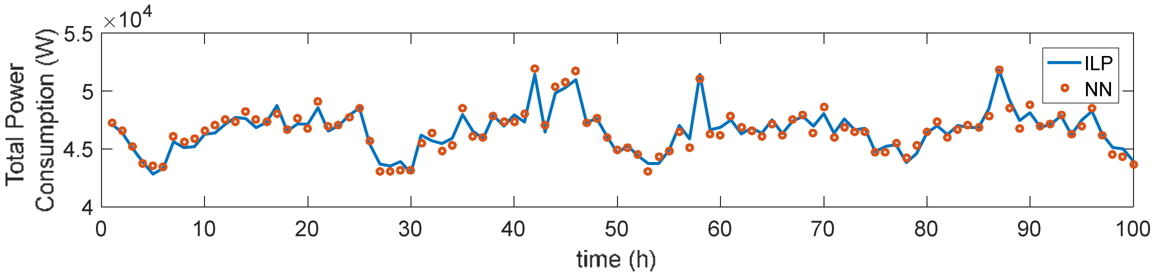


Figure 8: Total Power consumption when applying the ILP and the proposed NN scheme

strategies of every network element. This clearly shows that online optimal service provisioning can be achieved taking a practical low complexity approach adopting machine learning techniques that can be trained to take real time very close to optimal decisions. In this context, the training process plays a key role and can be performed taking advantage of the optimal decisions provided through offline tools based on ILP.

V. CONCLUSIONS

This study proposes a modeling framework to enable optimal online 5G service provisioning based on low computational complexity machine learning techniques such as NNs, exploiting optimal decisions taken through offline tools based on ILP for training purposes. The offline ILP model is first used to create a set containing the optimal design policies for converged 5G network environments. NNs then use the output of the ILP as a training set and after being trained, they can perform real time optimal decisions. To demonstrate the efficiency of the proposed technique, we consider the generic joint FH and BH service provisioning problem over a converged high capacity and flexibility optical transport aiming at minimizing the overall energy consumption of the 5G infrastructure. Our results indicate that the proposed approach adopting NN based real time service provisioning can provide very similar performance to the one derived adopting the high complexity but accurate ILP approach.

ACKNOWLEDGEMENT

This work has been financially supported by the EU Horizon 2020 project 5G-PICTURE under grant agreement No 762057 and the EU Horizon 2020 project IN2DREAMS under grant agreement No 777596.

REFERENCES

[1] Common Public Radio Interface:eCPRI Interface Specification, D01, Aug. 2017

[2] A. Gupta et al., |On service-chaining strategies using Virtual Network Functions in operator networks”, *Computer Networks*, Vo. 133, 14 March 2018,

[3] 5G Vision. The 5G Infrastructure Public Private Partnership: the next generation of communication networks and services. 2015. [Online]

[4] A. Tzanakaki et al., "Wireless-Optical Network Convergence: Enabling the 5G Architecture to Support Operational and End-User Services", IEEE Commun. Mag., pp. 184 – 192, Aug. 2017

[5] M. Anastasopoulos et al., “ICT platforms in support of future railway systems”, in proc. of TRA 2018, Apr. 2018

[6] F. Musumeci, C. Bellanzon, N. Carapellese, M. Tornatore, A. Pattavina, and S. Gosselin, "Optimal BBU Placement for 5G C-RAN Deployment Over WDM Aggregation Networks," J. Lightwave Technol. **34**, 1963-1970 (2016).

[7] T. Pfeiffer, "Next generation mobile fronthaul and midhaul architectures [Invited]," in *IEEE/OSA Journal of Optical Communications and Networking*, vol. 7, no. 11, pp. B38-B45, November 1 2015.

[8] M. Fiorani, S. Tombaz, J. Martensson, B. Skubic, L. Wosinska and P. Monti, "Modeling energy performance of C-RAN with optical transport in 5G network scenarios," in *IEEE/OSA Journal of Optical Communications and Networking*, vol. 8, no. 11, pp. B21-B34, Nov. 2016.

[9] A. Tzanakaki *et al*., "5G infrastructures supporting end-user and operational services: The 5G-XHaul architectural perspective," *2016 IEEE International Conference on Communications Workshops (ICC)*, Kuala Lumpur, 2016, pp. 57-62.

[10] M. Anastasopoulos, A. Tzanakaki and D. Simeonidou, "Stochastic Energy Efficient Cloud Service Provisioning Deploying Renewable Energy Sources," in *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 12, pp. 3927-3940, Dec. 2016.

[11] X. Chen et al., “Analyzing and modeling spatio-temporal dependence of cellular traffic at city scale,” in *proc. of IEEE ICC*, pp.3585-3591, 2015

[12] S. Haykin, Neural Networks and Learning Machines, New Jersey: Pearson, 2009.

[13] V. K. Tumuluru, P. Wang and D. Niyato, "A Neural Network Based Spectrum Prediction Scheme for Cognitive Radio," IEEE, 2010.

[14] Z. D. Zaharis, C. Skeberis, T. D. Xenos, P. I. Lazaridis and J. Cosmas, "Design of a Novel Antenna Array Beamformer Using Neural Networks Trained by Modified Adaptive Dispersion Invasive Weed Oprimization Based Data," IEEE, 2013.

[15] A. Omri, R. Bouallegue, R. Hamila and M. Hasna, "Channel Estimation for LTE Uplink System by Perceptron Neural Network," International Journal of Wireless & Mobile Networks (IJWMN), 2010.

[16] M. Baradas, G. Boanea and V. Dobrota, "Multipath Routing Management using Neural Networks-Based Traffic Prediction," The Third International Conference on Emerging Network Intelligence, 2011.

[17] C. Rottondi, L. Barletta, A. Giusti, and M. Tornatore, "Machine-Learning Method for Quality of Transmission Prediction of Unestablished Lightpaths," J. Opt. Commun. Netw. 10, A286-A297 (2018).

[18] S. Yan et al., “Field trial of Machine-Learning-assisted and SDN-based Optical Network Planning with Network-Scale Monitoring Database” ECOC 2017.

[19] A. Tzanakaki *et al*., "5G infrastructures supporting end-user and operational services: The 5G-XHaul architectural perspective," *2016 IEEE International Conference on Communications Workshops (ICC)*, Kuala Lumpur, 2016, pp. 57-62.

6. Invited papers

Open Marketplace and Service Orchestration for Virtual Optical Networks

(Invited Paper)

Shireesh Bhat*, George N. Rouskas†‡
*ECE Department, University of California Santa Barbara, USA
†Department of Computer Science, North Carolina State University, USA
‡Department of Computer Science, King Abdulaziz University, Saudi Arabia

***Abstract*—A key challenge in multi-vendor heterogeneous virtual optical networks is providing transparent access to network resources and virtual functions in a manner that enables users to combine them appropriately into meaningful end-to-end services. In this paper, we present a solution that consists of two components: an open marketplace where vendors and users of network resources and functions meet to establish economic relationships; and a planning service for creating end-to-end communication services from functional building blocks available in the marketplace. We also discuss algorithms for tackling variants of the network service orchestration problem.**

I. Introduction

In recent years the focus in the networking field has been on developing modular systems and moving away from the monolithic designs of the past. Specifically, network virtualization [1] decouples service functionality from the underlying resources (including network, compute, and storage) that are involved in delivering the service. Consequently, virtualization makes it possible to deliver end-to-end communication services that are composed from functional building blocks that (1) may be available at various locations strategically dispersed across the network, and (2) may be offered by different providers. By allowing for multiple service providers to co-exist on the same physical network substrate but separated by a virtualization layer, it is expected that network virtualization will lead to increased provider competition, more innovation, and more options/choices for users, as providers develop value-added services within their virtual network to stand out from the competition.

In general, network virtualization has been concerned with higher layers of the networking stack, and for the most part it has not touched the physical layer. In other words, the optical layer has typically been considered as a “black box:” sequences of bits are delivered to it for transmission, without the higher layers being aware of exactly how the transmission is accomplished. This separation of concerns imposed by the layering principle has allowed the development of upper layer protocols and services that are independent of the physical channel characteristics, but it has now become too restrictive as it prevents protocols or applications from taking advantage of additional functionalities that are increasingly available at the optical layer. In particular, in the past few years we have witnessed the development of optical layer devices that are *intelligent*, *self-aware*, and *programmable*, in that they can sense or measure their own characteristics and performance, and their behavior can be altered through software control.

The capabilities and functionality of these devices must somehow be exposed to higher layer applications and protocols, hence current network architectures cannot capture the full potential of the optical layer. For instance, the optical substrate increasingly employs various optical monitors and sensors, variable optical attenuators, bandwidth-variable transponders, distance-adaptive modulation, amplifiers and other impairment compensation devices. The monitoring and sensing devices are capable of measuring loss, polarization mode dispersion (PMD), or other signal impairments; based on this information, it should then be possible to use the appropriate impairment compensation to deliver the required signal quality to the application/user on demand. But such a solution cannot be accomplished within the current architecture, and has to be engineered outside of it separately for each application and impairment type; clearly, this is not an efficient or scalable approach. Reconfigurable optical add-drop multiplexers (ROADMs), flexible spectrum selective switches, and optical splitters with tunable fanout (for optical multicast) are additional examples of currently available devices whose behavior can be programmed according to the wishes of higher layer protocols. Based on current research trends one may anticipate further innovation in this field leading to the development of other sophisticated devices with programmable functionality that may be tailored to address specific requirements of higher applications.

Making this functionality available for delivering customized higher level end-to-end services to users presents two challenges. First, the device capabilities must be exposed to higher layer protocols, applications, and users in a manner that enables users and providers to form economic relationships around virtual optical network services that make use of these capabilities. Second, general purpose planning tools must be developed to stitch together lower-level functional blocks into meaningful end-to-end services. Therefore, in Section II we present a marketplace for the discovery, creation, and exchange of network services, and in Section III we discuss high-level algorithms for the orchestration of optical network resources. We carry out an evaluation of the algorithms in Section IV and we conclude the paper in Section V.

978-3-903176-07-2

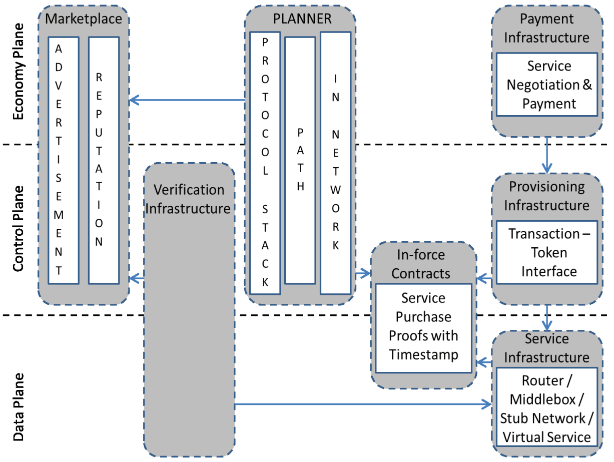

Economy Plane
Control Plane
Data Plane
Marketplace
ADVERTISEMENT
REPUTATION
PLANNER
PROTOCOL STACK
PATH
IN NETWORK
Payment Infrastructure
Service Negotiation & Payment
Verification Infrastructure
Provisioning Infrastructure
Transaction – Token Interface
In-force Contracts
Service Purchase Proofs with Timestamp
Service Infrastructure
Router / Middlebox / Stub Network / Virtual Service

Fig. 1. A Marketplace of network services within the ChoiceNet architecture

II. A Marketplace for Network Services

The Internet *edge* has thrived as an ecosystem where interactions among many stakeholders, acting as customers and/or providers of various hardware and software services for network, compute and storage functions, are mediated by economic considerations. However, this economic reality does not apply to any of the software or protocol interactions *within* the network, hence there is a lack of economic incentives for providing corresponding innovative services. To bridge this gap, as part of the ChoiceNet project [2], we introduced new mechanisms into the Internet architecture to enable an “economy plane” that allows the presentation of competing offerings for various networking services, the formation of contracts between users and providers, and determination of whether each provider meets its part of the contracts.

The ChoiceNet architecture, illustrated in Figure 1, introduces a platform for service providers to advertise their services and for customers to discover, negotiate and pay for these services. The main architectural component facilitating service advertisements is the “marketplace,” where service providers register their services and customers discover them via appropriate queries. Once a customer selects a service, further ChoiceNet interactions between customer and provider create a contract, with the customer receiving a token that is carried by its traffic as proof of purchase in order to receive the corresponding service. These customer/provider interactions constitute what we term the “economy plane” of ChoiceNet, shown on Figure 1 as separate from, but interacting with, the well-known control and data planes. We emphasize that the economy plane does *not* introduce new economic interactions; rather it empowers existing real-world interactions to take place (1) in-network (whereas today they take place outside the network architecture), and (2) at a range of short (e.g., in the order of a flow duration) to long time scales.

The marketplace is a virtual repository of network functions and services available to users. The repository provides interfaces for providers to list (advertise) their services, and for users (or their software agents) to obtain listings of service offerings that meet their requirements. ChoiceNet’s interfaces enable entities to realize complex service models where an entity may act as a provider to some customers and at the same time act as a customer to some providers. This feature is essential in a virtual network architecture as it enables service providers who lease physical network resources from infrastructure providers, in turn to lease their virtual network resources to other service providers.

Service advertisements in the ChoiceNet marketplace and protocol interactions in the economy plane are semantically enriched [3], [4] to allow automated composition, thus the marketplace is more like an ontology than a directory. As a result, although the ChoiceNet marketplace was conceived and demonstrated within a packet-switched network context [5], it may readily accommodate the optical layer functions and services we discussed in the previous section. For instance, consider an optical multicast service that is offered by deploying optical splitters in various nodes across the network. The service description in this case would include the address of the locations where the splitters are present, the maximum fanout of each, the spectrum range over which the splitters operate, the power loss due to splitting, and other relevant information that an orchestration algorithm may take into account in formulating a multicast communication service.

Realistically, orchestrating a set of services in the marketplace to create a complete service for a customer is expected to be a complex task for all but the simplest cases, thus the task must be automated and performed by software agents. The planner module in the ChoiceNet architecture of Figure 1 interacts with the marketplace over the economy plane and employs specialized algorithms to orchestrate marketplace services into end-to-end communication services for users. We note that the architecture allows for multiple planning services to co-exist in competition to each other: each planning service may query the marketplace repository to obtain the available services and hence focus on innovation in the design of orchestration algorithms to enhance customer experience. The reader may have noticed the analogy with a real-world application, the travel industry, a model that has guided our design of the ChoiceNet architecture. In the travel industry, service providers include the airlines, hotels, and rental car companies, whereas travel sites such as Orbitz or Priceline operate planning and orchestration services. These sites take as input traveler preferences and construct itineraries to ensure that users may access seamlessly all the services acquired across the various flight, accommodation, and car rental providers.

We assume that planners represent the services available in the marketplace in a graph format, such that service orchestration may be carried out using appropriately designed graph algorithms. We expect that such a graph will be highly dynamic as it must be updated whenever a user acquires or releases services. Also, the planner for a marketplace of virtual optical network services must consider offerings from multiple providers, including virtual operators who may lease resources from the same physical infrastructure or recursively lease services from other providers. Returning to the travel analogy, this is similar to a planner taking into consideration competing flights from multiple airlines between pairs of cities, as well as multiple hotels or rental car agencies within a city. Consequently, the graph of marketplace services is a superset of the underlying network topology. Specifically, nodes and edges in the services graph represent virtual entities rather than physical ones: a physical node may include multiple virtual nodes, each virtual node operated by a different service provider deploying a variety of network service instances. The graph may also include parallel edges between nodes that represent competing path services. Such a topology may be considerably larger than the underlying physical network topology, hence orchestration algorithms must scale to large graph sizes.

III. Service Orchestration

As we mentioned in the previous section, we expect that the deployment of marketplaces for network services will lead to innovation in the design and application of orchestration algorithms for the delivery of customized end-to-end communication services in virtual optical networks. In turn, the availability of planners that operate in short time scales (i.e., on par with the setting up of a service) is likely to generate further customer interest in specialized services, which will motivate virtual network providers to invest in novel services and more sophisticated orchestration algorithms to differentiate from the competition, creating a virtuous cycle similar to the one we have witnessed unfold at the edge of the network in the past twenty five years.

Broadly speaking, upon receiving a request from a user (customer), a planner must carry out three tasks as part of the service orchestration process [6], [7]:

- *Service Selection:* determine the set of virtual network services to satisfy the user request;
- *Service Ordering:* determine the order in which the selected services must be applied to the user's traffic; and
- *Service Concatenation:* construct path(s) from the source node to the destination node(s) that visit virtual nodes where instances of the selected services are deployed in the order determined by the service ordering step.

In previous work, we have considered the service orchestration problem in contexts where the three subproblems above are pairwise decoupled and may be carried out sequentially in the given order. In such situations, the service concatenation step will involve general algorithms that may applied to a broad set of services, as we discuss next.

Let us assume that multiple instances of each service k are deployed at various virtual nodes across the network, possibly operated by different service providers. Also let S_k denote the set of nodes where instances of service k are deployed. In earlier work [8] we considered the following general service orchestration problem for a user request that requires $K \geq 1$ services to be applied in a given order:

> Given the graph representing the union of the virtual network topologies represented in the marketplace, a source node s, a destination node d, and an ordering of K node sets $S_1, S_2, \ldots, S_K$, construct a path of minimum cost from s to d that visits one node in each set $S_k, k = 1, \cdots, K$, in the given order.

The above problem is equivalent to the shortest path tour problem (SPTP) that was first studied in a different context more than forty years ago [9], [10]. A shortest path tour is a path of minimum cost from s to d constructed as the concatenation of the $K+1$ path segments $[s, n_1], [n_1, n_2], \cdots, [n_K, d]$, where $n_k \in S_k, k = 1, \cdots, K$, and each segment may include nodes other than its two endpoints.

SPTP is solvable in polynomial time, but as we discussed in [8], earlier algorithms were designed for only certain classes of graphs (either sparse graphs or small dense graphs), and do not scale to graph instances we expect will arise in representing marketplaces of virtual network services. We also developed a new algorithm by introducing several novel modifications to Dijkstra's algorithm to construct the shortest path tour efficiently. This new algorithm, which we call depth-first tour search (DFTS), scales to graphs with thousands of nodes and large nodal degrees, and is appropriate for real-time service orchestration applications.

The DFTS algorithm, as well as earlier algorithms for the SPTP problem were developed for packet-switched networks, but they may certainly be applied in the context of virtual optical networks offering a range of services from the physical layer (including the ones we listed in Section I) to the application layer (as we have discussed in [6], [8]). For instance, consider a user request that requires services to be applied directly to the optical signal (e.g., amplification, dispersion compensation, etc.) carrying the user traffic, as well as transformation services (e.g., transcoding of application data, encryption or decryption, and more) to be applied to the data carried by the signal. As long as an ordered set of services is provided to the service concatenation step, then appropriate algorithms for the SPTP problem may be used to construct minimum cost paths that include nodes where the services are offered.

Often, however, the three subproblems of service orchestration (i.e., service selection, ordering, and concatenation), are not decoupled, hence solving them sequentially may not lead to an overall optimal solution (path) or even a feasible one. This may be especially true when optical layer services are part of the mix, due to cross-layer dependencies. For instance, for a given quality-of-service requested by the user, the selection subproblem may need to coordinate with the concatenation subproblem so as to take into consideration the length and other properties of the candidate optical path(s) in order to determine the modulation format or spectrum of the signal, or whether to include impairment compensation services. Although there has been considerable research in cross-layer optical network design [11], including routing algorithms that take into account physical layer impairments, to the best of our knowledge, the general service concatenation problem we defined above has not been studied when the three subproblems are tightly coupled.

A solution to the general service orchestration problem is outside the scope of this work and is the subject of ongoing research in our group. Nevertheless, we expect that more restricted variants of the problem will have applications in specific contexts. In this paper, we consider such a variant that arises in virtual networks offering multicast services at the optical layer. An optical layer multicast service may be used to implement point-to-multipoint connections directly at the physical layer by employing optical splitters that divide the power of an input signal into several output signals [12]. Let m denote the number of distinct destination nodes to which the signal must be delivered. Then, a multicast service with a fanout of at least m must be included in the service selection step of the orchestration process, along with any other services needed to satisfy the user request.

We consider a special case of the service orchestration problem wherein the service selection subproblem is independent of the other two subproblems, but the service ordering and service concatenation problems are coupled with respect to the order of the multicast service. More specifically, let $K, K > 1$, be the number of services, including the multicast service, determined by the selection step. Also assume that the relative order of the $K - 1$ services other than the multicast service has been decided (i.e., it remains fixed and is not subject to optimization), but that the multicast service may be placed in any position in that relative ordering. For instance, amplification may take place before or after splitting the optical signal, keeping in mind that in the latter case, the amplification service must be applied to all m output signals. Similarly for higher layer services, since, say, encryption or transcoding may be applied to the original traffic stream or the m streams produced as the result of splitting.

The problem of finding the shortest path tour from source s to the m destination nodes is a generalization of the SPTP problem we defined above, and we refer to it as the point-to-multipoint SPTP (P2MP-SPTP). Consider now a special case of the problem whereby the multicast service is placed as the k-th service in the ordering, $1 \leq k \leq K$. Recall that S_k is the set of nodes where the multicast service is offered, and let $|S_k| = L \geq 1|$. Further, let $S_k = \{n_1, \cdots, n_L\}$. We may obtain an optimal solution to this special case by following these steps:

1) Initialize $i = 1$.
2) Set $S'_k = \{n_i\}$.
   2a) Solve the SPTP problem from s to n_i with input $S_1, \cdots, S'_k$.
   2b) Solve the SPTP problem from n_i to the m destination nodes with input $S_{k+1}, \cdots, S_K$.
   2c) Concatenate the two tours to obtain the tour from s to the m destinations, and record its cost.
3) Increment i and repeat Step 2 while $i \leq L$.
4) Select among the L tours constructed the one with the minimum cost.

Step 2a ensures that the final tour consists of a single path from s to some node $n \in S_k$ where the multicast service is applied to split the input optical signal into m output signals. Performing Step 2 for each node in set S_k guarantees that the shortest tour is found in the last step.

We may now obtain an optimal solution to the original problem by repeating the above algorithm K times, each time with the multicast service as service $S_i, i = 1, \cdots, K$, in the order of services, and selecting the best overall solution. Note that the worst-case running time complexity of the DFTS algorithm we presented in [8] is $O(KE \log N)$, where E and N represent the number of edges and nodes, respectively, in the underlying graph. Therefore, the complexity of the above approach is $O(LK^2 E \log N)$, which for moderate values of L and K (e.g., in the order of 10-20) would be reasonable and appropriate for online operation. In particular, our experimental evaluation of DFTS has shown that the algorithm completes in well under one second even for dense graphs with up to $N = 5,000$ nodes. Therefore, even with the additional $O(LK)$ factor, the above algorithm for the P2MP-SPTP problem may be used at the time scales appropriate for setting up end-to-end flows in real time.

Consider now the general case of the joint service ordering and concatenation problem, and for simplicity assume point-to-point communication only, i.e., a single source s and a single destination d. A straightfoward approach to solving this problem would be to solve each of the $K!$ SPTP problems that arise for each possible permutation of the K selected services. Furthermore, not all $K!$ permutations may be valid, resulting in a smaller solution space for the original problem: for instance, note that encryption must precede decryption or that transcoding must precede encryption. Nevertheless, for larger values of K, enumerating all valid permutations of services is expected to be computationally infeasible, especially for applications that require results in real time. Further research is necessary to determine the complexity of this problem and to derive polynomial-time algorithms, perhaps by extending existing SPTP algorithms, including the DFTS algorithm we presented in [8].

As a final note, we conjecture that the most general service orchestration problem whereby all three subproblems (service selection, ordering, and concatenation) have to be solved jointly is computationally intractable. Even if the conjecture is true, efficient algorithms may exist under certain simplifying assumptions that may hold in practice. Such algorithms are essential so as to account for the cross-layer dependencies inherent in the delivery of end-to-end communication services in virtual networks with programmable optical layer capabilities. Therefore, we consider this an important research direction and one that may readily build upon the insights from recent and ongoing research in multilayer optical network design.

IV. Numerical Results

We evaluate our algorithm on random graphs generated using BRITE [13], a universal topology generator. We obtained undirected graphs by configuring BRITE to generate AS-Level Barabasi models; we then converted these graphs into directed ones that we used in our experiments. In generating random instances for the P2MP-SPTP problem, we considered the following parameters and varied their values as described below:

- The number N of nodes in the graph was varied from 1000 to 5000 in increments of 1000.
- The average nodal degree Δ of the graph was set to an integer in the range $[2,5]$.
- The number K of node sets in the tour was set to 4
- The number k for the relative order of the multicast service set took integer values in the interval $[1,4]$
- The number M of nodes in the multicast service set was varied from 2 to 8 in multiples of 2. The number of nodes in the non-multicast service sets was set to 25
- The number of destination nodes (multicast streams) was set to 10.

There are 240 unique combinations of the values of parameters N, Δ, k, and M that we considered in our experiments (refer to the top of this section). In Table I, we list the actual running time of our algorithm, for problem instances generated with each of these 240 parameter value combinations. Each entry in the table is the average running time over 1,000 problem instances generated from the stated values of the parameters. All experiments were performed on a High Performance cluster that included Dual Intel X5650 six core processors, with 48GB and infiniband interconnect.

We make the following observations:

- The running time increases linearly as a function of M when N, Δ, and k are kept fixed.
- The running time increases linearly as a function of $ElogN$ when M, Δ, and k are kept fixed.
- The running time increases slower than linearly as a function of Δ, when
- When N, M, and Δ are kept fixed, we observe that the relative order of the multicast service set produces some interesting results. In many instances when k is 1, we observe a marginally higher running time compared to the rest of the cases, but we do not see this pattern for all the instances. We infer that the inherent graph characteristics (N, E, Δ) influence k since we divide the P2MP-SPTP problems into multiple SPTP sub-problems and the value of k determines the size of the SPTP sub-problems.

To the best of our knowledge, this is the first work to address the P2MP-SPTP problem and develop an algorithm to solve it efficiently. We hope that this paper serves as a reference and paves the way for further investigation of the P2MP-SPTP problem.

Acknowledgments

This work was supported in part by the National Science Foundation under Grant number 1111088. We also acknowledge support from the Center for Scientific Computing from the CNSI, MRL: an NSF MRSEC (DMR-1720256) and NSF CNS-0960316.

V. Concluding Remarks

We envision virtual optical networks that empower end users to make informed choices in selecting among network services offered by competing providers within an ecosystem that promotes and rewards innovation. A marketplace that serves as repository of service building blocks and the meeting ground between customers and providers is the first component of such a vision. The second component consists of sophisticated orchestration algorithms that add value to the user experience by creating highly customized end-to-end communication services from the existing building blocks. We consider the development of orchestration algorithms that take into account cross-layer dependencies as a fruitful area of research for the optical networking community.

References

[1] A. Wang, M. Iyer, R. Dutta, G. N. Rouskas, and I. Baldine. Network virtualization: Technologies, perspectives and frontiers. *IEEE/OSA Journal of Lightwave Technology*, 31(4):523–537, February 2013.

[2] T. Wolf, J. Griffioen, K. Calvert, R. Dutta, G. N. Rouskas, I. Baldin, and A. Nagurney. ChoiceNet: Toward an economy plane for the Internet. *ACM SIGCOMM Computer Communication Review*, 44(3):58–65, July 2014.

[3] S. Bhat, R. Udechukwu, R. Dutta, and G. N. Rouskas. On service composition algorithms for open marketplaces of network services. In *Proceedings of EuCNC*, June 2017.

[4] R. Udechukwu, S. Bhat, R. Dutta, and G. N. Rouskas. Language of choice: On embedding choice-related semantics in a realizable protocol. In *Proceedings of IEEE Sarnoff Symposium*, September 2016.

[5] S. Bhat, R. Udechukwu, R. Dutta, and G. N. Rouskas. Inception to application: A GENI based prototype of an open marketplace for network services. In *Proceedings of IEEE INFOCOM Workshops*, April 2016.

[6] S. Bhat and G. N. Rouskas. On routing algorithms for open marketplaces of path services. In *Proceedings of IEEE ICC 2016*, May 2016.

[7] X. Huang, S. Shanbhag, and T. Wolf. Automated service composition and routing in networks with data-path services. In *Computer Communications and Networks (ICCCN), 2010 Proceedings of 19th International Conference on*, pages 1–8, Aug 2010.

[8] S. Bhat and G. N. Rouskas. Service-concatenation routing with applications to network functions virtualization. In *Proceedings of ICCCN 2017*, August 2017.

[9] C. P. Bajaj. Some constrained shortest-route problems. *Unternehmensforschung*, 15(1):287–301, 1971.

[10] A. Kershenbaum, W. Hsieh, and B. Golden. Constrained routing in large sparse networks. In *IEEE International Conference on Communications," pp. 38.14-38.18, Philadelphia, PA*, 1976.

[11] S. Subramaniam, M. Brandt-Pearce, and C. Vijaya Saradhi (Eds.). *Cross-Layer Design in Optical Networks*. Springer, 2013.

[12] G. N. Rouskas. Optical layer multicast: Rationale, building blocks, and challenges. *IEEE Network*, 17(1):60–65, January/February 2003.

[13] A. Medina, I. Matta, and J. Byers. Brite: A flexible generator of internet topologies. Technical report, Boston University, Boston, MA, USA, 2000.

TABLE I
AVERAGE RUNNING TIME (IN SECONDS) OF OUR ALGORITHM TO SOLVE P2MP-SPTP

		$k=1$			$k=2$			$k=3$			$k=4$		
N	Δ	$M=2$	$M=4$	$M=8$	$M=2$	$M=4$	$M=8$	$M=2$	$M=4$	$M=8$	$M=2$	$M=4$	$M=8$
1000	2	0.04324	0.08458	0.17306	0.0414	0.08166	0.16584	0.04056	0.08141	0.16102	0.0405	0.0797	0.16128
	3	0.0501	0.09792	0.19795	0.04751	0.09394	0.18942	0.04681	0.09387	0.18535	0.04676	0.09287	0.18596
	4	0.06994	0.13483	0.27256	0.06641	0.13228	0.26554	0.06552	0.13167	0.26078	0.06573	0.1303	0.26213
	5	0.07725	0.1528	0.30158	0.07404	0.147	0.29484	0.07389	0.14825	0.29393	0.07408	0.14719	0.29487
2000	2	0.17413	0.34667	0.68324	0.17126	0.33195	0.65452	0.16322	0.3212	0.64251	0.16298	0.32105	0.63134
	3	0.21837	0.43511	0.86546	0.21683	0.42536	0.84286	0.21247	0.4254	0.85354	0.21541	0.4222	0.83347
	4	0.26384	0.52766	1.04008	0.26234	0.51041	1.01864	0.25425	0.51412	1.01286	0.25942	0.50435	1.01448
	5	0.3053	0.60819	1.20743	0.30137	0.5911	1.17768	0.29254	0.5898	1.15671	0.29487	0.57646	1.16156
3000	2	0.4034	0.80138	1.58029	0.39145	0.77083	1.57162	0.39127	0.78169	1.55459	0.39346	0.76178	1.53388
	3	0.50435	0.98411	1.95159	0.48306	0.98064	1.94662	0.48391	0.96243	1.92723	0.488	0.9444	1.90324
	4	0.54207	1.11959	2.22108	0.55442	1.09921	2.21439	0.55444	1.12262	2.22104	0.56459	1.09202	2.19943
	5	0.61366	1.2748	2.5501	0.62496	1.2509	2.49379	0.63225	1.26157	2.50682	0.63836	1.25774	2.47685
4000	2	0.72241	1.39973	2.83001	0.69922	1.36145	2.66832	0.67913	1.32105	2.53718	0.67751	1.30977	2.62299
	3	0.89938	1.76487	3.58951	0.85358	1.73223	3.37386	0.89709	1.67892	3.24119	0.85957	1.68032	3.22716
	4	1.03653	2.10116	3.98855	0.9715	1.98775	3.83469	0.96385	1.87807	3.62008	0.95753	1.8564	3.68269
	5	1.25635	2.5306	4.80257	1.19211	2.54138	4.7413	1.19453	2.33547	4.48676	1.24731	2.32187	4.54973
5000	2	1.09231	2.15912	4.19843	1.07995	2.044	3.99808	1.06102	1.99967	4.05524	1.05971	2.02657	3.9953
	3	1.39378	2.70822	5.29489	1.37967	2.67659	5.11438	1.36784	2.66511	5.28403	1.35193	2.64737	5.23518
	4	1.41373	2.6871	5.32409	1.30994	2.65433	5.15203	1.33458	2.58562	5.16531	1.31004	2.57086	5.17844
	5	1.77974	3.53069	6.93992	1.72774	3.38791	6.72808	1.63398	3.32734	6.50819	1.61203	3.24615	6.37525

Experimental SDN Control Solutions for Automatic Operations and Management of 5G Services in a Fixed Mobile Converged Packet-Optical Network

(invited paper)

Ricardo Martínez, Ricard Vilalta, Manuel Requena, Ramon Casellas, Raül Muñoz and Josep Mangues
CTTC – Communication Networks Division
Av. Carl Friedrich Gauss, 7, 08860 Castelldefels, Spain
ricardo.martinez@cttc.es

***Abstract*— 5G networks will impose network operators to accommodate services demanding heterogeneous and stringent requirements in terms of increased bandwidth, reduced latency, higher availability, etc. as well as enabling emerging capabilities such as slicing. Operators will be then forced to make notable investments in their infrastructure but the revenue is not envisaged to be proportional. Thereby, operators are seeking for more cost-effective solutions to keep their competitiveness. An appealing solution is to integrate all (broadband) services including both fixed and mobile in a convergent way. This is referred to as Fixed Mobile Convergence (FMC). FMC allows seamlessly serving any kind of access service over the same network infrastructure (access, aggregation and core) and relying on common set of control and operation functions. To this end, FMC leverages the benefits provided by Software Defined Networking (SDN) and Network Function Virtualization (NFV). First, we discuss some of the explored FMC solutions and technologies, from both structural and functional perspectives Next, focusing on a Multi-Layer (Packet and Optical) Aggregation Network, we report two implemented and experimentally validated SDN/NFV orchestration architectures providing feasible FMC to address upcoming 5G challenges.**

Keywords— FMC, 5G, SDN and NFV, Multi-Layer Networks.

I. Introduction

The upcoming *5G networks* will bring advanced services with stringent requirements: increased data rate (100x compared to 4G data rates at cell edge), enhanced end-to-end latency (10 ms or less), enhanced energy efficiency, massive connectivity with strict quality of service (QoS) levels, etc. [1]. 5G services are sorted into three main communications types: i*) enhanced mobile broadband* (eMBB), ii) *massive machine type communications* (mMTC) and iii) *ultra-reliable low latency communications* (uRLLC). Fig. 1 (a) qualitatively illustrates the heterogeneity and impact on different requirements of such service types. The service being more bandwidth-hungry is eMBB; uRLLC requires connections with extremely low latencies used for delay-sensitive applications, such as gaming or automotive services. Finally, mMTC will impose handling a high connection density, that is a very large number of connections need to be handled in a reduced area.

The above connection types and service requirements of 5G will notably challenge network operators at the time of accommodating them in their infrastructures in a cost-effective manner. That is, operators are seeking for strategies aiming at reducing both CapEx (i.e., investments) and OpEx (control functions and operations). To do that, it is widely agreed that 5G services must be supported over the same infrastructure and managed by a common pool of control elements and functions (e.g., unified control and management, authentication, authorization and accounting, etc.). In light of this, converging (mostly mobile) 5G services with traditional broadband fixed services (like FTTH) is seen as a very plausible scenario. The rationale behind this is that despite operators see that fulfilling 5G service requirements entail a notable network transformation, the expected revenue for doing that will not be proportional. Thereby, the most economically-viable solution is to lower the Total Cost of Ownership (TCO) to keep operator's competiveness. Integrating all 5G, and especially broadband mobile and fixed services is essential and this is referred to as *Fixed Mobile Convergence* (FMC) [1][2].

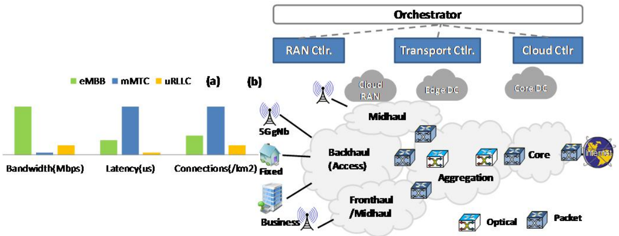


Fig. 1. (a) 5G service requirements; (b) FMC infrastructure supporting 5G

FMC embraces two convergence types: i) unifying their equipment and technologies in the different network segments (i.e., access, aggregation and core) and, ii) integrating the management as well as operations and business systems [3]. Both *convergence* approaches were recently explored and validated in the context of the EU FP7 COMBO project [4]. In this project, the first convergence type was termed as *structural convergence,* whilst the second one was named as *functional convergence.* Both FMC solutions can leverage the appealing features in terms of flexibility, agility, cost-efficiency, etc. brought by current networking trends (envisioned as essential for the 5G), such as the Software Defined Networking (SDN) and the Network Function Virtualization (NFV) [3].

SDN deals with a logically centralized control (relying on standard open interfaces) of the data plane infrastructure, i.e., Radio Access Network (RAN), Passive Optical Networks

978-3-903176-07-2 © 2018 IFIP

(PON) solutions, packet and optical switches, etc. This provides automatic and unified programmability of the underlying network relying on abstracted data plane information. Thus, the traditional lack of interaction between different network segments and technologies such as mobile and transport layers can be overpassed [5]. Indeed, mobile, fixed and transport layers has traditionally evolved independently which in turn does increase the TCO. Additionally, the cumbersome vendor lock-in can be removed favouring multi-vendor network scenarios.

On the other hand, NFV relies on exploiting cloud/IT virtualization techniques to enable network functions such as the mobile Evolved Packet Core (EPC), Border Network Gateway (BNG), Central Unit (CU), etc., which typically are allocated in dedicated hardware, to be run on the cloud as Virtual Network Functions (VNFs) [6]. This concept also favours reducing the TCO. Specifically, VNFs can be executed within Data Centers (DCs) (in Virtual Machines, VMs, or containers) and may be applicable to any data plane packet processing as well as control function comprised in fixed and/or mobile infrastructures [7][8].

Both SDN and NFV concepts allow operators offering emerging 5G capabilities, such as the *network slicing* [9][10]. Slicing provides, on the one hand, network virtualization, which exploits SDN abstraction capabilities to partition the physical infrastructure and compose multiple (logical) and isolated infrastructures (i.e., *multi-tenancy*). On the other hand, slicing also offers the allocation, tailoring and configuration of (virtualized) network functions required for a specific service relying on NFV. Required VNFs for a service could be deployed in DCs located at different FMC infrastructure locations depending on the service requirements. Therefore, slicing is also an enabler of FMC to accommodate over a common infrastructure heterogeneous services such as Mobile (Virtual) Network Operators (MVNO) or those related to the vertical industries (e.g., eHealth, Industry4.0, etc.) [11].

Herein we report some implementations and experimental validations we conducted addressing specific 5G and FMC objectives within an aggregation (backhaul) network segment. We consider a multi-layer network (MLN) infrastructure formed by both packet and optical switching and controlled by a transport SDN instance. This allows coping with both the envisaged tremendous growth of data traffic (specially eMBB), exploiting packet statistical multiplexing and huge optical capacity, as well as the expected high dynamicity and stringent requirements of both eMBB and uRLLC leveraging SDN flexibility and programmability. Finally, slicing is also explored where the abstraction and virtualization of the MLN is combined with NFV to specifically offer dynamic deployment of virtual backhaul transport for multiple MVNO demands.

The remainder of this paper is as follows. In Section II, we present a general FMC network architecture for 5G with special focus on access and aggregation segments, which will be the most impacted by 5G requirements. Section III addresses an implemented SDN-based orchestrator for dynamically serving both fixed and mobile broadband communications within an aggregation MLN. In Section IV, it is detailed an SDN/NFV orchestration solution providing dynamic composition of virtual backhaul infrastructures over a MLN for different MVNOs along with deploying VNFs (e.g., for EPC functions) at DCs. Finally, Section V concludes this work.

II. FMC Architectures in Support of 5G Services

Figure 1 (b) depicts a general SDN/NFV network architecture conceived to face up the challenges imposed by 5G. As mentioned, the goal pursued by network operators is to adopt a sufficiently flexible network solution satisfying both the dramatic growth and extreme dynamicity of 5G data traffic whilst reducing the TCO. In this scenario, FMC becomes very relevant, providing the integration of broadband communications via structural and functional convergence approaches. In the following, structural FMC solutions and technologies within both access and aggregation networks are reported. Next, SDN and NFV control and orchestration systems are discussed, paving the way not only for supporting 5G operations but also for addressing functional FMC approaches.

A. Access and Aggregation Networks for FMC

In the access domain, one of the most appealing convergence strategies relies on leveraging existing FTTH. Traditionally, such a deployment (in addition to the fiber to the cabinet, FTTC) were/are rolled out for delivering fixed broadband services. Nonetheless, to address the expected increase of densification of macro and small cells along with coping with the growth of 5G data traffic (eMBB, uRLLC and mMTC), the existing FTTH/FTTC infrastructure can be reused to foster structural FMC [3][12]. In this context, different RAN architectures (e.g., traditional backhaul, fronthaul, midhaul) have been proposed presenting their own functional split between Distributed Unit (DU) and CU This in turn impacts on the necessities (e.g., data rate and delay) to be dealt with by the optical fiber connectivity [13]. In all these RAN architectures, the purpose is to rely on FTTx technologies. Two of the most promising technologies to achieve that rely on the Wavelength Division Multiplexing (WDM) systems. This allows leveraging intrinsic WDM advantages, such as good scalability, low latency and commercial availability. Specifically, the current solutions being proposed are NGPON2 (using point-to-point WDM links) and wavelength-routed (WR) WDM PON [3].

