Optical Networks in Support of Open-RAN in 5G Systems and Beyond

M. Anastasopoulos(1), A. Pelekanou(1), A. Manolopoulos(1), A. Tzanakaki(1) and D. Simeonidou(2)

(1) National and Kapodistrian University of Athens, Athens, Greece, atzanakaki@phys.uoa.gr
(2) University of Bristol, Bristol, UK, dimitra.simeonidou@bristol.ac.uk

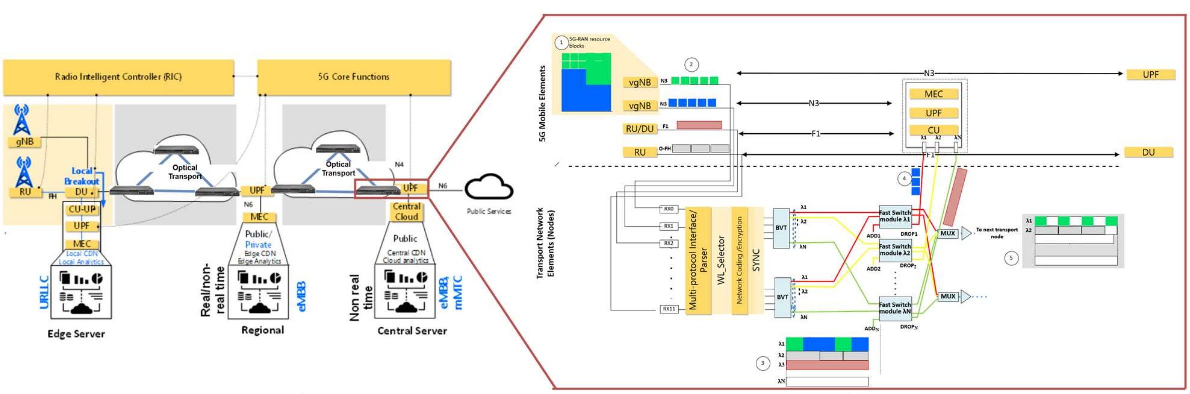
Abstract *This paper proposes an optical transport network suitable for 5G and beyond systems that is operated through a hybrid off-line optimisation and an on-line Machine Learning scheme executed at the Non-real time and the Real Time Radio Intelligent Controller respectively.*

Introduction

5G and beyond (B5G) systems are expected to operate in a highly heterogeneous environment supporting a large variety of applications with very stringent and greatly varying requirements in terms of bandwidth, latency, mobility and reliability [1]. These systems will need to flexibly interconnect a variety of network elements with general and specific purpose compute/storage resources. To achieve this, architectural and technological advancements such as hardware programmability and network softwarisation are needed. An example of a system that can be used to implement a variety of 5G and B5G deployment options is shown in Fig.1 These options combined can be used to support any service with highly variable Key Performance Indicators (KPIs) including Ultra Low Latency Communications (URLLC), massive Machine Type Communications) mMTC and enhanced Mobile Broadband) eMBB services.

In this environment, transport networks supporting jointly fronthaul (FH) and backhaul (BH) services need to satisfy a set of very stringent requirements in terms of bandwidth, delay and synchronisation. In this context, optical networking plays a key role due to its inherent high bandwidth as well as low latency and flexibility [2]. More specifically, optical transport nodes (Fig. 1), through appropriate control and management can address a wide range of transport network connectivity options such as point-to-point, point to multi-point and multi-point to multipoint. These can have increased requirements such as connectivity for Radio Access Networks (RANs) comprising advanced wireless access technologies i.e. massive MIMO, Reconfigurable Intelligent Surfaces etc as well as resilience of critical parts of the infrastructure against failures.

