Live Demonstration of Autonomous Link-Capacity Adjustment in Optical Metro-Aggregation Networks

Mihail Balanici, Pooyan Safari, Behnam Shariati, Aydin Jafari, Johannes Karl Fischer and Ronald Freund
Fraunhofer Institute for Telecommunications, Heinrich-Hertz-Institut, Einsteinufer 37, 10587 Berlin, Germany
Author e-mail address: mihail.balanici@hhi.fraunhofer.de

Abstract: We demonstrate a real-time ML-assisted network automation pipeline for dynamic, autonomous link-capacity allocation based on traffic-flow forecasting for optical metro aggregation networks. Its performance is compared to that of a classic, static bandwidth provisioning scheme.

1. Overview

The future 6G mobile technology is forecasted to bring a new era of Internet of Everything (IoE), in which things, humans, sensors, computing elements and processes, along with the robots and autonomous vehicles will all be interconnected within a converged physical-virtual ecosystem, generating Zettabytes of exchanged data [1]. As such, 6G will rely heavily on the underlying optical transport infrastructure to achieve high capacity, low latency, enhanced reconfigurability, and improved energy efficiency [1,2]. To satisfy all these requirements, optical networks will have to become intrinsically flexible, allocating channels dynamically on the fly and providing resources on demand. In traditional optical network architectures, the link-capacity is configured statically, remaining fixed over long periods of time. With fluctuating traffic patterns and flows, as well as constantly varying user demands, the static link-capacity provisioning often results in over- or under-provisioning of network resources, leading to suboptimal and inefficient network performance. With commercial availability and continuous development of tunable, flexible and programmable optical components, such as variable bit-rate coherent transceivers [3], the statically configured optical networks are slowly evolving into dynamic, reconfigurable and programmable infrastructures [4]. By enabling dynamic link-capacity allocation, optical networks can effectively adapt to changing traffic demands, ensuring that resources are allocated when and where they are needed, thus enhancing the network operation efficiency, maximizing the utilization of existing infrastructure, and reducing the resulting energy consumption.

One of the most effective ways to achieve autonomous, dynamic link-capacity allocation is by means of accurate forecasting of the future traffic behavior, namely the expected data rate and its variability over time. For that purpose, a large number of machine learning (ML) models for traffic-driven service provisioning has been previously proposed and investigated, and extensively reviewed [4]. A common aspect frequently seen in these works is that the high burstiness peculiar to real traffic flows is often neglected, resulting in traffic sets representing hourly-average data rates, when their actual traffic bursts surpass the average maxima [5]. This is a major reason why fine granular telemetry data with second or minute granularity is so essential in mitigation of under-provisioning risk [5-7]. Moreover, to the best of our knowledge, none of these works presents an actual demonstration of dynamic link-capacity allocation conducted on a real photonic testbed, but focus primarily on simulative analyses.

The main contributions of this work are three-fold: (1) we demonstrate a real-time ML-assisted control loop/pipeline for autonomous capacity provisioning on an optical metro-aggregation network testbed located at Fraunhofer HHI's premises in Berlin, (2) employ the novel Temporal Fusion Transformer (TFT) model [8] for traffic forecasting and subsequent dynamic capacity allocation, (3) and apply it on bursty, fine-granular traffic streams.

2. Innovation

In this work, we developed and deployed a dynamic capacity allocation control loop (Fig. 1 (a)) for autonomous link-capacity adjustment of an optical transponder, which is part of an optical metro-aggregation network (Fig. 1 (b)). The testbed consists of a hardware and a software layer of components, whose main interoperability is presented in a detailed workflow diagram (Fig. 1 (c)). The innovative aspect of this work is many-fold. First and foremost, although the concept of traffic-prediction-based dynamic capacity allocation is not new and has been researched in quite a few theoretical and simulative studies [4], to the best of our knowledge, none of them has attempted to demonstrate it live on a real optical testbed, comprising commercial or experimental hardware, as well as open-source and self-developed software tools. Secondly, we developed and deployed several Python-based tools, such as a customizable telemetry agent for telemetry data streaming with second-granularity, an online traffic forecasting module employing the latest ML-model (TFT) for time-series prediction, as well as a dynamic capacity allocation and a NETCONF-native network reconfiguration module that computes and reconfigures the optical transponder according to the expected traffic load. Lastly, we showcase the first prototype of the capacity adjustment control loop in real-time, and visualize its operation using specially customized Grafana dashboards (Fig. 2).