In the aggregation domain, the primary objective is to aggregate and transport traffic towards adjacent network segments or DCs (Edge DC in Fig. 1 (b)). Aggregation decisions must be made considering not only the efficient capacity usage, but also the service needs (e.g., latency). From the data plane perspective, an interesting aggregation approach is based on MLN. MLN takes advantage of both worlds: packet switching (e.g., MPLS) providing finer granularity and statistical multiplexing, and optical networks (fixed or flexi-grid WDM networks) offering huge transport capacity. Thus, packets flows arriving from mobile or fixed access networks are groomed and transported over a common aggregation infrastructure [14]. Consequently, detailed MLN capabilities lead to attain an efficient use of all the network resources (packet ports, link bandwidth, optical transceiver and spectrum) which in turn enables relaxing the increasing pressure to operators for accommodating the expected traffic growth.

B. SDN/NFV Control and Orchestration for FMC

Both SDN and NFV lead to speed up and to attain the functional FMC objectives, i.e., defining a set of generic network functions being applicable regardless of the access connectivity. Such functions cover [16]: i) a common mechanism to authenticate users / subscribers by managing the session regardless of the access network; ii) advanced interface selection and routing to provide enhanced data path decisions for controlled offloading, load balancing on multiple data paths and smooth handover between access technologies. The EU FP7 COMBO project [4] studied both functional FMC solutions, which were validated under the concept of the Universal Access Gateway (UAG) [16]. The UAG is a functional element (defined within the COMBO project) that allows terminating data flows from different access technologies. The implementation of the UAG used both SDN and NFV. Specifically, the unified authentication was deployed as a VNF hosted in the UAG cloud and referred to as universal authentication. Other instantiated UAG's VNFs supported multiple virtual EPC (vEPC) implementations addressing slicing capabilities. Finally, all flows in the UAG were programmed by an SDN controller.

In general, SDN/NFV control and orchestration provides the required framework to automate network service management involving heterogeneous physical and virtual functions and resources throughout different domains and technologies, as depicted in Fig. 1 (b). That is, a unified and coordinated system (relying on common interfaces and APIs) can dynamically accommodate any type of 5G service over the underlying FMC and Cloud infrastructure. To this end, the orchestration follows a modular *tree structure* (or hierarchical) where a dedicated controller programs a set of resources (network and cloud) within a domain [17]. The orchestrator using an abstracted view of the underlying resources commands those controllers to, from an end-to-end perspective, create, update and release services. Controllers are assigned on per segment/domain or technology basis [16]. In Fig. 1 (b), a dedicated controller handles separately the RAN, the MLN aggregation / transport, and the DCs (Cloud RAN, Edge and Core). This architecture is highly scalable and flexible and allows supporting cross-domain strategies for functional FMC goals: traffic offloading, re-optimization, load balancing. In the next two Sections we report two experimental validations led by CTTC demonstrating the feasibility of applying SDN/NFV orchestration to attain: i) unified transport of fixed and mobile data flows over an aggregation MLN; ii) dynamic deployment of virtual backhaul tenants in support of 5G slicing for MVNO demands.

III. SDN Orchestration of Aggregation MLN for Fixed and Mobile Services

This section addresses an implemented SDN orchestrator to automatically and seamlessly configure an aggregation MLN. First, it is presented the targeted scenario and how FMC objectives are achieved. Next, it is detailed the designed and deployed SDN orchestrator with the conducted validation.

A. Targeted Scenario

Figure 2 illustrates an FMC scenario focusing on an SDN-orchestrated aggregation MLN. Fixed and mobile services arrive to the access-aggregation bordering nodes. According to their service requirements (bandwidth and latency), they are groomed and transported towards either the core (e.g., Core DC) or forwarded to VNFs running at the Edge DC. In the example, both (blue) mobile and (green) fixed services, assumed to need similar requirements in terms of bandwidth (e.g., eMBB) are grouped and jointly transported over an optical tunnel towards their respective gateways (i.e., vEPC and vBNG) located at the Core DC. On the other hand, (red) mobile service with stringent latency requirement (uRLLC) is forwarded towards the gateway and application (vEPC and CDN) at the Edge DC exploiting the advantages of Multi-Access Edge Computing (MEC) [18].

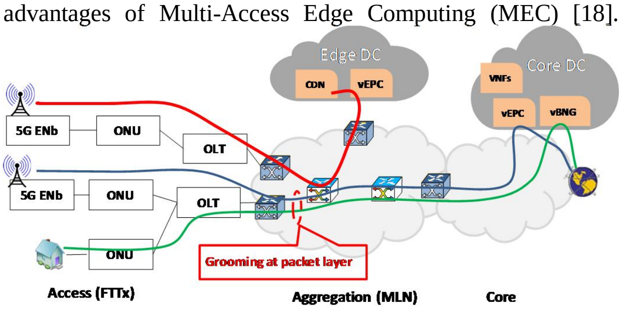


Fig. 2. Multiple fixed and mobile services in a converged MLN

To do the above in a dynamic fashion, the SDN orchestrator coordinates the configuration of all the network technologies forming the MLN. This requires awareness of: i) (abstracted) view of the resource status (i.e., packet ports, topology, optical link bandwidth, virtual packet link bandwidth); ii) the service requirements. Thus, the SDN orchestrator can accommodate and re-optimize requested and exiting services favouring grooming strategies to attain the most efficient use of all resources.

B. Deployed SDN Orchestrator

The architecture of the SDN orchestrator within the aggregation MLN is shown in Fig. 3 (a). The key architectural element is the Application Based Network Orchestrator (ABNO) [19]. The ABNO coordinates the set of controllers assigned for each technology (packet and optical) to provide end-to-end connectivity. Thus, the ABNO operates as a front-end for receiving and processing incoming (fixed or mobile) service requests. This element then coordinates/triggers the rest of the involved functions to eventually come up with the computation and programmability of the MLN [20].

In the example, the EPC's Mobility Management Entity (MME) after negotiating (out-of-band) the establishment of a new mobile service (Bearer), commands via a REST API the request for backhauling the service between the 5G ENb and the core gateway (SGW/PGW). The REST API message contains:

- *Endpoints*: IP addresses of both ENb and SGW/PGW
- *Transport Layer*: the requested packet service (MPLS)
- *Service Requirements*: bandwidth (bit/s) and latency (ms)
- *Tunnel Endpoint Identifiers* (TEID) identify the mobile service (Bearer) between the ENb and the core gateways.

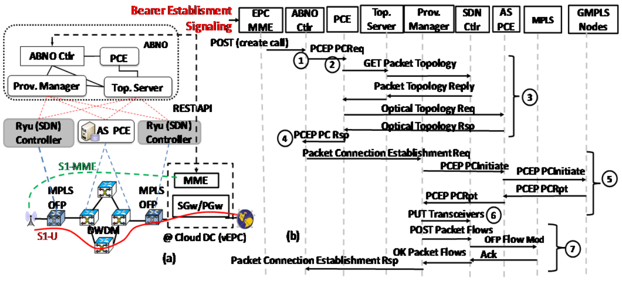


Fig. 3. (a) SDN orchestrator architecture for MLN supporting FMC services. (b) Workflow for dynamically creating a new mobile service.

The ABNO's Path Computation Element (PCE) computes MLN paths. As said, besides dealing with the demanded QoS requirements of the bearers the path computation is done to attain the most efficient use of network resources via grooming strategies. This requires the PCE having a (abstracted) view of network resources and topology being gathered by the Topology Server. This information is retrieved using a REST API for the packet domain and TCP/BGP-LS for the optical domain.

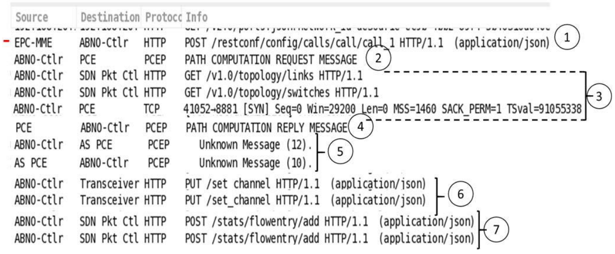
The Provisioning Manager coordinates via REST API the SDN controllers for each domain. In our setup, two SDN packet controllers (relying on Ryu implementation) and an Active Stateful (AS) PCE are used for handling packet and optical domains, respectively. The SDN controllers for the packet domains use OpenFlow protocol for the configuration. On the other hand, the optical domain is controlled combining the AS PCE (for computing and instantiating the connection) with a distributed GMPLS control plane. Last but not least, the Provisioning Manager also configures the optical transceivers (XFPs) at the MPLS nodes via a REST API selecting the nominal DWDM frequency. The experimental setup uses the LTE/EPC network provided by the ns-3 LENA emulator [21].

C. Experimental Validation

To create a new mobile service (Bearer), first the EPC's MME allocates the TEID and provides it to the 5G ENb in the connection establishment. This TEID carried into the GPRS Tunnelling Protocol (GTP-U) identifies a particular Bearer. In our approach, we associate the Bearer's TEID to a particular and unused MPLS tag at the packet nodes connected to both the cell site and the core gateways. The selected MPLS tags allows steering the traffic towards the Bearer's termination: Edge DC or Core DCs hosting vEPC user plane functions. For instance, for uRLLC services data traffic is forwarded to the Edge DC leveraging the MEC advantages.

Fig. 3 (b) depicts the workflow of the exchanged messages among: ABNO functions, packet SDN controllers, AS PCE and GMPLS control instances. The outcome is the end-to-end configuration of the whole MLN to carry a new mobile service. The triggering message (with the allocated TEID) is sent by the MME via a REST API (step 1 in Fig. 3 (b)). The message exchange is shown in Fig. 4. The message is processed by the ABNO controller which requests to the PCE (*PCEP PCReq*) the MLN computation (step 2). To do that, the PCE retrieves an updated view of the whole network (packet and optical layers) using the Topology Manager (step 3). If a feasible path is found fulfilling connection demands, a response (*PCEP PCRsp*) is returned to the ABNO controller (step 4).

The computed MLN path establishment is then triggered by the ABNO controller sending the *Packet Connection Establishment Req* message (REST API) to the Provisioning Manager. This message carries the computed path (i.e., nodes, links and resources) along with specific mobile service information, namely, the 5G ENB and SGW/PGW IPv4 addresses and TEIDs. In the MLN, the resulting packet paths transporting the mobile service may require first the establishment of optical tunnels to connect bordering MPLS switches. In this situation, an optical tunnel is configured (step 5) by combining the AS PCE and the distributed GMPLS control plane governing each involved optical node. Additionally, the optical transceivers at both endpoint packet nodes of the optical tunnel are configured (step 6) via REST API. Conversely, if the establishment of a new MPLS packet connection does not need to set up firstly an optical tunnel, the packet connection reuses existing optical tunnels with available bandwidth which favors grooming objectives. In other words, the path computation resorts on the virtual network topology derived from previously created optical connections. Consequently, steps 5 and 6 are not conducted. Finally, the computed MPLS path is established configuring the respective MPLS switches via OpenFlow [20].



```
Source      Destination  Protocc Info
EPC-MME     ABNO-Ctlr    HTTP    POST /restconf/config/calls/call/call_1 HTTP/1.1  (application/json)
ABNO-Ctlr   PCE          PCEP    PATH COMPUTATION REQUEST MESSAGE
ABNO-Ctlr   SDN Pkt Ctl  HTTP    GET /v1.0/topology/links HTTP/1.1
ABNO-Ctlr   SDN Pkt Ctl  HTTP    GET /v1.0/topology/switches HTTP/1.1
ABNO-Ctlr   PCE          TCP     41052→8881 [SYN] Seq=0 Win=29200 Len=0 MSS=1460 SACK_PERM=1 TSval=91055338
PCE         ABNO-Ctlr    PCEP    PATH COMPUTATION REPLY MESSAGE
ABNO-Ctlr   AS PCE       PCEP      Unknown Message (12).
AS PCE      ABNO-Ctlr    PCEP      Unknown Message (10).
ABNO-Ctlr   Transceiver  HTTP    PUT /set channel HTTP/1.1  (application/json)
ABNO-Ctlr   Transceiver  HTTP    PUT /set_channel HTTP/1.1  (application/json)
ABNO-Ctlr   SDN Pkt Ctl  HTTP    POST /stats/flowentry/add HTTP/1.1  (application/json)
ABNO-Ctlr   SDN Pkt Ctl  HTTP    POST /stats/flowentry/add HTTP/1.1  (application/json)
```

Fig. 4. Captured set of control messages exchanged between EPC and ABNO.

IV. SDN/NFV Orchestration Supporting Dynamic Deployment of MVNO

5G slicing allows operators owning the physical infrastructure to dynamically offer isolated and tailored tenants over it to accommodate a myriad of heterogeneous services such as, vertical industries or MVNO. In this section, it is described an SDN/NFV orchestration which processes, computes and deploys virtual packet backhaul networks for supporting MVNO's infrastructure. Each virtual backhaul is independently controlled by a virtualized SDN (vSDN) controller deployed in the cloud. The system is completed enabling the deployment of MVNO's EPC functions as VNFs into the Cloud DC.

A. Targeted Scenario

Figure 5 depicts an example of the targeted deployment of multiple independent (virtual) backhaul tenants over a common physical aggregation MLN. The virtual backhaul connects both the MVNO's RAN to the Core DC domain where VNFs for both vEPC and vSDN are instantiated. We assume that each MVNO owns its RAN, i.e. it is not part of the conducted slicing.

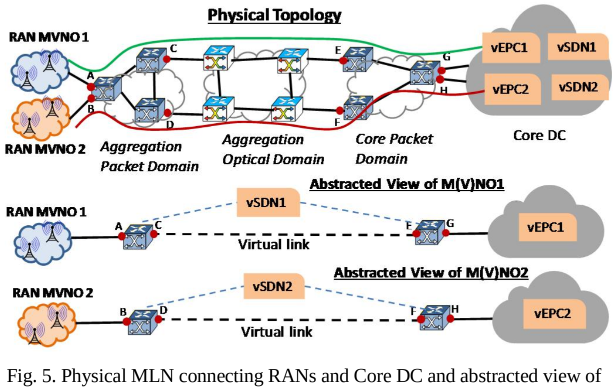


Fig. 5. Physical MLN connecting RANs and Core DC and abstracted view of backhaul network per MVNO.

An MVNO dynamically requests the creation and/or updating of its backhaul tenant according to the mobile traffic demands (e.g., envisaged eMBB) specifying the EPC needs deployed as VNFs. Besides computing and deploying that, a vSDN controller is instantiated enabling the requesting MVNO to control Bearers between the RAN and the vEPC. The vSDN controller has an (abstracted) view of the resulting backhaul tenant. The virtual backhaul is formed by a set of (virtual) packet nodes interconnected by virtual links on top of the physical MLN. In the example, for the MVNO1 its backhaul topology is made up of two virtual MPLS packet nodes (abstracted from the aggregation and core packet domains) which are connected by a virtual packet link over the physical aggregation optical domain.

B. *Deployed SDN/NFV Orchestrator*

The main building blocks forming the SDN/NFV orchestrator to automatically roll out MVNO's backhaul tenants with their respective DC's VNFs is represented in Fig. 6 (a). The NFV Orchestrator (NFVO) as front-end receives and processes MVNO requests. Such requests, as mentioned above, specify the requirements in terms of: network resources (link bandwidth and connectivity) as well as computing resources (VNFs for vEPC). Accordingly, the NFVO triggers the operations to allocate the demanded resources at both the cloud (DCs) and the MLN. In [6], such resources are aggregated and referred to as NFV Infrastructure (NFVI). For the cloud resources, when a VNF needs to be deployed, the NFVO requests it to the VNF Manager [6] which takes over the VNF lifecycle.

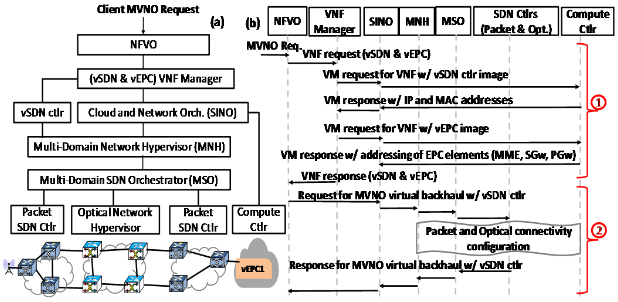

Client MVNO Request
(a)
NFVO
(vSDN & vEPC) VNF Manager
vSDN ctlr
Cloud and Network Orch. (SINO)
Multi-Domain Network Hypervisor (MNH)
Multi-Domain SDN Orchestrator (MSO)
Packet SDN Ctlr
Optical Network Hypervisor
Packet SDN Ctlr
Compute Ctlr
vEPC1
(b)
NFVO
VNF Manager
SINO
MNH
MSO
SDN Ctlrs (Packet & Opt.)
Compute Ctlr
MVNO Req.
VNF request (vSDN & vEPC)
VM request for VNF w/ vSDN ctlr image
VM response w/ IP and MAC addresses
VM request for VNF w/ vEPC image
VM response w/ addressing of EPC elements (MME, SGw, PGw)
VNF response (vSDN & vEPC)
Request for MVNO virtual backhaul w/ vSDN ctlr
Packet and Optical connectivity configuration
Response for MVNO virtual backhaul w/ vSDN ctlr
1
2

Fig. 6. (a) SDN/NFV orchestration for MVNO backhaul tenants; (b) Workflow creating both backhaul tenant and required VNFs (vEPC and vSDN).

In Fig. 6 (a) (bottom) there are the Core DC and the physical MLN. Observe that dedicated SDN controllers are deployed to configure the network elements of each particular domain: packet and optical and cloud/compute. Specifically, three network controllers are considered: 1) an SDN controller for the (MPLS) packet domain connected to the RAN; 2) an Optical Network Hypervisor (ONH) used for configuring the optical network; 3) an SDN controller for the packet network connected to the DC. Additionally, a Compute Controller is the responsible to create the VMs at the DC where the VNFs will be hosted. These controllers form part of the Virtual Infrastructure Manager (VIM) defined in the ETSI NFV architecture [6].

The Multi-Domain SDN orchestrator (MSO) is a unified network operating system enabling the end-to-end service provisioning across multiple domains. The MSO uses an abstracted view provided by each domain SDN controller. Thus, the MSO operates in a hierarchical way, as controller of controllers following the ABNO architecture (Section III.B).

The Multi-Domain Network Hypervisor (MNH) [22] partitions and aggregates the physical domain resources (i.e., nodes, links, optical spectrum, etc.) into virtual packet resources. Such resources are then interconnected to compose the MVNO's backhaul. Furthermore, the MNH provides to the vSDN controller the topology of each MVNO's backhaul.

The Cloud and Network Orchestrator handles the management of both cloud (VMs) and network resources. It uses a southbound interface (REST API) to basically retrieve (abstracted) network topology, serve connectivity requests, and perform end-to-end path computations. For the cloud resources (VMs and VNFs), the Cloud and Network Orchestrator communicates with the Compute controller. The Cloud and Network Orchestrator is aligned with the functionalities supported by the VIM in the ETSI NFV architecture [5] and hereink is termed as the SDN integrated IT and Network Orchestrator (SINO) [22].

C. *Experimental Validation*

Figure 6 (b) shows the implemented workflow among the functional blocks constituting the SDN/NFV orchestrator (SINO). The process is divided into two macroscopic steps:

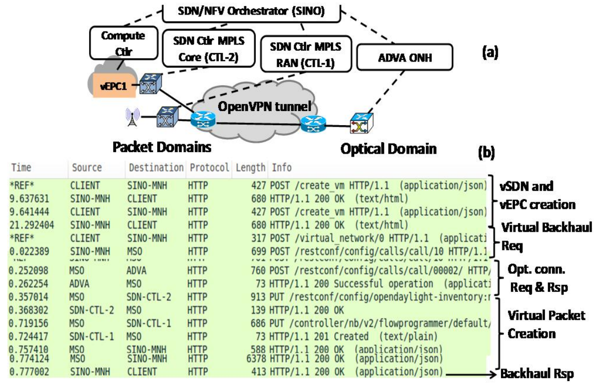
Step 1: The NFVO requests the VNF creation of the vSDN controller (to control via OpenFlow protocol the backhaul tenant) and the vEPC within the DC. These VNF requests are handled by the VNF manager. The VNF Manager communicates with the DC's Compute controller via a REST API, requiring the creation of the VMs (specifying CPU and memory) with the respective operating system image of the VNF implementations. The response determines the IP and MAC addresses of the involved elements and functions: vSDN and vEPC (including PGW, SGW, and MME).

Step 2: The backhaul tenant creation entails both allocating the network resources and enabling the connectivity to the created vSDN (in step 1). To do that, the MNH receives and processes the request (including the IP address of the vSDN). The MNH computes, using the abstract packet view of the MLN, the domain sequence to interconnect both the MVNO's RAN and vEPC within the DC. To this end, the service requirements (peak data rate or maximum tolerated latency) are considered. In the physical MLN, it is first necessary for the traversed packet domains to be interconnected via an optical connection triggered by the MSO. That is, the MNH computes a sequenced set of virtual packet nodes that in the physical infrastructure are connected via an optical domain. This configuration is coordinated by the MSO. When the optical connection is set up (using ONH controller), at the packet level, all domains are interconnected. For those packet domains the MSO subsequently requests packet flow provisioning specifying the ingress/egress links to derive the abstracted (virtual) packet node forming the virtual backhaul. This process is performed twice to support bidirectional packet communications within the backhaul tenant. Finally, a notification is sent to the NFVO. At that moment, the vSDN has a view of the virtual backhaul used to transport both mobile control and user plane traffic between the RAN and the vEPC.

Figure 7 (a) depicts the conducted validation through interconnecting (via OpenVPN tunnel) CTTC SDN/NFV orchestrator located at Barcelona, Spain, and the ADVA ONH in Meiningen, Germany. This work was reported in [23]. For the sake of completeness, the validation is only carried out at the control plane level. The SDN controllers for the packet domains were provided by CTTC as well as the vEPC implementation based on the ns-3 LENA emulator [21], whilst ADVA provided the controller (ONH) for the optical domain.

A capture of the exchanged messages following the workflow detailed above is illustrated in Fig. 7 (b). This set of messages starts with a CLIENT MVNO requesting (using a REST API) to the SINO the allocation of two VMs for hosting vEPC and vSDN VNFs. Once both VNFs are instantiated, another REST API message sent to the SINO-MNH triggers the virtual backhaul network computation and deployment which is then served by the MSO. In this message the IP addressing of both the vEPC and the vSDN within the DC are passed.

The MSO coordinates among the different packet and optical domains the end-to-end connectivity between both the MVNO RAN and the vEPC and the vSDN controller with the SINO-MNH. The connectivity entails the establishment of an optical tunnel between the two packet domains which is handled by the ADVA ONH (simply labelled ADVA in the following). Next, it is necessary to create the packet flows from the MNO's RAN and the deployed vEPC. To do that, the MSO entity communicates with the packet domains' controllers (SDN-CTL-1 and SDN-CTL-2) relying on REST API.



```
Time        Source      Destination  Protocol  Length  Info
*REF*       CLIENT      SINO-MNH     HTTP      427 POST /create_vm HTTP/1.1  (application/json)
9.637631    SINO-MNH    CLIENT       HTTP      680 HTTP/1.1 200 OK  (text/html)
9.641444    CLIENT      SINO-MNH     HTTP      427 POST /create_vm HTTP/1.1  (application/json)
21.292404   SINO-MNH    CLIENT       HTTP      680 HTTP/1.1 200 OK  (text/html)
*REF*       CLIENT      SINO-MNH     HTTP      317 POST /virtual_network/0 HTTP/1.1  (applicati
0.022389    SINO-MNH    MSO          HTTP      699 POST /restconf/config/calls/call/10 HTTP/1.1
0.252098    MSO         ADVA         HTTP      760 POST /restconf/config/calls/call/00002/ HTTP/
0.262254    ADVA        MSO          HTTP       73 HTTP/1.1 200 Successful operation  (applicati
0.357014    MSO         SDN-CTL-2    HTTP      913 PUT /restconf/config/opendaylight-inventory:r
0.368302    SDN-CTL-2   MSO          HTTP      139 HTTP/1.1 200 OK
0.719156    MSO         SDN-CTL-1    HTTP      686 PUT /controller/nb/v2/flowprogrammer/default/
0.724417    SDN-CTL-1   MSO          HTTP       73 HTTP/1.1 201 Created  (text/plain)
0.757410    MSO         SINO-MNH     HTTP      588 HTTP/1.1 200 OK  (application/json)
0.774124    MSO         SINO-MNH     HTTP     6378 HTTP/1.1 200 OK  (application/json)
0.777002    SINO-MNH    CLIENT       HTTP      413 HTTP/1.1 200 OK  (application/json)
```

Fig. 7. (a) Setup between CTTC SDN/NFV orchestrator and the ADVA ONH.; (b) Capture of the control messages for setting up the VNFs and backhaul.

V. Conclusions

This work has reported and justified candidate strategies within the FMC concept as a mean to reduce both CapEx and OpEx investments to efficiently address/support the upcoming and stringent requirements (increased bandwidth, short latency, etc.) as well as the expected requirements imposed by 5G services. FMC is basically attained using both a common infrastructure (specially on the access and aggregation segments) for seamlessly transporting any service type (fixed and mobile) and, adopting generic and unified control and operation functions. To this end, SDN and NFV appear as fundamental enablers. Focusing on an aggregation convergent MLN, we report two implemented SDN/NFV orchestration architectures. The first one allows the automatic accommodation of mobile (and fixed) data flows over the MLN exploiting the advantages of packet statistical multiplexing and optical transport capacity. The second implementation addresses the 5G slicing capability where the physical MLN can be partitioned to compose isolated backhaul tenants used for different MVNOs.

Acknowledgments

This work has been partially funded by the Spanish Ministry MINECO projects DESTELLO (TEC2015-69256-R) and 5G-REFINE (TEC2017-88373-R), and the EU H2020 5G TRANSFORMER project (grant no. 761536).

References

[1] NGMN Alliance, "5G White Paper", February 2015

[2] S. Gosselin, et. al., "Fixed and Mobile Convergence: Which Role for Optical Networks", IEEE/OSA J. of Optical Commun. and Networks (JOCN), vol. 7, no. 11, pp. 1705-1083, Nov. 2015.

[3] C. Behrens, et. al., "Technologies for Convergence of Fixed and Mobile Access: An Operator's Perspective", in IEEE/OSA J. of Optical Commun. and Networks (JOCN), vol. 10, no. 1, pp. A37-A42, Jan. 2018.

[4] EU FP7 ICT IP COMBO "COnvergence of Fixed and Mobile BrOadband access/aggregation networks", http://www.ict-combo.eu/

[5] R. Muñoz, et. al., "CTTC 5G end-to-end experimental platform: Integrating heterogeneous wireless/optical networks, distributed cloud, and IoT devices", IEEE Vehicular Technology Mag., Vol. 11, No. 1, pp. 50-63, Mach 2016.

[6] ETSI, "Network Function Virtualization (NFV)", Oct. 2014.

[7] C- Lange, D. Kosiankowski, and A. Gladisch, "Use-Case based Cost and Energy Efficiency Analaysis of Virtualization Concepts in Operator Networks", in Proc. of ECOC, Spain, Sept. 2015.

[8] M. R. Sama, et. al., "Software-defined control of the virtualization mobile packet core", IEEE Commun. Mag., vol. 53, no. 2, pp. 107-115, 2015.

[9] X. Li, et. al., "5G-Crosshaul Network Slicing: Enabling Multi-Tenancy in Mobile Transport Networks", IEEE Commun. Mag., Vol. 55, Aug. 2017.

[10] J. Ordone-Lucena, et. al., "Network Slicing for 5G with SDN/NFV: Concepts, Architectures, and Challenges", in IEEE Commun. Mag., vol. 55, no. 5, May 2017.

[11] EU H2020 project 5G TRANSFORMER "5G Mobile Transport Platform for Verticals", http://5g-transformer.eu/

[12] E. Weiss, et. al., "Assessment of Fixed Mobile Converged Backhaul and Fronthaul Networks", in proc. of International Conf. on Transparent Optical Networks (ICTON 2016), Trento, Italy, July 2016.

[13] T. Pfeiffer, "Next Generation Mobile Fronthaul and Midhaul Architectures", in IEEE/OSA J. of Optical Commun. and Networks (JOCN), vol. 7, no. 11, pp. B38-B45, Nov. 2015.

[14] A. Mathew, et. al., "Multi-Layer High-Speed Network Design in Mobile Backhaul using Robust Optimization", in IEEE/OSA J. of Optical Commun. and Networks (JOCN), vol. 7, no. 4, pp. 3528-367, April 2015.

[15] J. M. Fabrega, et. al., "Experimental Validation of Mobile Front-/Back-Haul Traffic Delivering with OFDM Transmission and Direct Detection in Elastic Metro/Access Networks using Sliceable Transceivers", in Proc. of ECOC), Gotheborg, Sweden, Sept. 2017

[16] P. Olaszi, et. al., "Fixed-Mobile Convergence: Architecture and Functionality", in Proc. of IEEE High Performance Switching and Routing (HPSR 2015), (Tutorial), Budapest, Hungary, July 1-4,

[17] A. Rostamin, et. al., "Orchestration of RAN and Transport Networks for 5G: An SDN Approach", IEEE Commun. Mag., vol. 55, no. 4, pp. 64 – 70, April 2017.

[18] ETSI GS MEC 002, "Mobile Edge Computing (MEC); Technical Requirements", 2016

[19] D. King and A. Farrel, "A PCE-based architecture for Application-based Network Operations", IETF RFC 7491, March 2015.

[20] R. Martínez, et. al., "Experimental Validation of a SDN orchestrator for the Automatic Provisioning of Fixed and Mobile Services", in Proc. of European Conf. on Optical Commun. (ECOC), Spain, Sept. 2015.

[21] CTTC LTE-EPC Network Emulator (LENA), http://networks.cttc.es/mobile-networks/software-tools/lena/

[22] R. Vilalta, et. al., "Multidomain network hypervisor for abstraction and control of OpenFlow-enabled multi-tenant multi-technology transport networks" IEEE/OSA J. of Optical Commun. and Networks (JOCN), vol. 7, no. 11, pp. B55-B61, 2015.

[23] R. Martínez, et. al., "Integrated SDN/NFV Orchestration for the Dynamic Deployment of Mobile Virtual Backhaul Networks over a Multilayer (Packet/Optical) Aggregation Infrastructure", IEEE/OSA J. of Optical Commun. and Networks (JOCN), vol. 9, no. 2, pp. A135-A142, 2017.

Can OTN be replaced by Ethernet?

A network level comparison of OTN and Ethernet with a 5G perspective

Steinar Bjørnstad
TransPacket AS Oslo, Norway/Simula Research Laboratory Oslo, Norway/Norwegian University of Science and Technology (NTNU)
Institute of Information security and communication
Trondheim, Norway, Steinar.Bjornstad@ntnu.no

***Abstract*—Ethernet has evolved from a protocol for local area network transport to advanced carrier class metro transport as new features are brought in. Recently, industrial, automotive and 5G mobile fronthaul network applications have been addressed. Several new mechanisms are proposed and standardized, e.g. enabling deterministic latency. In light of 5G requirements, this paper reviews and discusses differences between Ethernet and ITU-T G.709 - Optical Transport Network OTN, and analyses Ethernet as an alternative to OTN for optical transport and access network applications.**

Keywords—OTN; carrier Ethernet; deterministic Ethernet; RAN; fronthaul

I. INTRODUCTION

The optical network is constantly evolving into an increasingly number of application areas. While starting in the transport network, it has now evolved into the access network with Fiber To The Home (FTTH) and is also the preferred choice for transport in the mobile Radio Access Network (RAN). SDH/SONET was originally developed for the purpose of transporting voice and data traffic across the optical network. The need for supporting the growing data-traffic and Wavelength Division Multiplexing (WDM) motivated the need for the Optical Transport Network (OTN) G.709 [1] protocol. When OTN was designed, the amount of data-traffic had increased beyond the amount of voice-traffic and a variety of transport protocols were used simultaneously in the network, like e.g. ATM, SDH, PDH and Ethernet. Thus, OTN was designed for transporting all these protocols and is currently the preferred physical layer protocol for optical transport networks.

Ethernet started out as a Local Area Network (LAN) protocol over a shared coaxial cable medium. Since then it has constantly evolved and now stands out as an alternative for telecom networks, especially for metro and mobile RAN transport applications. While the old operators still offer circuit switched services like e.g. PDH and SDH, a heritage from the past, building pure Ethernet packet based transport networks is especially attractive for operators established in a time when data-traffic transport is the dominant service. If some of their customers require transport of circuit services over the packet network, circuit emulation over packet may be applied.

The quality of packet services varies and depends on the load of the network and if Quality of Service (QoS) mechanisms are applied. Packet delay varies with load, and packet loss may occur if the network is congested. Circuit services on the other hand are known for always offering high performance, i.e. low and fixed latency and no packet loss. While packet services today are the dominant service offering compared to circuit services, the performance and reliability requirements are becoming stricter than ever. For the 5G networks, the demand for meeting low latency applications is put forward as one of the main differentiators from previous generations of networks. Furthermore, the vision of the 5G networks includes building the network with a high density of short-range radio access points for achieving high capacity and low latency. For cost efficient operation and network design, this motivates the use of disaggregated RAN, centralizing functionality in a Base Band Unit (BBU), feeding several Remote Radio Heads (RRH) through a so called “fronthaul network” [2, 3]. Recently, a new specification, eCPRI [4] targeting fronthaul networks, was released. While it does not specify a protocol for its transport, two candidate protocols for this are OTN and Ethernet.

In this paper we compare OTN and Ethernet for use in optical transport and access networks in general. Additionally, we discuss the strict delay requirements of the 5G network and how these can be met in optical mobile front and backhaul networks.

II. OPTICAL TRANSPORT AND ACCESS NETWORK REQUIREMENTS

A. *Mobile fronthaul and backhaul delay requirements*

There are two main drivers putting strict delay requirements on mobile fronthaul for 5G networks: the delay sensitive services targeted by the 5G network, and the fronthaul design itself. Figure 1 illustrates the maximum tolerable delay of some delay-sensitive applications that will need to be supported by both future backhaul and fronthaul networks.

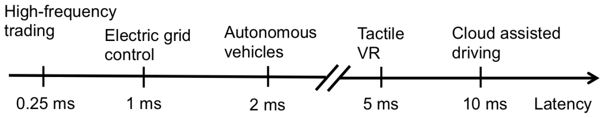


Figure 1. Delay sensitive applications, adapted from [5]

As seen in the figure, there are applications tolerating delays of 1 ms or less among the 5G target applications. The delay requirements in eCPRI-based fronthaul are even stricter. For fronthaul transport with split in the physical layer, as found in CPRI over Ethernet [6] and in eCPRI option “D” and “E” [4],

978-3-903176-07-2 © 2018 IFIP

the Hybrid Automatic Retransmit reQuest (HARQ) protocol sets restrictions on maximum delay between the RRH and BBU. In [7], a one-way delay of 123 μs is found as the maximum. In [4] and [8] an even stricter delay requirement of 100 μs one-way delay is set as a requirement.

B. Optical transport network requirements

While mobile backhaul and fronthaul requires transport over moderate distances, typically below 100 km, the optical backbone network offers transport over several hundred or thousands of kilometers. Hence, a dominant delay-component in the backbone network is the delay in the fibre itself, given as 5 μs/km. A key feature for long distance transport is the Forward Error Correction (FEC), enabling correction of bit-errors due to physical impairments along the optical transport path. Furthermore, Operations Administration and Management (OAM) functionality is of high importance for detecting and communicating errors, as well as characterizing performance of the network. While FEC is of highest importance for long distance transport, where physical impairments have the greatest impact, OAM capabilities are important also for metro and access networks. Furthermore, when carriers are offering services, bandwidth isolation between these services for avoiding interference between traffic of different customers is desirable. In addition, carriers with a history in offering SDH/SONET services are still offering transport of legacy protocols like e.g. SDH, PDH, InfiniBand, ATM and Ethernet. While the SDH/SONET network was natively synchronous, today mobile networks also demand frequency and time synchronization [8]. This may be supported locally by synchronizing using GPS. However, a GPS signal may be difficult to distribute, especially to base-stations located e.g. within large buildings. Furthermore GPS may be disturbed by jamming or e.g. solar storms. Hence, because of security and reliability reasons distributing time and frequency in the network is desirable.

III. OTN Functionality

OTN has inherited many functions from SDH/SONET. The data streams to be transported are framed into containers of fixed length, encapsulating the payload together with fields containing additional information. This information enables e.g. OAM for wavelengths, universal container supporting any type of service, communication channels for control traffic, end-to-end optical transport transparency of customer traffic and multi-level path OAM [9]. The OAM functionality has a number of features. This includes e.g. end-to-end path monitoring using parity check: Bit Interleave Parity (BIP) for finding bit errors in the Optical Payload Unit (OPU). Furthermore, the Tandem Connection Monitoring (TCM) is a powerful tool enabling monitoring across different networks and operator domains by using up to six dedicated fields for error checking. Six independent tandem connections may then be monitored, allowing both overlap and nesting of the connections [10]. The TCM allows carriers to define their own path layers for monitoring, enabling paths to go across different networks and operator domains. As an example, a connection belonging to operator A, but crossing three operator networks on its way: A, B and C is illustrated in Figure 2. Carrier A uses the end-to-end path monitoring for monitoring the customers signal from the entry to the exit of the network. Carrier A additionally uses TCM2 for monitoring the signal when crossing carrier B and C. Carrier B uses TCM1 to perform path monitoring at the entry and exit points of its networks. Likewise, Carrier C may re-use TCM1 to perform path monitoring on the signal as it enters and exits the Carrier C network [9].

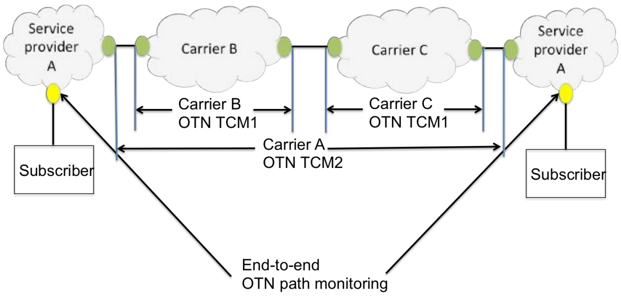


Figure 2. Example of TCM and end-to-end path monitoring in OTN.

The smallest container defined in OTN is the Optical Data Unit 0 (ODU0), operating at 1.25 Gb/s [1]. Hence, this defines the smallest channel rate in OTN, resulting in waste of bandwidth if trying to map a channel of lower bitrate into an OTN channel. OTN is suited for multiplexing client signals of 1 Gb/s bitrate and beyond into higher bitrate line-signals. On the line-side, OTN supports 2.5 Gb/s, 10 Gb/s, 40 Gb/s and 100 Gb/s. A standard multiplexing hierarchy exists enabling a mapping structure defining how to map client signals of different bitrates into higher bitrate line signals. For all channels transported in OTN yields the same benefit of full transparency and bandwidth isolation.

Furthermore, while OTN was originally a point-to-point transport and grooming technique, OTN switching is now available. Transparent switching of the client channels independent of type of service and the transported protocol is achieved. Hence, switching of fully monitored virtual links is enabled since performance and alarm monitoring capabilities are preserved end-to-end.

The General Communication channels (GCC1 and GCC2) allow communication between two network elements having access to the ODU frame structure. Since the communication channel is based on using reserved fields within the frames, both bandwidth and communication is guaranteed independent of payload content and network load.

Forward Error Correction (FEC) enables detection and correction of errors in an optical link caused by physical impairments in the transmission medium. When using FEC, a lower signal quality in the link can be accepted, e.g. by adding a 7 % FEC overhead, a gain in power level of approximately 5 dB is achieved [9]. FEC is a powerful tool in OTN. A higher gain in power level than the FEC first defined for OTN is now available. This is especially attractive for sub-sea systems where power margins are a scarce resource.

IV. Ethernet Functionality

In Carrier Ethernet, new functions has been brought in for making Ethernet more suitable for network operators building Metropolitan Area Networks (MAN) and Wide Area Networks (WAN) [11]. Ethernet is not only applicable for point-to-point transport, like OTN. Carrier Ethernet also defines point to multipoint and multipoint to multipoint transport. Furthermore, while OTN is based on framing data into fixed length frames in fixed data-rate channels, Ethernet allows framing of data of variable bitrate into variable length frames.

A main difference between OTN and Ethernet is how multiplexing is performed. OTN always applies static multiplexing of lower bitrate channels into higher bitrate channels. Ethernet typically applies statistical multiplexing, allowing efficient multiplexing of variable bitrate channels with statistically distributed packet arrival patterns. While this allows for efficient multiplexing using buffers for smoothing out packet bursts, buffering adds a delay depending on the traffic patterns. This is a challenge for some applications, like e.g. mobile fronthaul, having very strict requirements to packet delay and packet delay variations. However, Ethernet allows a number of different ways of doing multiplexing since a single method is not explicitly defined in the IEEE 802.1Q [12] standard defining Ethernet. As an example, for each output interface, one output queue may be assigned per input interface. A multiplexing method is then to go round-robin on queues, scheduling packets from the queues one-by-one to the output interface. Hence, if packets arrive simultaneously at the input interfaces but destined for the same output, one or more packets must stay in their queues before being scheduled to the output. Because the buffering delay then varies according to how many packets are arriving simultaneously at the inputs, this causes packet delay variation (PDV). Furthermore, if the volume of traffic being multiplexed to an output interface is larger than the bandwidth of the interface, queues will fill up resulting in packet loss and high delays. While this may be sufficient for e.g. Internet applications like web-browsing applying TCP for retransmission, it is not sufficient for time and loss -sensitive applications.