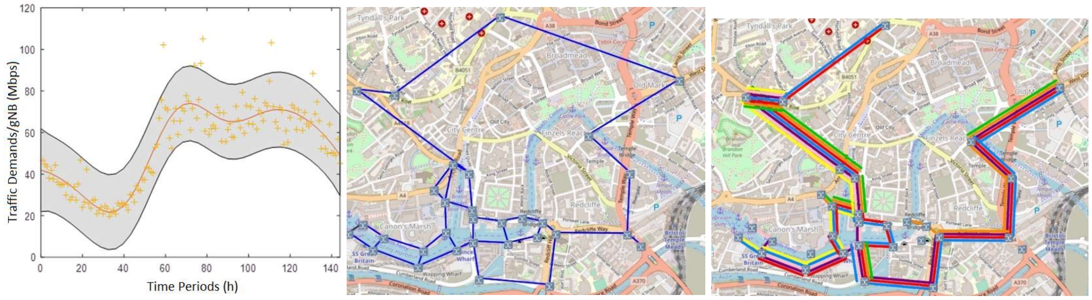
To achieve this, the optical transport nodes require a generic edge interface that enables seamless integration of any RAN technology with the transport network domain. This interface has to be programmable both at the transport protocol level and the network function level and should be able to host latency sensitive network functions. In addition, this interface needs to support both FH and BH services. FH services are supported through standardised protocols (e.g. CPRI, eCPRI) as well as the Open-RAN (O-RAN) FH (O-FH) and F1 stream (defined by NGFI-I) to facilitate the notion of BBU processing functional splits and the disaggregated RAN architecture [3]. In accordance to this architecture, the remote unit (RU), the Distributed Unit (DU) and the Central Unit (CU) of gNodeB (gNB) can be either collocated or placed at different physical locations. Key backhaul interfaces including N3, N6 and N9 will need to



a) b)

Fig 1: a) 5G Deployment Options including, the monolithic gNB deployment and the disaggregated gNB deployment, where the RAN function is split across different sites. b)TSON extension to support transport for 5G-RAN and 5G-CORE (UPF interfaces)

978-1-6654-3868-1/21/$31.00 ©2021 IEEE



a) b) c)

Fig 2 a) Estimation of Traffic Demands applying the GPR method, b) Bristol topology and c) optimal paths/wavelengths used for the interconnection of the edge nodes where are RUs are attached with the CUs

be also supported, while compatibility with the O-RAN radio Intelligent Controller (RIC) through its control and management capabilities will be also necessary.

This paper proposes an optical transport network solution that can support the requirements of most advanced and flexible 5G network architectures and protocols. The proposed solution is operated through the combination of an off-line optimisation scheme based on Integer Linear Programming (ILP) that is executed at the Non-Real Time (Non-RT) Radio Intelligent Controller (RIC) and a Machine Learning (ML) scheme that runs at the Real Time (RT) RIC.

Time Shared Optical Network

The proposed optical network solution offering these features is the Time Shared Optical Network (TSON) that has been proposed as a dynamic 5G transport network and was adopted as the future proof approach to support jointly FH and BH services by the 5G architecture vision of the 5G PPP Architecture Work Group [2]. TSON is an active WDM network providing high bandwidth and low latency connectivity in support of 5G services and beyond, through sub-wavelength switching granularity and elastic allocation of optical bandwidth. In this context TSON can support the most advanced O-RAN implementation options mentioned above. More specifically, in order to address the requirements of 5G service slices, TSON is able to allow allocation of network resources in the optical transport taking a two-stage approach exploiting the features of the Non-RT and RT RIC elements of O-RAN. The Non-RT RIC module is adopted to allocate network resources and establish the required connectivity through the assignment of optical network resources (wavelengths) in a deterministic manner. The RT-RIC module is used to take real time decisions by statistically multiplexing flows at a timeslot level. At a hardware level this is achieved utilizing fast optical switching integrated at TSON node [4], [5].