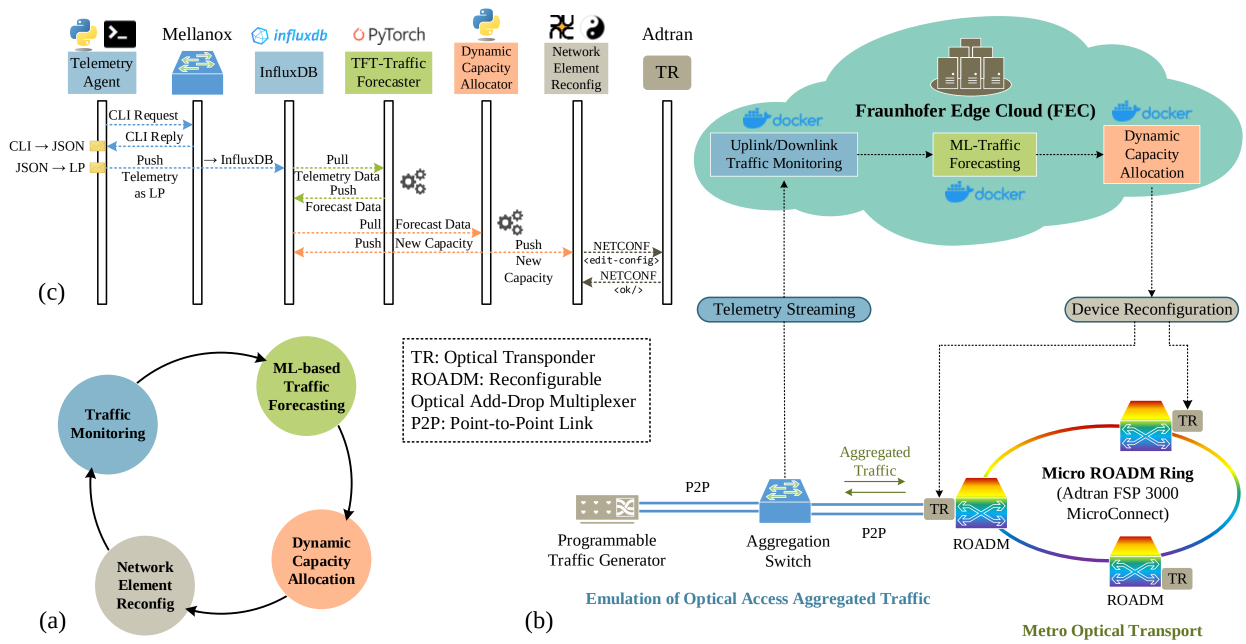

Fig. 1. (a) Demonstrated dynamic link-capacity allocation control loop. (b) The photonic testbed on which the demonstration is carried out, comprising the optical hardware layer, and the containerized control software layer running on a VM within the Fraunhofer Edge Cloud (FEC). (c) Detailed workflow of the dynamic link-capacity allocation/(re)configuration presenting all the involved hardware and software components.

3. OFC Relevance

Our demonstration presents an ML-assisted network automation pipeline for real-time capacity provisioning using a metro optical transport testbed. We use state-of-the-art open-source software tools (e.g., InfluxDB, PyTorch, TFT, ncclient) to build an autonomous, cloud-native network automation pipeline for commercially available hardware. Given the highest level of expertise of the attending academic and industrial communities, OFC is by far one of the best environments to showcase our current prototype, as its attendees represent a large part of our target audience.