A. Making Ethernet deterministic

Recently, a number of mechanisms have been proposed enabling zero packet loss and a low and even fixed delay in Ethernet. This has especially been attractive for industrial applications of Ethernet, named “deterministic Ethernet”. In Integrated Hybrid (hybrid as in packet and circuit) Optical Networks (IHON) [13], mechanisms addressing optical transport with zero packet loss and fixed delay are proposed and explored for Ethernet. In the IEEE 802.1 Ethernet standardization group, mechanisms ensuring zero congestion packet loss, as well as bounded delay and PDV are proposed. Recently, main drivers for the evolvement in standardization include industrial control and automotive applications, with mobile fronthaul as the most recent.

1) Deterministic delay

In the IEEE 802.1 work, Time Sensitive Network (TSN) mechanisms include both mechanisms for minimizing delay and for controlling the delay variation, ensuring that all priority packets receive low and bounded delay. The IEEE 802.1Qbu [14] defines a preemption mechanism enabling minimized delay on deterministic traffic when mixed with best-effort traffic within the same network. By disrupting the transmission of best-effort packets when a time-sensitive high priority packet arrives, packet delay caused by packet contention is lower than e.g. the strict priority mechanism where maximum delay corresponds to the duration of a best-effort Maximum Transfer Unit (MTU) packet. Preemption is only performed if at least 60 bytes of the pre-emptable frame are already transmitted and at least 64 of the frame remain to be transmitted, resulting in a worst case delay of 155 bytes and best-case zero delay [8]. Hence, PDV correspond to the duration of transmitting 155 bytes. Preemption works hop-by-hop, reassembling incoming and fragmenting outgoing packets at every hop. Since fragments do not contain e.g. MAC address-headers, forwarding of fragments through bridges is not supported, i.e. preemption may only be activated with bridges supporting the IEEE 802.1Qbu standard.

The IEEE 802.1Qbv [15] (enhancement for scheduled traffic) defines how a set of queues, destined for an output port, may be served by a round-robin mechanism; As opposed to a round-robin scheduling, where delay depends on the number of queues populated with packets, it allows each queue to be served within a dedicated timeslot. One-by-one in a cycle of timeslots, one or more packets are scheduled in bursts from each of the queues into their designated time-slot. The duration, and hence, start of the time-slots, is deterministic. Moreover, time-synchronization is required, e.g. using the IEEE 1588 protocol [16]. The maximum delay on a packet is given by the duration of the scheduling cycle.

A mechanism not relying on packet preemption, while enabling a mix of deterministic traffic and best-effort traffic in a network, is a time-window based priority mechanism described for IHON. The mechanism eliminates PDV on the time-sensitive traffic by adding a fixed delay corresponding to the MTU of the best effort traffic. Best effort packets are scheduled in between time-sensitive packets whenever a gap is available that is equal to- or larger than- the packet waiting in a best effort queue. Thus, any interference and PDV on the time-sensitive traffic caused by best effort traffic is eliminated. As opposed to preemption, the mechanism allows packets to be transmitted also through bridges not supporting the time-window mechanism, achieving lowered PDV in the network for each node that it is applied to.

Furthermore, IHON describes an aggregation and scheduling mechanism where PDV from contention is avoided. The mechanism relies on preserving the packet gaps between packets in the individual deterministic packet streams.

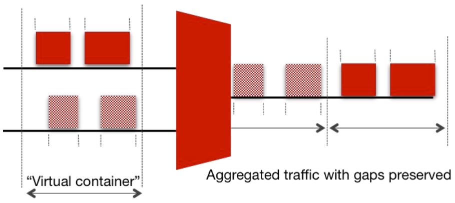


Figure 3. Aggregation of multiple deterministic packets streams into virtual containers while preserving packet gaps.

Packet streams being aggregated are scheduled into time-slots in a cycle synchronized across the network using a control packet at the start of each time-slot. However, the packet streams aggregated are allowed to be asynchronous with variable length packets and still transferred with no added PDV. As illustrated in Figure 3, the streams being aggregated are divided up into virtual containers, fitted into time-slots, before being scheduled to the output. A minimum fixed delay corresponding to one cycle time is added to each of the packet streams.

2) Ethernet performance examples

An IHON field-trial demonstrates deterministic aggregation of 1 Gb/s into 10 Gb/s Ethernet, transmission through 3.25 km of fibre and de-aggregation back to 1 Gb/s with load independent end-to-end delay of 67.22 µs and PDV of 160 ns [13]. Furthermore, experiments have been performed demonstrating combined fronthaul and backhaul traffic in a 100 Gb/s Ethernet wavelength [21]. The fronthaul traffic receives a low latency and ultra low PDV independent of load, while the less time-critical backhaul traffic experiences a higher latency and PDV. For the emulated fronthaul traffic, delay through one node was measured to 1.3 µs and PDV to 0.2 µs, independent of fronthaul and backhaul traffic loads. Hence, at 100 Gb/s speeds even tens of hops can be allowed, still meeting the 100 µs fronthaul delay limit.

3) Avoiding packet loss by controlling bandwidth

For carrier Ethernet applications, policing mechanisms for controlling the bandwidth into the network are defined [12]. A policer is a mechanism limiting the bandwidth into and/or out of a queue, enabling a service provider to offer sub-rate bandwidth services with a lower bitrate than the physical bitrate of the interface being offered. I.e. policing allows the bandwidth offered being any bandwidth equal to- or lower than- the bandwidth of the interface. A guaranteed offered bandwidth is defined as a Committed Information Rate (CIR), where packet loss in the network due to contention and full buffers should not occur. An Excess Information Rate (EIR) defines additional traffic being allowed, but that may be dropped in a congested network. Traffic exceeding the EIR is always dropped.

B. Framing legacy formats in Ethernet

The Ethernet standard [12] does not define framing of legacy formats like TDM and ATM into Ethernet frames. Circuit emulation techniques do exist for Ethernet, but applying an MPLS layer on top of Ethernet for multiservice transport is a more common approach [9]. For MPLS, circuit emulation techniques are defined enabling transport of legacy signals, sharing the links with the IP/Ethernet based data. Ethernet networks may therefore not be an efficient choice if e.g. mainly legacy services are to be transported. However, when the amount of legacy services are minor, Ethernet with circuit emulation support is likely to be the best choice for the future.

C. OAM in Ethernet

Both link OAM [17] and end-to-end service monitoring, service OAM [18], are defined for Ethernet. Different administrative levels are defined allowing different user types accessing different Service OAM capabilities. These levels are called Maintenance Entity Groups (MEGs) in the ITU-T Y.1731 [17] standard. Eight levels of MEGs are defined, allowing different levels to be applied across different service providers and between subscribers. This is applied for monitoring Ethernet Virtual Connections (EVC) or Operator Virtual Connections (OVC) defined by their Maintenance Endpoint (MEP). Maintenance Intermediate Points (MIPs) are placed between MEPs and used at internal interfaces in the carriers for additional troubleshooting purposes. This is illustrated in Figure 4; all parties are capable of individual monitoring of their service: the subscriber, the carrier delivering the service across the network (service provider A), and the individual carriers involved, carriers B and C. Performance parameters being monitored are packet loss, packet delay and packet delay variation.

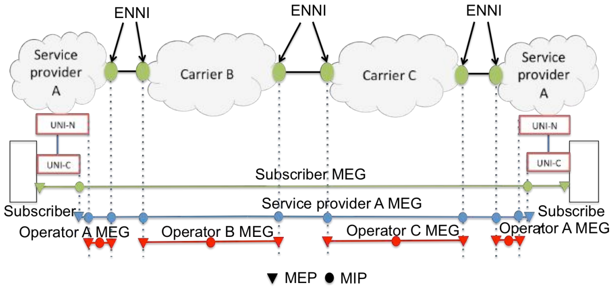


Figure 4. Different MEG levels applied between different service providers and the subscriber. ENNI: Ethernet Network-Network interface. UNI-N: User Network Interface Network side. UNI-C: User Network Interface Customer side.

D. Ethernet fault management

End-to-end connectivity Fault management for Ethernet is defined in [18]. Two important tools are continuity check and link trace. For continuity check, Continuity Check Messages (CCMs) are exchanged between MEPs. The rate of the CCM messages may be set high, enabling availability being measured every 10 ms, or even more frequent. Link trace sends Link Trace Messages (LTM) over EVCs or OVCs. The MEP and MIPs along the EVC/OVC return a Link Trace Respond message, confirming the MEP/MIP points availability. Hence, this enables a precise fault location within the network.

E. Ethernet and FEC

FEC has not been a part of Ethernet until recently. As bitrates are increasing to 100 Gb/s and beyond, FEC becomes a requirement for achieving sufficiently long reach. For 100 Gb/s, IEEE 802.3bm [19], and for 200 Gb/s and 400 Gb/s, IEEE 802.3bs-2017 [20], FEC is defined as a feature.

V. COMPARISON OTN AND ETHERNET

In this section we compare the properties and functionality described for OTN and Ethernet. An overview of the properties is given in table 1. As can be observed both OTN and Ethernet support most of the listed features. There are however some major differences, mostly related to OTN being a circuit type of transport only. I.e. OTN only supports static multiplexing with 1.25 Gb/s as a minimum bandwidth on point-to-point services. Ethernet on the other hand, supports both static and statistical multiplexing of any bandwidth and both point-to-point and point-to-multipoint transport services. While the OTN static multiplexing is known to enable absolute guarantees on the services: zero packet loss, fixed and low delay, Ethernet may achieve the same properties using IHON mechanisms, which enables static multiplexing. In addition, IHON may be used for inserting packets in gaps between packets in a static multiplexed packet stream. Statistical multiplexing may then be combined with static multiplexing, increasing throughput without imposing delay variations or packet loss on the static multiplexed packet stream. For a mobile fronthaul application, the bandwidth granularity of OTN is sufficient for transport of eCPRI rates and statistical multiplexing may not be required for this purpose. However, if fronthaul and backhaul is combined within the same link, the throughput of the backhaul transport may benefit from the statistical multiplexing capability since strict delay guarantees may not be required for the majority of the backhaul traffic volume. While for metro and backbone network transport purposes the bandwidth granularity of OTN may be sufficient, offering enterprise and residential services typically requires higher bandwidth granularities in the order of tens of Mb/s that can be satisfied by Ethernet. Furthermore, especially in the metro and access network, carriers may benefit from the statistical multiplexing of Ethernet efficiently aggregating traffic at the edge of the network.

Looking into OAM and fault handling capabilities, both OTN and Ethernet are equipped with a powerful set of tools for ensuring and documenting delivery of customer services crossing multiple carrier network domains. A major difference is however that OTN monitors errors at a bit-level while Ethernet monitors at a packet level. This makes OTN OAM more suitable for characterizing physical link quality while Ethernet OAM is more suitable for revealing congestion in Ethernet nodes. While OTN monitors bit-errors only, Ethernet OAM may be used for documenting packet loss, delay and delay variation of services.

Table 1. Comparison of features for OTN and Ethernet

Feature	**OTN**	**Ethernet**
Legacy service transport	Framing any service	Additional circuit emulation protocol required
Packet service transport	Fixed rate circuit	Native packet - variable rate statistical multiplexing
Connectivity type	Point-to-point	Point-to-point Point-to-multipoint Multipoint-to-multipoint
Granularity of bandwidth	Min. 1.25 Gb/s (ODU0)	Any bandwidth
Time-sensitive application support	No buffering for contention: Low and fixed latency	Low and fixed latency using IHON. Bounded delay using IEEE TSN mechanisms.
Multiplexing type	Static	Static and/or statistical
Switching capability	Circuit switching, 1.25 Gb/s granularity.	Packet switching with packet granularity.
Operation and Maintenance	End-to-end Path and 6 levels of TCM	End-to-end Service monitoring and 8 MEG levels for EVC monitoring
Parameters monitored	Bit errors	Packet loss, delay, delay variation
Fault management	Monitor mode TCM	Continuity check and link trace
Error correction	Correction of bit errors using FEC	FEC available for 100 Gb/s and beyond.

Furthermore, OTN is always equipped with a FEC code, enabling a high tolerance to signal quality degradation in the link. This especially comes in handy on long distance optical links where signal quality is degraded by noise from optical amplifiers and non-linear physical transmission impairments in the fibre. For mobile fronthaul and backhaul, as well as metro and access network distances, amplifier noise and transmission impairments are less of a problem, and FEC and physical link monitoring therefore typically become less important. For very high bitrates of 100 Gb/s and beyond, Ethernet also defines FEC as a feature. However, long distance transport beyond 10 km is currently not defined for these bitrates. Hence, OTNs FEC enables benefits for long distance transport networks while Ethernet will be sufficient for metro, access and mobile transport network purposes.

VI. Summary and Conclusion

In this paper OTN and Ethernet network functionality has been compared with respect to applications including longhaul, metro, access and mobile fronthaul and backhaul. Because OTN natively defines how to frame a number of different protocols into OTN frames, it is more suitable than Ethernet for transport of legacy services. We expect however this to become less relevant for future networks. We find that using the functionality added to Ethernet through Carrier Ethernet, it now offers the same level of OAM functionality as OTN. Furthermore, OTN with static multiplexing supports a zero packet loss, low and fixed latency transport with full isolation between services. This is however also achieved in Ethernet using the IHON mechanisms. Furthermore, while providing the same level of deterministic service as OTN, Ethernet may additionally allow higher throughput utilization through statistical multiplexing using IHON mechanisms. OTNs Forward Error Correction capability is known to extend the reach of long-haul transport and is available for all OTN rates. For high Ethernet rates, 100 Gb/s and beyond, FEC is added, opening up for the same benefits as earlier only found for OTN. For these bitrates current maximum distance defined for Ethernet is 10 km.

OTN therefore shows benefits for legacy service and long-haul transport. For network segments less sensitive to physical transmission impairments, including metro, access and mobile backhaul and fronthaul, we find Ethernet to deliver the same level of service quality and availability while supporting a higher throughput efficiency than OTN. Hence, our conclusion is that today Ethernet is a beneficial choice for mobile transport, access and metro networks while only OTN is defined for high bitrate long-haul transport. Furthermore, as Ethernet today also contains FEC, up to now the prime OTN benefit for longhaul, it may replace OTN in the future for long-haul if IEEE chooses to define long-haul Ethernet interfaces.

Acknowledgment

This work is part of research project 5G-PICTURE supported by the European Horizon 2020 initiative.

References

[1] ITU-T, G. 709, Interfaces for the optical transport network. June 2016.

[2] A. Pizzinat *et al*., "Things You Should Know About Fronthaul", *IEEE/OSA J. Lightwave Technol*., vol. 33, 2015, pp. 1077-1083.

[3] A. Checko et al., "Cloud RAN for Mobile Networks—A Technology Overview", *IEEE Comm. Surveys & Tutorials*, vol. 17, no. 1, 2015, pp. 405-426.

[4] Common Public Radio Interface (CPRI) Specification, Sept. 2017, http://www.cpri.info/spec.html.

[5] Nokia, "5G ultra-low latency infographics", Aug. 2017; https://resources.ext.nokia.com/asset/201030.

[6] IEEE Std. P1914.3, "Radio Over Ethernet Encapsulations and Mappings," Sept. 2009; http://sites.ieee.org/sagroups-1914/p1914-3/

[7] H. J. Son, and S.M. Shin, "Fronthaul Size: Calculation of maximum distance between RRH and BBU", Sept. 2017; http://www.netmanias.com/en/post/blog/6276/c-ran-fronthaul-lte/fronthaul-size-calculation-of-maximum-distance-between-rrh-and-bbu.

[8] IEEE Std. P802.1CM, "Time Sensitive Networking for fronthaul", Sept. 2017; http://www.ieee802.org/1/pages/802.1cm.html. .

[9] Fujitsu white-paper: "The key benefits of OTN", available online: https://www.fujitsu.com/us/Images/OTNNetworkBenefitswp.pdf, accessed 27/01-2018.

[10] Viavi white-paper: Andreas Schubert "G.709 – The optical transport network (OTN)", available online: https://www.viavisolutions.com/en-us/literature/g709-optical-transport-network-otn-white-paper-en.pdf, Accessed 27/1-2018.

[11] Fujitsu white-paper: "Carrier Ethernet essentials", available online: https://www.fujitsu.com/us/Images/CarrierEthernetEssentials.pdf, accessed 27/1-2018.

[12] IEEE 802.1Q standard "IEEE 802.1Q bridges and bridged networks"

[13] R. Veisllari *et al*., "Field-Trial Demonstration of Cost Efficient Sub-wavelength Service Through Integrated Packet/Circuit Hybrid Network [Invited]", *IEEE/OSA J. Opt. Comm. Net*., vol. 7, no. 3, Mar. 2015, pp A379-A387.

[14] IEEE Std. P802.1Qbu, "Standard for Local and Metropolitan Area Networks-Media Access Control (MAC) Bridges and Virtual Bridged Local Area Networks - Amendment: Frame Preemption", Jul. 2015; http://www.ieee802.org/1/pages/802.1bu.html.

[15] IEEE Std. 802.1Qbv, "Standard for Local and Metropolitan Area Networks-Media Access Control (MAC) Bridges and Virtual Bridged Local Area Networks Amendment: Enhancements for Scheduled Traffic", Mar. 2016; http://www.ieee802.org/1/pages/802.1bv.html

[16] IEEE Std. 1588-2008, "Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems", Jul. 2008.

[17] ITU-T recommendation Y-1731 "Performance monitoring in a service provider network"

[18] IEEE 802.1ag: IEEE 802.1ag (also CFM) (IEEE Standard for Local and Metropolitan Area Networks Virtual Bridged Local Area Networks Amendment 5: Connectivity Fault Management"

[19] IEEE 802.3bm: "IEEE Standard for Ethernet Amendment 3: Physical Layer Specifications and Management Parameters for 40 Gb/s and 100 Gb/s Operation over Fiber Optic Cables - IEEE Standard for Ethernet - Amendment 3: Physical Layer Specifications and Management Parameters for 40 Gb/s and 100 Gb/s Operation over Fiber Optic Cables."

[20] IEEE 802.3bs-2017: "IEEE Std 802.3bs-2017 (Amendment to IEEE 802.3-2015 as amended by IEEE's 802.3bw-2015, 802.3by-2016, 802.3bq-2016, 802.3bp-2016, 802.3br-2016, 802.3bn-2016, 802.3bz-2016, 802.3bu-2016, 802.3bv-2017, and IEEE 802.3-2015/Cor1-2017) - IEEE Standard for Ethernet Amendment 10: Media Access Control Parameters, Physical Layers, and Management Parameters for 200 Gb/s and 400 Gb/s Operation."

[21] R. Veisllari *et al*., "Experimental Demonstration of 100 Gb/s Optical Packet Network for Mobile Fronthaul with Load-independent Ultra-low Latency", In proceedings of *ECOC 2017*.

Machine Intelligence in Allocating Bandwidth to Achieve Low-Latency Performance

Lihua Ruan and Elaine Wong

Department of Electrical and Electronic Engineering, The University of Melbourne,
VIC 3010, Australia. ewon@unimelb.edu.au

***Abstract*— In this work, we present a complete rethink of the decision-making process in allocating bandwidth in a heterogeneous Fiber-Wireless network with machine intelligence. We highlight the use of an artificial neural network (ANN) at the central office to learn the uplink latency performance using multiple network and packet features. In turn, the trained ANN enables the central office to facilitate flexible bandwidth allocations under diverse network scenarios in meeting low-latency communication demands.**

***Index Terms*—Artificial neural network; dynamic bandwidth allocation; fibre-wireless networks, machine learning; low latency; Tactile Internet.**

I. Introduction

The Tactile Internet era is igniting an explosion of real-time, remotely controlled human-to-machine (H2M) and machine-to-machine (M2M) applications [1]-[4]. To support low latency (in the order of milliseconds, ms) and highly-reliable delivery of control/sensor-oriented traffic typical of such applications, we have previously considered the delivery of traffic over converged Fiber-Wireless (FiWi) networks along with the relocation of control servers closer to the end users [5]-[6] to expedite feedback and response.

An illustration of the heterogeneous FiWi network considered in our work is shown in Fig. 1. In the FiWi network, uplink bandwidth is shared amongst many optical network units (ONUs) that support aggregated wireless local area traffic from multiple end users. The process of allocating bandwidth to and scheduling transmission from each of these end users thus influence the overall latency. In this respect, the decision making process in allocating bandwidth and scheduling transmission is critical in meeting strict latency requirements and thus warrants attention.

Dynamic bandwidth allocation (DBA) schemes in fiber access networks are commonly centralized at the central office (CO) to schedule bandwidth resources for uplink transmissions. The bandwidth allocated to individual ONUs is typically determined based on the requested bandwidth in the REPORT message sent from each ONU [7]. Efforts in reducing uplink latency have been previously reported in [7]-[10], by predicting bandwidth demand based on the information in the REPORT messages and on arrival traffic characteristics. Statistical prediction methods such as constant credit and linear credit [7], arithmetic average [8], exponential smoothing [9] and Bayesian estimation [10], have been used in DBA schemes to predict bandwidth demand and subsequently to facilitate bandwidth allocation decisions. However, the limitation of these existing algorithms lies in their use of single traffic/network features, e.g. packet arrival rate or aggregated traffic load, to predict bandwidth demand. When network traffic load, packet length, and/or network configuration such as CO-to-ONU distance vary, the effectiveness of these algorithms in predicting bandwidth demand is compromised. Research in [11] and [12] explicitly reported on the challenge in determining the appropriate bandwidth to be allocated when network/traffic parameters vary.

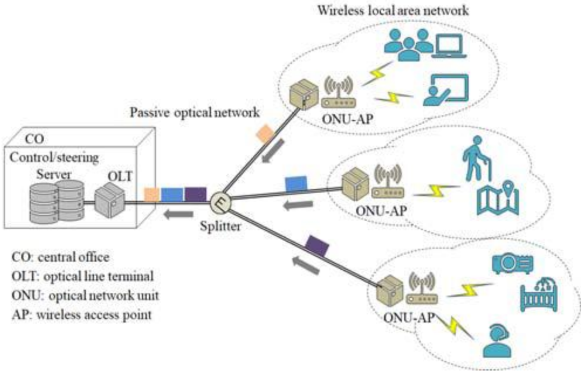

Wireless local area network
Passive optical network
ONU-AP
CO
Control/steering Server
OLT
Splitter
ONU-AP
ONU-AP
CO: central office
OLT: optical line terminal
ONU: optical network unit
AP: wireless access point

Fig. 1. An illustration of a heterogeneous wireless local area and optical access network architecture for converged service delivery

In this work, we present a complete rethink of the decision-making process of allocating bandwidth with machine intelligence. Although machine learning (ML) techniques have been recently adopted in traffic routing, post-processing of signals, network failure prediction, the capability of machine intelligence in benefiting bandwidth resource allocation still remains an open question. For illustrative purposed, we show

978-3-903176-07-2 © 2018 IFPF

in this work, the exploitation of an artificial neural network (ANN) in (a) learning network uplink latency performance using diverse and multiple network features and in turn, (b) facilitating flexible bandwidth allocation decisions that effectively reduce the uplink latency under various network scenarios.

II. Artificial Neural Network Facilitated Dynamic Bandwidth Allocation (DBA)

A. DBA in Heterogeneous Networks

In a typical DBA algorithm, the CO grants an amount of bandwidth through a GATE message to each ONU upon receiving the REPORT messages from ONUs in the previous polling cycle(s). A polling cycle is defined as the time interval between consecutive transmissions from an ONU. Early works on predicting bandwidth demand have used a limited-service approach whereby the CO would use the requested bandwidth BW_{req} from REPORT messages to estimate the bandwidth demand, BW_{dem}. As discussed in Section I, statistical prediction methods such as constant credit and linear credit [7], arithmetic average [8], exponential smoothing [9] and Bayesian estimation [10], have been used to estimate BW_{dem}. In these early works, once BW_{dem} is obtained, the CO would subsequently grant $\min\{\mathrm{BW}_{dem}, \mathrm{BW}_{max}\}$ to the ONUs in the next polling cycle. Here, BW_{max} is the maximum bandwidth that can be allocated by the CO to the ONUs.

A major challenge of limited-service DBA algorithms lies in estimating an accurate BW_{dem} since bandwidth over-granting or likewise under-granting due to an inaccurate bandwidth prediction, may potentially increase uplink latency. Compounding the issue is that to the accuracy of $\mathrm{BW}_{\mathrm{dem}}$ depends on multiple network features, e.g. statistics of packet length, network traffic load and network configuration. It is also complex to derive BW_{dem} using conventional mathematical or analytical methods.

B. ANN Learning and Decision-Making Model

Here, we present an ANN learning and decision-making model and show how machine intelligence can be used to predict BW_{dem} with high accuracy. BW_{dem} can be resolved into two bandwidth components as shown below:

$$\mathrm{BW}_{dem} = \mathrm{BW}_{req} + \lambda T_{POLL}(\alpha S_{min} + (1-\alpha)S_{max}) \quad (1)$$

where the first term on the right hand side, BW_{req} is the requested bandwidth in the REPORT message from each ONU, and the second term on the right hand side is the predicted bandwidth. T_{POLL} is the polling cycle duration of an ONU. $S_{\max}$ and $S_{\min}$ are the maximum and minimum packet length, respectively. The parameters λ and α ($0 \leq \alpha \leq 1$) are arrival rate and the defined prediction coefficient, respectively. Our ANN learning and decision-making model predicts the second term on the right-hand side of (1) and hence $\mathrm{BW}_{\mathrm{dem}}$, to yield the lowest uplink latency through selection of α.

An ANN comprises an input layer, an output layer and some hidden layers in between, and learns by iteratively adjusting its weight and bias associated with the neurons in each layer to yield a desired output. An ANN learns complex nonlinear relationships between the input features, and yields a target output. A schematic of our proposed ANN learning and decision-making model is presented in the top diagram of Fig. 2. It must be noted that the average uplink latency over heterogeneous networks is impacted not only by α but also by diverse network features. Therefore, to train our ANN, we use the following key *input features*:

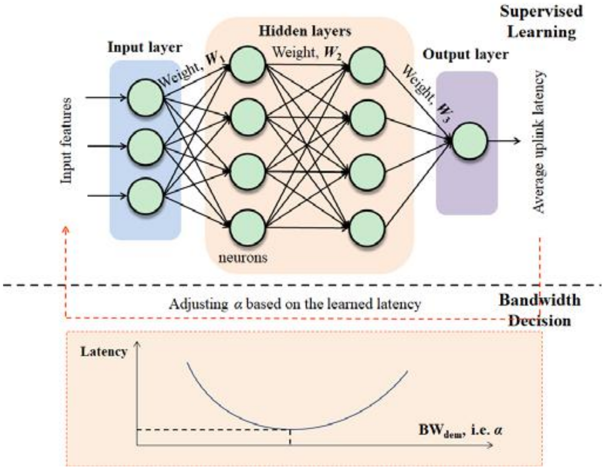


Fig. 2. An illustration of the proposed ANN learning and bandwidth decision-making model.

- $S_{\max}/S_{\min}$ — maximum/minimum packet length
- $S_{\mathrm{avg}}/S_{\mathrm{var}}$ — mean/variance packet length
- λ — packet arrival rate in the wireless local area network
- α — prediction coefficient
- N — The number of ONUs
- $D_{\max}$ — The maximum CO-to-ONU distance
- R_{PON} — Data rate of the passive optical networks (PON)
- R_{WLAN} — Data rate of the wireless local area networks (WLAN)

The target output is the average uplink latency of the network. As such, we train an ANN to learn the latency performance associated with different BW_{dem} decisions through varying α. When supervised learning is complete, the trained ANN predicts the average uplink latency for any α value that can possibly be selected (refer to bottom diagram of Fig. 2), thereby enabling α that yields minimum latency to be solved. The CO then allocates bandwidth with the BW_{dem} solution corresponding to the selected α. In the following section, we show how the supervised training can be implemented and highlight latency improvements achieved by a DBA algorithm facilitated by the trained ANN. This DBA algorithm is termed ANN-DBA for clarity.

III. Latency Performance Improvement

A. Supervised Training

We use a training set generated with varying input features to train an ANN with three hidden layers. The number of neurons of the three hidden layers is 5, 10, and 5, respectively. The target output of a training sample is the average uplink latency over a 1000-ms network running time (approximately 1000 polling cycles times), corresponding to a given input network feature in an event-driven packet-level simulation environment. With the knowledge of the dependence between uplink latency performance and the selection of α learnt by the trained ANN, the bandwidth allocation decision in (1) can be performed by finding α that minimizes latency.

For illustrative purposes, we report on the training process and decision-making outcome of a 16-ONU PON-WLAN network when λ and α change. We first use a training set containing 100 samples generated in an event-driven packet-level MATLAB simulation environment. The target output of a training sample is the average uplink latency, D_{uplink}, over a 1000-ms network running time, i.e. around 1000 polling cycles times. A network configuration comprising 16-ONUs with 10-km CO-to-ONU distance, packet lengths that are uniformly-distributed between 64 and 1518 bytes, and data rates of 1 Gbps and 100 Mbps for the optical and wireless segments respectively [7], are considered for illustrative purpose.

Another 250-sample test set was generated to validate the training outcome. The input features of the test set was fed to the trained ANN. The ANN predicted latency values were compared with the target latency values provided by the test set. Fig. 3 illustrates the prediction error arising from our use of the trained ANN, the mean square error of which is 6.6041. Next, the training set was increased from 100 to 300 samples with the training outcome validated using the same 250-sample test set. As shown in Fig. 3(b), with a MSE reduced to 2.2589 the performance of the ANN is significantly improved. As expected, the training outcome improves with an increased size of the training samples.

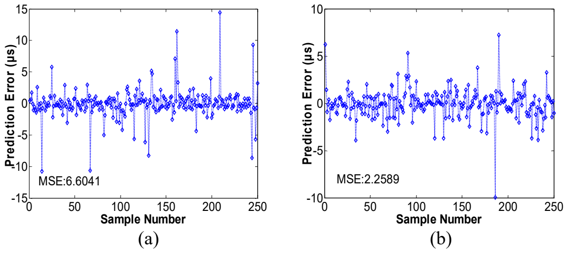
With the trained ANN, we are then able to analyze how uplink bandwidth allocation decisions, BW_{pre}, will impact uplink latency performance. Table I lists the selected α values and the corresponding minimum uplink latency as a function of traffic load. Note that the aggregated traffic load listed is normalized by $\lambda N S_{\text{avg}}/R_{\text{PON}}$. Table I highlights that after supervised training, the ANN can flexibly adjust bandwidth allocation decisions when the aggregated network load changes in the 16-ONU network.

TABLE I
OPTIMAL PREDICTION COEFFICIENT α
(16-ONU network, 10 km CO-to-ONU distance)

Traffic load	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
α	0	0	0	0.10	0.37	0.48	0.56	0.58
Latency (μs)	45.52	40.60	40.01	45.81	55.43	71.47	104.85	196.64

B. Latency Performance

The effectiveness of ANN-DBA in making flexible bandwidth allocation decisions that minimizes uplink latency, is highlighted in Fig. 4. The ANN-DBA allocates bandwidth in accordance to the decisions listed in Table I. As shown in Fig. 4, for all network loads, the uplink latency in a network using ANN-DBA (simulation) agrees with the ANN predicted latency (theory). Additionally, ANN-DBA results in latency performance improvement as compared to using the conventional limited-service DBA. A comparison of the uplink latency performance between ANN-DBA and the limited-service DBA as a function of varying CO-to-ONU distance, is shown in Fig. 3. Once again, the proposed ANN-DBA makes bandwidth allocation decisions that minimizes latency and does so irrespective of varying CO-to-ONU distances.



(a) (b)

Fig. 3. Prediction error of the trained ANN. (a) Training with 100 samples; and (b) training with 300 samples.

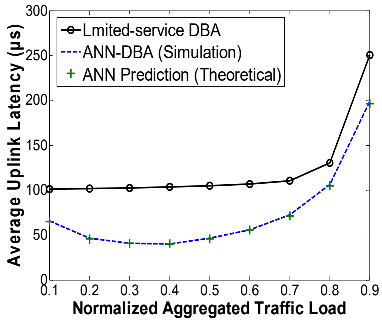


Fig. 4. Latency performance comparison as a function of traffic load.

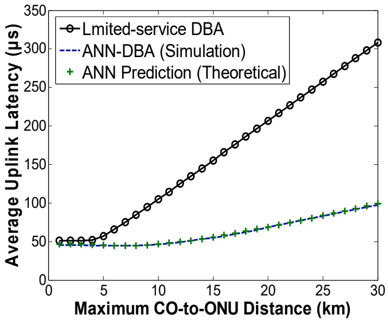


Fig. 5. Latency performance comparison as a function of CO-to-ONU distance.

Our results in Figs. 4 and 5 show that the ANN is capable of learning and predicting network latency performance with diverse network features as compared to the conventional limited-service DBA that relies on a singular traffic feature. In practice, training sets can be collected during network operation. Computation will be mainly spent in the supervised learning process, when the optimal weight matrix for each ANN layer is determined. Once training has ended, the CO needs only to store the weight matrix for each ANN layer. Mapping of input features to the target output value can be done without further computation.

IV. Conclusions

In this work, we investigated the applicability of an ANN in learning uplink latency, thereby in achieving flexible bandwidth allocation decisions that reduce latency. We highlighted the ANN's capability in predicting latency utilizing multiple network features. With the trained ANN, we showed that flexible bandwidth allocations under diverse application scenarios can be achieved and low-latency communication demands can therefore be met.

References

[1] M. Simsek, *et a*l, "5G-Enabled Tactile Internet," *IEEE J. Sel. Areas Commun.,* vol. 34, no. 3, pp. 460–473, Mar. 2016.

[2] G. P. Fettweis, "The Tactile Internet: Applications and Challenges," *IEEE Vehic. Techn. Mag.,* vol. 9, no. 1, pp. 64-70, Mar. 2014.

[3] M. Maier, *et al*, "The tactile internet: vision, recent progress, and open challenges," *IEEE Commun. Mag.,* vol. 54, no. 5, pp. 138-145, May 2016.

[4] J. Liu, et. al, "New Perspectives on Future Smart FiWi Networks: Scalability, Reliability, and Energy Efficiency," *IEEE Commun. Surveys Tuts.,* vol. 18, no. 2, pp. 1045-1072, 2nd quarter 2016.

[5] E. Wong, M. P. I. Dias, L. Ruan, "Predictive Resource Allocation for Tactile Internet Capable Passive Optical LANs," *J. Lightw. Technol.,* vol. 35, no. 13, pp. 2629-2641, Jan. 2017.

[6] S. Mondal, G. Das and E. Wong, "A Novel Cost Optimization Framework for Multi-Cloudlet Environment over Optical Access Networks," in *Proc.of. IEEE GLOBECOM*, 4-8 Dec, Singapore, 2017, pp. 1-7.

[7] G. Kramer, Ethernet Passive Optical Networks. McGraw-Hill Professional, 2005.

[8] R. Kubo*, et. al*, "Adaptive Power Saving Mechanism for 10 Gigabit Class PON Systems," *IEICE Trans. Commun.,* vol. E93–B, no. 2, 2010.

[9] M. Fiammengo, *et. al*, "Experimental Evaluation of Cyclic Sleep with Adaptable Sleep Period Length for PON," *in Proc. 37th Eur. Conf. Exhib. Opt. Commun.,* 18-22, Sep, Geneva, pp. 1-3, 2011.

[10] M. P. I. Dias, B. S. Karunaratne, E. Wong, "Bayesian Estimation and Prediction-Based Dynamic Bandwidth Allocation Algorithm for Sleep/Doze-Mode Passive Optical Networks," *J. Lightw. Technol.,* vol. 32, no. 14, pp. 2560-2568, 2014.

[11] R. Bushra, M. Hossen and M. M. Rahman, "Online Multi-thread Polling Algorithm with Predicted Window Size for DBA in Long Reach PON," in *Proc.of. ICEEICT*, 22-24, Sep, Dhaka, pp. 1-5, 2016.

[12] A. Dixit, *et. al*, "Delay models in ethernet long-reach passive optical networks," in *Proc. of. INFOCOM*, 26. Apr – 01. May, Kowloon, pp. 1239-1247, 2015.

An Open Controller for the Disaggregated Optical Network

Marc De Leenheer, Yuta Higuchi, Guru Parulkar

***Abstract*— The Open and Disaggregated Transport Network is an operator-led initiative to build data center interconnects using disaggregated optical equipment, common and open standards, and open source software. We discuss the project objectives, roadmap, and design.**

***Index Terms*—Open source software, Optical fiber networks, Software defined networking, Standards**

I. Introduction

Service providers are constantly trying to simplify and automate network operations. The combination of hardware disaggregation and the use of advanced software tools to control, configure and observe networks is expected to be a major driver towards this goal. To this end, we founded the ODTN project (short for Open and Disaggregated Transport Networks), an industry-wide and operator-led initiative to build an open source reference platform and deliver production-ready solutions using an innovative supply chain model.

The Open Networking Foundation has launched this project to rally service providers, hardware vendors and system integrators around the following objectives:

1. Build a reference implementation using (a) open source software, (b) open and common data models, and (c) disaggregated hardware devices.
2. Perform lab and field trials using the reference implementation.
3. Identify supply chain gaps and propose solutions.

In particular the project will focus on disaggregated DWDM systems, including but not limited to transponders and Open Line Systems, amplifiers, multiplexers, all-optical switches and ROADMs.

II. Objectives and Roadmap

The project addresses increasingly complex network scenarios, starting with relatively simple point-to-point open line systems and ending with a meshed network consisting of disaggregated optical equipment.

Initial focus in Phase 1 is on what is commonly referred to in the industry as the *Open Line System* [1]. Typically in the form of a simple point-to-point topology, such deployments are increasingly popular for cloud and content providers. These companies operate at extreme limits of networking performance and scale, and require high levels of flexibility and programmability. We expect the popularity of the OLS will only increase over time. Indeed, many traditional telco companies are starting to deploy and operate their own telco cloud infrastructure. Concrete initiatives such as CORD (Central Office Re-architected as a Data center) [2], and the more general multi-access edge computing and edge cloud projects reflect the growing importance of this trend. As such, interconnecting these Central Offices is the prime use case for our software stack.

Phase 2 will introduce integrated ROADM devices, leveraging open APIs to control and configure such equipment. Such deployments are necessary to support meshed topologies. The final Phase 3 will perform disaggregation of the ROADM itself; it is as of yet to be decided what *granularity of disaggregation* will be performed. The complete disaggregation into basic optical components as shown in the figure is unrealistic for reasons of performance and complexity of integration. It is expected that some intermediate form of disaggregation will offer acceptable performance while greatly improving flexibility and programmability. This is an area of active research in both academia and industry; guidelines and best practices will be folded into the project as these become available.

Each phase produces a set of deliverables that consist of the following:

1. Reference implementation composed of complete open source software stack, set of applications, and device drivers.
2. Integration report detailing measured, expected and potential gaps for performance levels and feature sets. Of particular interest are limitations brought about by a multi-vendor environment.
3. Detailed plan for operator-led deployment (lab or field trial), including system design, gap analysis, vendor selection, and resource commitments.

Paper first submitted on February 6th, 2018.

Marc De Leenheer is with the Open Networking Foundation (e-mail: marc@ opennetworking.org).

Yuta Higuchi is with NEC Corporation America (e-mail: yuta.higuchi@necam.com).

Guru Parulkar is with the Stanford University and the Open Networking Foundation (e-mail: parulkar@stanford.edu).

978-3-903176-07-2 © IFIP 2018

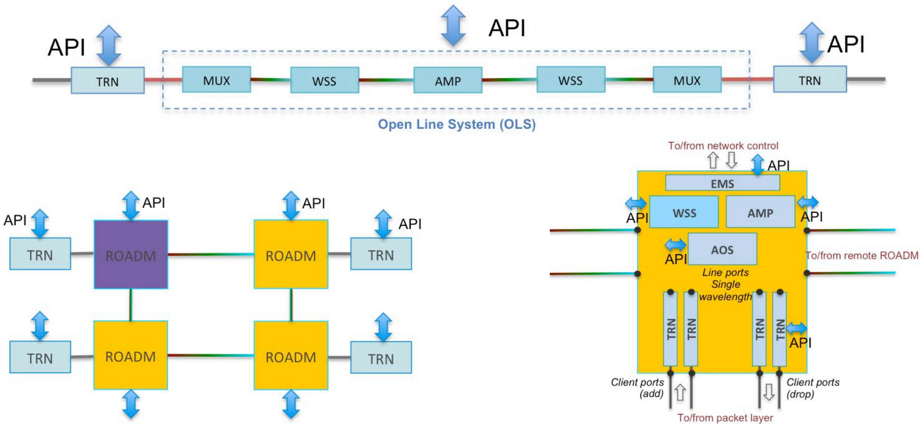


Fig. 1. Three phases to go from simple point-to-point deployments to fully disaggregated, meshed optical networks. Phase 1 addresses the OLS, initially controlling the terminal equipment (Phase 1.0), afterwards adding explicit control of the line system itself. Phase 2 introduces control of integrated ROADM nodes. Phase 3 disaggregates the ROADM further into more basic components.

III. Open Source Software Stack

The high-level design of our reference implementation is shown in Figure 2. It essentially breaks down into three parts: the northbound or service layer API, the network OS, and the southbound or device layer API.

On the northbound side, Transport API is an open standard for configuration and control of transport networks, offering a variety of services such as topology, connectivity, path computation and others [3]. The standard has achieved significant industry traction – in the form of field trials and interoperability testing – with numerous service providers and the support of a variety of equipment vendors. Many options are available in terms of northbound protocol specification; our initial implementation supports RESTCONF-based interaction between service orchestrator and the network operating system.

In turn, the southbound protocol is foremost based on NETCONF due to the limited availability of mature implementations on hardware devices. Of greater interest is the choice of standard device models. The importance of using open and common standards for optical equipment cannot be overstated, as it is the chief way to finally diminish vendor lock-in. OpenConfig is a set of vendor-neutral network management models, driven by actual operational use cases from network operators [3]. Another option is the OpenROADM multi-source agreement, which focuses on disaggregated models for optical ROADMs, with a strong emphasis on interoperability in multi-vendor environments. For Phase 1 at least, ODTN is developing a solution based on the OpenConfig models, while Phases 2 and beyond will evaluate the applicability of OpenROADM [4].