Model Description

To evaluate the performance of the architectural and technology approach described above and to propose a solution for the required control, the following model has been developed. We consider a 5G network, modelled as an undirected graph $G(\mathcal{N},\mathcal{E})$ where $\mathcal{N}$ is the set of nodes and $\mathcal{E}$ the set of links. This 5G system comprises both RAN and Core elements. The 5G-RAN segment comprises a set $\mathcal{R}$ of R gNBs used to provide connectivity services for a set $\mathcal{U}$ of U mobile users. gNBs are assumed to adopt the disaggregated architecture and the concept of functional split, according to which RUs, DUs and CUs are separated. In 5G networks, CUs and DUs can be implemented in software and run as Virtualized Network Functions (VNFs) in commercial off-the-shelf Mobile Edge Compute (MEC) servers. MEC servers also host Core elements (i.e. such as the User plane Function - UPF) necessary for the establishment and the provisioning of the end-user services (Fig 1). Traffic demands between the mobile elements are described through a demand vector $D_r = (r, d, t_{Sr}, t_{Fr}, \Delta_r)$ with r being the source nodes i.e. $r \in \mathcal{R}$, d the destination nodes i.e a MEC server. t_S denotes the earliest possible timeslot to schedule demand D_r and t_F the latest possible time slot for D_r. Δ_r is the time duration of the connection. The request can be scheduled starting with any start time slot $t_i \in [\, t_{Sr}, t_{Fr} - \Delta_r + 1]$ [6]. End-to-end connections can be successfully established if optical network resources (i.e., wavelengths, timeslots) are available across the end-to-end path. This problem can be formulated as an ILP model with the objective to identify the timeslots where these demands (slices) need be scheduled as well as the Routing and Wavelength Assignment (RWA) policy for each admitted request by maximizing a utility function considering both the network and compute elements. Given that the RWA problem is computationally intensive, we split it into simpler sub-problems. For the first sub-problem, initially we estimate the total number of demands to be served within a specific time frame. To forecast traffic demands per gNB we adopt the Gaussian Process Regression (GPR) method. Unlike supervised machine learning algorithms that can estimate specific values for every parameter in a function, the Bayesian approach determines the probability distribution function for these parameters (Fig.2a). Based on the

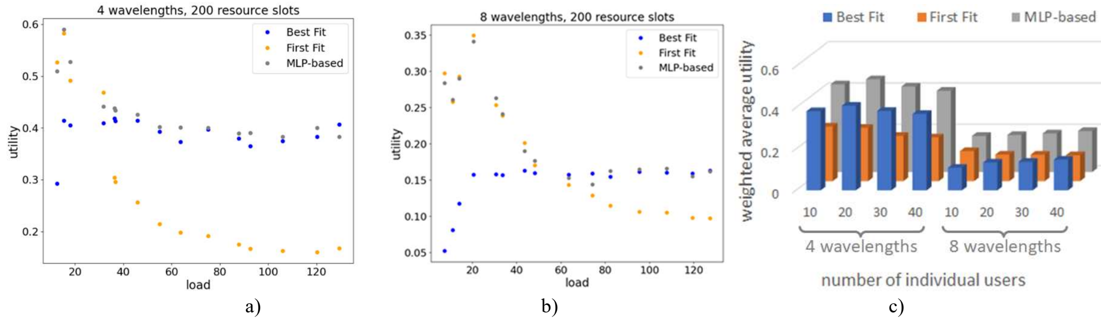


Fig 3: a) Utility function of the BF, FF and MLP-based scheduler for the 4x200 frame, b) Utility function of the BF, FF and MLP scheduler for the 8x200 frame, c) Weighted utility for the BF, FF and MLP scheduler for U {10, 20, 30, 40} and W {4, 8}.

forecasted demands, the Non-RT RIC controller solves the RWA wavelength per connection and identifies the optimal network path consisting of links and switches along which the flows are transferred from the gNB to the destination (i.e. UPF, MEC) Fig.2b shows the 5G-Bristol optical transport topology over which a 5G-RAN is deployed whereas an example of the output of the ILP allocating the wavelengths and establishing the end-to-end connections is shown in Fig. 2c.