4. Demo content & implementation

The underlying hardware layer of the photonic testbed comprises a programmable optical network tester (ONT)/traffic generator from VIAVI, connected through 2xQSFP28 transceivers to a Mellanox aggregation switch, generating an aggregated traffic flow consisting of multiple bursty and cyclic traffic streams with a cumulative data rate ranging between 0 to 200G (Fig. 2 (a), green curve). The aggregation switch is connected to an Adtran QuadFlex-Transponder through 2x100G client ports. The transponder is also connected through its two configurable line/network ports to an Adtran MicroConnect ROADM. The two network ports are NETCONF-native, with tunable DWDM line rates ranging between 100G (QPSK) – 150G (8QAM) – 200G (16QAM). Three pairs of transponder-ROADM form the optical metro-ring network testbed presented in Fig. 1 (b). Worth noting is that the ONT and aggregation switch are meant to emulate the infrastructure and aggregated traffic originating from a variety of optical access networks, such as XGS-PONs and passive WDM, which are currently being deployed to extend the photonic testbed.

The software layer of the shown testbed consists of a Docker-containerized capacity allocation pipeline, running on an Ubuntu VM within the Fraunhofer Edge Cloud (FEC). Its operation can be summarized with the closed control loop depicted in Fig. 1 (a), with the detailed operation presented in the workflow diagram (Fig. 1 (c)). The main operation steps of the control loop comprise: (1) fine granular traffic monitoring and telemetry reporting enabled by a dedicated and customizable telemetry framework written in Python that is querying the aggregation switch on the ingress and egress traffic rates; followed by (2) a real-time TFT-based traffic forecasting. The TFT is an attention-based deep neural network (DNN) optimized for multi-horizon (multi-step ahead) prediction – a feature particularly important for optical link-capacity adjustment, considering that the transponder needs up to 150 s for its different hardware settings such as modulation format, transmission rate, or optical frequency to be reconfigured to the new link-throughput. In this context, it is essential to emphasize that during the transponder reconfiguration phase, the traffic has to be rerouted at the aggregation switch level using load balancing and dedicated protection links, in order to prevent any traffic loss. This is a matter of current and future work, and will not be explicitly considered in this demo. Based on the forecasted traffic values, (3) the new capacity is computed as: $C(t_{i+n}) = \lceil \tilde{R}(t_{i+n})/\Delta C \rceil$, where