The translation between north- and southbound models occurs inside the network OS, in this particular case the Open Networking Operating System or ONOS. ONOS is an open-source SDN operating system architected to meet the stringent availability, scalability and performance demands of service provider networks. It was built from the ground up with a distributed architecture and introduces innovative state management techniques necessary to build a robust distributed SDN controller.

ONOS provides APIs that enable applications to view the network topology and to inspect and control the devices that compose it. However, many application developers would rather operate at a network-level, rather than a device-level, and do not need specific control of all parameters. To this end, ONOS championed the *intent abstraction*, which frees application developers from the need to specify and control low-level parameters, leaving them for the network operating system to optimize as network conditions change. Additionally, ONOS offers a converged topology abstraction, in which multiple network layers and technologies are presented as a single logical graph. Next to these network-centric abstractions, ONOS also offers a sub-system that can span both network and device-level state management. Indeed, the Dynamic Configuration Sub-system or DCS allows application developers to interact with a configuration database, offering advanced features such as distributed state management, transactions, and automated protocol handlers. Finally, although ONOS comes with a built-in multi-layer PCE, the modular system design allows external PCEs to be plugged in and out on-demand.

A concluding remark is that the complete reference implementation, including network OS, applications and drivers, is being made available under an Apache 2 software license. This is an open source software license that is relatively liberal in how. This deliberate choice increases the chance for further adoption by the industry, and ultimately improves the community to take ownership and maintenance of the code base.

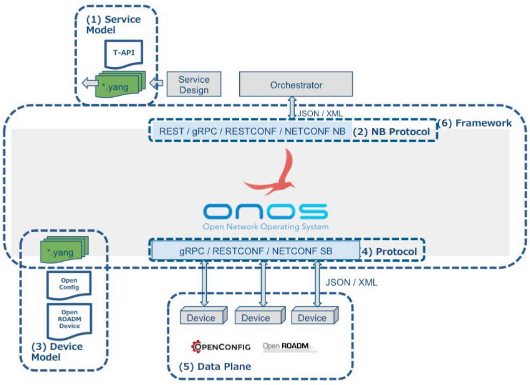


Fig. 2. Software stack based on the Open Networking Operating System (ONOS), detailing both north- and southbound models.

IV. Service Application

As a network OS, ONOS' first mission is to manage network hardware and offer common services for network applications. In ODTN, the core functionality to translate between north- and southbound operations is handled by the service application depicted in Figure 3. It shows how different ONOS sub-systems work together to take (1) a network-level service request, (2) semantically interpret the request, (3) compute a set of abstracted device configurations, and finally (4) generate concrete device control and config instructions. The intermediate abstract device configurations are necessary to allow flexibility in terms of device models, both in terms of differing/competing models or revisions of existing ones.

V. Conclusions

The ODTN project is an industry-wide effort to bring openness and innovation to the optical networking space. It does this by putting service providers in the driver's seat, allowing them to prioritize use cases and hopefully maximize the opportunity for production deployments. The project is built on three core principles: (1) disaggregated hardware, (2) open and common data models, and (3) open source software. All of these are essential to rally the industry eco-system, rebuild the supply chain, and bring about much-needed disruptive changes. It is our sincere hope that the community will step up and help us realize this vision.

References

[1] Mark Filer, Hacene Chaouch, and Xiaoxia Wu, "Toward Transport Ecosystem Interoperability Enabled by Vendor-Diverse Coherent Optical Sources Over an Open Line System," Journal of Optical Communications and Networking, 10(2):A216–A224, February 2018.

[2] Larry Peterson, Ali Al-Shabibi, Tom Anshutz, Scott Baker, Andy Bavier, Saurav Das, Jonathan Hart, Guru Parulkar, William Snow, "Central office re-architected as a data center," IEEE Communications Magazine, 54(10):96–101, October 2016.

[3] Open Transport Configuration and Control, https://www.opennetworking.org/projects/open-transport/

[4] OpenConfig, http://www.openconfig.net/

[5] OpenROADM MSA, http://openroadm.org/home.html

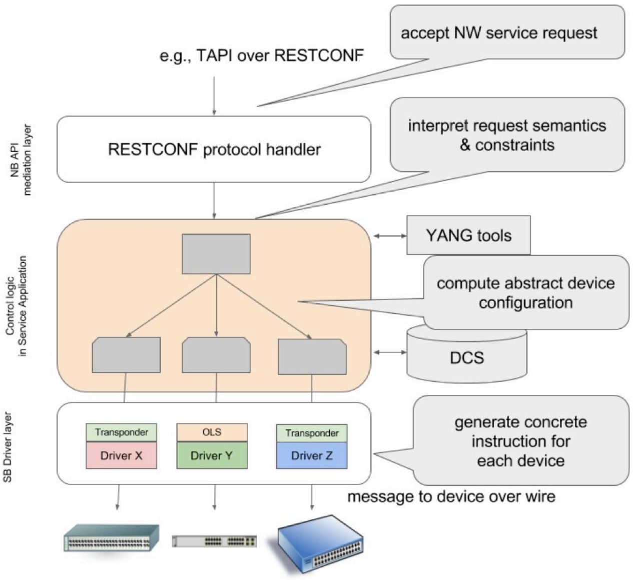

accept NW service request
e.g., TAPI over RESTCONF
NB API mediation layer
RESTCONF protocol handler
interpret request semantics & constraints
YANG tools
Control logic in Service Application
compute abstract device configuration
DCS
SB Driver layer
Transponder
Driver X
OLS
Driver Y
Transponder
Driver Z
generate concrete instruction for each device
message to device over wire

Fig. 3. Service application that deals with translating service-level requests into device operations. The application relies extensively on key ONOS sub-systems, such as the Dynamic Config Sub-system (DCS), ONOS YANG tools, Intent Framework and the Driver & Behaviour mechanisms.

Marc De Leenheer is a Member of Technical Staff at the Open Networking Foundation, an operator-led consortium that is transforming networks into agile platforms for service delivery. Marc leads the teams for packet/optical convergence and Enterprise CORD, and his software platforms are being trialed by some of the largest service providers on the planet. He received the PhD degree in computer science from Ghent University, Belgium in 2008, and completed post-doctoral stays at UCDavis and Stanford University.

Yuta Higuchi received the Master's degree in Information Science from Nagoya University. He joined NEC Corporation in 2007, where he has been undertaking development of a platform framework for operation management software stacks. He is currently with the Optical IP Development Division, NEC Corporation of America. His research interests include architecture design of scalable frameworks for Software-Defined Networking.

Guru Parulkar is Executive Director of Open Networking Foundation (ONF), Stanford Platform Lab, and Consulting Professor of EE at Stanford University. At ONF he leads open source projects ONOS and CORD.

Guru has been in the field of networking for over 25 years. He joined Stanford in 2007 as Executive Director of its Clean Slate Internet Design Program. At Stanford Guru helped create three programs: OpenFlow / Software-Defined Networking, Programmable Open Mobile Internet 2020, and Stanford Experimental Data Center Laboratory.

Prior to Stanford, Guru spent four years at the National Science Foundation (NSF) and worked with the broader research community to create programs such as GENI, Future Internet Design, and Network of Sensor Systems. Guru received NSF Director's award for Program Management excellence.

Before NSF Guru founded several startups including Growth Networks (acquired by Cisco) and Sceos (IPO'd as Ruckus Wireless). Guru served as Entrepreneur in Residence at NEA in 2001 and received NEA's Entrepreneurship Award.

Prior to this Guru spent over 12 years at Washington University in St. Louis where he was a Professor of Computer Science, Director of Applied Research Laboratory and the head of research and prototyping of high performance networking and multimedia systems.

Guru received his PhD in Computer Science from the University of Delaware in 1987. Guru is a recipient of the Alumni Outstanding Achievement award and the Frank A. Pehrson Graduate Student Achievement award.

Data-driven network analytics and network optimisation in SDN-based programmable optical networks

Shuangyi Yan
High Performance Networks Group
University of Bristol
Bristol, UK
Shuangyi.Yan@bristol.ac.uk

Reza Nejabati
High Performance Networks Group
University of Bristol
Bristol, UK
Reza.Nejabati@bristol.ac.uk

Dimitra Simeonidou
High Performance Networks Group
University of Bristol
Bristol, UK
Dimitra.Simeonidou@bristol.ac.uk

***Abstract*—5G, IoT and other emerging network applications drive the future optical network to be more flexible and dynamic. Fully awareness of current network status is critical for better network programming in short timescale. In this paper, the centralized network database with network monitoring data and network configuration information enables network analytics application to support the future dynamic and programmable optical network.**

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

Traditional static optical networks have been evolving continuously to offer more point-to-point link capacities by introducing coherent detection technology, sophisticated digital signal processing algorithms and space division multiplexing technologies [1]. However, the recent emerging network applications, such as big data, Internet of things (IoT) and 5G networks, require not only high-capacity optical links, but a dynamic or even programmable optical network. For example, the future 5G applications with peak network speed up to 10 Gbps will bring dynamic mobile traffic with significant bandwidths [2]. The dynamic and mobile network traffic requires network management in an end-to-end approach with ultra flexible network functions [3]. On the other side, software-defined networks (SDN) have been extended to optical networks to enable programmable, automatic and disaggregated optical networks [4]. The centralized SDN controller enables network programmability in the network controller layer and brings traffic engineering to multi-domain networks [5], [6]. Therefore, it's time to reconsider the architecture of the traditional static optical networks and bring programmability and flexibility for future dynamic optical networks.

In the future optical networks, network dynamics suggest two key points about network flexibility. Firstly, in dynamic optical networks, both the optical hardware and the control software should be flexible, configurable or even programmable. Each individual network function can be configured in a fine granularity to satisfy variable requirements in the dynamic optical networks. Furthermore, the whole optical node can deploy network functions in an on-demand manner [7]. Secondly, the dynamic optical networks also mean network connections will serve in a short timescale, not the same as the set-and-go link setup in the traditional static networks. In dynamic optical networks, network reconfigurations will happen more frequently. Thus, a short configuration/install time is required in the dynamic optical networks. In addition, network planning should consider the short period of the dynamic network services, which may lead to a margin-reduced optical link [8]. The network dynamics will bring many advantages to optical networks. However, it also raises new challenges for network management and network operation.

In this paper, a network-scale centralized network configuration and monitoring database (NCMdB) is implemented over an SDN-enabled optical network. The centralized NCMdB stores both the current and historical network configurations and network performance monitoring information. The NCMdB enables data-driven network analytics applications and support network optimization through the SDN interface in programmable optical networks. The network analytics applications can analyze the data in the NCMdB and offer suggestions for network planning and network optimization through the SDN controller. In this paper, a machine-learning OSNR prediction application that run over the NCMdB has been developed to assist both the network planning and network optimization. The NCMdB-powered network analytics and network optimization forms a new network-management approach for the dynamic optical network and bring new network functions.

II. Control loop for dynamic optical networks

Driven by the dynamic bandwidth requests, dynamic optical networks require a new network-service-deployment mechanism. The network functions/services should be deployed automatically. The time-consuming test and optimization operations need to be eliminated and to be replaced with instant

This work is supported by the UK EPSRC through the TOUCAN project (EP/L020009/1) and the INSIGHT project (EP/L026155/2), MetroHaul project.

978-3-903176-07-2 © 2018 IFIP

network responses for deployment verifications. To achieve this, both the network monitoring device and novel control softwares are required. It's worthy noting that the network responses are not only for the latest-deployed services, but also for the previous deployed network services. The reason is that the latest-deployed network function may affect the current network services. In some extreme cases, the existing network service may fail due to the new network functions. Especially, in the margin-reduced optical network, the failure may happen more frequently and the network replanning, which is totally forbidden in the traditional networks, need to be introduced in the dynamic optical networks.

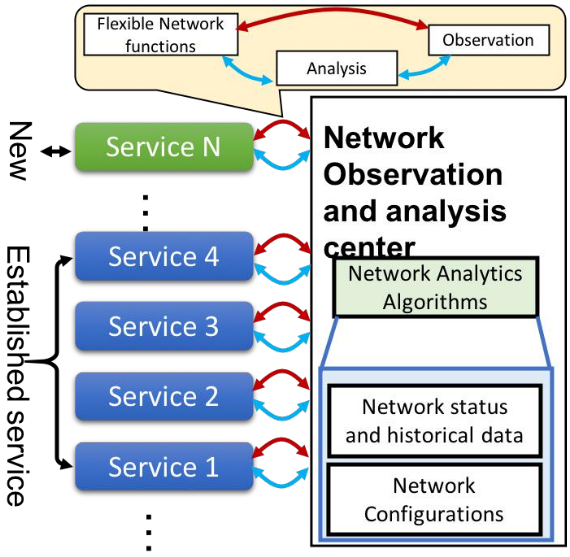

Flexible Network functions
Observation
Analysis
New
Service N
Network Observation and analysis center
Established service
Service 4
Service 3
Service 2
Service 1
Network Analytics Algorithms
Network status and historical data
Network Configurations

Fig. 1. Network control loop for dynamic optical networks.

Figure 1 shows the proposed network control loop in dynamic optical networks. For each service, deployment, observation and optimization will form a continuous operation circle. As shown in the inset of Fig. 1, a closed service operation loop will be established for each service. Firstly, the flexible network service or function is deployed according to the traffic request. Then, the network observation and analysis center will allocate the monitoring and related network analytics resources for the deployed services. The established operation circle provides instant responses the new deployed services to confirm the successful deployment. Then, the circle will continuously monitor the service. The network observation and analysis center will provide the monitoring functions to the deployed services. The monitoring functions are implemented with both network performance monitoring and the network analytics algorithms. In most case, the monitoring data are shared among several services. Thus, the monitoring data are uploaded to the cloud and processed jointly under different perspectives. As introduced first in [9], Fig. 2 presents the implementation of the proposed network observation and analysis center. Several key technologies are introduced as follows:

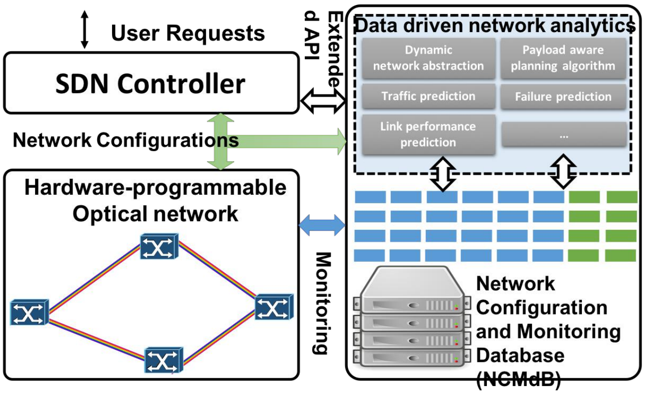

User Requests
SDN Controller
Extended API
Data driven network analytics
Dynamic network abstraction
Payload aware planning algorithm
Traffic prediction
Failure prediction
Link performance prediction
...
Network Configurations
Hardware-programmable Optical network
Monitoring
Network Configuration and Monitoring Database (NCMdB)

Fig. 2. Implementation of centralized network observation and analysis center.

A. Programmable node architecture and hitless reconfigurable network functions

Dynamic optical networks require flexible network configurations for the diverse network requests. In addition, network function reconfiguration are required for possible network replanning. Thus, network functions should be flexible and can be reconfigurable without any data loss, i.e., hitless operation. Part of the network reconfigurability can be achieved in the network control software with network function virtualization technologies (NFV) [10]. In this part, we mainly focus on the flexibility offered by the hardware implementation.

Regarding optical node functions, architecture-on-demand (AoD) based optical node architecture has been introduced to optical network by modularizing key network functions and deploying network function programmability [7], [11]. For the first time, the AoD concept is used to provide an SDM/WDM ROADM solution for future multi-core fiber based multi-dimensional optical networks [12].

For optical network functions, bandwidth variable transmitter (BVT) is one of the key enabling technologies for elastic optical networks. From the perspective of the optical hardware, BVTs require the hitless reconfigurations in operation baud rate and spectral efficiency. BVT with variable baud rates can be configured to accommodate variable spectrum slots in elastic optical networks. In [13], BVT is used to tolerate the filtering effect of the legacy wavelength selective switching (WSS) device in a fixed-grid and flex-grid coexisted network. Possible hitless operation of the BVT can be achieved for the elastic interface in flex-grid optical networks [14]. On the other side, BVTs with variable spectral efficiencies could deliver variable transmission capacities based on the link distance. In [15], real-time modulation-adaptable transmitter is reported to offer quick switching between QPSK and 16QAM signal formats. An spectral-efficiency adaptable transmitter with a fine-granularity is implemented based probabilistic shaping for network planning [16].

B. Network configuration and monitoring database (NCMdB)

In dynamic optical networks, network status becomes critical to plan, deploy, and configure network functions. Understanding the current network status could lead to an intelligent network planning. After the deployment of network functions or services, instant network responses will confirm the success of functions deployments. Then, the deployed network functions or services requires continuous network performance monitoring, which also rely on the awareness of the network status. In addition, network analytics applications would require the history network status. Therefore, the network status data is at the heart of the dynamic optical networks. On the other side, the centralized network controller in SDN makes the collecting of network configurations possible and easier. So, a well-structured network configuration and monitoring database is one of the key technologies to implement the closed network control loop in dynamic optical networks.

In [16], we built a network-scale NCMdB, which collects all the network configurations from the centralized SDN controller, network performance monitoring data from all the optical performance monitoring devices, and device operation information from the used electrical devices. Any event that will change the network status will trigger a new record to store the new data in the NCMdB. The NCMdB stores the raw data from all the optical performance monitors through dedicated links. All the physical parameters can be monitored and stored in the database. The monitoring data are linked to the network configurations. In principle, the NCMdB could record all the operation information of the current networks and the previous network status.

The NCMdB collects all the raw monitoring data to a centralized network space. Therefore, the monitoring data can be processed by different applications, i.e., network analytics applications. As shown in Fig. 2, network analytics applications can be developed over the NCMdB.

A parallel database schema is used to manage the data. In our design, the MongoDB, which has a hierarchical document-based data model design using JavaScript Object Notation (JSON) as the file format, is used to store all the collected information. The MongoDB based solution is able to support the complex data structures recorded throughout the experiments. Furthermore, the MongoDB based NCMDB could easily provide network interfaces for other applications.

C. SDN enabled network analytics application

On top of the NCMdB, multiple network analytics application can be developed.The NCMdB that includes monitoring and configuration data from all the network devices and links enables end-to-end connection analysis. Variable network analytics applications can access the monitoring and configuration data simultaneously. As shown in Fig. 2, a myriad of network analytics applications can run on top of the NCMdB to offer new network functions.

In SDN-enabled networks, the SDN controller can talk to the NCMdB rather than the real network monitoring device to access the monitoring data. The network analytics application can access more data with more computing resources, therefore can offer more information than the monitoring data. In addition, the network analytics applications could offer extra network functions. To facilitate the communications with the SDN controller, the developed network analytics applications should also support SDN protocol. Then, the SDN controller can use the network analytics applications more efficiently.

In the SDN controller, extra SDN plug-ins for the network analytics application should be developed to inform the extra network functions offered by the network analytics applications.

In [16], a machine-learning (ML) based OSNR predictor is developed based on a multilayer perceptron (MLP) artificial neural network (ANN) trained using various link and signal parameters extracted from the NCMdB. A supervised learning method i.e. Levenberg-Marquardt (LM) backpropagation (BP) is used for the offline training of ANN. During the training process, vectors p comprising of different link/signal parameters (such as launched power, EDFAs' gains, EDFAs' input and output powers, EDFAs' noise figures (NF) etc.) are applied at the input of ANN while the known OSNR values o at a given node corresponding to these parameters are used as targets. All the link parameters are retrieved from the NCMDB. After training, the OSNR predictor can predict the OSNR penalty of the unestablished optical path.

III. Use Cases

In this section, SDN-based network planning with ML-based OSRN prediction was demonstrated successfully over a field-trail testbed. The experimental demonstration is reported in [16]. The field-trial testbed consists of parts of the national dark fiber infrastructure service (NDFIS) from Bristol to Froxfield. Several extra nodes are located in our lab.

Figure 3 shows the workflow of the experimental demonstration. User requests are emulated and submitted to the SDN controller. Each user requests from the SDN controller to connect a source to a destination at a specific bandwidth. The SDN controller leverages the path computation application to determine a suitable path for the user request. If a path is successfully found, the SDN controller then finds a set of available wavelengths for the transmission. The SDN controller afterwards queries the trained ML application to predict the link performance for all the available wavelengths. Based on the predictions, the controller calculates the possible available link bandwidth and the optical modulation to use for the available wavelengths. In case the predicted bandwidth is lower than the one requested from the user, then the algorithm terminates, and the user request is not accommodated. If not, then the first available wavelength that meets the user bandwidth requirement, across the selected path that connect the user-selected endpoints, is chosen for the transmission. Following that, the SDN controller configures the optical switches appropriately using the OpenFlow protocol.

In the demonstration, the ML algorithm predicts the link performance and return the OSNR at the receiver side around 21 dB and the suggested spectral efficiency is 3.9987. The

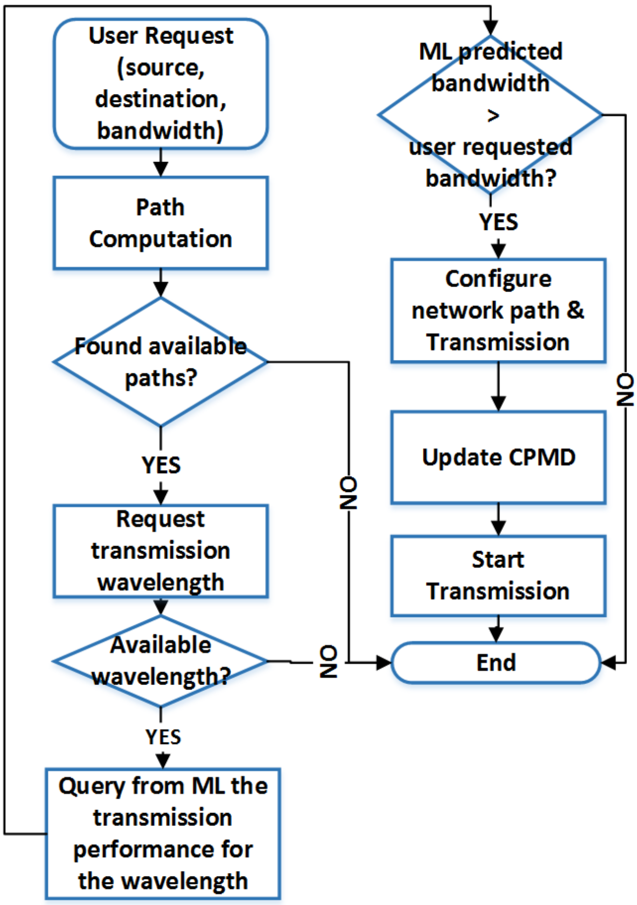

User Request (source, destination, bandwidth)
Path Computation
Found available paths?
YES
NO
Request transmission wavelength
Available wavelength?
YES
NO
Query from ML the transmission performance for the wavelength
ML predicted bandwidth > user requested bandwidth?
YES
NO
Configure network path & Transmission
Update CPMD
Start Transmission
End

Fig. 3. Workflow of the experimental demonstration.

probabilistic shaping-based transmitter can offer optical signals with spectral efficiencies of 2.8, 3.2, 3.6 and 4 bits per polarization. Based on the QoT prediction, the transmitter is configured with SE at 3.6. Thus, network link can be setup based on the QoT prediction. The recovered constellations after 336.4-km fiber transmission is shown in Fig. 4.

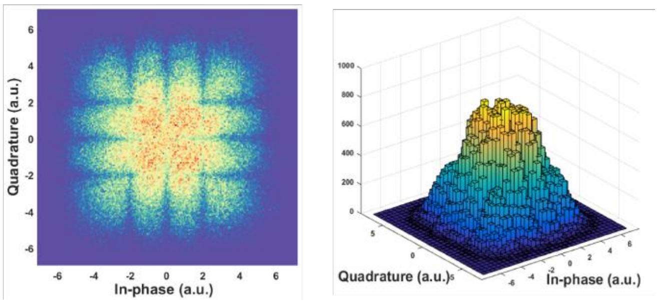


Fig. 4. (a) Recovered constellation diagram with entropies 3.6 after 336.4-km transmission; (b) The received constellation distribution.

IV. Conclusion

In this paper, novel network operation loop is introduced for the dynamic network services. The network-scale NCMdB that stores network monitoring and configuration information provides novel network functions through network analytics application. In this paper, we reported an machine-learning based QoT predictor, which help the SDN controller to plan the network efficiently. The introduced NCMdB and the network analytics applications open new possibilities for future dyanmic optical networks.

References

[1] E. Agrell, M. Karlsson, A. R. Chraplyvy, D. J. Richardson, P. M. Krummrich, P. Winzer, Kim Roberts, J. K. Fischer, S. J. Savory, B. J. Eggleton, M. Secondini, F. R. Kschischang, A. Lord, J. Prat, I. Tomkos, J. E. Bowers, S. Srinivasan, Mat Brandt-Pearce, and N. Gisin, “Roadmap of optical communications,” *Journal of Optics*, vol. 18, no. 6, p. 063002.

[2] P. Fan, J. Zhao, and C. L. I, “5G high mobility wireless communications: Challenges and solutions,” *China Communications*, vol. 13, pp. 1–13, N 2016.

[3] M. Ruffini, “Multidimensional Convergence in Future 5G Networks,” *Journal of Lightwave Technology*, vol. 35, no. 3, pp. 535–549.

[4] M. Channegowda, R. Nejabati, and D. Simeonidou, “Software-Defined Optical Networks Technology and Infrastructure: Enabling Software-Defined Optical Network Operations [Invited],” *Journal of Optical Communications and Networking*, vol. 5, no. 10, p. A274.

[5] H. Yang and G. F. Riley, “Traffic engineering in the peer-to-peer SDN,” in *2017 International Conference on Computing, Networking and Communications (ICNC)*, pp. 649–655.

[6] J. M. Fbrega, M. S. Moreolo, A. Mayoral, R. Vilalta, R. Casellas, R. Martnez, R. Muoz, Y. Yoshida, K. Kitayama, Y. Kai, M. Nishihara, R. Okabe, T. Tanaka, T. Takahara, J. C. Rasmussen, N. Yoshikane, X. Cao, T. Tsuritani, I. Morita, K. Habel, R. Freund, V. Lpez, A. Aguado, S. Yan, D. Simeonidou, T. Szyrkowiec, A. Autenrieth, M. Shiraiwa, Y. Awaji, and N. Wada, “Demonstration of Adaptive SDN Orchestration: A Real-Time Congestion-Aware Services Provisioning Over OFDM-Based 400G OPS and Flexi-WDM OCS,” vol. 35, no. 3, pp. 506–512.

[7] Y. Shuangyi, H.-S. Emilio, O. Yanni, N. Reza, and S. Dimitra, “Hardware-programmable Optical Networks (Invited),” *SCIENCE CHINA Information Sciences*, vol. 59, no. 10, p. 102301:1102301:12.

[8] Y. Pointurier, “Design of low-margin optical networks,” *IEEE/OSA Journal of Optical Communications and Networking*, vol. 9, no. 1, pp. A9–A17.

[9] Shuangyi Yan, Alejandro Aguado, Yanni Ou, Rui Wang, Reza Nejabati, and Dimitra Simeonidou, “Multi-Layer Network Analytics with SDN-based Monitoring Framework [Invited],” *Journal of Optical Communications and Networking*, vol. 9, no. 2, Feb. 2017.

[10] D. Cho, J. Taheri, A. Y. Zomaya, and L. Wang, “Virtual Network Function Placement: Towards Minimizing Network Latency and Lead Time,” in *2017 IEEE International Conference on Cloud Computing Technology and Science (CloudCom)*, Dec. 2017, pp. 90–97.

[11] N. Amaya, G. Zervas, and D. Simeonidou, “Architecture on demand for transparent optical networks,” in *2011 13th International Conference on Transparent Optical Networks (ICTON)*, 2011, pp. 1–4.

[12] Yanlong Li, Shuangyi Yan, Nan Hua, Yanni Ou, Fengchen Qian, Reza Nejabati, Dimitra Simeonidou, and Xiaoping Zheng, “Hardware Programmable SDM/WDM ROADM,” in *OFC 2017*, Los Angeles, Mar. 2017.

[13] S. Yan, E. Hugues-Salas, A. Hammad, Y. Yan, G. Saridis, S. Bidkar, R. Nejabati, D. Simeonidou, A. Dupas, and P. Layec, “Demonstration of Bandwidth Maximization between Flexi/Fixed Grid Optical Networks with Real-Time BVTs,” in *ECOC 2016*, Dsseldorf, Germany, Sep. 2016.

[14] A. Dupas, E. Dutisseuil, P. Layec, P. Jennev, S. Frigerio, Y. Yan, E. Hugues-Salas, G. Zervas, D. E. Simeonidou, and S. Bigo, “Real-Time Demonstration of Software-Defined Elastic Interface for Flexgrid Networks,” in *Optical Fiber Communication (OFC), collocated National Fiber Optic Engineers Conference, 20105Conference on (OFC/NFOEC)*. Anaheim, CA: OSA, 2015, p. Paper M3A.2.

[15] Shuangyi Yan, Arash Farhadi Beldachi, Fengchen Qian, Koteswararao Kondepu, Yan Yan, Chris Jackson, Reza Nejabati, and Dimitra Simeonidou, “Demonstration of Real-Time Modulation-Adaptable Transmitter,” in *ECOC 2017*, Gothenburg, Sep. 2017, p. Paper TH.1.A.

[16] Shuangyi Yan, N. Khan Khan, Alex Mavromatis, Dimitrios Gkounis, Qirui Fan, Foteini Ntavou, Konstantinos Nikolovgenis, Changjian Guo, Fanchao Meng, Emilio Hugues Salas, Chao Lu, Alan Pak Tau Lau, Reza Nejabati, and Dimitra Simeonidou, “Field trial of Machine-Learning-assisted and SDN-based Optical Network Planning with Network-Scale Monitoring Database,” in *ECOC 2017*, Gothenburg, Sep. 2017, p. TH.PDP.B4.

Spatial Division Multiplexing for Optical Data Center Networks

Rui Lin[1,5], Joris Van Kerrebrouck[3], Xiaodan Pang[1,4], Michiel Verplaetse[3], Oskars Ozolins[4], Aleksejs Udalcovs[4], Lu Zhang[1], Lin Gan[5], Ming Tang[5], Songnian Fu[5], Richard Schatz[1], Urban Westergren[1], Sergei Popov[1], Deming Liu[5], Weijun Tong[6], Timothy De Keulenaer[7], Guy Torfs[3], Johan Bauwelinck[3], Xin Yin[3] and Jiajia Chen[1,2]

[1]KTH Royal Institute of Technology, Electrum 229, 164 40 Kista, Sweden,
[2]SCNU South China Normal University,
[3]Ghent University - imec, IDLab, Department of Information Technology, Belgium
[4]Networking and Transmission Laboratory, RISE Acreo AB, Kista, Sweden
[5]Huazhong University of Science and Technology, Wuhan, China, tangming@hust.edu.cn
[6]Yangtze Optical fiber and Cable Joint Stock Limited Company (YOFC), Wuhan, China
[7]BiFAST, spin-off of IDLab, Ghent University–imec, Ghent, Belgium

Email: jiajiac@kth.se

***Abstract*—Emerging mobile and cloud applications drive ever-increasing capacity demands, particularly for short-reach optical communications, where low-cost and low-power solutions are highly required. Spatial division multiplexing (SDM) techniques provide a promising way to scale up the lane count per fiber, while reducing the number of fiber connections and patch cords, and hence simplifying cabling complexity. This talk will address challenges on both system and network levels, and report our recent development on SDM techniques for optical data center networks.**

I. INTRODUCTION

CURRENT demand for supporting data center applications has been posing stringent requirements on short-reach transmission techniques. To address the high capacity demand in data center networks (DCNs), scaling up the fiber capacity by spatial division multiplexing (SDM) approach has been proposed to boost the single fiber capacity, which can greatly reduce the number of fiber connections and patch cords and consequently, simplify cabling complexity [1]. Transporting data through different spatial channels in a single fiber, e.g., different modes in few mode fiber (FMF), independent cores in multicore fiber (MCF) or a hybrid way, are widely studied[1][2]. In FMF based optical communications, complex digital signal processing (DSP) is needed to address the modal interference and differential mode dispersion, which hinders its deployment in DCNs. On the other hand, single mode (SM) MCF-based SDM system is appreciated for providing high capacity as well as good performance in DCNs, where low cost and low power transceivers are one of the main concerns.

Similarly, for the cost and complexity consideration in DCNs, intensity modulation and direct detection (IM/DD) systems are preferable for deployment. With the availability of high bandwidth optical transceivers [3], the simple non-to-zero (NRZ) and partial response modulation format, such as electrical duo-binary (EDB) are attractive for high-speed DCNs. With extremely low complex transceiver design, especially when the communications are achievable with analog equalization, real-time communication can be realized [4]. Many research efforts have been also put on advanced modulation formats, such as 4-level pulse amplitude modulation (PAM4) [5] and discrete multi-tone (DMT) [6] to support high-speed single lane transmission with relaxed requirement on the system bandwidth. For DCN applications for which distances of more than a few kilometres are required, recent works on high data rate advanced modulation signals transmission with 1.5-μm single mode vertical cavity surface emitting laser (VCSEL) and single mode fiber (SMF) [6] also shows promising results to match the requirement on cost, energy consumption, footprint, and potential for seamlessly extension to long-term parallelism enabled by SDM techniques.

In this paper, we report our recent work [7-9] to address the high-speed requirement in SDM-enabled DCNs. Using 7-core fiber, the demonstration of i) real-time 100 Gbps/λ/core NRZ and EDB transmission over 10 km MCF with optical dispersion compensation with BiCMOS based tranmistter and receiver chip and a monolithic C-band 100 GHz distributed feedback electro-absorption modulated laser (DFB-EAM) [7]; with a 1.5-μm SM-VCSEL ii) 50Gbaud/λ/core PAM-4 over 1 km dispersion-uncompensated and 10 km dispersion-compensated MCF links [8]; and iii) total net rate over 700 Gbps DMT optical signals over 2.5 km dispersion-uncompensated MCF [9]. In all of the demonstrations above, bit error rate (BER) below the 7%-overhead hard decision forward error correction code limit (7%-OH HD-FEC) is achieved.

II. EXPERIMENTAL SETUP

The experiment setup of the 100 Gbps/λ/core demonstration of real-time NRZ/EDB and VCSEL-based PAM4 and DMT transmission is illustrated in Fig.1. For the NRZ and EDB signal

978-3-903176-07-2 © 2018 IFIP

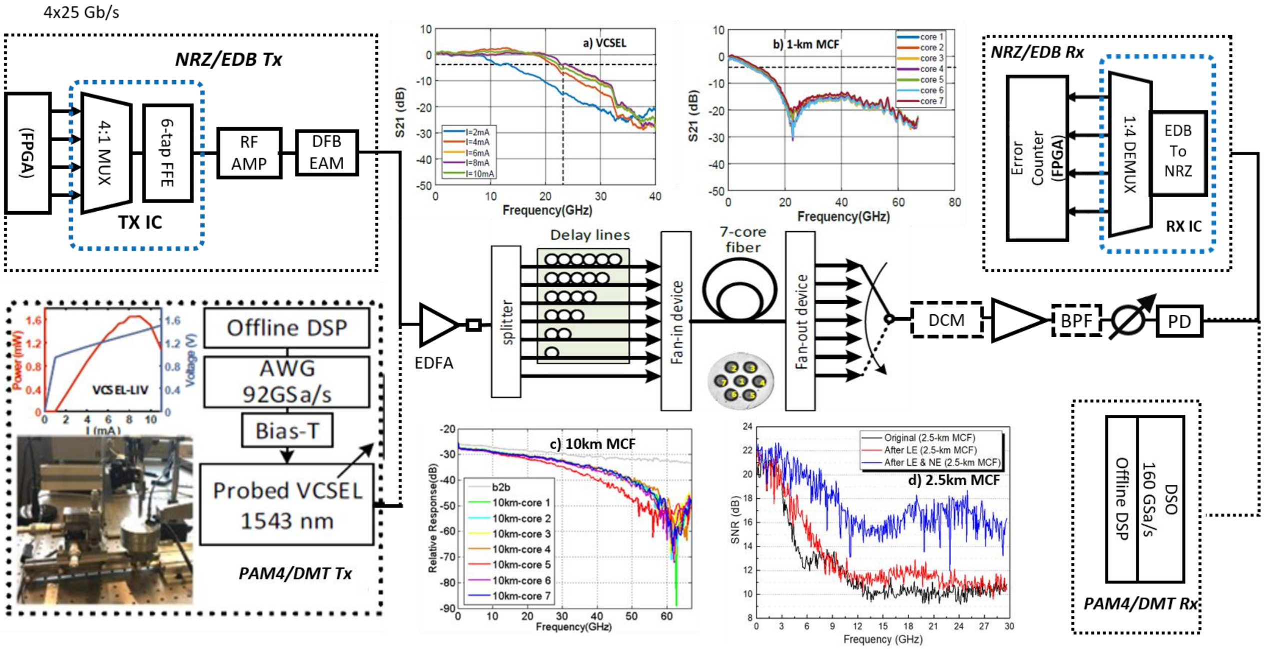


Fig. 1 The experimental setup of the high-speed transmission based on MCF enabled SDM systems for optical interconnects.

generation, four electrical 2^7-1 pseudo-random bit streams (PRBS) at 25 Gbps is generated by a Xilinx Virtex Ultrascale FPGA and multiplexed into a 100 Gbps NRZ signal by the transmitter (TX) IC. A six-tap analog feedforward equalizer (FFE) at the TX side is used to compensate the frequency roll-off and dispersion induced inter-symbol-interference in the link. A 50 GHz RF amplifier is used to drive a C-band 100 GHz distributed feedback electro-absorption modulated laser (DFB-EAM) [3]. The DFB-EAM is driven by 117.0 mA and biased at −1.85 V, emits at 1548.7 nm.

The advanced modulation formats, i.e., PAM4 and DMT signals are generated offline. Specifically, the PAM-4 symbols are offline generated, before being up-sampled and filtered with a raise cosine filter of 0.15 roll-off factor. Based on the characterized end-to-end channel frequency response pre-equalization is performed. In the DMT signal generation the length of the inverse fast Fourier transform (IFFT) and cyclic prefix were set to 1024 and 16 and the first subcarrier is set to null. The linear channel equalization (LE) pilot ratio was 3.3%. Volterra nonlinear equalization (NLE) was used to compensate the VCSEL. Clipping ratio is set to 0.7 to improve the output signal to noise ratio (SNR) from the DAC. The VCSEL is operated at room temperature without active cooling, and the optimal driving current is found to be 7.8 mA, considering both bandwidth and output power.

The generated signals from the transmitters are amplified and split into 7 branches using a 1x8 splitter and further de-correlated with different delays and launched into the 7 cores of the MCF via a low-loss and low crosstalk fan-in (FI)/fan-out(FO). The FI/FO module is fusion spliced to the MCF at one end, split and connected to seven individual single mode fiber (SMF) pigtails at the other end. The frequency response of the link in back to back (b2b) configuration and MCF transmission in different lengths, i.e., 1km, 2.5 km and 10 km can be found in Fig.1. In the case of 10 km MCF, a fixed dispersion compensation module (DCM) of −159 ps/nm is used for coarse dispersion compensation. Approximate homogeneous characteristics of the cores can be found due to small difference between the responses. It should be noted that, in the VCSEL-MCF link the mux and demux process in practice should be more effectively achieved with a VCSEL/PD array butt-coupled to each core of the MCF.

At the receiver side, an EDFA and a variant optical attenuator (VOA) are used to adjust the received optical power. The optical signal is detected by an InP-based PIN-PD (>90 GHz 3-dB BW) with a responsivity of 0.2 A/W. The received 100 Gbps data is deserialized by the receiver (RX) IC into 4x25 Gbps streams for real-time error detection. For receiving PAM4 and DMT signal detection, offline DSP is carried out, while for the case of NRZ/EDB real-time transmission is implemented.

III. Results and Analysis

In both DFB-EAM and VCSEL based SDM systems, 7%-OH HD-FEC is reached for all the investigated modulation schemes. In particular, some of the performance evaluation results can be found in Fig. 2, the performance of the received EDB after 10 km MCF transmission is shown in Fig. 2(a). Thanks to the broadband optical transceiver, BER=2.8E-6 is reached for both NRZ and EDB signals with the received power of around 7 dBm in b2b configuration. Except for core 5, all the other spatial channels have similar receiver sensitivity to reach the BER=3.8E−3 after 10km MCF transmission. The sensitivity variation is mainly due to the end face reflection in the MCF link. With higher received optical power, in which all the spatial channels are well below the KP4 FEC limit (BER=2.0E−4).