In the second stage, once the end-to-end connections have been established, the RT-RIC controller solves the optimal scheduling problem at timeslot level adopting ML. This is implemented through a scheduler based on a Multilayer Perceptron (MLP)-based neural network (NN). The MLP periodically decides on the most efficient resource allocation strategies (i.e., selection of the appropriate heuristic to be onboarded in the TSON edge node) taking into consideration availability of resources, type and characteristics of services and loading conditions. Once the MLP has been executed it is able to identify the appropriate timeslot allocation strategy. The timeslot allocation policy scheme is continuously evaluated and if an alternative option has been recommended by the MLP, it is onboarded at the edge node. To keep the analysis tractable, we assume that resource scheduling in the network can be performed using either the Best Fit (BF) or the First Fit (FF) algorithm. The FF has been selected as it requires no global knowledge and performs well in terms of blocking probability and it is preferred in practice because of its small computational and communication overhead, and its low complexity. The BF scheme introduces additional communication overhead and has higher computational cost and complexity than the FF. However, it slightly outperforms FF in terms of resource utilization. The analysis can be easily extended to also include additional scheduling policies. We implement these heuristics with the following constraints: (a) The set of slots used by a user must be contiguous on one wavelength. (b) Multiple requests for a given service session must be served over the same wavelength. To train the MLP-NN to identify the optimal scheduling policy a simulation environment has been developed that evaluates the performance of FF-based and BF-based schedulers by comparing their utility functions. These results are then used to create the dataset used for training of the MLP NN. The model has been trained for a wide range of parameters i.e. number of input traffic statistics and network configuration parameters (number of input wavelengths and links per node, number of frames/timeslots etc).

Numerical Results

To quantify the benefits of the developed MLP-based resource scheduler, we repeated the previously described simulations. Specifically, we assumed two different frames with 4 and 8 wavelengths and 200 resource slots each. In addition, for each frame, we performed simulations for 10, 20, 30 and 40 individual users and the average offered load ranged between 5 and 150. For all simulation environments, the trained MLP periodically chose the most efficient resource allocation method, between FF and BF, taking into account the size of the frame, the number of users and the average load of the network. Then, the selected heuristic assigned the resources to users until the MLP decides again which heuristic should be used. The utilities of the three different resource schedulers were calculated after each simulation. In Fig. 3 we compare the utilities of the BF, FF and the MLP-based schedulers for the 4x200 and 8x200 frame case, respectively. The MLP selects the heuristic with the best performance, in terms of utility, depending on the corresponding load.

In Fig.3c, the three schedulers are compared in terms of weighted average utility (combining network and compute cost). By employing the MLP-based scheduler, the average utility is increased for every combination of network parameter compared with the case when either the BF or the FF are selelected.

Acknowledgements

This work has been financially supported by the 5G PPP EU projects 5G-COMPLETE and 5G-VICTORI.

References

[1] White Paper:“5G PPP use cases and performance evaluation models,” https://5g-ppp.eu/white-papers/, 2016

[2] A. Tzanakaki et al, “Wireless-Optical Network Convergence: Enabling the 5G Architecture to Support Operational and End-User Services”, IEEE Communications Magazine, August 2017.

[3] 3GPP TS 36.211, Evolved Universal Terrestrial Radio Access (E-UTRA); Physical channels and modulation

[4] H2020 5G-PPP, 5G-PICTURE. D3.1 Initial report on Data Plane Programmability and infrastructure components. https://www.5g-picture-project.eu/publication_deliverables.html

[5] A. F. Beldachi, A. Tzanakaki, R. Nejabati and D. Simeonidou, “Time shared optical network (TSON): A programmable network edge solution for multi-access support”, Applied Sciences (Switzerland), 2019, Iss. 9, vol. 22, 4786

[6] D. Andrei, H.-H. Yen, M. Tornatore, C. U. Martel and B. Mukherjee, "Integrated Provisioning of Sliding Scheduled Services Over WDM Optical Networks [Invited]," Journal of Optical Communications and Networking, vol. 1, no. 2, pp. A94-A105, July 2009.