Fig. 2(b) shows the BER performance after the 1 km MCF

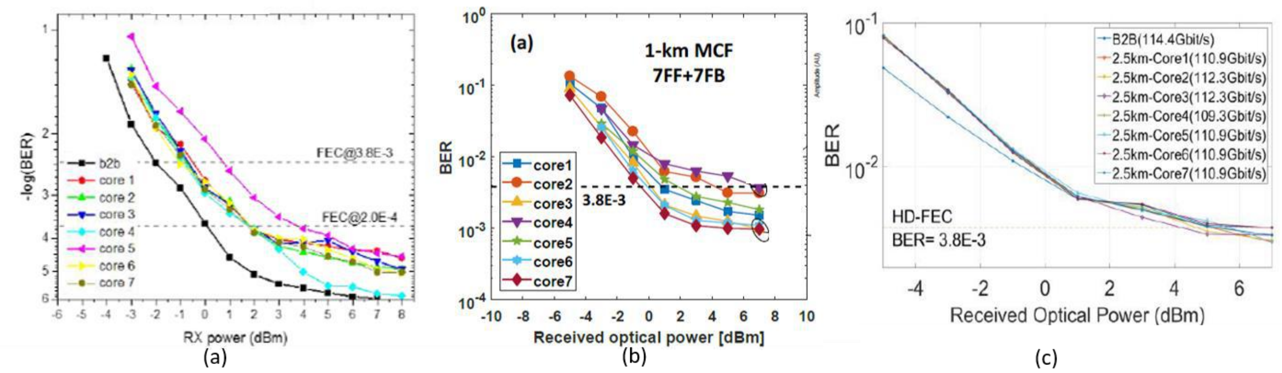


Fig. 2 BER performance of (a) 100Gbps EDB transmission over 10 km dispersion compensated MCF, (b) 50 Gbaud PAM-4 signal after 1 km, and (c) the DMT signals transmission over 2.5 km MCF.

transmissions for PAM-4 signal at 50 Gbaud. For the uncompensated 1 km MCF cases, a DFE of 7-feed forward (FF)-tap + 7-feedback (FB)-tap can successfully recover the received signals in all the cores to reach BER performance below the 7%-OH-HD-FEC limit. Additionally, different performances between different cores can be seen which is in accordance with the differences in the characterized frequency response. It should be noted that due to added complexity, we did not apply any specific amplitude-level dependent symbol decision method or nonlinear equalization techniques to mitigate the eye skew. Therefore, due to narrower horizontal eye opening, better BER performance and/or higher baud rates can be potentially achieved after eye skew correction.

As for the DMT case, the measured BER in function of the received optical power (RoP) at the PD input is shown in Fig. 2(c). The 7% HD-FEC limit is achieved with NLE. The achieved gross bit rate is 114.4 Gb/s for the B2B case. For 2.5 km MCF transmission, we have achieved gross-rates of 110.9-Gb/s, 112.3-Gb/s, 112.3-Gb/s, 109.3-Gb/s, 110.9-Gb/s, 110.9-Gb/s, and 110.9-Gb/s, respectively. The total system capacity with 2.5-km MCF is about 777.5 Gb/s (net rate 726.7 Gb/s).

IV. Conclusions

Based on the MCF enabled SDM, the IM/DD 7x100Gbps/λ/core communication and beyond are achieved. Real-time NRZ/EDB transmission are realized with monolithic DFB-EAM and BiCMOS based transceiver chips while high-speed PAM4 and DMT signal transmission are carried out with 1.5 um single mode-VCSEL. The presented SDM system proves its potential in providing high-speed, low cost, and low power optical interconnect for DCNs.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (61331010, 61722108, 61775137, 61671212, and 61550110240), Swedish Research Council Swedish Research Council projects PHASE (2016-04510) and Go-iData (2016-04489), the Swedish Foundation for Strategic Research (SSF), the Göran Gustafsson Foundation, the Swedish ICT-TNG, the Celtic-Plus sub-project SENDATE-EXTEND & SENDATE FICUS funded by Vinnova, the Flemish FWO & IWT, IOF, EU H2020 projects STREAMS (688172), TeraBoard (688510), PICTURE (780930), and the Natural Science Foundation of Guangdong Province.

References

[1] R. G. H. van Uden, et al., Ultra-high-density spatial division multiplexing with a few-mode multicore fibre, Nat. Photonics. 8 (2014) 865–870.

[2] D. J. Richardson, et al., Space-division multiplexing in optical fibres Nat. Photonics. 7 (2013) 354.

[3] M. Chaciński, et al., Monolithically integrated 100 GHz DFB-TWEAM, J. Light. Technol. 27 (2009) 3410–3415.

[4] M. Verplaetse, et al., Real-time 100 Gb/s transmission using three-level electrical duobinary modulation for short-reach optical interconnects, J. Light. Technol. 35 (2017) 1313–1319.

[5] J. Verbist, et al., DAC-less and DSP-free PAM-4 transmitter at 112 Gb/s with two parallel GeSi electro-absorption modulators, in ECOC PDP 2017.

[6] C. Kottke, et al., High speed 160 Gb/s DMT VCSEL transmission using pre-equalization, in OFC.2017.

[7] R. Lin, et al., Real-time 100 Gbps/λ/core NRZ and EDB IM/DD transmission over 10 km multicore fiber, in OFC, 2018.

[8] X. Pang, et al., 7 × 100 Gbps PAM-4 transmission over 1-km and 10-km single mode 7-core fiber using 1.5- μm SM-VCSEL, in OFC, 2018.

[9] J. Van Kerrebrouck, et al.,Single-mode VCSEL discrete multi-tone transmission over 2.5-km multicore fiber, in OFC, 2018.

Hardware-supported Softwarized and Elastic Optical Networks

Hiroaki Harai, Hideaki Furukawa, Yusuke Hirota
National Institute of Information and Communications Technology
Koganei, Tokyo 184-8795, Japan
harai@nict.go.jp

***Abstract*—We present elasticity and agility in softwarized optical network construction, service continuation, and service update. Programmability among multiple network protocols and multiple classes of transmission and processing speeds is a necessary solution for lower CAPEX and agile network setup. We present beyond 100 Gbps hardware-support programmability in optical edge. Existing services should be kept transient quality against sudden traffic changes and failures. We also show proper optical power management using burst-tolerable EDFAs in network protection for service continuation of in-service paths.**

Keywords—optical networks; softwarization; elasticity; agility

I. Softwarization, Elasticity and Agility

Optical switching and transmission technology is a key enabler for provision of huge, long-distance, and energy-aware communication in 5th generation mobile networks (Fig. 1), where low-latency (< 1 msec in wireless), peak data rate (20 Gbps), and energy efficiency (100x) are expected capabilities [1]. The capabilities come from diverse and extreme application requests such as vehicle communications and video entertainment with greening environment. A high-quality network is generally costly so desired quality with cost-efficiency is different among different services. Network softwarization or virtualization is a concept for sharing but isolating a common network and computing infrastructure with QoS satisfaction [2]. Figure 2 is an example of a softwarized and elastic network where each service is provided on a separated edge-cloud network for QoS. Optical resources are also shared and isolated among different services.

Traditionally, optical core networks, which consist of a number of optical paths, are likely static. The optical paths (or wavelength paths) are provisioned according to the increase in the amount of traffic at the related communication sites. Topological change is slow. On the other hand, in network softwarization era, the optical paths are dynamically and agilely set up or torn down according to the birth/emerging and death/environmental change of services (e.g., a lot of human movement) [3]. An example case is illustrated in Fig. 2, where computing resources are elastically expanded in respond to the scene change of an event proactively or reactively. The optical networks also elastically provide different-rate communication channels between edge and cloud computing resources. For this purpose, elastic optical network technology provides different spectrum, modulation, and symbol rate wavelengths in an optical fiber to meet a predefined service quality [4]. Future softwarized optical network should be tolerated to the dynamic and elastic behavior of optical signals [5]. Softwarized and elastic optical networking are intensively studied for application to optical transport networks (e.g., [6]), optical access (e.g., [7]), mobile fronthaul networks (e.g., [8]), and data center networks (e.g., [9]).

Kitayama *et al.* [10] identified the capability to synthesize desired switching and transmission functions by software control as a key solution while mentioning following situation of the network operators:

- The cost reduction is the crucial issue to be profitable under a strict price cap of their services with coming capacity crunch.
- Current network operation and management (OAM) is labor intensive. The operating expenditure (OPEX) can be saved and the time for service delivery can be minimized if the module or card of the OTPs can be automatically

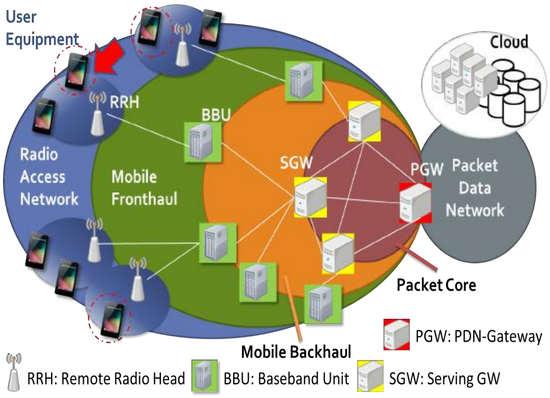


Fig. 1. A mobile network.

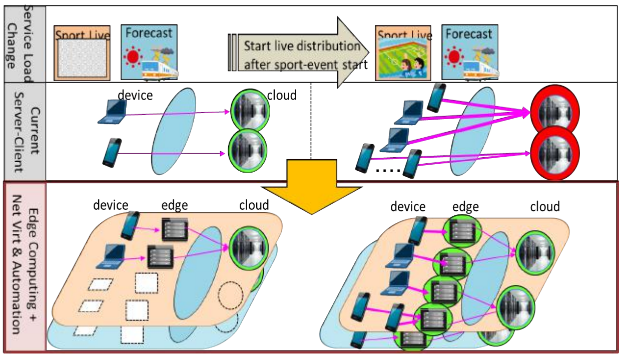


Fig. 2. A softwarized and elastic network. Each service is allocated a separated edge-cloud network with optical technology (green: stable, red: congested).

978-3-903176-07-2 © 2018 IFIP

upgraded by updating software from a remote site, and if the switch can be automatically configured.

These points show the necessity of softwarization and elasticity in the optical network. Moreover, optical edge nodes as well as optical transmission and computing environment are also desired target components for softwarization and elasticity. In a service level, the optical edge nodes accommodate various network protocols such as Ethernet, IP and MPLS in client networks with low capital expenditure (CAPEX) and OPEX. Programmable hardware, which provides different protocols time-by-time in a single component (e.g., line card) based on reconfiguration of electronic circuit and/or software, is powerful for not only service continuity and quick service launch but also CAPEX reduction. However, the requested performance is diverse and higher speed components like 400 Gbps capacity are not yet programmable. Accordingly, optical edge nodes are much costly.

The network operators cannot predict the volume of service requests perfectly. Towards smooth service launch, they prepare some redundant resource pools. The size of the pools is likely proportional to the number of speed grades and number of network protocols. The solution is the *equipment communization* and *hardware elasticity*. In other words, we need reconfigurable technology about *network functions* such as switching, processing and management, *protocols* and *performance* in terms of capacity (e.g., 25 to 400 Gbps with 25 Gbps granularity), availability and manageability in common hardware, in response to the demands from the edge services. Thus, sharp cut of CAPEX and OPEX is achieved.

Network operators dynamically configure appropriate functions such as packet framing and packet processing with requested performance by using an appropriate number of hardware resources like FPGAs, network processors, and CPUs. Collaborative processing functions for meeting a capacity, availability and manageability are also facilitated.

When the optical networks are agile and softwarized, optical signals should be managed carefully. A bulk of optical paths are setup at the initial construction of a network service and many backup paths are activated in the case of a failure event. Here, the input optical power to EDFAs suddenly changes. The gain transient of conventional EDFA causes optical power fluctuation for multiple wavelengths [11]. Generally, the setting of EDFAs and variable optical attenuators (VOA) are readjusted to suppress the power fluctuation. However, this operation takes long time and insufficient for prompt path provision or quick restoration. We need agile and stable power management for such sudden power changes to avoid service interruption.

In this paper, we present elasticity and agility in softwarized optical network construction, service continuation, and service update. Optical hardware is the key to all the situations. Beyond 100 Gbps optical reconfigurable edge nodes and burst tolerable optical amplifiers are presented toward this purpose. These are also beneficial to quick service launch and cost reduction.

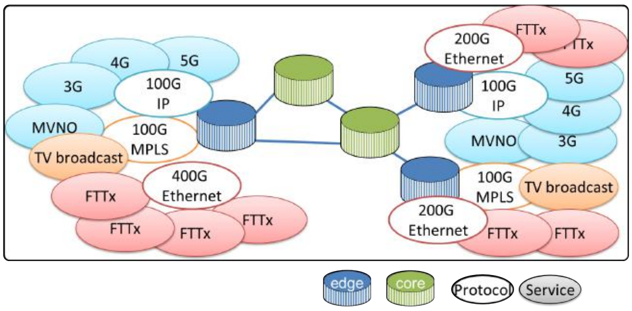


Fig. 3. Optical core and edge nodes for accommodation of diverse protocols and services.

II. Softwarized and Elastic Network Architecture

Our designed network consists of optical core nodes, reconfigurable optical edge nodes and access networks. The core nodes provide a circuit-switched based optical network and are tolerant to power fluctuation against sudden simultaneous addition and deletion of a set of lightpaths. The optical edge nodes support multiple access network protocols and data is encapsulated into (decapsulated from) optical network format such that optical network separately transmits data on each of access network by using WDM and other multiplexing technologies. Figure 3 shows an optical network that accommodates multiple access services on different network protocols.

The network is designed with aware of the following general requirement of optical networks.

- Save existing service interruption. New additional service accommodation and failure should not affect to continuation of the optical paths for in-progress services.
- Save unused equipment stock and/or increase utilization of equipment. CAPEX should be minimum.
- Hasten the service launch and update.

III. Reconfigurable Optical Edge Nodes

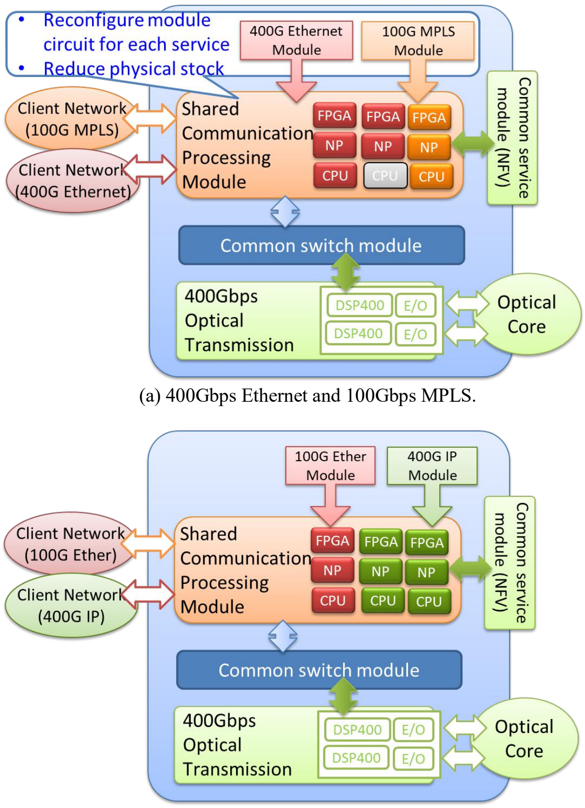
In the beyond 100 Gbps era, cost of equipment modules is higher and keeping redundancy with low CAPEX is difficult. Programmability of optical edge nodes presented in this section saves CAPEX and OPEX of optical networks and time for service launch and update.

Figure 4 shows a reconfigurable optical edge node, which consists of a shared communication processing module, a common switching module and an optical transmission system. The communication processing module co-exists multiple network protocols and sets the protocols accordingly by properly reconfiguring the FPGAs, network processors, and CPUs. By sharing the communication processing modules in different speed and protocols, CAPEX and OPEX can be saved.

Reconfigurable Communication Processor (RCP) over Lambda Project [12][13] designs beyond 100Gbps programmable edge nodes. The node is intended to have resource pools consisting of reconfigurable service modules (RSM) such as filtering, DPI, OAM and IPsec, and reconfigurable processing modules (RPM) for network protocols such as IP, MPLS, and Ethernet. Figure 5 shows the RCP development module. An edge node has a Tbps class switching module and a number of pools of NPs, and FPGAs. Appropriate network protocols can be reconfigured into a portion of RSMs and RPMs.

Toward efficient accommodation of wide variety of different bandwidth client demands, it is expected that optical core network provides different bandwidth links to client networks based on the client requests. The optical core network elastically changes link bandwidth by organizing multiple beyond 100Gbps physical links. Flexible Ethernet (FlexE) is considered as the promising technology for realizing the bandwidth-variable multi-links [14]. FlexE accommodates wide variety of client MAC flows efficiently in the client side while it creates scalable multi-link in the optical network side [15].

FlexE lacks monitoring function for identifying flows that contain bit errors. Tanaka *et al.* [16] developed a function on electronic circuit hardware in which flows containing bit errors are detected by using 25 Gbps FlexE flows. We hope this function will be mapped into flexible channelized links in Fig. 5 toward elastic bandwidth provisioning at the client side. I becomes a promising monitoring technology for 100 Gbps and beyond reconfigurable hardware.


• Reconfigure module circuit for each service
• Reduce physical stock
400G Ethernet Module
100G MPLS Module
Client Network (100G MPLS)
Client Network (400G Ethernet)
Shared Communication Processing Module
FPGA
NP
CPU
Common service module (NFV)
Common switch module
400Gbps Optical Transmission
DSP400
E/O
Optical Core
100G Ether Module
400G IP Module
Client Network (100G Ether)
Client Network (400G IP)

(a) 400Gbps Ethernet and 100Gbps MPLS.

(b) 100Gbps Ethernet and 400Gbps IP.

Fig. 4. Reconfigurable optical edge node. Multiple different network protocols and different speeds supported. Appropriate number of FPGAs, Network Processors and CPUs is used for a network service and it is reconfigured from a remote site.

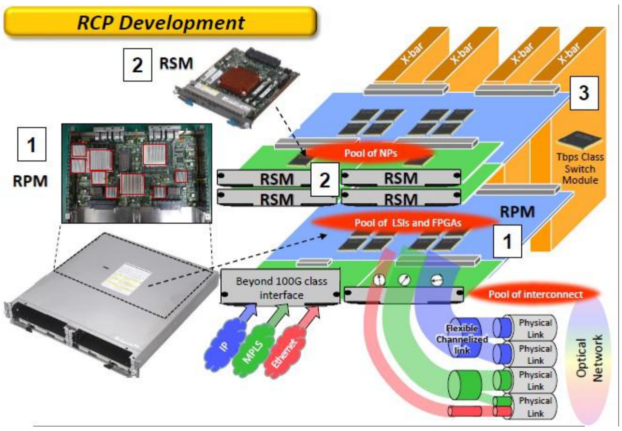


Fig. 5: A reconfigurable optical edge node [12]. Multiple network protocols and diverse network functions are configured into RPM and RSM, respectively. Beyond 100 Gbps interfaces are provided to client networks. Arbitrary rate (e.g., multiple of 25 Gbps) are reserved at optical core network using Flex Ethernet.

IV. Power Management in Agile Optical Networks

A. *Optical Power Fluctuation*

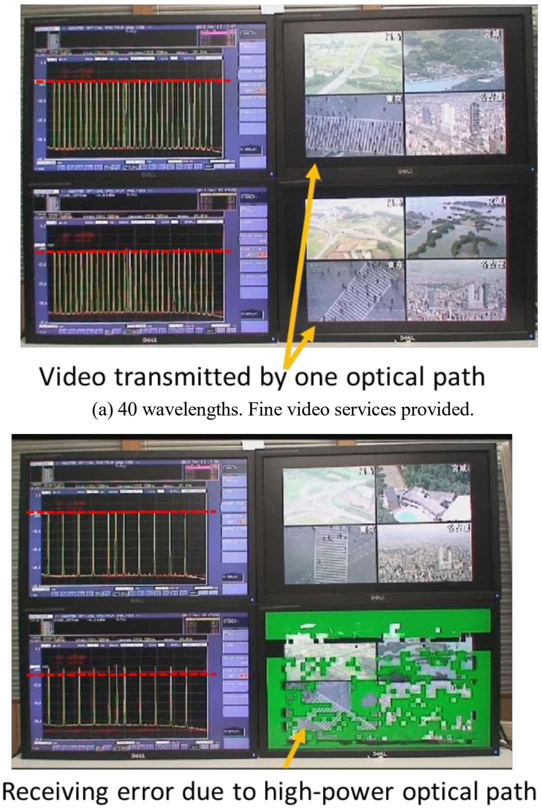
A problem in the agile optical networks is the optical power management of the amplifier-assisted optical networks in dynamic path setup and release. Usually optical paths are setup one-by-one with careful power management with gain-controlled optical amplifier and variable optical attenuators (VOAs) so as not to cause disruption of other paths due to power fluctuation. Several minutes may be acceptable for multi-path setup. On the other hand, in the case of a fiber cut failure on 1:1 protection, we have to recover optical signal reach immediately. At an upstream node of a failed link, we switch the direction of the optical signals to a different optical fiber to deliver the optical signals to the proper destinations.

Assume that 80 wavelengths are multiplexed in a fiber. Up to 80 wavelengths are switched to the different fibers and go into the optical amplifier systems simultaneously. Dynamic gain controlled function does not respond properly and quickly. Existing services multiplexed on a fiber may experience transient signal degradation. Figure 6 shows an image of a service disruption by simultaneous loss of 30 wavelengths. Figures 6(a) and 6(b) have wavelength spectrum and received video on a single wavelength. In each subfigure, top is a result of using burst-tolerable EDFA and bottom is that of conventional EDFA. Through transition from Fig. 6(a) to Fig. 6(b), we observe that the received video quality of the bottom degrades due to slow gain control of the conventional EDFA.

A straightforward way is path-by-path recovering. However, recovery time is too long. For example, recovering time is proportional to the number of damaged optical paths.

B. *Burst-Tolerated Optical Amplifiers for Protection*

We proposed a simultaneous protection framework in wavelength switched optical network (WSON), where burst-tolerated (burst-mode) erbium-doped fiber amplifiers (EDFA) are installed for transient optical power management [11]. This is beneficial to the dynamic environmental change, where a set of lightpaths are setup and released as well as recovery from a failure.


Video transmitted by one optical path
Receiving error due to high-power optical path

(a) 40 wavelengths. Fine video services provided.

(b) 10 wavelengths (top) service continues (bottom) service is disrupted.

Fig. 6. Sudden change of the number of multiplexed wavelengths in a fiber and received video quality on a single wavelength.

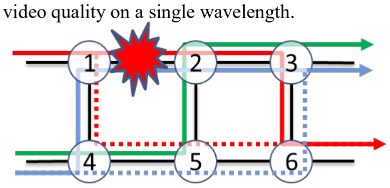


Fig. 7. A link failure and rerouted paths (solid lines are working paths and dashed are backup).

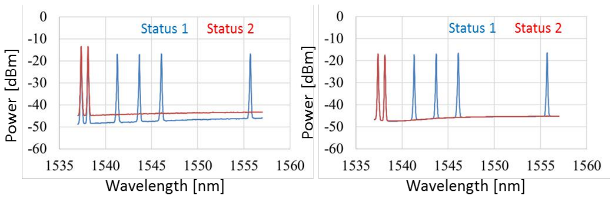


Fig. 8. Spectral waveform. (left) conventional EDFA (right) burst-mode EDFA.

Figure 7 shows a long-distance WSON, where EDFAs are embedded into optical nodes. Denote $l_{i,j}$ for link between nodes i and j. Working wavelength paths are drawn in solid lines and backup ones are dashed lines. 1:1 protection (not 1+1 protection) is assumed for red and blue working paths. In the case of link $l_{1,2}$ cut, backup paths on links $l_{1,4}$, $l_{4,5}$, $l_{5,6}$, and $l_{3,6}$ become active and link $l_{2,3}$ loses two wavelength paths simultaneously. If a number of wavelengths paths in a link are active or disappear in a very short period of a time, optical power at the link may steeply reduce and/or increase by the gain fluctuation due to time dependent saturation effect of the optical amplifier [17] and this phenomenon may give ill influences to the optical systems and the working paths for different services. For example, the green-colored working path from node 4 to node 3 remains at link $l_{1,2}$ cut so automatic gain controller (AGC) in an EDFA in node 2 may increase power of the green path suddenly by the effect of sudden loss of working paths (red and blue lines). On the other hand, the green path may decrease its power at link $l_{4,5}$ due to sudden appearance of backup paths (dashed lines). Then, a network system to prevent such power fluctuations is necessary.

Problem in the traditional system is that optical nodes sequentially process the path setup for multiple wavelengths due to the signal power stability of optical network even though the prompt activation of backup paths is required. As a result of the sequential processing, the processing time increases in proportion to the number of the channels to handle by one wavelength each. Therefore, the parallel processing method is necessary to reduce the processing time.

We proposed to use the parallel path processing with burst-tolerant EDFA [11] so that optical nodes can setup of multiple wavelength paths by only one command. For dynamic optical paths, transient-suppressed burst-mode EDFAs are key components for the power management and the system stability. Conventional EDFAs with burst input signals due to setup and release of multiple wavelength paths may cause optical power surges which damage optical components or impose gain transients which impair the signal quality. Here, we introduced the parallel path processing method with burst-tolerant EDFA into the protection framework to cope with sudden loss and appearance of optical signals on multiple wavelength paths. We showed the framework works well experimentally and achieved 9-sec path setup for 4 wavelengths with no transient signal degradation. Notably, a remaining wavelength path can keep high-quality data and video transmission. Figure 8 shows the wavelength spectrum after a link failure in the experiment, where 6 paths (blue) are reduced to 2 paths (red) [18]. We observed that conventional EDFA raised optical power in the status change while burst-tolerant one did not. This framework is extendible to M:N protection, where N backup paths are prepared for M working paths.

C. Signal Fluctuation Propagation

A link failure causes degradation of quality of transmission (QoT) to the paths which do not transit the failed link as well as the disconnected paths at downstream nodes, as we discussed in the previous subsection. That is, optical signals are simultaneously disappeared at the later-hop nodes. More impressively, the QoT (Quality of Transmission) degradation propagates over a wide range of the network. We identified that a link failure causes temporal QoT degradation propagation of roughly 40 % existing paths in a whole network by conventional EDFA [19], which will be described below. Note that paths in the failure link are called disconnected paths in this paper. Paths of which quality is degraded by a link failure is called QoT damaged paths.

Figure 9 shows QoT degradation propagation in a network. There are three provisioned wavelength-multiplexed optical paths. After link 7 failed, optical signals disappeared at links 18 and 22 due to disconnection of path A. Then, paths B and C are excessively amplified due to the gain transient of EDFAs. It means that the QoT of provisioned paths B and C are degraded even if the paths do not transit the failure link. In this paper, we call these paths as QoT damaged paths. In addition, the QoT of paths which share links with the damaged paths is also degraded. Thus, QoT degradation is widely propagated. For example, in Fig. 9, after link 7 failed, path D is also insufficiently amplified due to the increased power of paths B and C. The QoT of path D is also degraded in spite of no sharing with path A via QoT damaged paths B and C.

The QoT degradation would be recovered by readjusting EDFAs and VOAs at each link, however, dynamic gain controlled function does not respond properly and quickly. Moreover, this gain control function is operated sequentially along the paths. During these sequential operations, many existing services, regardless of whether they transit through a failed link or not, suffer from signal quality degradation.

We confirmed network-scale impact of the QoT degradation propagation by computer simulation when conventional EDFAs are used [19]. We firstly adopted JPN12 network (12 nodes, 34 unidirectional links, Fig. 9) as evaluated topology. Optical paths are provisioned between all source-destination pairs. Each simulation trial, a link failure occurs and the optical paths through the failed link are disconnected. Figure 10 shows the number of affected links by link failure event. The horizontal axis means disconnected link ID. The vertical axis shows the number of affected links. From this figure, a link failure event forces other neighbor links to suffer optical signal power reduction. In the worst case, a link failure event (i.e., link 4) affects 10 links. This is roughly 30% of links, where the QoT of optical paths are damaged. Figure 11 plots the maximum, average, and minimum ratio of damaged paths per link in arbitrary one link failure cases. Note that this result does not include disconnected paths. Edge side links such as link 1 has no damaged paths. From this figure, when a link failure occurs, roughly 25% paths are encountered QoT degradation on average even if the paths do not transit a failure link. We suppose 7.5 dB or more variation of power gives severe influence on the paths through the conventional EDFAs from our previous empirical knowledge. Paths on 4 links among working 32 unidirectional links are damaged by a failure of a single link due to QoT degradation propagation.

From these results, optical paths on a wide range of networks are impacted even if a single link failure occurs. On the other hand, because burst tolerable EDFA can suppress transient optical power quickly, it is more effective for mitigating QoT degradation propagation compared with conventional EDFAs.

We experimentally demonstrated the signal power behavior of remaining optical paths at downstream in a link failure event [19]. 44 wavelengths with 100 GHz spacing are multiplexed at a single fiber, where 8 wavelengths come from a single source, *x* and (36-*x*) wavelengths come from other different sources. 36 wavelengths are multiplexed at first and remaining 8 wavelengths are then multiplexed at a different link.

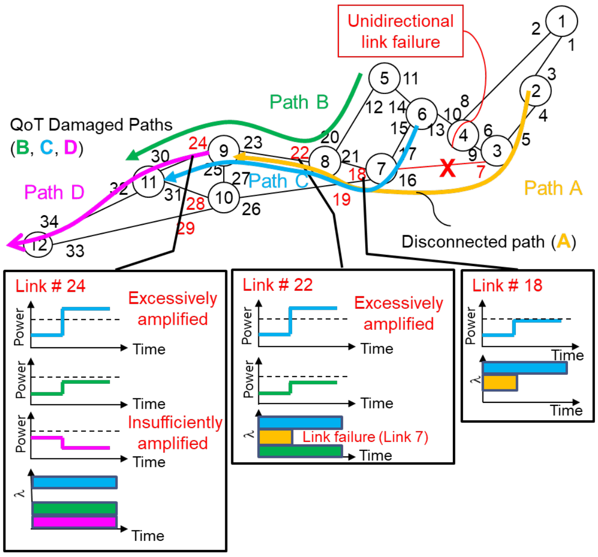


Fig. 9. QoT degradation propagation.

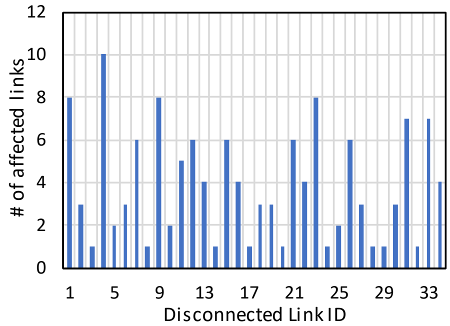


Fig.10: The number of affected links by one link failure in JPN12 topology.

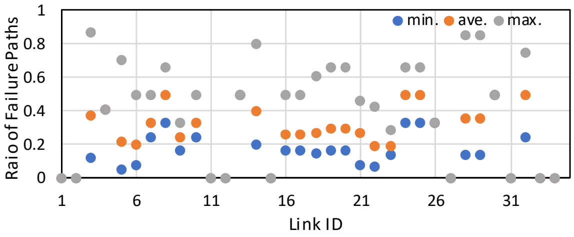
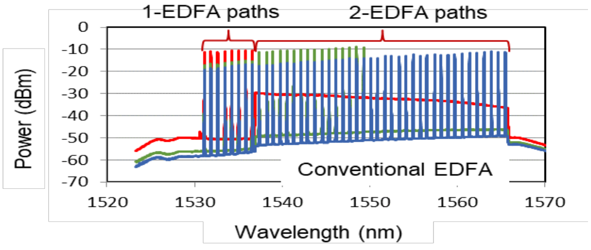
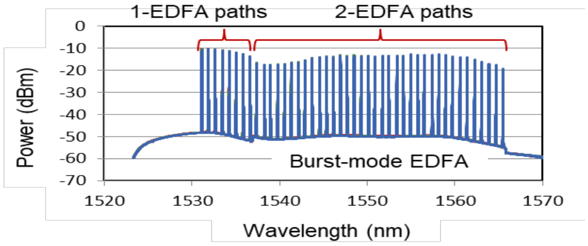


Fig. 11: Affected link ID v.s. Ratio of damaged paths against transit paths per link under arbitrary link failure case.



(a) with conventional EDFAs.



(b) with burst tolerated EDFAs.

Fig. 12: Spectra of optical paths measured.

At outputs of 36 and 44 wavelengths multiplexed, conventional EDFAs or burst tolerant EDFAs are inserted for compensating loss given by VOA. Therefore, optical paths on 36 wavelengths from two sources are transmitted through two-stage EDFAs and other optical paths on 8 wavelengths from the remaining source are transmitted through an EDFA. In each condition, *x* wavelengths are lost by the failure of the link.

Here, we investigated the variation of optical signal power of working optical paths. Figure 12 shows the spectra of optical paths measured in the case of conventional EDFAs and burst tolerable EDFAs. Three spectra acquired in 3 conditions ($x = 0$, 20, and 36) were superimposed. In conventional EDFA (Fig. 12(a)), we confirmed that disconnected optical paths due to a link failure gave the power change into remaining optical paths at multiple downstream nodes of the failure link. In addition, the power change became bigger as the number of disconnected paths increased. On the other hand, burst tolerant EDFAs did not affect the spectra and the signal power of remaining optical paths in all conditions (see Fig. 12(b)). When this experimental data is fed into simulation using JPN12 in Fig. 9, we observed that link 7 failure gives damage to paths on 4 links due to QoT degradation propagation [19].

V. Concluding Remarks

Elastic and agile features for softwarized optical networks are promising to accommodation of diversified services. We have presented elastic and agile technologies for network construction, service continuation, and service update. Optical edge nodes that provide programmability among multiple network protocols and multiple classes of transmission and processing speeds for lower CAPEX and agile network setup, has been presented. Optical burst mode EDFAs for proper optical power management in network protection for service continuation of in-service paths have also been presented.

Acknowledgment

The authors wish to thank to Drs. Yoshinari Awaji, Masaki Shiraiwa, and Naoya Wada for continuous discussion of this work. This work includes direction and current outcome of Reconfigurable Lambda Project conducted by ALAXALA Corporation, NTT Corporation and Keio University.

References

[1] Recommendation ITU-R M.2083-0 "IMT Vision. Framework and overall objectives of the future development of IMT for 2020 and beyond," Sep 2015.

[2] 5GMF White Paper "5G mobile communications systems for 2020 and beyond," (2016). http://5gmf.jp/en/

[3] H. Harai, "Softwarized, Elastic and Agile Optical Networks for Dynamic Environmental Change and Failure Recovery," Optical Fiber Communication Conference (OFC 2018), M3A.4, Mar 2018 (invited).

[4] M. Jinno, H. Takara, B. Kozicki, Y. Tsukishima, Y. Sone, and S. Matsuoka, "Spectrum-Efficient and Scalable Elastic Optical Path Network: Architecture, Benefits, and Enabling Technologies," IEEE Commmunications Magazine, Vol.47, No. 11, pp. 66-73, 2009.

[5] D. C. Kilper and Y. Li, "Optical physical layer SDN: Enabling physical layer programmability through open control systems," Optical Fiber Communication Conference (OFC 2017), W1H.3, Mar 2017.

[6] R. Martínez, R. Casellas, R. Vilalta and R. Muñoz, "Experimental Evaluation of a PCE Transport SDN Controller for Dynamic Grooming in Packet over Flexi-Grid Optical Networks," ECOC 2017, P2.SC7.45, Sep 2017.

[7] A. Elrasad and M. Ruffini, "Frame Level Sharing for DBA Virtualization in Multi-Tenant PONs," ONDM 2017, May 2017.

[8] A. Tzanakaki, A. Markos and D. Simeonidou, "Optical networking: An important enabler for 5G," ECOC 2017, M.2.A1, Sep 2017.

[9] C. Jackson, K. Kondepu, Y. Ou, A. F. Beldachi, A. P. Cruz, F. Agraz, F. Moscatelli, W. Miao, V. Kamchevska, N. Calabretta, G. Landi, S. Spadaro, R. Nejabati and D. Simeonidou, "COSIGN : A Complete SDN Enabled All-Optical Architecture for Data Centre Virtualisation with Time and Space Multiplexing," ECOC 2017, W.2.A.4, Sep 2017.

[10] K. Kitayama, A. Hiramatsu, M. Fukui, T. Tsuritani, N. Yamanaka, S. Okamoto, M. Jinno and M. Koga, "Photonic Network Vision 2020—Toward Smart Photonic Cloud," IEEE/OSA Journal of Lightwave Technology. Vol. **32**, No. **16**, pp. 2760-2770, Aug 2014.

[11] M. Shiraiwa, H. Furukawa, T. Miyazawa, Y. Awaji and N. Wada, "High-speed wavelength resource reconfiguration system concurrently establishing/removing multiwavelength signals," IEEE Photonics Journal, Vol. **8**, No. **2**, Apr 2016.

[12] Reconfigurable Communication Processor over Lambda Project, http://www.pilab.jp/ipop2017/exhibition/panel/iPOP2017_ALAXALA_panel.pdf, iPOP 2017 (web, Jan 29, 2018 accessed).

[13] NICT web, http://www.nict.go.jp/collabo/commission/k_189.html (web, Jan. 29, 2018 accessed)

[14] S. J. Trowbridge, "Flex Ethernet Implementation Agreement 1.0," OIF, March 2016.

[15] T. Tanaka, S. Kuwabara, T. Inui, Y. Yamada and S. Kobayashi, “A High-Availability Scheme for Bandwidth-Variable Multi-Links with Flex Ethenet," iPOP 2017, June 2017.

[16] T. Tanaka, S. Kuwabara, T. Inui, Y. Yamada and S. Kobayashi, “State Monitoring of Flexible Channelized Links in Flex Ethernet,” IEICE Communications Society Conference (B-10-53), p. 169, Sep 2017. (in Japanese)

[17] C. Tian and S. Kinoshita, "Analysis and control of transient dynamics of EDFA pumped by 1480- and 980-nm lasers," IEEE/OSA Journal of Lightwave Technology, Vol. **21**, No. **8**, pp. 1728-1734, Aug 2003.

[18] H. Furukawa, M. Shiraiwa, H. Harai and N. Wada, "Softwarized dynamic optical switching network suppressing transient optical power in link failures," Photonics in Switching (PS 2017), PTu2D.2, Jul 2017.

[19] Y. Hirota, M. Shiraiwa, H. Furukawa, H. Harai and N. Wada, “Demonstrating Network-scale Gain Transient Impact of Multiple Series EDFAs in Link Failure Cases,” Optical Fiber Communication Conference (OFC 2018), Tu3E.5, Mar 2018.

Resource Allocation in Slotted Optical Data Center Networks

K. Kontodimas[1], K. Christodoulopoulos[1], E. Zahavi[3], E. Varvarigos[1,2]
[1]School of Electrical and Computer Engineering, National Technical University of Athens, Greece
[2]Department of Electrical and Computer Systems Engineering, Monash University, Australia
[3]Mellanox Technologies Ltd., Yokneam, Israel
{kontodimas, kchristo}@mail.ntua.gr, eitan@mellanox.com, vmanos@central.ntua.gr

Abstract— **The introduction of all-optical switching in data center interconnection networks (DCN) is key for addressing several of the shortcomings of state-of-the-art electronic switched solutions. Limitations in the port count and reconfiguration speed of optical switches, however, require novel DCN designs offering network scalability and dynamicity. We present the NEPHELE DCN which relies on hybrid electro-optical top-of rack (TOR) switches to interconnect servers over multi-wavelength all-optical rings. We described in detail the NEPHELE control cycle which follows the SDN paradigm. We evaluate the performance of NEPHELE regarding the effect of the control plane delay under realistic traffic.**

Keywords— Time-Wavelength-Space division multiplexing; slotted and synchronous operation; dynamic resource allocation, scheduling; matrix decomposition

I. Introduction

The widespread availability of cloud applications to billions of end-users and the emergence of platform- and infrastructure-as-a-service models rely on concentrated computing infrastructures, the Data Centers (DCs). As traffic within the DC (east-west) is higher than incoming/outgoing traffic, and both are expected to continue to increase [1], DC networks (DCN) play a crucial role. High throughput, scalable and energy/cost efficient DCN networks are required to fully harness the DC potential.

State-of-the-art DCNs are based on electronic switches connected in fat-tree topologies using optical fibers, with electro-opto-electrical transformation at each hop [2]. However, fat-trees tend to underutilize resources, require a large number of cables and switches, suffer from poor scalability and upgradability (lack of transparency), and they result in very high energy consumption [3], [4]. Application driven networking [5], [6], an emerging trend, would benefit from a network that flexibly allocates capacity where needed.

The introduction of optical switching in DCN is a key for solving these shortcomings. Many recent works proposed hybrid electrical/optical DCN, a survey of which is presented in [7]. The authors of [8] and [9], proposed a DCN in which heavy long-lived (elephant) flows are selectively routed over an optical circuit switched (OCS) network, while the rest of traffic goes through the electronic packet switched (EPS) network. The identification of elephant flows is rather difficult on the fly, while it was observed that such flows are not very typical [4], making it difficult to sustain high OCS utilization. Instead, a high connectivity degree is needed [4]. To enable higher connectivity, [10] proposed and prototyped a very dense hybrid DCN that also supports multi-hop connections. The total delay, including control plane and OCS hardware reconfiguration (micro electro-mechanical system – MEMS – switches), was measured to be in the order of hundreds of msec. Multi-hop routing was exploited as *shared* circuits in [11] controlled via extended OpenFlow [12], showing that circuit sharing compensates for slow OCS reconfigurations.

Other proposed DC interconnects completely lack electrical switches. Proteus, an all-optical DCN architecture based on a combination of wavelength selective switches (WSS) and MEMS was presented in [13]. Again, multi-hop is used to improve the utilization and hide the low reconfiguration speed of MEMS. [14] introduced hybrid OCS and optical packet/burst switching (OPS/OBS) architectures, controlled using SDN. Various other architectures based on OPS/OBS were proposed in [7], [15] and references therein. However, OPS/OBS technologies are not yet mature, so their current target could be only small-scale networks with limited upgradability potentials.

The authors in [16] presented a hybrid DCN architecture called Mordia, which uses WSS to achieve switching times of 11.5 μs. Mordia operates in a dynamic slotted manner to achieve high connectivity. However, the scalability of Mordia is limited as it uses a single wavelength division multiplexing (WDM) ring that can support traffic for a few racks, while resource allocation algorithms exhibit high complexity and cannot scale to large DCs.

The European project NEPHELE developed an optical DCN that leverages hybrid electrical/optical switching with SDN control to overcome current datacenter limitations [17]. To enable dynamic and efficient sharing of optical resources and collision-free communication, NEPHELE operates in a synchronous slotted manner. Timeslots are used for rack-to-rack

978-3-903176-07-2 © 2018 IFIP

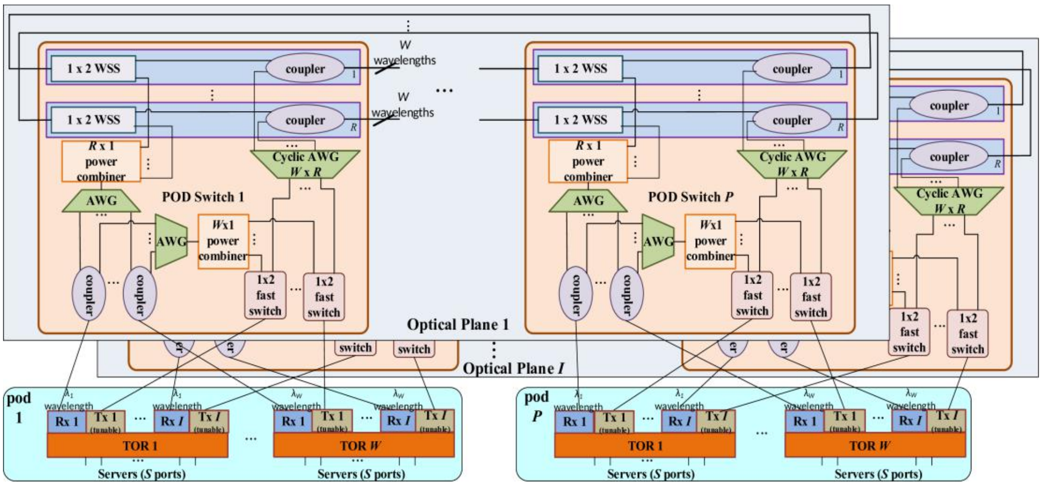


Fig. 1. NEPHELE Resource allocation and Data cycles.

communication and are assigned dynamically, on a demand basis, so as to attain efficient utilization, leading to both energy and cost savings. Moreover, multiple wavelengths and optical planes are utilized for a scalable and high capacity DC network.

The NEPHELE network relies on WSSs, which are faster than MEMS used in [8]-[11], and more mature than OPS/OBS used in [14]-[15]. The fast switching times, along with the dynamic slotted operation, provide high and flexible connectivity. Compared to Mordia [16], which also relies on WSS, NEPHELE is more scalable: it consists of multiple WDM rings, re-uses wavelengths, and utilizes cheap passive routing components and scalable scheduling schemes. We present fast scheduling algorithms for NEPHELE DCN to meet dynamic reconfiguration requirements and evaluate their effect along with the control plane overhead on the performance of realistic applications.

In the following we shortly present the NEPHELE data plane (Section II) and then we describe in detail its control cycle (Section III) and the scheduling approaches (Section IV). Finally, using a packet level simulator we evaluate the effect of the control plane delay under realistic traffic (Section V).

II. NEPHELE Network Architecture

The NEPHELE DCN, shown in Fig. 1, is divided into P pods of racks and is built out of hybrid electrical/optical top-of-rack (TOR) switches and all-optical POD switches. A pod consists of I POD switches and W TOR switches, interconnected as follows: each TOR switch has I ports with tunable transmitters (Tx) and each one is connected to a different POD switch (among I). A rack consists of S servers connected through S corresponding ports to the TOR switch. The POD switches are interconnected through WDM rings to form optical planes. An optical plane consists of a single POD switch per pod (for a total of P POD switches) connected with R fiber rings. Each fiber ring carries WDM traffic of W wavelengths (by design equals the number of racks/pod). There are I identical and independent optical planes. In total, there are $I \cdot P$ POD switches, $W \cdot P$ TOR switches and $I \cdot R$ fiber rings.

The key routing concept is that each TOR switch listens to a specific wavelength and wavelengths are re-used among pods. Each TOR employs Virtual Output Queues (VOQ) per TOR destination to avoid head-of-line blocking. The NEPHELE TORs employ tunable Tx that are tuned according to the desired destination. Thus, the tunable Tx selects the destination/ wavelength to route traffic. The 1 × 2 space switch inside the corresponding POD switch is set according to whether the destination is in the same pod with the source or not. Intra-pod traffic is forwarded to an AWG that passively routes it towards the selected destination. Inter-pod traffic is routed via the 1 × 2 switch towards a $W \times R$ CAWG and then to one of the R fiber rings (according to the input port/source and the wavelength used). The traffic propagates in the ring passing through intermediate POD switches and is dropped at the destination pod, by setting appropriately the related WSSs. The drop port is connected to an AWG that again passively routes the traffic. Finally, NEPHELE operates in a *slotted and synchronous manner* as discussed in the next section. A parallel EPS network can also be utilized to handle high priority traffic and/or ACK TCP packets which, according to simulations presented in this paper, seem to play a major role in the network performance. A more detailed description of the NEPHELE data plane is provided in [17][18]. Also [17] presents some basic techno-economic results.

III. NEPHELE Control Cycle

NEPHELE exploits the SDN concept that decouples the data and control planes through open interfaces, enabling programmability of the networking infrastructure [17]. A key functionality of NEPHELE SDN controller is the coordination of the resource usage, including the timeslot/plane dimension [19]. Thus, an important building block of the SDN controller is the *scheduling engine*, which allocates resources for TOR communication in a periodic and on demand manner.

Two scheduling approaches are envisioned in NEPHELE.

We assume that long and medium term traffic variations can be solved with offline scheduling algorithms. The offline scheduling algorithms that run periodically or on demand (triggered by significant application/traffic changes) can calculate the optimum resource allocation (schedule) since their running time is not crucial. Then sort-time traffic variations or failure events are treated by faster online scheduling algorithms that calculate incremental changes in the running schedule.

Moreover, we also envision two traffic identification modes: (i) application-aware and (ii) feedback-based. The former mode [5], [6] assumes that applications communicate to the NEPHELE SDN controller (or via the DC orchestrator) their traffic requirements. The latter, feedback-based, mode assumes that the central controller collects (monitors) data from the TOR queues [9]. Hybrid versions of these two modes are also applicable.

In the NEPHELE network we divide the time in slots and we have periods of T timeslots. In all scheduling and traffic identification cases, we assume that the controller creates the *queue matrix* $\mathbf{Q}(n)$ (of size $(W \cdot P) \times (W \cdot P)$) for period n. We denote by $\mathbf{A}(n)$ the matrix of arrivals at the queues during period n, and by $\mathbf{S}(n)$ the schedule calculated for period n.

The NEPHELE network operates in *two parallel cycles*: a) Data communication cycles of T timeslots (also referred to as a *Data period*), where actual communication between TORs takes place and b) Control plane cycles of duration C (measured in Data periods), where control information is exchanged. Control plane cycle n corresponds to Data period n, and computes the schedule $\mathbf{S}(n)$ to be used during that period. Note, however, that the schedule is computed based on information that was available C periods earlier than the Data period the control plane cycle is applied to. Thus, $\mathbf{S}(n)$ is a function of $\mathbf{Q}(n-C)$, i.e.,

$$\mathbf{S}(n) \ = f\ (g[\mathbf{Q}(n-C)]), \qquad (1)$$

where $\widehat{\mathbf{Q}}(n) = g[\mathbf{Q}(n-C)]$ is the function that creates the *estimated queue matrix* $\widehat{\mathbf{Q}}(n)$ from $\mathbf{Q}(n-C)$, upon which the schedule is calculated, and f is the scheduling algorithm. When $C\ > 1$ period (control delay is larger than the Data period), a new Control plane cycle still starts every Data period. So, there are C Control plane cycles running in parallel.

The queues evolution is described by $\mathbf{Q}(n+1) = \mathbf{Q}(n) + \mathbf{A}(n) - \mathbf{S}(n)$, where $\mathbf{S}(n)$ is calculated as in Eq. (1). The control plane delay C depends on many factors, including the execution time of the scheduling algorithm, the delay of the control protocol carrying information from TORs to the SDN controller and from the SDN controller to the data plane devices. Both delays depend on the network size and the choice of the Data period T. For the scheduling decisions to be efficient, the scheduling matrix $\mathbf{S}(n)$ computed based on an estimated queue matrix $\widehat{\mathbf{Q}}(n)$ [which in turn is calculated from $\mathbf{Q}(n-C)$] should be a *"good"* scheduling to be used during Data interval n. This is true when $\widehat{\mathbf{Q}}(n)$ is a good approximation of $\mathbf{Q}(n)$. In case of slowly or moderately changing traffic, we expect calculations made for previous periods to be valid.

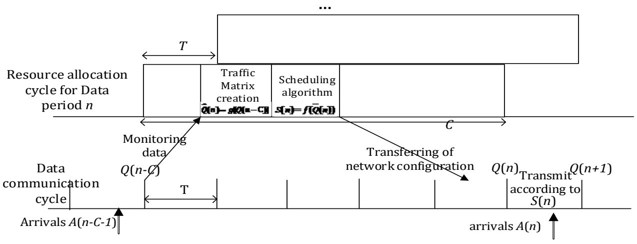

...
T
Resource allocation cycle for Data period n
Traffic Matrix creation
$\widehat{Q}(n)=g[Q(n-C)]$
Scheduling algorithm
$S(n)=f[\widehat{Q}(n)]$
C
Monitoring data
$Q(n\text{-}C)$
Transferring of network configuration
$Q(n)$
Transmit according to $S(n)$
$Q(n\text{+}1)$
Data communication cycle
T
Arrivals $A(n\text{-}C\text{-}1)$
arrivals $A(n)$

Fig. 2. NEPHELE Data and Control plane cycles.

In estimating $\widehat{\mathbf{Q}}(n)$ from $\mathbf{Q}(n-C)$, it is possible to use statistical predictions, filters, and other (notably, predefined cluster application communication patterns) methods to improve performance. Moreover, it is possible for the scheduler to *void fill* the unallocated resources in $\mathbf{S}(n)$ to enable opportunistic transmissions. Finally, the overall scheme is "self-correcting": if some queues are not served for some periods due to poor scheduling and their size grows due to new arrivals, this will be communicated with some delay to the controller, and they will eventually be served.

IV. Scheduling Algorithms

We now focus on the scheduling problem in the NEPHELE network. In any traffic identification mode (application-aware or feedback-based), we start from the estimated queue matrix $\widehat{\mathbf{Q}}(n)$ and devise algorithms to calculate the schedule $\mathbf{S}(n)$ (function f in Eq. (1)). For reference, we can assume that we calculate the estimated queue matrix (function g in Eq. (1)) as $\widehat{\mathbf{Q}}(n) = \mathbf{A}(n-C-1) + \widehat{\mathbf{Q}}(n-1) - \mathbf{S}(n-1)$, where in the expression we acknowledge that due to control plane delay C, the central scheduler has access to (delayed) arrival information $\mathbf{A}(n-C-1)$ instead of $\mathbf{A}(n)$. This corresponds to the case where the schedule $\mathbf{S}(n)$ calculated on $\widehat{\mathbf{Q}}(n)$ serves the arrived traffic $\mathbf{A}(n-C-1)$, plus a correction equal to traffic not served in the previous period $\widehat{\mathbf{Q}}(n-1) - \mathbf{S}(n-1)$.

The scheduling algorithm provides the schedule $\mathbf{S}(n)$, which identifies the TOR pairs that communicate *during each timeslot and for each optical plane* for Data period n. Note that wavelengths and rings are dependent resources; the selected wavelength is determined by the destination and the ring depends on the source and destination [18]. So, in NEPHELE, the allocated resources are the timeslots and the optical planes ($I \cdot T$ in total), also called *generalized slots*. The scheduling algorithm takes the estimated queue matrix $\widehat{\mathbf{Q}}(n)$ and decomposes it (fully or, if not possible, partially) into a sum of $I \cdot T$ *permutation matrices.* These identify the source and destination TORs that communicate at each generalized slot. The scheduling algorithm takes into account the constraints under which a TOR can transmit/receive to/from a single TOR.

As discussed earlier, there are two scheduling approaches: offline and incremental, trading off execution time for optimality.

A. *Offline Scheduling*

Offline scheduling pertains to the optimal decomposition of matrix $\widehat{\mathbf{Q}}(n)$. We define the *critical sum* $H[\widehat{\mathbf{Q}}(n)] = h$ as the maximum of the row sums and column sums of matrix $\widehat{\mathbf{Q}}(n)$. The decomposition of $\widehat{\mathbf{Q}}(n)$ can be performed in an optimal manner following the well-known Hall's theorem (an integer version of the Birkhoff-Von Neumann theorem). A more detailed analysis of these techniques is discussed in [18].

The column sums will be *on the average* $\leq S \cdot T$, if the destinations of packets are uniformly distributed, or with *high probability*, if the network operates at less than full load. Also, a flow control mechanism can be applied to smoothen the traffic going to a given destination and enforce this constraint. In such a ("typical") case, the column sums of the arrival matrix $\mathbf{A}(n)$ will be $\leq S \cdot T$ and so will also be its critical sum, and thus the schedule $\mathbf{S}(n)$, that is calculated based on $\widehat{\mathbf{Q}}(n) = \mathbf{A}(n - C)$, assuming $S \leq I$, can be chosen so as to completely serve all the arrivals in $\mathbf{A}(n - C)$ in the available $I \cdot T$ generalized slots. Thus, in the typical case, NEPHELLE provides both *full throughput and delay guarantees*.

In the worst case, the optimal algorithm executes a maximum matching algorithm $I \cdot T$ times. Finding a maximum matching with e.g. the well-known Hopcroft–Karp algorithm exhibits complexity of $O(M(\widehat{\mathbf{Q}}) \cdot \sqrt{W \cdot P})$, where $M(\widehat{\mathbf{Q}})$ is the number of nonzero elements in $\widehat{\mathbf{Q}}$, which in the worst case equals $(W \cdot P)^2$.

B. *Incremental Scheduling*

It is evident from the above discussion and related results [18] that offline scheduling is not suitable to serve short-term varying traffic. Measurements in commercial data centers indicate that application traffic can be relatively bursty, with flows activating/ deactivating within ms [4]. However, the traffic tends to be highly locally persistent: a server tends to communicate with a set of destinations that are located in the same rack or the same cluster/ pod [4]. Note that TOR switches in NEPHELE aggregate the flows of the servers in a rack, smoothening out the burstiness of individual flows, especially considering locality persistent traffic.

A detailed definition of locality persistency is given in [18]. We denote by $\mathbf{D}(n) = \mathbf{A}(n) - \mathbf{A}(n-1)$ the arrival matrix difference, and by $\delta(\cdot)$ the density of a matrix. Then, the *Locality Persistency Property* holds if $\delta(|\mathbf{D}(n)|) \ll 1$. We also define the estimated queue matrix difference as $\mathbf{D}_{\widehat{\mathbf{Q}}}(n) = \widehat{\mathbf{Q}}(n) - \widehat{\mathbf{Q}}(n-1)$. Note that when arrivals have the locality persistency property, then we also expect $\delta(|\mathbf{D}_{\widehat{\mathbf{Q}}}(n))|) \ll 1$.

Motivated from the high locality observation, we investigated incremental scheduling, i.e. rely on the previous schedule to calculate the new one. The expected benefit is that we need to update only changed elements of the permutation matrices of the decomposition of $\widehat{\mathbf{Q}}(n+1)$, with no need to modify the rest. A number of incremental scheduling algorithms are presented in [18] where we also present a greedy incremental heuristic with complexity of $O(\delta(|\mathbf{D}_{\widehat{\mathbf{Q}}}|) \cdot I \cdot T \cdot (W \cdot P)^2)$, where $\delta(|\mathbf{D}_{\widehat{\mathbf{Q}}}|) \ll 1$ in view of the persistency property.

This heuristic achieves throughput that is close to optimal and running time in the order of hundreds of ms [18], using Matlab and an Intel® Core™ i5 laptop. A parallel implementation of the heuristic algorithm on an FPGA was presented in [20] and showed that the schedule can be computed in tens of ms even for dense input matrices $(\delta(|\mathbf{D}_{\widehat{\mathbf{Q}}}| < 0.25)$ using incremental algorithms. This implies that we can calculate the schedule within 1 Data period, which is quite promising for the performance of the NEPHELE architecture. However, the control plane overhead C depends also on the signaling overhead: monitoring (in feedback based traffic estimation mode) and transferring the schedule to the data plane devices, the NEPHELE POD and TOR switches. The effect of the total control plane overhead is examined in the next section.

V. PERFORMANCE EVALUATION

A. *Simulation Model and parameters*

To evaluate the performance of the NEPHELE architecture, we developed a packet level network simulator. The simulator is an extension of OMNET++ 4.3.1 with INET 2.4.0, a framework that contains implementations for various real-life network components and protocols. We evaluated the network performance using an application that simulates MapReduce, which was implemented by Mellanox.

In our simulation model, we consider that the control plane delay, which includes the time to gather monitoring information (if we operate the network in feedback based, would be zero in application-aware mode), to calculate the schedule (which as previously discussed is fast, within 1 Data period [20]) and to distribute the schedule to the data plane devices, is described through the parameter C. This in turn defines the number of multiple identical (virtual) schedulers that work in parallel. We also assume that each parallel scheduler knows the C previous schedules (feasible, as the schedule is computed in 1 Data period).

In the simulated network we run a number of MapReduce jobs simultaneously. Each MapReduce job requires a number of worker nodes: *mappers*, *reducers* and *storage servers* and runs for a number of iterations. The communication pattern for each particular MapReduce job, regarding the server where each worker node resides, the size of the MapReduce data produced in each phase, the number of MapReduce iterations and the computational delay for *map* and *reduce* operations, are described using appropriate semantics in an input file. In the simulations the assignment of the worker nodes to the servers was random. This means that a server could host simultaneously multiple types of worker nodes for the same or different jobs.

The communication between the worker nodes is achieved via Ethernet packets over TCP/IP. We assumed full-duplex 10G Ethernet from a server to the corresponding NEPHELE TOR switch. For the TOR to TOR communication we rely on NEPHELE TDMA operation. The Ethernet packets are stored in

Virtual output queues (VOQ) and served in slots according to the computed schedules.

We study the impact of various parameters, such as the Control cycle delay C, the number of MapReduce jobs, or the cluster size $(P \cdot W)$, on the throughput, in terms of total makespan. The makespan is defined as the time it takes for all MapReduce jobs to finish. TABLE I. summarizes the NEPHELE network parameters, as well as the TCP-related parameters. Note that a target for the NEPHELE network would be to have 1600 racks with 20 servers each, while each timeslot (of duration 200μs) aggregates the traffic of all servers residing in a rack. Since it is not possible to simulate a fully-fledged NEPHELE network, but only smaller clusters with fewer servers per rack, the NEPHELE parameters are also scaled down accordingly. We assumed $I = 2$ optical planes, and the scheduling period T took values so that the generalized slots/resources equals to the number of racks $(T \cdot I = P \cdot W)$.

The key parameters that we examine are the Control cycle delay C, the number of MapReduce jobs that run simultaneously in the cluster and the number of cluster's racks; their default values are 4, 5 and 8, respectively. In all scenarios, the ratios of the MapReduce worker nodes types remained the same: the number of mappers equals to half, while the number of reducers and storage servers equals to a quarter of the available servers. A parallel (dual) network (utilizing 1 Gbps capacity) is also used to route the TCP ACKs.

TABLE I. SIMULATION PARAMETERS

<table>
<tr><th colspan="3">Parameter</th><th colspan="3">Value</th></tr>
<tr><td colspan="3">Number of servers in each rack (S)</td><td colspan="3">2</td></tr>
<tr><td colspan="3">Number of planes (I)</td><td colspan="3">2</td></tr>
<tr><td colspan="3">Link capacity per plane (each direction)</td><td colspan="3">10Gbps</td></tr>
<tr><td colspan="3">Timeslot duration</td><td colspan="3">200μs</td></tr>
<tr><td colspan="3">Maximum segment size (MSS)</td><td colspan="3">625 bytes</td></tr>
<tr><td colspan="3">TCP window size</td><td colspan="3">65000 bytes</td></tr>
<tr><td>Storage server</td><td>Mapper</td><td>Reducer output</td><td>5</td><td>10</td><td>5 Mbytes</td></tr>
<tr><td colspan="3">Mapper processing time</td><td colspan="3">25μs</td></tr>
<tr><td colspan="3">Reducer processing time</td><td colspan="3">20μs</td></tr>
<tr><td colspan="3">Number of MapReduce iterations</td><td colspan="3">3</td></tr>
</table>

We examine three queue matrix estimation policies. The first estimation policy assumes static uniform traffic under which no traffic identification mode (monitoring or application awareness) is assumed and the resource allocation is evenly distributed among the TOR pairs (round-robin scheduling). The second policy assumes that $\widehat{\mathbf{Q}}(n)$ (described in Sections III and IV) is computed based on the most recent known arrivals $\mathbf{A}(n-c-1)$. The third policy is a simplistic prediction mechanism that assumes that the arrivals for the next C Data periods will be equal to the latest $\mathbf{A}(n)$. It then virtually applies the latest C known schedules and computes an estimation for the remainder in the queues when the schedule will be applied (after C Data periods). The above queue estimation policies are combined with the incremental scheduling algorithm which is extended with a greedy randomized void filling heuristic. Void filling is used to fill the unallocated slots left empty by the scheduling algorithm. In particular, a randomized greedy heuristic greedily computes a set of matchings in order to fill the free slots in an uniform way, taking into account the previously allocated slots and the transmission constraints that they yield.

B. Simulation Results

We initially examine the effect of utilizing i) a parallel packet switched network over which we sent TCP ACK packets and ii) a randomized void filling heuristic to fill the empty slots/ permutations of the schedules on slot (network capacity) utilization over time. As it can be observed in Fig. 3, both the effect of the parallel network and the randomized void filling heuristic is quite significant. Since, TCP features congestion control, the TCP window limits the traffic load the servers transmit. This has a major impact to the overall slot utilization and thus to the throughput and the makespan of the network. These two techniques improve the TCP window pipelining resulting to improved slot utilization and reduced makespan. In particular, we observed a reduction of the makespan for the 4 MapReduce jobs from ~27,4 s in the case of *no parallel/no void filling* to ~27,2 s in the case of *parallel/no void filling* and to ~14 s in the case of *no parallel/void filling*. The combination of *parallel/ void filling* achieves a substantially lower makespan of ~10,3 s. In the following we will assume that the NEPHELE network uses both *parallel/ void filling*.

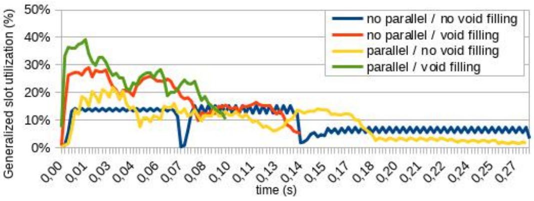


Fig. 3. Impact of the parallel network and randomized void filling heuristic on slot utilization.

We now examine the effect of the control delay C which was varied from 0, 5, 10, 20, 50 to 200 Data periods. As it is shown in Fig. 4, the makespan for the case of the static round-robin policy remains constant at about 0.36 s, regardless of the Control cycle delay. Meanwhile, the other two policies seem to perform better for at most 19%, given that they take into account the traffic (monitoring or application awareness) and carry out scheduling based on $\widehat{\mathbf{Q}}(n)$ estimates. This performance improvement decreases as the Control cycle delay increases, and eventually in the sample of Control cycles equal to 200 Data periods, it gets worse than the static round-robin for at most 13%. This is expected, since the longer control delay results to an increased chance the actual traffic at the queues to substantially differ from the calculated schedule. It can also be observed that in small numbers of Control cycles, utilizing prediction also improves the performance. However, this improvement fades out from 20 Control cycles and on.

In the next scenario, we consider the cases where we have 1, 4, 7 and 10 MapReduce jobs simultaneously running on the cluster. It is expected that as the number of jobs increases, the network load increases, but also the traffic dynamicity decreases, given that the assignment of the worker nodes with the servers is done randomly and uniformly. As shown in Fig. 5, the makespan increases with the job number in all queue matrix estimation policies, since the network load increases. However, especially in the case of 1 job, where only certain parts of the network are utilized in each Mapreduce phase, we can see that the static round-robin policy performs much worse than the other two policies for about 32%. This difference is reduced for larger numbers of jobs to at least 16%.

In the last considered scenario, we have different cluster sizes, namely of 4, 8, 16 and 32 racks (8, 16, 32 and 64 servers, respectively). Fig. 6 shows the performance of the three queue matrix estimation policies. In particular, we can observe that the policies that take into account the traffic have a much better performance than the static round-robin that ranges between 12-48% and increases with the increase of the cluster size.

In our tests we compared NEPHELE with a fat-tree topology. We observed that when Control cycle was set to 0, NEPHELE performed similarly to a fat-tree, in terms of makespan.

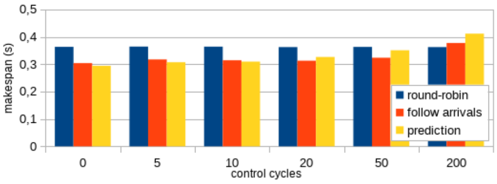


Fig. 4. Impact of control cycle (in Data periods) on makespan.

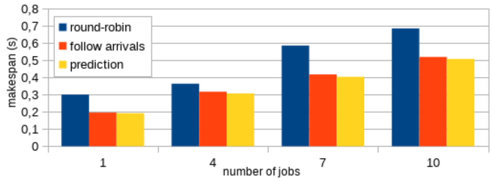


Fig. 5. Impact of the number of MapReduce jobs on makespan.

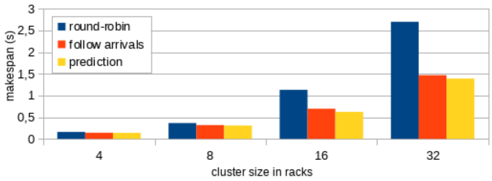


Fig. 6. Impact of cluster size on makespan.

VI. Conclusions

We presented the NEPHELE DCN architecture and described the related resource allocation problem. In NEPHELE, a centralized SDN controller allocates slots and optical planes to communicating pairs, and thus coordinates over time, space and wavelength to avoid collisions and achieve efficient operation. We described the NEPHELE control cycle, including the importance of the policy used to obtain good queue matrix estimates that approximate the traffic pattern after the control cycle delay. We conducted simulations using OMNET++ under MapReduce realistic traffic. We examined the effect of utilizing a parallel network for TCP ACKs, and of a void filling heuristic. We observed that both these techniques, improve the makespan. We considered the case of applying a static round-robin policy and two policies that take into account the traffic. We observed that when the control cycle delay is high, a static round-robin policy seems preferable. The policies that take into account the traffic induce a significant improvement to the total makespan that can reach 48% when the short-term load dynamicity is high.

Acknowledgment

Partially funded by EC through NEPHELE project (g.a. 645212).

References

[1] Cisco Global Cloud Index: Forecast and Methodology, 2016-2021.

[2] Al-Fares, A. Loukissas, A. Vahdat, “A Scalable, Commodity Data Center Network Architecture”, ACM SIGCOMM, 2008

[3] T. Benson, A. Akella, D. Maltz, “Network Traffic Characteristics of Data Centers in the Wild”, ACM SIGCOMM, 2010.

[4] A. Roy, H. Zeng, J. Bagga, G. Porter, A. Snoeren, “Inside the Social Network's (Datacenter) Network”, ACM SIGCOMM, 2015.

[5] J. Follows, D. Straeten, “Application driven networking: Concepts and architecture for policy-based systems”, IBM Corporation, 1999.

[6] X. Zheng, Z. Cai, J. Li, H. Gao, “An application-aware scheduling policy for real-time traffic”, International Conference on Distributed Computing Systems (ICDCS), 2015.

[7] C. Kachris, I. Tomkos, “A Survey on Optical Interconnects for Data Centers”, IEEE Communications Surveys & Tutorials, 14 (4), 2012.

[8] N. Farrington, G. Porter, S. Radhakrishnan, H. H. Bazzaz, V. Subramanya, Y. Fainman, G. Papen, A. Vahdat, “Helios: a hybrid electrical/optical switch architecture for modular data centers”. ACM SIGCOMM, 2010.

[9] G. Wang et al, “c-through: part-time optics in data centers”, ACM SIGCOMM, 2010.

[10] K. Christodoulopoulos, D. Lugones, K. Katrinis, M. Ruffini, D. O’Mahony, “Performance Evaluation of a Hybrid Optical/Electrical Interconnect”, IEEE/OSA Journal of Lightwave Technology, 2015.

[11] Y. Ben-Itzhak, C. Caba, L. Schour, S. Vargaftik, “C-Share: Optical Circuits Sharing for Software-Defined Data-Centers”, arXiv, 2016.

[12] N. McKeown et al, “OpenFlow: Enabling Innovation in Campus Networks”, ACM Computer Communication Review, 2008.

[13] A. Singla et al, “Proteus: a topology malleable data center network,”, ACM SIGCOMM Workshop on Hot Topics in Networks, 2010.

[14] S. Peng, et al, “Multi-Tenant Software-Defined Hybrid Optical Switched Data Centre”, IEEE/OSA Journal of Lightwave Technology, 2015.

[15] N. Calabretta, W. Miao, “Optical Switching in Data Centers: Architectures Based on Optical Packet/Burst Switching”, Optical Switching in Next Generation Data Centers, pp.45-69, Springer, 2017.

[16] G. Porter et al, “Integrating microsecond circuit switching into the data center,” ACM SIGCOMM, 2013.

[17] P. Bakopoulos, et. al. “NEPHELE: an end-to-end scalable and dynamically reconfigurable optical architecture for application-aware SDN cloud datacenters”, IEEE Communications Magazine, 2018.

[18] K. Christodoulopoulos et al, "Efficient bandwidth allocation in the NEPHELE optical/electrical datacenter interconnect", IEEE/OSA Journal of Optical Communications and Networking, 9(12), pp. 1145-1160, 2017.

[19] G. Landi, M. Capitani, K. Christodoulopoulos, D. Gallico, M. Biancani, M. Aziz, “An Application-Aware SDN Controller for Hybrid Optical-Electrical DC Networks”, ICN 2017.

[20] I. Patronas, V. Kitsakis, N. Gkatzios, D. Reisis, K. Christodoulopoulos, E. Varvarigos, "Scheduler Accelerator for TDMA Data Centers", PDP 2018.

Exploring the Potential of VCSEL Technology for Agile and High Capacity Optical Metro Networks

[Invited]

Michela Svaluto Moreolo, Josep M. Fàbrega, Laia Nadal, and F. Javier Vílchez
Optical Networks and Systems Dept., Communication Networks Division,
Centre Tecnològic de Telecomunicacions de Catalunya (CTTC/CERCA)
Castelldefels, Spain
michela.svaluto@cttc.es

Abstract— **In this paper, vertical cavity surface emitting laser (VCSEL) technology is presented as potential prominent performer to address key challenges and novel functionalities of future optical metro networks. The adoption of VCSEL-based modules is particularly attractive for the implementation of programmable (SDN-enabled) transceiver architectures, targeting a radical reduction of cost, power consumption and footprint. Different flavours of these architectures are presented to enable agile, scalable and high capacity optical metro networks. Furthermore, advanced functionalities and programmability aspects are analysed to explore the potential of adopting solutions based on this technology.**

Keywords— vertical cavity surface emitting laser (VCSEL); optical metro networks; transceiver architecture; sliceable bandwidth variable transceiver (S-BVT); defragmentation.

I. Introduction

The future optical metro network scenario is evolving towards a highly-dynamic paradigm where multiple, new and bandwidth-hungry 5G services should be supported, minimizing the cost and power consumption. As an example, it is envisioned that content delivery networks (CDNs) will carry 71% of Internet traffic by 2021 (up from 52% in 2016) and that 35% of end-user Internet traffic will be delivered within a metro network by 2021 (22% in 2016) [1]. Actually, CDNs will carry traffic closer to the end user and, although at present much CDN traffic is onto regional core networks, the metro-delivered traffic is growing faster than core-delivered traffic according to the predicted percentage increase [1]. Thus, future optical metro networks should be able to accommodate the envisioned increase of speed and volume of traffic, as well as low-rate service connections, while managing adaptive bit rates and peak of traffic according to the available bandwidth.

In this context, programmable adaptive transceivers supporting great capacity and dynamicity arise as key enabler to address these challenges, such as offering ultrabroadband 5G services and managing high peak-average traffic ratio. Particularly, bitrate and bandwidth variable transceivers (BVTs) allow software defined optical transmission adaptive to different network conditions and requirements. Sliceable BVTs (S-BVTs), supporting multi-flow transmission, are able to give service to a higher number of sites and to provide greater capacity per site, resulting a crucial element for supporting any type of traffic within an evolutionary metro network scenario. Novel programmable – to be integrated in a software defined network (SDN) based control plane - transceiver architectures, transparent to service and with a wide range of granularities, are required to facilitate metro network operation, supporting increased bandwidth demand and network dynamicity. A modular design is also desirable for targeting scalable architectures allowing to grow-as-needed. Furthermore, a cost-effective implementation should be targeted to be suitably adapted to the metro segment needs.

Cost-effective data plane solutions for flexible metro networks using S-BVTs have been proposed in [2]. They are based on orthogonal frequency division multiplexing (OFDM) technology. Actually, multicarrier modulation (MCM), as OFDM or discrete multitone (DMT), enables dynamic and flexible adaptation to traffic/channel conditions with fine granularity, when bit and power loading (BL/PL) algorithms are implemented at the digital signal processing (DSP) [3]. Alternative architectures for metro/regional scenario have been assessed and analyzed also from a techno-economic point of view in [4]. It has been shown that multiple low bit rate flexgrid connections can be supported over regional optical paths, adopting centralized S-BVTs shared among multiple end-points with cost-effective transceiver schemes. In [5] transparent/dynamic delivery of mobile front-/back-haul in a converged optical metro architecture has been experimentally demonstrated, employing SDN-enabled S-BVTs based on adaptive MCM. The proposed architecture is specifically tailored for a transparent services delivery across the access and metro network segments. Network testbed

This work has been partially supported by the EU-H2020 PASSION Project (780326) and the Spanish DESTELLO project (TEC2015-69256-R).
978-3-903176-07-2 copyright 2018 IFIP

experiments show successful connectivity at distances up to 175 km and capacities beyond 30 Gb/s per flow.

In line with these previous works, we propose to leverage the concept of S-BVT to design and implement modular transceiver architectures able to support the evolutionary scenario of future metro networks, considering a centralized S-BVT of high-capacity (e.g. able to support high peak rate) shared among distributed (S)-BVTs of lower capacity. More precisely, S-BVTs would be located at the metro/aggregation nodes, either metro/regional or metro/access (exchange) nodes, suitably sizing the number of modules and the slice-ability according to the need; while simpler and even more cost-effective BVT architectures are located closer to the end-user.

In this framework, energy- and cost-efficient photonic technologies should be explored to design and implement novel modular S-BVT architectures, to meet the requirements of the metro segment, while providing high capacity [6]. Particularly, in order to reduce the transmitter cost, direct modulation of the laser source can be considered instead of using external modulation. The adoption of vertical cavity surface emitting diode (VCSEL) technology could be considered as an attractive option to provide energy-efficient, low-cost solutions with small footprint. In fact, VCSEL manufacturing cost is substantially lower than other common options (e.g. DFB lasers), and the transceiver devices can be integrated in the same photonic platform. Thus, this technology is worth to be explored for the design and implementation of S-BVTs especially tailored for the metro segment [7].

VCSEL technology is usually considered for short-reach and low-data-rate applications at 850 nm. Highest bitrates have been achieved adopting 4-PAM, carrierless-amplitude-phase (CAP) modulation and DMT modulation [8]. Coherent detection has been proposed to flexibly extend the achievable reach, showing the potential of VCSELs for 100G and beyond metro network applications at 1550 nm and using polarization division multiplexing [9]. DMT, as OFDM, enables spectral manipulation with very fine granularity (subwavelength level), and is considered a promising candidate to support future 1T-class transceivers [10].

Actually, MCM with BL/PL enables spectral manipulation, with spectrum granularity even finer than elastic optical networks (EONs) (12.5GHz) such as ultra-dense wavelength switching networks [11]. Flexible transport, with subwavelength granularity and variable capacity per channel, enables optimal resource usage and advanced functionalities [3]. Particularly, this allows a soft migration of fixed grid networks towards a flexible architecture and also enables a subcarrier-based defragmentation of the spectrum, without requiring a network re-optimization [3].

In this paper, VCSEL technology is explored to design the fundamental building block and/or specific module(s) of the (S)-BVT, as well as to implement advanced functionalities, which can be particularly suitable for future optical metro networks.

The paper is organized as follows. In Sec. II, different flavors of programmable S-BVT architectures adopting VCSEL-based modules are presented to enable agile, scalable, cost-effective and high capacity optical metro network. Different types of VCSEL are introduced and their capabilities presented/discussed, reporting recent results. Sec. III deals with advanced functionalities enabled by programmable transceivers equipped with VCSELs; Sec. III A is particularly devoted to the capability of mitigating the spectrum fragmentation. Conclusions are drawn in Sec IV, where future work on this topic is also discussed.

II. Programmable Transceiver Architectures Adopting VCSEL Technology

Different architectures and modulation formats can be used to implement S-BVTs and programmable transceivers [12]. Coherent transceivers combined with suitable DSP and advanced complex modulation formats provide high capacity/reach and

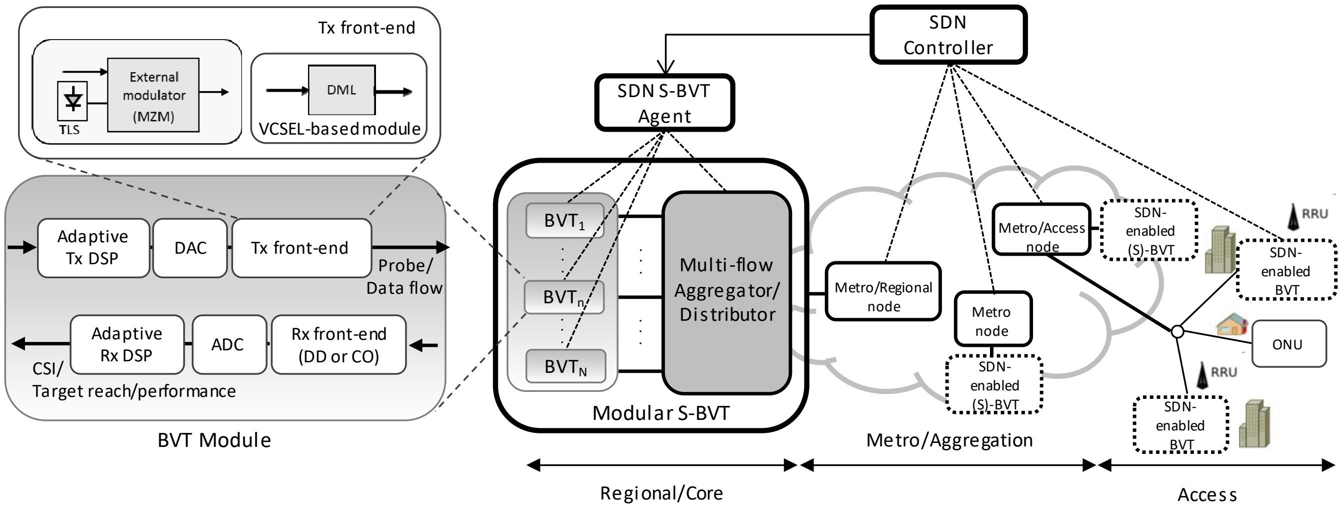


Fig. 1. Schematic architecture of an SDN-enabled modular S-BVT to be suitably sized according to the network/node needs and related scenario.

flexibility at the expense of cost and energy-efficiency, which can result not appropriate/affordable for the metro segment. Alternative approaches more suitable and specifically tailored for metro networks combine simple intensity or amplitude modulation at the transmitter, while adopting direct detection (DD) or keeping coherent (CO) detection at the receiver [3, 4, 6].

As mentioned in Sec. I, a VCSEL-based transmitter scheme can be combined with either a simple low-cost detection scheme based on DD [8, 13] or CO detection to achieve ultimate performance in terms of reach and capacity [9].

To address very high capacity link target (e.g. up to 100 Tb/s), both spectrum and space dimensions should be considered [14-15].

Figure 1 shows a schematic architecture relying on a modular approach to allow suitably sizing the S-BVT according to the network needs or to the specific node, where the (S)-BVT is located. The programmable (SDN-enabled) S-BVT consists of an array of *N* BVT modules and an aggregator/distributor of the multiple flows (slices). This aggregation/distribution element can be implemented with a bandwidth-variable spectrum selective switch (SSS). The multiple slices are transmitted over the network as a single high-capacity flow or can be split into lower capacity flows routed towards independent paths to reach different destination nodes. Alternative choices for the transceiver optoelectronic front-ends are indicated in Fig. 1. Direct modulation of the laser source (DML), using a VCSEL-based module, at the BVT transmitter (Tx), is proposed instead of adopting external modulation (e.g. using a Mach Zehnder modulator, MZM) with a tunable laser source (TLS). Tunable VCSEL modules can be also envisioned. The adaptive DSP of the BVT array allows implementing DMT or OFDM, including margin adaptive (MA) or rate adaptive (RA) BL/PL algorithms, to suitably adapt the transmission at a target capacity over a certain reach, according to the channel state information (CSI) and the variable rate request ensuring a target performance. The CSI, and thus the information related to the signal to noise ratio (SNR) corresponding to each DMT/OFDM subcarrier, is retrieved at the receiver, transmitting a probe signal with uniform loading (e.g. adopting 4-QAM format for each subcarrier).

Recently, VCSELs potential has been shown at 1550 nm with advanced modulation formats for passive optical networks (PONs) and metro/access EONs [16-17]. Particularly, in [17], transparent service delivery at variable data rates has been demonstrated in an SDN converged elastic metro/access optical network with cost-effective programmable transceiver based on VCSEL technology. In that experiment, a directly modulated VCSEL of 4.5 GHz bandwidth working at λ=1539.61 nm was used for implementing the transmitter module. Up to 16 Gb/s with 9 GHz maximum spectral occupation was achieved, adopting DMT with BL/PL and DD, considering a minimum power budget of 20 dB in the access segment. A target bit error ratio (BER) of $4.62 \cdot 10^{-3}$, corresponding to a hard decision forward error correction (HD-FEC) of 7% overhead, has been taken into account. The performance has been assessed in the ADRENALINE testbed network, shown in Fig. 2. Specifically, it consists of four nodes - two OXCs (optical cross-connects) and two ROADMs (reconfigurable optical add-drop multiplexers) - and five amplified links ranging from 35 km to 150 km [18]. The 35 km link between OXC-2 and ROADM-1 is flexgrid as the nodes are equipped with programmable SSS modules. A maximum transmission distance of 200 km has been successfully covered, considering 50 km PON tree and a 150 km single hop path of the ADRENALINE network. The achieved rate was about 8 Gb/s (half of the maximum rate) at the target BER.

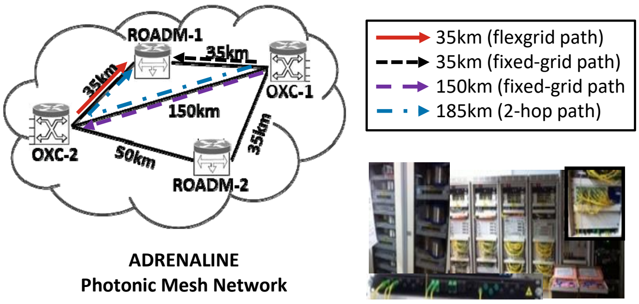


Fig. 2. ADRENALINE network paths and testbed picture.

Adopting larger bandwidth VCSELs and high-performance digital-to-analog and analog-to-digital converters (DAC/ADC), higher capacities can be achieved. A single VCSEL with 3-dB bandwidth of 15 GHz, directly modulated with DMT and using DD, has been demonstrated for short-reach applications at 100Gb/s [8]. Up to 95 Gb/s (net data rate of 79.2 Gb/s) has been achieved over 4 km of standard single mode fiber (SSMF). If CO detection is considered, the achievable reach can be further extended to target the metro/regional segment [15]. An S-BVT architecture based on this approach can be implemented integrating a bank of VCSELs at different operating wavelengths. The multiple modules can be suitably enabled/disabled for wavelength selection and bandwidth-variable adaptation.

Widely tunable micro-electro-mechanical system (MEMS) VCSELs allow adapting the operating wavelength over a spectrum range of more than 60 nm [19]. This type of VCSELs, with 3-dB bandwidth of 7 GHz, was successfully demonstrated for DMT transmission at 26 Gb/s over 40 km of SSMF, targeting converged WDM-PON applications [16].

This VCSEL technology can be further explored for implementing the building block or specific modules of an S-BVT able to cover extended reach, targeting metro networks [7]. To this extent, single sideband (SSB) OFDM modulation is required to make the direct modulated signal more robust against chromatic dispersion. Thus, a bandwidth variable optical filter, which can be implemented at the S-BVT aggregator stage, should be used. If the network nodes are equipped with suitable SSS, the optical filtering could be performed at any node, not only at the sliceable transceiver.

The capacity performance of a BVT module, adopting a widely tunable MEMS VCSEL, has been assessed in [7]. SSB-OFDM is compared to pure DMT transmission over different paths of the ADRENALINE network. The BVT capacity in back-to-back (B2B) with optimized BL/PL assignment and 10.5 GHz bandwidth occupancy is above 30Gb/s. Bit rates greater than 20 Gb/s up to a 2-hop path of 185km are supported by adopting OFDM with SSB modulation, due to its higher robustness against chromatic dispersion. In addition, in this case the spectral efficiency is doubled, since only half of the optical bandwidth occupancy is required with respect to the DMT signal. Particularly, the spectral occupancy in terms of assigned flexgrid frequency slots of 12.5 GHz for the case of DMT is two (25 GHz), while for the SSB-OFDM slice/flow is only one (12.5 GHz).

Table I summarizes the results reported in this section for the different types of VCSEL-based transmitter adopting MCM (either DMT or OFDM) and DD.

III. Advanced Functionalities of VCSEL-based Transceivers

Programmable S-BVTs, relying on an SDN-based control plane paradigm, provide advanced functionalities according to the specific architecture, modules, devices and/or technology adopted for the target design/application.

In this section, we explore the functionalities that can be enabled if the S-BVT is equipped with VCSEL-based module(s). Particularly, we would like to devote special attention to the capability of filling spectral gaps provided by S-BVTs equipped with tunable VCSEL at the transmitter. This feature is particularly relevant to show their potential ability of mitigating the spectrum fragmentation, without requiring a network re-optimization.

Actually, the BVT array of the S-BVT (see Fig. 1) can be composed of subsets of BVT modules, which can also be based on different technologies. This enables a scalable pay-as-you-grow model based on pluggable units and photonic integration. As shown in Fig. 1, the BVTx front-end can be based on either a DM VCSEL or an external modulator driven by a TLS. Each slice generated by the sub-modules of an S-BVT can potentially support a maximum data rate and bandwidth, which can be varied and adapted according to the network channel, target reach and traffic demand. Similarly, by tuning and/or enabling/disabling the laser source(s) as well as the individual subcarrier loading, the optical carrier of each flow can be adapted to the available channel, while suitably allocating and/or squeezing the bandwidth for optimal spectral usage [3]. These features enable rate/distance adaptive functionalities with unique granularity (even finer than EON) and grid adaptation. A soft migration of fixed grid networks towards a flexgrid paradigm is possible, thanks to the use of S-BVTs, enhancing the network flexibility even in fixed-grid networks as demonstrated in [3].

Another important functionality is the slice-ability: multiple adaptive flows are generated/received and aggregated/distributed at the S-BVT. The aggregated flow can be sliced into data flows with less capacity, concurrently serving multiple destination nodes at variable rate over different paths. This functionality also allows enabling inverse multiplexing, as the traffic demand can be split into multiple flows routed via multiple independent paths to the same end-node [3].

A. *Spectrum Fragmentation Mitigation Capability*

The reconfigurability of multiple parameters by software defined networking allows the SDN-enabled S-BVT adapting the transmission to the dynamic variation of optical metro networks.

In particular, the suitable configuration of DSP modules, optoelectronic frontends and the aggregator/distributor, also make the S-BVT capable of coping with spectrum fragmentation, as shown in [3]. On one hand, the wavelength tunability and the fine granularity of MCM enable a subcarrier-based spectrum defragmentation; on the other hand, as one or more flow(s) can

TABLE I. Performance of direct modulated VCSEL-based transceivers adopting MCM and DD.

MCM	VCSEL-based Tx	Operating λ	3-dB Bandwidth	Performance			Target Application	Reference
				Capacity	*Link*	*FEC limit*		
DMT	Large-bandwidth VCSEL Tx	1550 nm	15 GHz	95 Gb/s	4 km	$1.5 \cdot 10^{-2}$	Short-reach	[8]
DMT	MEMS-VCSEL Tx	Tunable[a]	7 GHz	26 Gb/s	40 km	$2.26 \cdot 10^{-3}$	Converged WDM-PONs	[16]
DMT	VCSEL-based BVT	1539.61 nm	4.5 GHz	8 Gb/s	200 km[b]	$4.62 \cdot 10^{-3}$	Metro/Access	[17]
DMT	MEMS-VCSEL (S)-BVT	Tunable[c]	7 GHz	12 Gb/s	185 km[d]	$4.62 \cdot 10^{-3}$	Metro Networks	[7]
SSB-OFDM	MEMS-VCSEL (S)-BVT	Tunable[c]	7 GHz	20 Gb/s	185 km[d]	$4.62 \cdot 10^{-3}$	Metro Networks	[7]

[a] Tuning range of 30 nm.

[b] 50 km SSMF (PON tree) and 150 km single-hop path of ADRENALINE.

[c] Configured to operate at 1550.12 nm.

[d] 2-hop ADRENALINE path: OXC1-OXC2-ROADM1.

be directed towards the same end-node through different paths, an optimal spectral/resource usage can be performed without requiring a network re-optimization.

When the adaptive DSP is combined with cost-effective and energy-efficient VCSEL-based modules, the programmable transceiver allows suitably filling spectral gaps, thanks to its tunability and adaptive narrow bandwidth. Thus, this technology option results attractive to be explored for mitigating the spectrum fragmentation. Particularly, the adoption of widely tunable MEMS VCSEL configured on-demand by the SDN controller, by means of an SDN S-BVT agent, facilitates this functionality.

In [7], this has been demonstrated for optical metro networks, with fine spectrum granularity. The analysis has been conducted, considering an S-BVT flow/slice generated by a VCSEL-based module in presence of adjacent slices generated by S-BVT modules based on external modulators. In fact, we assume that the S-BVT is composed of subsets of BVT modules as pluggable units, based on either direct or external modulation, equipped with widely-tunable VCSEL or TLS, respectively.

Negligible bitrate penalty has been found for a guard-band of 12.5 GHz between the transmitted channels. In addition, even for lower value of the guard-band, a capacity increase has been evidenced, when the VCSEL module is correctly enabled by the SDN controller. With a guard-band of 6.25 GHz, a maximum capacity penalty of less than 10% is experienced by the VCSEL flow, with respect to the capacity obtained when the adjacent slices are disabled, and less than 30% penalty is found without any guard-band. Even lower penalties have been measured for the (larger bandwidth) adjacent channels. It has been also assessed that this solution can be used in both flexible and fixed-grid scenarios (soft-migration), even if the latter limits the performance of the S-BVT advanced features [7].

Thus, the proposed S-BVT architecture equipped with such module(s) represents a promising candidate for a hitless spectrum defragmentation, filling spectral gaps with fine granularity and without requiring any network re-optimization.

IV. Conclusions and Future Work

VCSEL technology has been presented as an attractive option to design and implement programmable S-BVT for future agile and high capacity optical metro networks. Starting from previously proposed alternative S-BVT solutions relying on external modulation, the adoption of direct modulated VCSEL-based modules has been explored to radically reduce cost, power consumption and footprint. The presented architectures aim to address the challenges of an evolutionary metro network scenario, where a centralized S-BVT of high-capacity (e.g. able to support high peak rate), able to serve multiple endpoints, is shared among distributed (S)-BVTs of lower capacity closer to the end-user.

Recent results have been reported to show the potential of VCSEL technology at 1550nm to target metro network distance/capacity. The modularity is key to promote a grow-as-needed paradigm and correctly size the S-BVT to the node/network need, as well as to enable advanced functionalities. Particularly relevant for agile metro network is the ability of VCSEL-based S-BVT modules of filling spectral gaps with fine granularity and coping with the spectrum fragmentation without requiring a network re-optimization.

The proposed architectures based on VCSEL technology, especially if combined with dense photonic integration, represent a very promising solution to achieve ultimate performance in terms of cost/power-efficiency and small footprint. Future work envisions the design and assessment of scalable and modular S-BVT architectures able to adaptively generate multiple flows with enhanced capacity to support the high traffic demand of metro networks. The use of large bandwidth VCSELs, to be densely integrated on a same photonic platform, and the adoption of CO reception will be explored to enhance the S-BVT performance while extending the achievable reach [15]. The programmability and advanced features provided by this approach will be also extensively investigated, towards the integration of these S-BVT architectures in an SDN-based control plane to continue exploring the potential of VCSEL technology.

Acknowledgment

The authors would like to thank Christian Neumeyr (Vertilas GmbH), Alberto Gatto, Paola Parolari and Pierpaolo Boffi (Politecnico di Milano), as well as Juan-Pedro Fernández-Palacios (Telefónica I+D) for the fruitful collaboration and discussions that inspire this paper.

References

[1] CISCO White Paper: “The Zettabyte Era: Trends and Analysis,” June 2017.

[2] M. Svaluto Moreolo, J. M. Fabrega, L. Nadal, F. J. Vilchez, V. López, J. Pedro Fernández-Palacios, “Cost-Effective Data Plane Solutions Based on OFDM Technology for Flexi-Grid Metro Networks Using Sliceable Bandwidth Variable Transponders,” Proc. ONDM 2014, Stockholm (Sweden), 19-22 May 2014.

[3] M. Svaluto Moreolo et al., "SDN-Enabled Sliceable BVT Based on Multicarrier Technology for Multiflow Rate/Distance and Grid Adaptation," J. Lightwave Technol., vol. 34, no. 6, pp. 1516-1522, March 2016.

[4] M. Svaluto Moreolo, J. M. Fabrega, L. Martin, K. Christodoulopoulos, E. Varvarigos, J. Pedro Fernández-Palacios, “Flexgrid Technologies Enabling BRAS Centralization in MANs,” IEEE/OSA J. Opt. Commun. Netw., vol. 8, no. 7, pp. A64-A75, July 2016.

[5] J. M. Fabrega, M. Svaluto Moreolo, L. Nadal, F. J. Vílchez, R. Casellas, R. Vilalta, R. Martínez, R. Muñoz, J. P. Fernández-Palacios, L. M. Contreras “Experimental Validation of a Converged Metro Architecture for Transparent Mobile Front-/Back-Haul Traffic Delivery using SDN-enabled Sliceable Bitrate Variable Transceivers” IEEE/OSA J. Lightwave Technol., vol. 36, no.7, pp. 1429-1434, Apr. 2018.

[6] M. Svaluto Moreolo, J. M. Fabrega, and L. Nadal, "S-BVT for next-generation optical metro networks: benefits, design and key enabling technologies," in Proc. SPIE 10129, San Francisco, CA (USA), Jan. 2017.

[7] M. Svaluto Moreolo, et al., "Modular SDN-enabled S-BVT Adopting Widely Tunable MEMS VCSEL for Flexible/Elastic Optical Metro Networks," Proc. OFC 2018, S. Diego, CA (USA), March 2018.

[8] C. Xie, P. Dong, S. Randel, D. Pilori, P. Winzer, S. Spiga, B. Kögel, C. Neumeyr, and M.-C. Amann, "Single VCSEL 100-Gb/s short reach system using discrete multi-tone modulation and direct detection," Proc. OFC 2015, paper Tu2H.2, 2015.

[9] C. Xie, S. Spiga, P. Dong, P. Winzer, M. Bergmann, B. Kögel, C. Neumeyr, and M.-C. Amann, "Generation and Transmission of 100-Gb/s PDM 4-PAM Using Directly Modulated VCSELs and Coherent Detection," Proc. OFC'2014, PDP Th5C.9, 2014.

[10] H. Isono, "Recent standardization activities for client and networking optical transceivers and its future directions," Proc. SPIE 10131, Next-Generation Optical Networks for Data Centers and Short-Reach Links IV, 101310G, Jan. 2017.

[11] X. Zhou, W. Jia, Y. Ma, N. Deng, G. Shen, A. Lord, "An Ultradense Wavelength Switched Network," IEEE/OSA J. Lightwave Technol., vol. 35, no. 11, pp. 2063–2069, 2017.

[12] N. Sambo, et al., "Next generation sliceable bandwidth variable transponder," IEEE Communications Magazine, vol. 53, no. 2, pp. 163-171, Feb. 2015.

[13] A. Gatto, D. Argenio, P. Boffi "Very high-capacity short-reach VCSEL systems exploiting multicarrier intensity modulation and direct detection," Optics Express, vol. 24, p. 12769-12775 (2016).

[14] M. Svaluto Moreolo, et al., "Towards Advanced High Capacity and Highly Scalable Software Defined Optical Transmission," Proc. ICTON 2017, Girona (Spain), July 2017.

[15] www.passion-project.eu.

[16] C. Wagner et al., "26-Gb/s DMT Transmission Using Full C-Band Tunable VCSEL for Converged PONs," Photonic technology Lett., vol. 29, no. 17, pp. 1475 – 1478, 2017.

[17] L. Nadal, et al., "Transparent Service Delivery in Elastic Metro/Access Networks with Cost-Effective Bandwidth Variable Transceivers," Proc. ICTON 2017, Girona (Spain), July 2017.

[18] R. Muñoz, L. Nadal, R. Casellas, M. Svaluto Moreolo, R. Vilalta, J. M. Fabrega, R. Martínez, A. Mayoral, F. J. Vilchez, "The ADRENALINE Testbed: An SDN/NFV Packet/Optical Transport Network and Edge/Core Cloud Platform for End-to-End 5G and IoT Services," in Proc. EuCNC 2017, Jun. 2017

[19] S. Pau et al., "10-Gb/s direct modulation of widely tunable 1550-nm VCSEL," J. Selected Topics Quantum Electron., vol. 21, no. 6, p. 1700908, Nov/Dec. 2015.

Monitoring and Data Analytics: Analyzing the Optical Spectrum for Soft-Failure Detection and Identification [Invited]

B. Shariati, A. P. Vela, M. Ruiz, and L. Velasco*

Optical Communications Group (GCO).
Universitat Politècnica de Catalunya (UPC), Barcelona, Spain
*lvelasco@ac.upc.edu

***Abstract*—Failure detection is essential in optical networks as a result of the huge amount of traffic that optical connections support. Additionally, the cause of failure needs to be identified so failed resources can be excluded from the computation of restoration paths. In the case of soft-failures, their prompt detection, identification, and localization make that recovery can be triggered before excessive errors in optical connections translate into errors on the supported services or even become disrupted. Therefore, Monitoring and Data Analytics (MDA) become of paramount importance in the case of soft-failures. In this paper, we review a MDA architecture that reduces remarkably detection and identification times, while facilitating failure localization. In addition, we rely on Optical Spectrum Analyzers (OSA) deployed in the optical nodes as monitoring devices acquiring the optical spectrum of outgoing links. Analyzing the optical spectrum of optical connections, specific soft-failures that affect the shape of the spectrum can be detected. A workflow consisting of machine learning algorithms, designed to be integrated in the aforementioned MDA architecture, will be studied to analyze the optical spectrum of a given optical connection acquired in a node and to determine whether a filter failure is affecting it, and in such case, what is the type of filter failure and its magnitude. Exhaustive results are presented allowing to evaluate the proposed method.**

***Keywords*—Failure Detection and Identification, Failure Magnitude Estimation, Elastic Optical Networks.**

I. Introduction

Hard failures detection at the optical layer, e.g. a fiber cut, can be easily detected, e.g. by the end transponders of optical connections. Even though the proper identification of the failed element is not an easy task, e.g. a failure in an intermediate optical amplifier, determining that the failure is in a topological element (node or link) is enough for excluding such element when restoration routes are computed [1].

However, the scenario is more difficult when we face soft-failures, such as laser drift and filter problems in the optical layer, whose presence is indirectly revealed, e.g., by observing bit error rate (BER) variations [2]. Such BER degradation, although not very high at first, could evolve toward high values and even cause disconnections. This is the very reason behind continuously monitoring the network, so such degradations can be anticipated. Nonetheless, it is not enough to monitor the BER evolution in the end transponders of optical connections, but also in intermediate nodes, so localization algorithms can determine the failed resource thus, facilitating proactive restoration strategies. In our previous work in [3], we used Optical Spectrum Analyzers (OSA) to analyze the shape of the optical spectrum of a signal to determine whether the signal was affected by a soft-failure. By installing OSAs in every optical node, such soft-failures can be easily identified, i.e., the cause of the failure is identified, and localized, being thus their output, the perfect input for restoration algorithms.

Machine learning (ML) algorithms can help in the process of detecting and identifying soft-failures. However, those algorithms should be placed closed to the network nodes aiming at reducing the amount of monitoring data to be conveyed from *Observation Points* (OP), as well as increasing the frequency of measurements, so as to reduce detection times [4]. For this very reason, the authors in [5]-[7] proposed a distributed Monitoring and Data Analytics (MDA) architecture, where data analytics capabilities are placed closed to the network nodes. OPs are configured from a centralized system to perform measurements that are immediately exported to the local MDA system, where ML algorithms are in charge of aggregating and analyzing the received data and, in case of detecting any anomaly or degradation, send a notification to the central system in charge of localizing its cause.

In this paper, we assume such distributed MDA architecture and study different ML-based methods for filter failure detection and identification. The rest of the paper is organized as follows. Section II is devoted to introduce the basic concepts about the optical spectrum and the features that allows its shape analysis. Filtering failures that might change its shape are also introduced. Next, useful ML approaches are briefly introduced and then, the distributed MDA architecture considered in this paper is summarized, so as to clearly identify where the spectrum analytics for failure detection and identification should be placed. Section III focuses on the proposed method studied in this paper; the method is based on the combined application of a classifier and failure magnitude estimators. Section IV presents representative results from realistic scenarios, where the performance of the proposed method is evaluated. Finally, Section V concludes the paper.

This work was partially supported by the EC through the METRO-HAUL (G.A. nº 761727) project, from the AEI/FEDER TWINS project (TEC2017-90097-R) and from the Catalan Institution for Research and Advanced Studies (ICREA).

978-3-903176-07-2 © 2018 IFIP

II. Optical Spectrum Analysis and MDA Architecture

In this section, we first introduce the features that are used to analyze the optical spectrum of a given signal and that support detecting and identifying filter failures. Next, useful ML algorithms for failure detection and identification are briefly introduced. Finally, the MDA architecture supporting data analytics distribution is presented highlighting the placement of the module for filter failure detection and identification.

A. *Optical Spectrum and Filter Failures*

Fig. 1 shows an example of the optical spectrum of a 100Gb/s DP-QPSK modulated signal. By inspection, we can observe that a signal is properly configured when: *i*) its central frequency is around the center of the allocated frequency slot; *ii*) its spectrum is symmetrical with respect to its central frequency; and *iii*) the effect of filter cascading is limited to a value given by the number of filters that the signal has traversed. However, when a filter failure occurs, the spectrum is distorted, and the distortion can fall into two categories: *i*) the optical spectrum is asymmetrical as a result of one or more filters are misaligned with respect to the central frequency of the slot allocated for the signal (filter shift, *FS*) and *ii*) the edges of the optical spectrum look excessively rounded compared to the expected considering the number of filters; it is a consequence of the filter's bandwidth being narrower than the slot width allocated for the signal (filter tightening, *FT*).

In order to detect the above distortions, an optical signal (which formally consists of an ordered list of frequency-power (<*f*, *p*>) pairs) can be processed to compute a number of relevant signal points that facilitate its diagnosis. Before processing an optical spectrum acquired by an OSA, the spectrum is equalized by setting its maximum power to 0 dBm. Next, a number of signal features are computed as follows:

- equalized noise level, denoted as *sig* (e.g., -60dB + equalization level),
- edges of the signal computed using the derivative of the power with respect to the frequency, denoted as ∂,
- the mean (μ) and the standard deviation (σ) of the central part of the signal computed using the edges from the derivative (*fc*_∂±Δ*f*),
- a family of power levels computed with respect to μ minus $k\sigma$, denoted as $k\sigma$,
- a family of power levels computed with respect to μ minus a number of dB, denoted as *dB*.

Using these power levels, a couple of cut-off points can be generated and denoted as $f_1(\cdot)$ and $f_2(\cdot)$ (e.g., f_{1sig}, $f_{1\partial}$, f_{1dB}, $f_{1k\sigma}$). Besides, the assigned frequency slot is denoted as f_{1slot}, f_{2slot}. Combining the above, other features are computed as linear combinations of the relevant point focus on characterizing a given optical signal; they include:

- *bandwidth*, computed as $bw_{(\cdot)}=f2_{(\cdot)}-f1_{(\cdot)}$,
- *central frequency*, computed as $fc_{(\cdot)}=f1_{(\cdot)}+0.5*bw_{(\cdot)}$,
- *symmetry* with respect to a reference (frequency slot or derivatives), computed as $sym_{(\cdot)\text{-ref}}=(f1_{(\cdot)}-f1_{\text{ref}})-(f2_{\text{ref}}-f2_{(\cdot)})$.

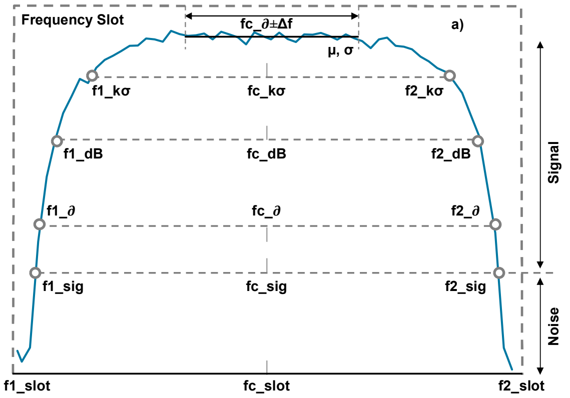


Fig. 1. Example of optical spectrum and signal features.

These features are used as input for the subsequent failure detection and identification modules.

Although relevant points have been computed from an equalized signal, note that signal distortion due to filter cascading effect has not been corrected yet. As abovementioned, this effect might induce to a wrong diagnosis of a filter problem for a normal signal. In order to overcome this drawback, we apply a *correction mask* to the measured relevant points to correct such distortions. Correction masks can be easily obtained by means of the theoretical signal filtering effects or experimental measurements taken for a distinct number of cascaded filters. Every time a diagnosis is started, the specific correction mask considering the actual number of cascading filters that the signal traverses is used to correct the relevant points.

These two different filter failures are illustrated in Fig. 2, where the solid line represents the optical spectrum of the normal signal expected at the measurement point and the solid area represents the optical spectrum of the signal with failure. Note that the expected signal is the signal used for the correction mask. In the case of filter shift, a 10 GHz shift to the right was applied (Fig. 2a), whereas the signal is affected by a 20GHz FT (Fig. 2b).

B. *ML algorithms for failure detection and identification*

Generally speaking, the term machine learning (ML) denotes a computer science field grouping algorithms for data analysis able to learn and make predictions from data [8]. In supervised learning, the ML algorithm is first trained with labeled data to learn a general rule that maps inputs to outputs.

Two useful ML algorithms for failure detection and identification are *classification* and *regression*. In

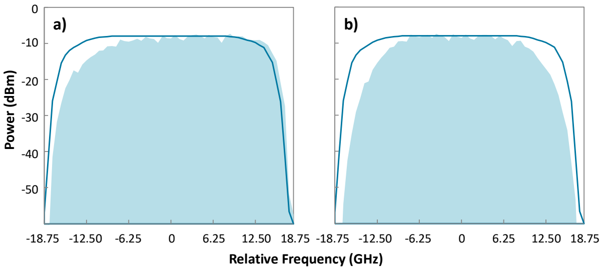


Fig. 2. Example of filter failures considered in this paper: FS (a) and FT (b).

classification, the objective is to classify unknown received data, e.g., an optical signal, and decide whether the signal belongs to the normal class, the FS class, or the FT class. In regression, the objective is to predict a behavior; e.g., regression can be used to estimate the magnitude of a failure.

Although several ML algorithms are suitable for the same output, choosing the best one is a problem-dependent decision where their performance needs to be studied for the specific case. Regarding classification, different ML methods are available in the literature, e.g., decision trees (DT) and support vector machines (SVM). A DT is a hierarchical tree structure that models the relationships between the features and the potential outcomes. DTs use a structure of branching decisions and leaves that represent the different class labels. An SVM is a binary classification technique; in the training phase, the input data is separated into groups of similar features by the computation of a boundary, called *hyperplane*, that better separates the two considered classes. As for prediction, one of the most popular algorithms is *linear regression*, which uses observations to find the best polynomial fitting for predictions.

C. Monitoring and Data Analytics Architecture

Let us now present the distributed MDA architecture considered in this paper, which consists of two components: the *MDA agent* and the *controller* (Fig. 3). The architecture is based on UPC's MDA platform named CASTOR [5], [6].

The MDA agent is directly connected to one or more local network nodes through an interface for configuration and another for monitoring and telemetry. The agent includes a local Knowledge Discovery from Data (KDD) module to enable local data analysis thus, reducing anomaly or failure detection times [4]; to this end, the KDD module contains KDD applications in charge of handling and processing data records. A KDD manager is the entrance point for KDD applications; it receives monitoring data records and delivers them to the corresponding KDD application. In addition, the MDA agent includes a local configuration module that enables local control loops implementation, i.e., applying local node re-configuration/re-tuning based on the results of the data analysis.

The MDA controller is a centralized system that collects monitoring data and notifications from the MDA agents and connects to the SDN controller to keep a synchronized a local copy of operational databases, e.g., topology and connections, as well as to keep it informed about any event detected in the network. The MDA controller exposes an IPFIX interface to the MDA agents so as to collect monitoring data records and notifications; received data is stored into a collected repository based on a scalable multi-master database. A process manager module is notified, and the corresponding KDD process is executed. Additionally, the MDA controller manages the configuration of the MDA agents, including KDD applications and OPs.

The role of the MDA agent is many-fold; apart from OP management, monitoring data received from active OPs can be aggregated before being sent toward the MDA controller. On the other hand, KDD applications continuously analyze measurements collected from OPs and send notifications toward the MDA controller in case of detecting anomalies, degradations or failures. Therefore, the MDA agent is the perfect place where the algorithm for filter failure detection and identification can be placed. In such case, the optical spectrum acquired by the local OSA(s) are periodically received through the IPFIX interface of the MDA agent. The algorithm for filter failure detection and identification receives the spectrum for every particular signal in the optical band and, in the case of detecting a failure, its class together with its estimated magnitude is reported to a hypothetical failure localization algorithm located in the MDA controller.

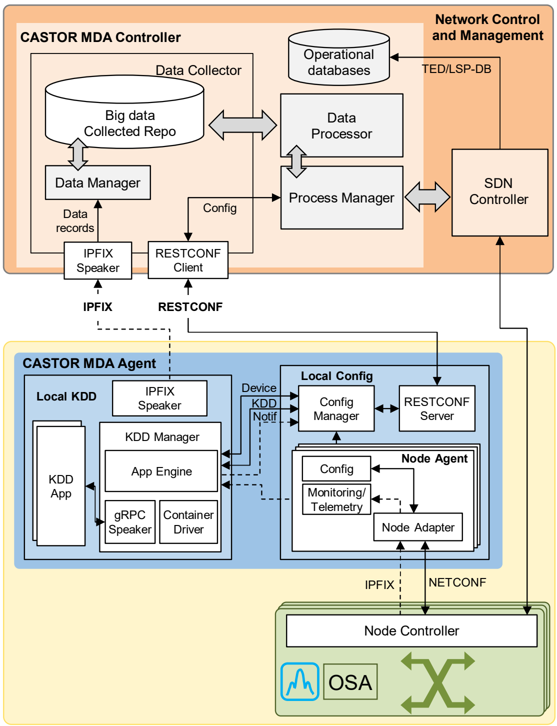


Fig. 3. CASTOR MDA architecture and interfaces.

III. Proposed Methods for Filter Failure Detection and Identification

In this section, we define two alternative classifiers for filter failure detection and identification based on the features defined in the previous section. Additionally, we study whether transforming features would improve classification accuracy. Once the optical spectrum of a signal has been acquired in an OP, the features are extracted and corrected applying the specific correction mask that corrects filter cascading effects for the number of filters that the signal has traversed from the transmitter to the OP. Next, failure analysis can be carried out; Fig. 4 summarizes the workflow that returns the detected class of the failure (if any) and its magnitude.

The first alternative classifier is based on DTs, whereas the second one selects SVMs. Both classifiers aim at identifying whether a filter failure is affecting a connection and if so, which is the type of failure: FS or FT. In the case that a failure

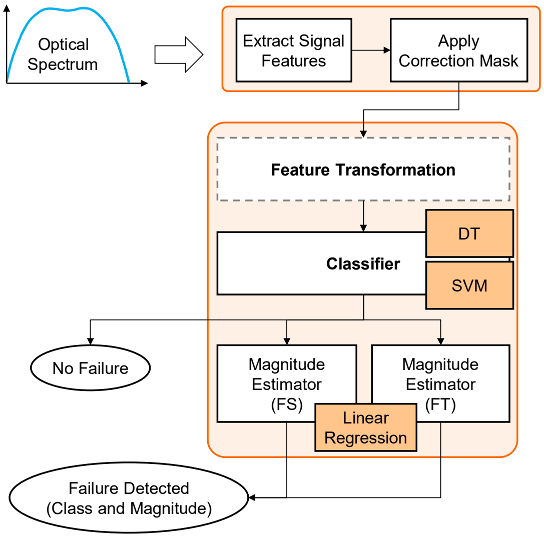


Fig. 4. Considered workflow for failure detection and identification.

Table I. General classification training algorithm pseudocode

```
INPUT dataset, Configs, maxSplits
OUTPUT model
 1:  dataset←balanceClassesByReplication(dataset)
 2:  for each config in Configs do initialize GoC[config]
 3:  for i=1..maxSplits do
 4:    <trainingSet,testingSet>←randomSplit(dataset)
 5:    initialize configParams
 6:    for each config in Configs do
 7:        model ← fit(trainingSet, config)
 8:        errorTraining ←predict(model, trainingSet)
 9:        errorTesting ←predict(model, testingSet)
10:        gocConfig ← GoC[config].addNew()
11:        gocConfig.configParams ← config.params
12:        gocConfig.errorTraining ← errorTraining
13:        gocConfig.errorTesting ← errorTesting
14:  bestConfig←computeBestConfig(GoC)
15:  return fit(dataset, bestConfig)
```

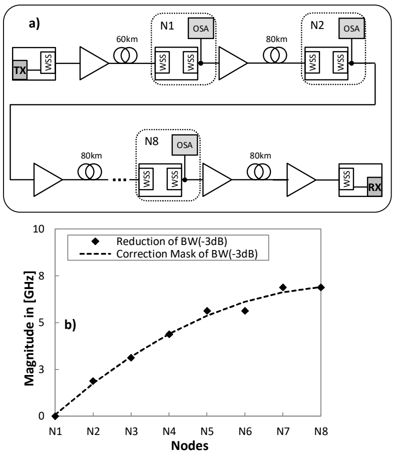


Fig. 5. VPI setup (a) and correction mask of bw_{-3dB} of the setup (b)

has been detected, its magnitude needs to be estimated. Two filter failure magnitude estimators can be called depending on the detected failure; both are based on linear regression.

Regarding the feature transformation block, for the sake of simplicity, we consider the magnitude of the failures as additional features for training the classifiers, so we use the magnitude estimator before a failure has been detected. In this way, original features are linearly combined to create new ones that might aggregate information in the hope of improving classification accuracy.

Let us now get insight about the training process of the classifiers (see pseudocode in Table I). The algorithm receives a dataset of labeled examples that is firstly balanced by adding copies of instances from the under-represented class to have the considered classes (normal, FS, FT) equally represented (line 1 in Table I). A set of configurations that contain specific parameters for the classification algorithm selected will be used during the training process. The parameters considered to fit DTs are the number n of observations per leaf, for every n a DT model is obtained. As for SVM fitting the parameters are the degree of the polynomial kernel (*kernelDegree*) for complexity control and the cost of misclassifying (*misClassCost*) for the size of the SVM. For every configuration, a number of randomly-generated splits of the data set for training and testing will be performed. To store the goodness of each configuration, the *GoC* array will be used in the rest of the algorithm and it is firstly initialized (line 2). Next, a new dataset split is generated, where the training set is used for fitting a model for the classifier with the specific selected configuration (lines 3-7). Once a model is computed, predictions using the training and testing data set are carried out (lines 8-9); the training and testing errors between the model prediction and the actual values are stored in the *GoC* array together with the current configuration parameters (lines 10-13). Finally, the results obtained for the different configurations and training/testing data splits are evaluated to select the configuration with minimum error (line 14). Such configuration is eventually used to fit a model using the whole dataset to improve the algorithm performance (line 15).

IV. Illustrative Results

In this section, we numerically compare the different failure identification classifiers and feature transformation described in the previous section. Let us begin by describing the transmission test-bed modeled in VPI Photonics (shown in Fig. 5a) that we use to generate the optical spectrum database required for training and testing the proposed algorithms. In the transmitter side, a 30 GBd DP-QPSK signal is generated. The signal passes through 8 single mode fiber spans. After each span, an optical amplifier compensates for the accumulated attenuation of the fiber. Each node is modeled with two 2nd order Gaussian filter emulating optical switching functionality performed by WSSs; filters bandwidth is set to 37.5 GHz, leaving 7.5 GHz as a guard band for the lightpath. Finally, the DP-QPSK signal ends in a coherent receiver that compensates for the impairments introduced throughout the transmission. One OSA per node, configured with 625 MHz resolution, is considered to monitor the optical spectrum of the lightpath. As previously discussed, a correction mask should be considered for the features affected by filter cascading (these features get reduced/increased while passing through

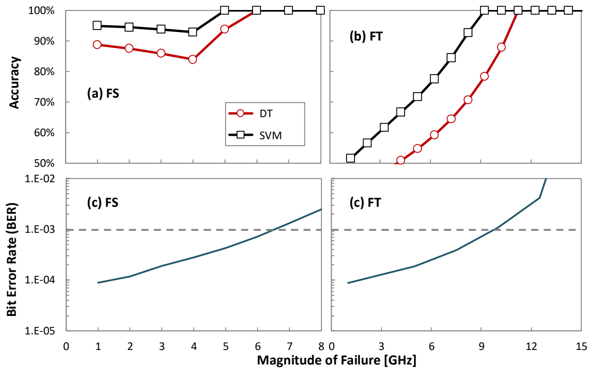


Fig. 6. Accuracy of the proposed methods for identifying FS (a) and FT (b), and BER in terms of failure magnitude of FS (c) and FT (c).

WSSs). Fig. 5b shows the amount of reduction in the magnitude of bw$_{-3dB}$ feature and the corresponding correction mask, obtained by fitting a 2nd order polynomial.

In this study, we focus on the cases where failure happens just at the 1st node. Then, in order to emulate failure scenarios, we modify the characteristics of the 2nd WSS of the 1st node; its bandwidth and central frequency are modified to model FT and FS failures, respectively. Utilizing this setup, we collect large database of failure scenarios with different magnitude (magnitude of 1 to 15 GHz for FT and 1 to 8 GHz for FS, both with 0.25 GHz step-size.). Let us firstly focus on detecting the failure at the node where it happens, which requires one OSA per node. We use accuracy (defined as the number of correctly detected failures over the total number of the failures) as a metric to compare the performance of the different options in the workflow.

Fig. 6a and Fig. 6b show the accuracy of identifying FS and FT in terms of the magnitude of the failure, respectively; every point in Fig. 6a-b is obtained by considering all the observations belonging to that particular failure magnitude and above. This representation reveals the accuracy of the proposed classifiers (without the feature transformation block) while considering failures with magnitude above certain thresholds. For instance, the accuracy of detecting FS in a dataset comprises observations larger than 1 GHz (in our case it comprises of failures up to 8 GHz in which there are equal number of observations per each magnitude) is around ~96% for SVMs, while it hardly approaches 89% for DTs. On the other hand, the accuracy of SVMs becomes 100% for failures larger than 5 GHz, while this level of accuracy for DTs is achieved for failures larger than 6 GHz.

To better relate this accuracy to the performance of the optical transmission system, we can look into the details of the BER of the signals. Fig. 6c shows the BER evolution for increasing magnitude of FS. It is shown that the proposed methods can perfectly detect a FS problem ahead of exceeding the FEC-threshold of BER (10^{-3}). Now let us focus on the FT case. Note that the magnitude of filter tightening is defined as the difference between the ideal bandwidth of the filter (equal to 37.5 GHz) and its bandwidth during the failure. As shown in Fig. 6b, the best accuracy of the proposed classifiers for low magnitudes (below 6 GHz) is around 80% (achieved for SVMs), which is due to the fact that the shape of the optical spectrum is quite similar to the normal scenario, making it very challenging for the classifier to distinguish. This is in contrast to the case of FS, which its impact is more evident from the very beginning due to its asymmetric behavior with respect to the optical spectrum. As shown in Fig. 6b, for the magnitudes above 9 GHz, exploiting SVMs, the classifier perfectly detects the failed lightpaths. However, this level of perfect accuracy is achieved for magnitudes above 10.5 GHz for the DTs. If we look at the BER performance of the system, it can be understood that the FEC-threshold limit of BER is exceeded for magnitude around 10 GHz, meaning that yet by exploiting SVMs, we are able to detect a failure ahead of disruption of the connection. Note that, the BER values reported in the paper are obtained in our VPI setup including 18 WSSs without power tuning of the components to increase the OSNR. In practice, better BER performance can be achieved with the help of more sophisticated DSP techniques and at the expense of some level of OSNR penalty [9], meaning that the detection threshold of 100% accuracy can be even further away from the FEC-limit of BER.

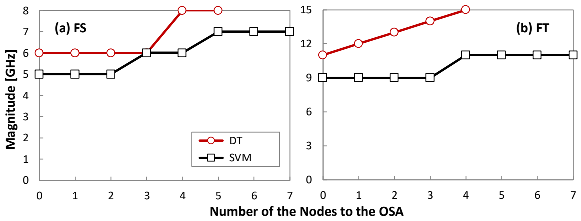


Fig. 7. Smallest failure magnitude for 100% classification accuracy vs. the OSA location for FS (a) and FT (b).

In the second part of the analysis, we focus on detecting the failures in some nodes after the point where the failure happens, showcasing the robustness of the proposed methods with respect to the evolution of the optical signal along the transmission line. In addition, by following this approach, the number of utilized OSAs in the network can be reduced. Fig. 7 shows the minimum magnitude after which the accuracy of classifiers remains 100% in terms of the location of OSA compared to the point that failures happen; 0 on the x-axis means that OSA is placed at the node where the failure happens (N1 in Fig. 5), while 7 means that is placed 7 nodes away from the location of the failure (N8 in Fig. 5). It can be understood that the SVM-based classifier is robust regardless of the location of the OSA and it perfectly detects the failures above a magnitude threshold where the FEC-limit of BER is not yet exceeded. Even though the DT-based classifier shows an acceptable performance for FS failures up to 3 nodes distance from the location of the failure, it fails when considering FT failures.

Once the failures are detected, filter shift estimator (FSE) and filter tightening estimator (FTE) can be launched to return the magnitude of the failures; estimators are based on linear regression. Estimated values of FS and FT with respect to their expected values are illustrated in Fig. 8a and Fig. 8b, respectively. As shown, the estimators can predict the magnitude of failures with very high accuracy, with mean

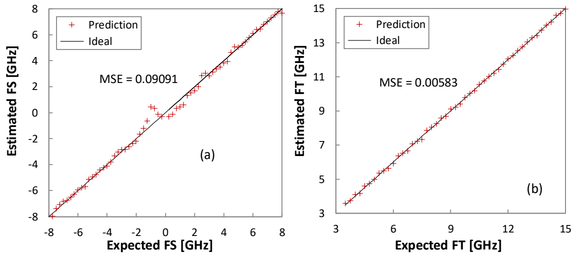


Fig. 8. Prediction accuracy of FS (a) and FT (b) estimators.

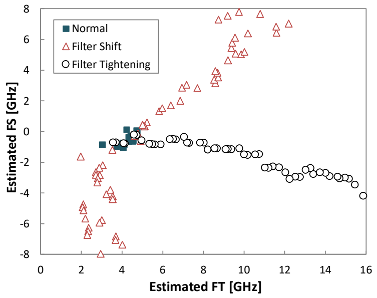


Fig. 9. Estimated FS vs estimated FT as a 2D vector space.

square error (MSE) equal to 0.09091 and 0.00583 for FSE and FTE, respectively.

In addition to the use of these estimators to explore the magnitude of the failures, they can be used in the feature transformation step as anticipated in Section III. In fact, the output of FSE and FTE can be considered as two principal components of an imaginary two-dimensional (2D) vector space as shown in Fig. 9. In such space, FS and FT failures evolve in different directions of the vector space. As illustrated in Fig. 9, the observations belonging to normal operation and the small magnitudes of the failures coincide. However, they become perfectly distinguishable as the magnitude of failures increase.

Let us evaluate the benefits of exploiting the outputs of FSE and FTE estimators as additional features for training the classifiers; Fig. 10 presents the obtained results. For the sake of conciseness, we group the magnitudes into three groups of low (L), medium (M), and high (H) magnitudes, instead of reporting all of them independently. Regarding the location of the OSAs, we report just three locations. Analyzing Fig. 10a, one can realize that the accuracy of DTs can be substantially improved, notably for low and medium magnitude, when using the estimations of FSE and FTE as new features. Additionally, it makes the classifier based on DT more robust while using OSAs far away from the location of the failures. However, it yet cannot outperform the classifier based on SVMs, even with these additional features. We also see that adding these new features does not enhance the performance of SVMs, revealing that such classifier can internally exploit the primary features to the maximum extend (Fig. 10b). Therefore, the classifier based on SVMs does not require an additional preprocessing step to generate more features; note that the magnitude of the failures are just linear combinations of primary features. Conversely, the substantial improvement seen in the DT classifier reveals that DTs cannot maximally

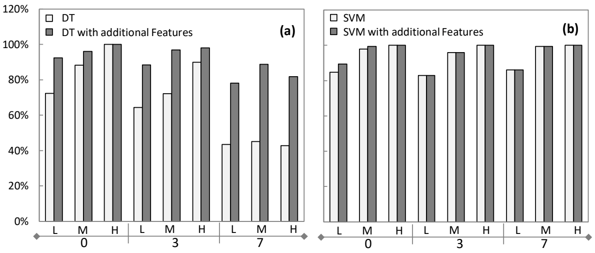


Fig. 10. Comparison of the classifiers with and without additional features.

exploit the information carried by the primary features and requires some pre-processing to grasp more information, which is a weakness of DTs compared to SVMs.

V. Concluding Remarks

In this work, we have studied the benefits of exploiting OSAs for identification and localization of filter related failures. Considering the classification methods, we have compared the performance of two different algorithms, based on DTs and SVMs, in terms of their accuracy in the detection of the failures. Additionally, we have evaluated their robustness with respect to the evolution of the optical signal along the transmission line.

The DT-based approach shows a reasonable performance if it follows a pre-processing step, aiming at generating more useful features. Otherwise, it performs weakly and lacks robustness. On the other hand, the classifiers based on SVMs have shown significant performance in detecting critical filter related failures at any point along the route of a lightpath without any pre-processing step. In addition, it benefits from very high robustness with respect to the evolution of signal along the transmission line. This robustness relaxes the requirement of deploying one OSA per node for spectrum monitoring, which in turn results in a significant reduction of the number of OSA used in the network.

References

[1] A. Castro *et al.*, “Experimental Assessment of Bulk Path Restoration in Multi-layer Networks using PCE-based Global Concurrent Optimization,” IEEE Journal of Lightwave Technology (JLT), vol. 32, pp. 81-90, 2014.

[2] A. P. Vela *et al.*, “BER Degradation Detection and Failure Identification in Elastic Optical Networks,” IEEE Journal of Lightwave Technology (JLT), vol. 35, pp. 4595-4604, 2017.

[3] A. P. Vela *et al.*, “Soft Failure Localization during Commissioning Testing and Lightpath Operation [Invited],” IEEE Journal of Optical Communications and Networking (JOCN), vol. 10, pp. A27-A36, 2018.

[4] A. P. Vela, M. Ruiz, and L. Velasco, “Distributing Data Analytics for Efficient Multiple Traffic Anomalies Detection,” Elsevier Computer Communications, vol. 107, pp. 1-12, 2017.

[5] L. Velasco *et al.*, “An Architecture to Support Autonomic Slice Networking [Invited],” IEEE Journal of Lightwave Technology (JLT), vol. 36, pp. 135-141, 2018.

[6] Ll. Gifre *et al.*, “Autonomic Disaggregated Multilayer Networking,” IEEE Journal of Optical Communications and Networking (JOCN), vol. 10, pp. 482-492, 2018.

[7] L. Velasco *et al.*, “Building Autonomic Optical Whitebox-based Networks,” IEEE Journal of Lightwave Technology (JLT), 2018.

[8] C. Bishop, *Pattern Recognition and Machine Learning*, Springer-Verlag, 2006.

[9] T. Rahman *et al.*, “On the Mitigation of Optical Filtering Penalties Originating from ROADM Cascade,” IEEE Photonics Technol. Letters, vol. 26, no. 2, pp. 154-157, 2014.

Analog Radio-over-Fiber Solutions in Support of 5G

D. Apostolopoulos, G. Giannoulis, N. Argyris, N. Iliadis, K. Kanta and H. Avramopoulos

School of Electrical and Computer Engineering, National Technical University of Athens, Athens, Greece

apostold@mail.ntua.gr

***Abstract*— We introduce Centralized (C-RAN) architectures combined with analog optical transport schemes to realize high-speed lanes between BBU pool and RRHs. The DSP-enabled architecture relies on the use of a powerful digital engine supporting data plane functions for both fiber-wireless parts. The presented concept is also supported through preliminary experiments demonstrating the proof-of-concept operation of a Fiber-Wireless link. IFoF/mmWave transmission of single-band radio signals at 60-GHz is demonstrated for Downlink directions using QAM-modulated signals at 1 Gbaud.**

Keywords—5G, Centralized-RAN, Mobile Fronthaul, Analog RoF, Millimeter Wave Radio, DSP-enabled BBU, Digital Pre-compensation/Equalization, M-QAM

I. Introduction

We have reached a critical point where the first wave of 5G compliant technologies and architectures have escaped research laboratories and are now on the verge of commercial exploitation [1]. In this context, 4G-Long Term Evolution (4G-LTE) mobile ecosystem is in the start of a transformation journey, triggered by 5G hallmarks such as millimeter waves, massive Multiple-Input and Multiple-Output (MIMO) and beamforming [2]. This transitionary phase of the entire deployed ecosystem, imposes also the process of mutation for the optical solutions which currently support the data transport within LTE networks [3]. In this way, traditional optical transport concepts supporting the Mobile Fronthaul (MFH) are expected to be replaced with new ones, capable to meet 5G requirements [4]. In more detail, current fronthaul exploits digitized optical transmission, a technique that results in a 10x bandwidth demand than the respective wireless bit-rates [5]. This digital approach can simply not scale with the increasing radio bit-rates and number of antennas required by 5G network densification approaches. Industrial consensus seems currently to agree that short- to mid-term solutions can come within digital MFH, but the required maximization of bandwidth utilization, offered by Analog Radio-over-Fiber (A-RoF) transmission, renders the transition towards concepts based on analog optics an attractive longer term 5G solution. [6].

In this work, the concept of Digital Signal Processing (DSP)-assisted optical transmission capable to support analog MFH is discussed. This RoF concept aims to alleviate the bandwidth limitations of the 5G MFH through the use of analog optics, which can carry native wireless data signals via installed fibers. We introduce this ambitious analog concept within the 5G landscape emphasizing on the structural changes and challenges that analog MFH attempts to address. In the next paragraphs, we thoroughly discuss the architectural shift towards Centralized Radio Access Network (C-RAN) topologies, which put the traditional digital MFH transport on the question. The DSP-enabled Analog architecture supporting the MFH is then presented, focusing on the digital functions undertaken from a powerful centralized DSP engine. Preliminary experiments that provide a proof-of-concept validation of our A-RoF concept are also discussed.

II. Mobile Network Expansion and Centralized Topologies

A. *The transition towards C-RAN architectures*

The exponential growth of the number of femto-cells to meet the demands of mobile traffic is one of the prominent features of 5G architectures that are still under investigation. To support low latency, high capacity, cost-effectiveness and low energy consumption, the entire end-to-end network should be overhauled. Currently, many Mobile Network Operators (MNOs) operate using a Distributed Radio Access Network (D-RAN), in which the 4G radio at the macro site tower consists of a collocated Baseband Unit (BBU) at the base of the tower [7]. The main advantage of D-RAN is the efficient use of backhaul bandwidth which can be achieved through various well-established technologies (Ethernet, Passive Optical Network (PON), etc.) [8]. However, dense 5G cellular topologies apply significant pressure on the static nature of D-RAN where BBUs are assigned statically to a number of cells. The spatial and temporal volatility of mobile traffic, makes the static D-RAN topologies suboptimal and new flexible topologies are needed to obtain energy and cost savings for the MNOs [9].

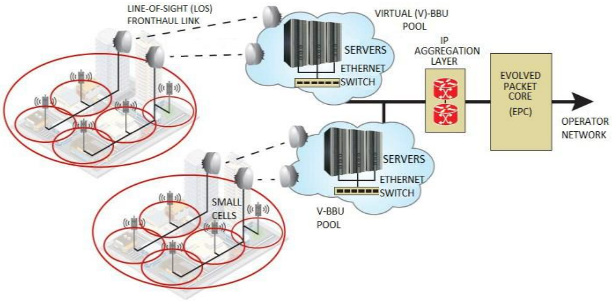


Figure 1: C-RAN architecture that has been proposed from Fujitsu for mobile scenarios within dense urban environment [7].

978-3-903176-07-2 © 2018 IFIP

Following this rational, the idea of centralization has been re-invented in the era of 5G networks, since it presents significant offerings compared to traditional D-RAN topologies [10]. The C-RAN approach favors the separation of radio elements of the base station (called Remote Radio Heads, RRH) and the elements processing the base band signal (called BBUs), which are centralized in a single location or even virtualized into the cloud. This approach, which has become a hot research topic in both academia and industry [11], benefits from simpler radio equipment at the network edge, easier to operate and cheaper to maintain, while the main RAN intelligence (BBUs) is centralized in the operator-controlled premises. The centralization is further enhanced with cloud computing [9], providing elasticity, and virtualization with possibility for multitenancy among MNOs. Several main 5G use cases, such as Virtual Reality (VR) applications, which require the real-time processing of massive amount of data, can push much of the processing from a local server to the cloud [12]. This practically means that the computational resources can be pooled and dynamically allocated to a virtual BS, which brings cost-effective hardware and software design [13]. Figure 1 provides a practical implementation of the above architecture showing a small-cell based C-RAN approach proposed from Fujitsu in a dense urban environment [7]. Beyond the software-centric solutions, the C-RAN approach needs also a paradigm shift on the hardware side to meet the challenges for the centralized baseband processing that serves a large number of RRHs. At the heart of this change, powerful Field Programmable Gate Array (FPGA) boards can offer the capability to implement high-throughput 5G transceivers for the data plane while they can also realize the Software Defined Network (SDN) functions described above for the software-centric architecture [14]. It should be noted that since the D-RAN remains the dominant deployed architecture of the antenna sites, research efforts should be considered that target a smooth transition towards 5G RAN ecosystem, supporting the coexistence of D-RAN and future C-RAN topologies [15].

B. *Optical transport for MFH*

The fronthaul mobile traffic transport between the BBUs and RRHs seems to be a significant challenge for 5G topologies, since it needs to address several issues related to the convergence of optical channels with complex radio interfaces. In the current C-RAN architecture designed for the Long-Term Evolution-Advanced (LTE-A) mobile network, the fronthaul interface is based on Common Public Radio Interface (CPRI). This Digitized-RoF (D-RoF) interface which relies on a link transmitting In-phase and Quadrature (IQ) data of the baseband signal components, suffers from low bandwidth efficiency since it uses the available bandwidth to send IQ data samples, decreasing thereby the effective data rate of the transport [16]. In this context, several solutions have been proposed to overcome this bandwidth wall, which mainly focus on compressed CPRI techniques with minimal impact on the optics and fiber network. The emerging need for fronthaul compression has been addressed in the current LTE-A fronthaul links, through a large set of CPRI compression algorithms [17]. Compressed CPRI links in a high-speed Pulse-Amplitude-Modulation-4 (PAM-4) are actively investigated, offering a 2x rewards on bandwidth efficiency of the D-RoF approach [18].

The CPRI compression techniques offer remarkable bandwidth gains without any structural shift on the current C-RAN architectures. However, the cost of increased complexity at the BBU side needs to be considered, while bandwidth limitations still come from digital electronics and their interfaces at the BBU side [19]. To overcome this challenge, Physical (PHY) functional split has been proposed as a possible solution for the MFH by shifting some DSP operations from BBUs to RRHs. This functional split between BBU and RRH relaxes the BBU digital overloading and lowers the fronthaul bandwidth requirements on the optical link. However, it faces great challenges when advanced coordination functionalities are required for a large number of RRHs while the latency budget is also affected through extra processing burden [20] [21]. An alternative D-RoF concept can be implemented through Open Base Station Architecture Initiation (OBSAI) which is also implemented through a packet-based interface [16]. Since the mapping methods of CPRI are more efficient than OBSAI [22], most global vendors and MNOs have chosen CPRI for deployed C-RAN topologies.

A-RoF revolutionizes the MFH landscape by fully releasing the bandwidth capabilities of mmWave bands, requiring only simple functions to exploit the offered bandwidth of the fronthaul part. Moreover, Analog MFH implementations for 5G services can harmonically co-exist over the installed fiber infrastructure supporting PON topologies of fixed wireline services [23][24]. These unique benefits come at a cost of increased hardware complexity at the BBU since it hosts the entire set of DSP functions. In addition, the optical distribution of radio signals over Intermediate Frequency/Radio Frequency (IF/RF) carriers is susceptible to a number of generation and transmission impairments, which in turn add noise and distortion due to channel nonlinearities [25].

The A-RoF approach described in the current document aims to address the above challenges through the use of powerful DSP engines implemented at ultra-high-speed FPGA boards. The proposed DSP-assisted A-RoF solution, proposes a structural shift in the current MFH deployments since it aims to combine the implementation of DSP-based functions for ultra-broadband radio signals (covering the entire unlicensed 57-64 GHz band) with the electro-optic conversion and transmission through installed fiber infrastructure.

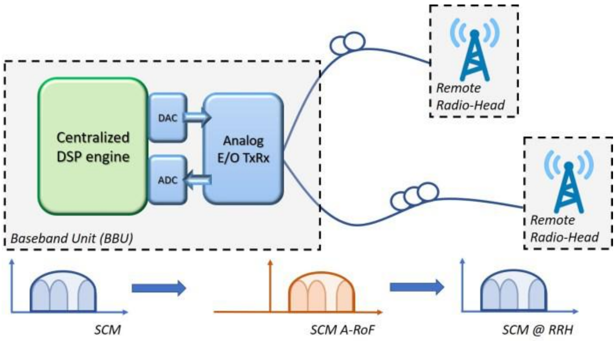


Figure 2: DSP-enabled Analog-RoF concept.

C. DSP assisted A-RoF fronthaul

Several studies explore the advantageous approach of DSP assisted A-RoF fronthaul approach. In the technique presented in [26] a number of IQ data channels, each of them corresponding to a single CPRI stream, are multiplexed and transmitted through the fronthaul link in an Intermediate Frequency-over-Fiber (IFoF) scheme. Such implementations combine commodity optical transceiver modules, carrying low bandwidth components (~10 GHz), with high speed Digital to Analog Converters (DACs) and Analog to Digital Converters (ADCs) [27]. Build upon this concept, an A-RoF architecture with DSP functionalities is presented in Fig. 4. The core of the proposed BBU architecture is a centralized DSP engine, which is responsible for implementing the physical layer functionalities for the fronthaul link. The set of these functions covers all the necessary coding/decoding, modulation and MIMO processing of the wireless channel signals. These radio signals are generated by high-performing DACs, first transmitted through the installed fiber and eventually over the air interface at the mmWave frequency band. Such an A-RoF/mmWave Fronthaul approach realizes actual centralized-RAN, since the complete set of baseband operations are digitally performed in the BBU, removing thereby any processing stage from the RRHs. A first advantage of this approach is the advanced implementation of inter-cell coordination. As the baseband processing for the radio signals to/from different RRHs is done at the same engine, tighter coordination of neighboring antennas becomes more feasible. As an example of advanced inter-cell cooperation, it is possible that two (or more than two) radiowave signals are jointly received and processed in the BBU pool so that so-called network MIMO can be achieved. Moreover, it ensures scalability since many RRHs can be placed when the capacity demands are increased in a plug-and-play manner. Besides, advanced inter-cell coordination enables better management of interference between adjacent cells, a critical point for 5G ultra-dense cellular networks [28]. Finally, the centralization of DSP engine in the BBU pool can also offer significant energy and cost savings using coordination schemes among them [29], while the hardware resources at the RRH side are practically minimized.

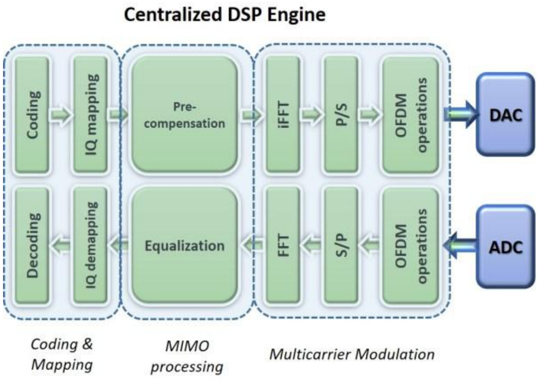


Figure 3: Centralized DSP engine for 5G-compatible Analog MFH.

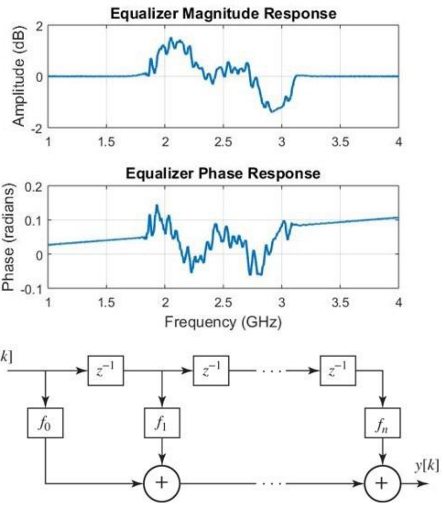


Figure 4: FIR equalizer is implemented by the Frequency Response of the A-RoF link.

III. A-ROF FOR MFH USING A CENTRALIZED DSP ENGINE

A. Centralized DSP engine

Looking into more detail on the core blocks of the centralized DSP engine, the one that lies closer to the RRH is the Modulation and Channel Mapping block. In this stage, the digital sequences are mapped into the appropriate waveforms that will be transmitted over the Fiber/Wireless link. Digital modulation techniques support the generation of any Single-Carrier (SC) or Multi-Carrier (MC) scheme (e.g. Orthogonal Frequency-Division Multiplexing (OFDM)-like waveforms), thus allowing for compatibility with the current LTE standards, future upgrade to 5G candidate waveforms (eg. Universal-Filtered Multi-Carrier (UFMC), Filter Bank Multi-Carrier (FBMC), General Frequency-Division Multiplexing (GFDM) [30]) as well as more forward-looking approaches thanks to the arbitrary waveform generation capabilities of the engine. A higher bandwidth efficiency is achieved compared to D-RoF approaches of CPRI and Physical Layer Split (PLS) where the bandwidth is utilized for serial transmission of digitized IQ waveforms. In the A-RoF scheme low Intermediate Frequencies (~5 GHz) are employed to carry the modulated radio signals, resulting in bandwidths which are accommodated by typical low-cost transceivers.

Moreover, the use of Digital Sub-Carrier Multiplexing (SCM) techniques can also be adopted to further increase the bandwidth efficiency. For the SCM generation, digital upconversion schemes are employed to obtain the appropriate IF frequencies for A-RoF transmission, eliminating the need of external analog mixers and local oscillators. An additional advantage of employing SCM schemes, we fully utilize the bandwidth offered by the A-RoF components.

As the “DSP-free” RRH units are not capable of baseband signal processing, the centralized DSP engine serves on a two-fold dimension. For the downlink direction, digital pre-distortion based on the fiber/wireless channel response is performed. Channel estimation methods based on training sequences determine the magnitude and phase response of the Fi-Wi link. Thus, the response of electrical and optical components such as DACs, RF drivers, modulators, filters and photoreceivers of the optical part and RF mixers, up/down converters of the antenna subsystem is reversed using linear equalizers implemented with Finite Impulse Response (FIR) filters. Fig. 4 illustrates a channel response extracted by a real experimental testbed with both optical and RF frontends. In the uplink, equalization stages are enhancing the demodulation and detection of the received radio signals after Wireless/Fiber transmission.

Since the A-RoF based MFH scheme appears to be significantly more efficient D-RoF approaches, accommodation of multiple RRHs' traffic can be achieved. In the special case where two or more RRHs serve the same small cell or coverage area, the DSP engine can be employed to perform equalization utilizing the spatial channel characteristics and antenna diversity, thus realizing a DSP MIMO system. Such MIMO processing capabilities along with robust coding schemes offer significant reduction in operational margins in terms of required Signal-to-Noise-and-Interference-Ratio (SNIR) and received power levels [31].

The implementation of the above rich digital portfolio within the centralized BBU will be undertaken by powerful FPGA boards. The use of these powerful FPGAs is to accelerate the critical functions of the baseband chain and sustain the necessary throughput to meet the bitrate and latency requirements within 5G. Since the FPGA boards becomes the essential part of the envisaged digital engines, the design, implementation and validation of real-time testbeds has became a significant point of interest for the 5G hardware research community [32][33]. Through the literature, several works have been conducted setting key specifications for FPGA implementation and proposing efficient implementations to meet the 5G network goals [34][35].

IV. Preliminary experiments & results

In this section we present preliminary experiments based on the DSP-assisted A-RoF concept described above. Figure 5(a) depicts a standard Intensity Modulation/Direct Detection (IM/DD) testbed where the A-RoF link is emulated.

An Arbitrary Waveform Generator (AWG) was used to provide the DSP functionalities for the downlink part of the Analog MFH. The programmable data source allowed Transmitter (Tx)-side DSP operations such as pulse shaping (using Root Raised Cosine (RRC), implementation of digital filters and digital pre-distortion of the optoelectronic components as described in Section III. To this end, channel estimation has been performed, prior the actual data transmission, where an amplitude and phase channel response was estimated on the frequency domain. Through the AWG, the pre-distorted data signals were digitally upconverted to the selected IF. This mixerless scheme was based on the use of embedded DAC of the AWG without the need of any external analog RF signals. A single-drive electro-optic modulator was used to generate the IFoF signal carrying the radio bands with the intensity modulator biased at the quadrature point ($V_\pi/2$). The Mach-Zehnder Modulator (MZM) was biased at the quandrature point to ensure its operation at the linear regime of transfer function.

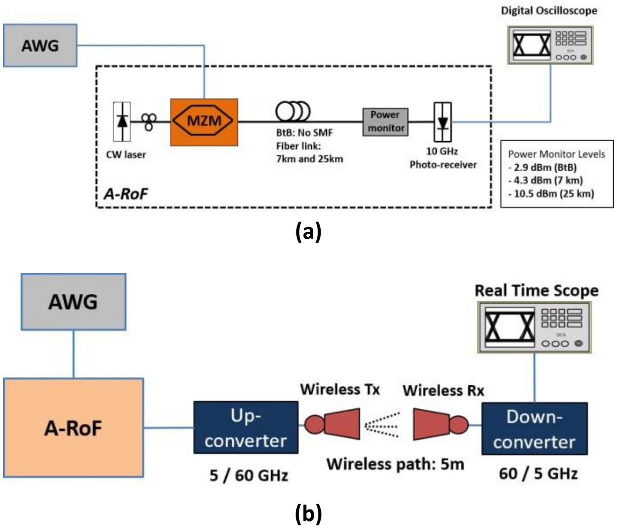


Figure 5: (a) A-RoF testbed on an IM/DD link using laboratory equipment (b) Fiber-Wireless link

Fiber links of Standard Single-Mode Fiber (SSMF) up to 25 km were used in the testbed to emulate the fiber connection between the BBU and RRHs. At the receiver side, commercial off-the-shelf 10 GHz photo-receiver allowed for the optical-to-electrical conversion and the efficient interface with the radio hardware. Figure 5(b) illustrates the above emulation of a converged Fiber-Wireless link where the A-RoF output is connected directly to V-band radio hardware. The received IF signal was up- and down-converted via commercially available V-band upconversion modules with Local Oscillator (LO) operated at 58-63GHz. Standard pyramidal gain horn V-band antennas of 23 dBi gain and 10° beamwidth were employed. These directional antennas together with the up- and downconversion units, were mounted on wooden tripods and kept fixed at a height of 1.4 m above the floor located in a 5 m horizontal distance in an indoor laboratory environment.

Single-Band and Multi-band transmission experiments were carried out considering initially only the fiber part and then, the Fiber-Wireless link. An IF frequency of 5 GHz was selected in order to meet the specifications of the mmWave Up- and Down-converter units at the radio part. Such selection easily accommodates the entire bandwidth (~7GHz) within the targeted unlicensed radio band and provides resilience against the power fading due to Chromatic Dispersion (CD) for fiber lengths up to 25 km [36]. Moreover, since IF frequencies were employed to balance the bandwidth gap between the electronic and photonic devices, the need for high speed optical frontends is eliminated.

Exploiting once more the DSP capabilities on generation of complex waveforms, four different subcarriers have been digitally synthesized, before feeding a single DAC board in order to generate the desired multiband radio signal. The 4 sub-bands were assigned at 0.625 GHz, 1.875 GHz, 3.125 GHz and 4.375 GHz center frequencies (around 2.5 GHz) and each of them was modulated at 1 Gbd symbol rate, pulse-shaped with a RRC filter ($\alpha = 0.2$), utilizing thereby a total 5 GHz bandwidth. Different modulation types were investigated for the sub-bands. Figure 6(a) shows that in the case of back-to-back measurement, all sub-bands have almost the same performance and the modulation type does not affect the measured Error Vector Magnitude (EVM) values. After fiber transmission over the 25-km fiber link, the effect of dispersion-induced power fading is evident, since the higher frequency components suffer from severe distortion compared to lower. The use of a higher order modulation format seems

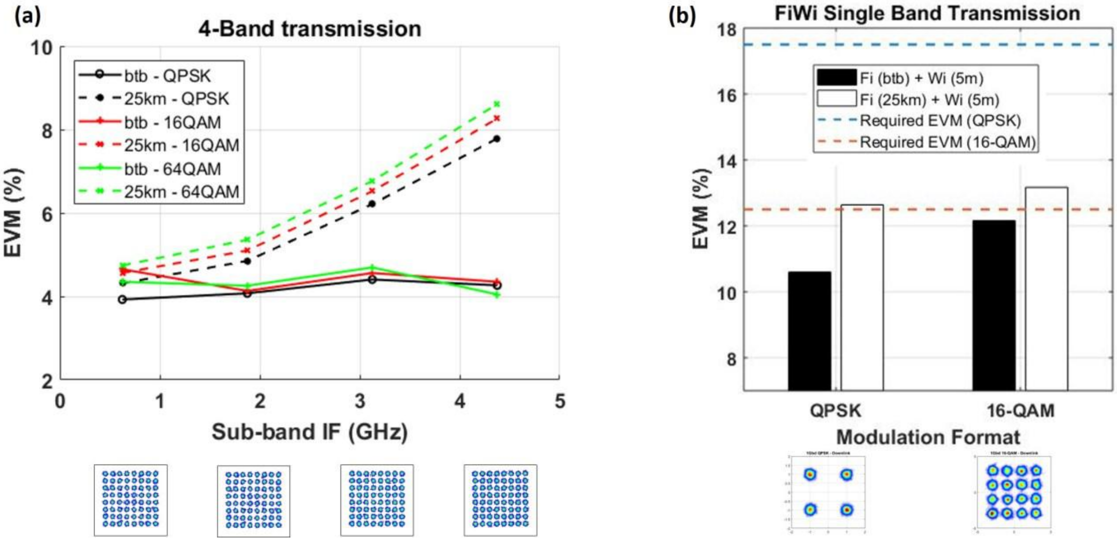


Figure 6: (a) EVM measurements for each sub-band of the SCM A-RoF link (a) 4-bands, Constellation diagrams for 64-QAM after 25km, (b) EVM bar-diagram measurements for A-RoF/mmWave transmission in downlink directions. Constellation diagrams for QPSK and 16-QAM after Rx-side DSP are also included.

to slightly increase the EVM value, which does not hamper the demodulation of this formats in this specific case. EVM values below 8.5% were achieved for all QPSK, 16-QAM and 64-QAM schemes in all the allocated spectrum bands. A total data transport rate for 64-QAM modulation type of 24 Gb/s can be achieved.

Comparing the EVM performance between the single band and multiple band IFoF transmission, it is evident that the signal distortion is increased for increased number of radio bands as it is expected. Moving towards wide-band, multi-carrier approaches, the analog photonic links suffer from cross modulation distortion (XMD) which introduces severe distortion in the modulated in-band of each signal [37].

The next evaluation step comprises the characterization of fiber-wireless link including both fiber and wireless part. Combining the proposed A-RoF IM/DD optical fronthaul topology with commercial V-band radio hardware (operated at 57-64 GHz unlicensed band), indoor wireless measurements using directional antenna elements were performed.

For the DL operation the signal after the transmission through the A-RoF link was received by a photoreceiver (Avalanche Photodiode Photoreceiver (APD) + Transimpedance Amplifier (TIA)). This modulated IF output is fed into a mmWave upconverter to obtain the mmWave transmission through the air. After the digitization of the received signals, digital downconversion from the IF frequency to baseband was performed. There, matched filtering, resampling and proper timing synchronization was applied in order to extract a single-sample per symbol sequence. For the equalization stage, a static 5-tap Radius Directed Constant Modulus Algorithm (CMA) algorithm was employed for off-line equalization of both fading effects stemming from both fiber-air transmission and to remove the frequency response from mmWave components. Finally, a Carrier Phase Recovery stage compensated for the phase noise due to local oscillators' mismatches. Statistical constellation analysis and error counting was employed for estimating and measuring the transmission quality. In this experiment Single-CarrierQuadrature Phase Shift Keying (QPSK) and 16-QAM) at 1 Gbd symbol rate were employed as modulated radio signals. A RRC pulse shaping filter with 20% excess bandwidth resulted a total 1.2 GHz to be transmitted through the fiber/wireless link. In Figure 6(b), we present an EVM bar diagram for both uplink and downlink operation using the above modulation types. In order to investigate the role of fiber transmission, the long fiber-part of 25 km was also used for the measurements. The reported results show that the fiber part of 25 km introduces severe signal distortion which leads to higher EVM values compared to back-to-back measurements, as it was originally expected. The calculated EVM value after the fiber wireless transmission was measured to be just above the limit value, according to 3GPP standards, for successful 16-QAM signal demodulation [38]. As a future step, the use of channel estimation techniques and improved equalization algorithms could offer even lower EVM values, to meet the 3GPP requirements for 16-QAM modulation type.

V. Conclusion

This work presented a DSP-enabled concept that can efficiently support the MFH of C-RAN architectures. Through literature review, the shift towards 5G centralized topologies was introduced while the optical transport was addressed via bandwidth efficient A-RoF concept. The DSP engine, responsible for the data plane functions, was thoroughly discussed focusing on the efforts to host the complete set of digital functions at the BBU side, real-time implemented via powerful FPGA boards.

In this work, results from preliminary experiments are also included aiming at the verification of the proposed A-RoF solution for realistic fronthaul scenarios. The optical transport of multiband radio signal carrying 64-QAM modulated signals was experimentally demonstrated showing EVM values below 8.5%. The wireless transmission was achieved by connecting radio equipment at 60GHz. We experimentally demonstrated fiber-wireless transmission link up to 4 Gb/s using QPSK and 16 QAM modulation types which exhibited EVM values below 13% for both cases.

Acknowledgment

This work has been supported by the European Commission through H2020 5GPPP Phase II project 5G-PHOS, Contract Number 761989.

References

[1] A"5G Cell Service Is Coming. Who Decides Where it Goes?", article appeared at online version of www.nytimes.com, posted at 2018-03-02 by Allan Holmes.

[2] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for next generation wireless systems, " IEEE Commun. Mag., vol. 52, no. 2, pp. 186–195 (2014).

[3] N.J. Gomes et al., "The new flexible mobile fronthaul: digital or analog, or both?", in Proc. of International Conference on Transparent Optical Networks (ICTON 2016), 10-14 July 2016, Trento, Italy.

[4] M. Presi et al., "Optical Solutions supporting 5G and Beyond", Presentation slides uploaded in the 5G-PPP repository, 23 March 2016.

[5] T. Pfeiffer, "Next generation mobile fronthaul and midhaul architectures," J. Optic. Comm. Netw., Vol. 7, Issue 11, pp. B38-B45 (2015).

[6] J. Kani, J. Terada, K.-I. Suzuki, A. Otaka, "Solutions for Future Mobile Fronthaul and Access-Network Convergence," IEEE Journal of Lightwave Technology, Vol.35, Issue 3, pp.527-534 (2017).

[7] "The Benefits of Cloud-RAN Architecture in Mobile Network Expansion", Application Note from Fujitsu Network Communications Inc. (2014).

[8] Agata and K.Tanaka, "NG-EPON for Mobile Access Network", presented in IEEE 802 Plenary Session, March 2014, Beijing, China.

[9] A. Checko et al., "Cloud RAN for Mobile Networks – A Technology Overview", IEEE Communications Surveys & Tutorials, Vol.17, Issue 1, pp. 405-426 (2015).

[10] China Mobile Research Institute, "C-RAN: The Road Towards Green RAN," White Paper, 2013. Available: http://labs.chinamobile.com/cran/

[11] Manli Qian et al, "A super base station based centralized network architecture for 5G mobile communication systems", Digital Communications and Networks (2015).

[12] M. Koziol, "Mobile World Congress 2018: Don't Expect 5G Service Anytime Soon", posted in spectrum.ieee.org, 2 March 2018. https://spectrum.ieee.org/tech-talk/telecom/wireless/mobile-world-congress-2018-5g-isnt-for-you

[13] D. Wubben et al., "Benefits and impact of cloud computing on 5G signal processing: Flexible centralization through cloud-RAN," IEEE Signal Process. Mag., vol. 31, no. 6, pp. 35–44, Nov. 2014.

[14] M. Milosavljevic, "FPGAs for Reconfigurable 5G and Beyond Wireless Communication", presented in NMI FPGA Network: "Safety, Certification and Security", University of Hertfordshire, 19 May 2016, Hatfield, UK.

[15] C. Raack, J.M. Garcia, R. Wessaly, "Centralized versus Distributed Radio Access Networks: Wireless intergation into Long Reach Passive Optical Network", in Proc. of CTTE 2015, 9-10 November 2015, Munich, Germany.

[16] A. Olivia et al., "An Overview of the CPRI specification and its application to C-RAN based LTE scenarios", IEEE Communications Magazine, Vol. 54, Issue 2, pp. 152-159 (2016).

[17] White paper from Altera. "The Emerging Need for Fronthaul Compression", June 2016.

[18] F. Lu et al. "Adaptive Digitization and Variable Channel Coding for Enhancement of Compressed Digital Mobile Fronthaul in PAM-4 Optical Links", IEEE Journal of Lightwave Technology, Vol.35, No.21,pp.4714-4720 (2017).

[19] N. Carapellese et al, "BBU placement over a WDM aggregation network considering OTN and overlay fronthaul transport,," European Conf. on Optical Com. (ECOC), pp. 1-3, 2015

[20] K. Miyamoto, S. Kuwano, J. Terada, and A. Otaka, "Split-phy processing architecture to realize base station coordination and transmission bandwidth reduction in mobile fronthaul," presented at the Optical Fiber Communications Conf. Exhib., Los Angeles, CA, USA, 2015, Paper M2J.4.

[21] J. Armstrong, "OFDM for Optical Communications," Journal of Lightwave Technology, vol. 27, no. 3, pp. 189-204, Feb.1, 2009.

[22] M. Nahas, A. Saadani, J. Charles, and Z. El-Bazzal, "Base stations evolution: Toward 4G technology," in Telecommunications (ICT), 2012 19th International Conference on. IEEE, 2012, pp. 1–6.

[23] X. Hu, C.Ye and K.Zhang, "Converged Mobile Fronthaul and Passive Optical Network Based on Hybrid Analog-Digital Transmission Scheme", in Proc. of Optical Fiber Communications Conf (OFC) 2016, paper No. W3C.5, 20-24 March 2016, Anaheim, CA, USA.

[24] G. Kalfas et al., "Non-Saturation Delay Analysis of Medium Transparent MAC Protocol for 60 GHz Fiber-Wireless Towards 5G mmWave Networks," J. Lightwave Technol. 35, 3945-3955 (2017).

[25] C. Lim et al., "Mitigation strategy for transmission impairments in millimeter-wave radio-over-fiber networks", Journal of Optical Networking, Vol.8, No.2, pp.201-214 (2009).

[26] X. Liu, H. Zeng, N. Chand and F. Effenberger, "Efficient Mobile Fronthaul via DSP-Based Channel Aggregation," in Journal of Lightwave Technology, vol. 34, no. 6, pp. 1556-1564, March, 15 2016.

[27] X. Liu, H. Zeng, N. Chand and F. Effenberger, "CPRI-compatible efficient mobile fronthaul transmission via equalized TDMA achieving 256 Gb/s CPRI-equivalent data rate in a single 10-GHz-bandwidth IM-DD channel," 2016 Optical Fiber Communications Conference and Exhibition (OFC), Anaheim, CA, 2016, pp. 1-3.

[28] Nurul Huda Mahmood et al, "A centralized inter-cell rank coordination mechanism for 5G systems", 13th International Wireless Communications and Mobile Computing Conference (2017).

[29] B.J.R. Sahu, S. Dash, N.Saxena, A. Roy, "Energy-Efficient BBU Allocation for Green C-RAN", IEEE Communications Letters, Vol.21, Issue 7, pp. 1637-1640 (2017).

[30] X. Zhang, M.Jia, L.Chen, J. Ma, J. Qiu, "Filtered-OFDM – Enabler for Flexible Waveform in the 5th Generation Cellular Networks", in Proc. of IEEE Global Communications Conference (GLOBECOM 2015), 6-10 December 2015, San Diego,CA, USA.

[31] C. Masterson, "Massive MIMO and Beamforming: The Signal Processing Behind the 5G Buzzwords", Analog Devices White Paper, June 2017

[32] P. Harris, et al. "Performance characterization of a real-time massive MIMO system with LOS mobile channels." IEEE Journal on Selected Areas in Communications 35.6 (2017): 1244-1253.

[33] M. Wu, B.Yin, G. Wang, C.Dick, J.R. Cavallaro, and C.Stuper, "Large-Scale MIMO Detection for 3GPP LTE: Algorithms and FPGA Implementations", IEEE Journal of Selected Topics in Signal Processing, Vol.8, Issue 5, pp. 916-929 (2014).

[34] T.H. Pham, S.A. Fahmy, and I.V. McLoughlin. "An End-to-End Multi-Standard OFDM Transceiver Architecture Using FPGA Partial Reconfiguration." IEEE Access 5 (2017): 21002-21015.

[35] S. Malkowsky et al., "The World's First Real-Time Testbed for Massive MIMO: Design, Implementation and Validation", IEEE Access, Vol.5, pp.9073-9088 (2017).

[36] R. Hui et al., "Subcarrier Multiplexing for High-Speed Optical Transmission", Journal of Lightwave Technology, Vol.20, No.3, pp.417-427 (2002).

[37] X. Liang, et al., "Digital suppression of both cross and inter-modulation distortion in multi-carrier RF photonic link with down-conversion", Optics Express, Vol.22, No.23, pp.28247-28255 (2014).

[38] 3rd Generation Partnership Project, "Technical Specification Group Radio Access Network – NR - Base Station (BS) radio transmission and reception", V15.0.0 (2017-12).