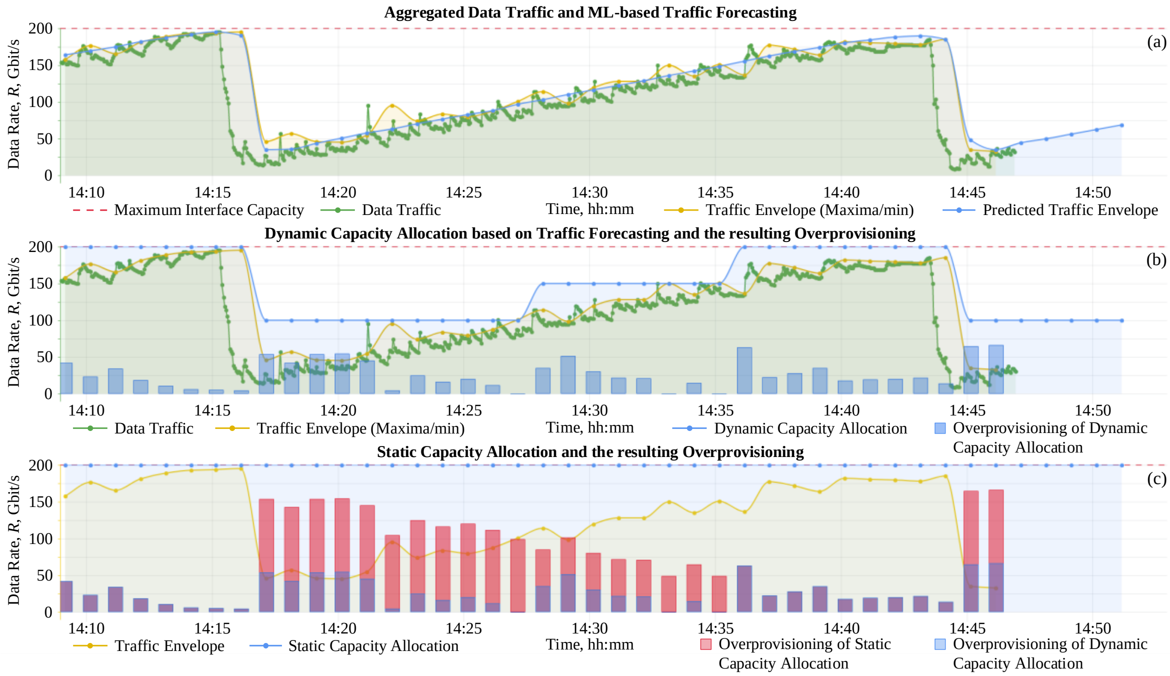

Fig. 2. Grafana dashboard for demonstration of dynamic link-capacity allocation: (a) Example of aggregated bursty and cyclic traffic with a 3s traffic sampling granularity, its traffic envelope, and predicted envelope/traffic evolution. (b) Dynamic link-capacity allocation and the resulting over-provisioning. (c) Static/fixed link-capacity at 200G, and the comparison of over-provisioning amounts for dynamic vs. static capacity allocation schemes.

$\tilde{R}(t_{i+n})$ is the n-th predicted traffic rate value relative to the current time instance t_i, and $\Delta C = 50$G is the transponder capacity adjustment step. Phase (3) is finalized by sending a NETCONF `<edit-config>` RPC-request, and eventually concluded by phase (4) the actual line port-capacity reconfiguration, confirmed by a NETCONF `<ok/>` RPC-reply.

A first implemented, operational example of the running pipeline is presented in Fig. 2. The generated traffic is monitored at the switch level, with a granularity of 3 s, and the traffic maxima are calculated each minute, resulting in a traffic envelope (Fig. 2 (a)). The TFT has been trained offline using a collected traffic envelope data, and then the trained model integrated into the online pipeline for real-time traffic forecasting. The forecaster predicts 5 traffic maxima ahead corresponding to the future 5 minutes, based on the last 30 maxima (i.e, past 30 min). The traffic forecaster is able to correctly predict the overall traffic evolution, however it still presents a challenge to correctly foresee some of the occurring traffic bursts, as these normally have a random nature specific to traffic outliers. Based on the forecasted traffic envelope (Fig. 2 (a)), the new required capacity is computed as presented above, with the over-provisioning value calculated every minute (Fig. 2 (b)). For visual performance comparison, Fig. 2 (c) presents the over-provisioning values of a dynamic (blue bars) versus a traditional static (red bars) link-capacity allocation.

Note that the showcased traffic profile, increasing ramp-wise over time, is merely used to present the concept and pipeline operation clearly, when the capacity levels change. In the actual demo, more realistic traffic profiles will be deployed, such as the real, bursty Gaussian bell-shaped profile to emulate the daily tidal effect of traffic variation [5].

5. Acknowledgements

This work was partly funded by the German Ministry of Education and Research (BMBF) in the framework of the project 6G-RIC (FKZ 16KISK020K) and GreenICT@FMD (16ME0495). The authors would also like to thank Jasper Müller and Dr.-Ing. Achim Autenrieth from Adtran for their valuable inputs and guidance related to this demo.

6. References

[1] 5G Infrastructure Association, "European vision for the 6G network ecosystem," White Paper, pp. 1-46, 2021.
[2] C. Ranaweera, *et. al.*, "4G to 6G: disruptions and drivers for optical access," *J. Opt. Commun. Netw.*, vol. 14, no. 2, pp. A143-A153, 2022.
[3] D. Welch, *et. al.*, "Digital Subcarrier Multiplexing: Enabling Software-Configurable Optical Networks," *J. Lightw. Technol.*, vol. 41, 2023.
[4] T. Panayiotou, M. Michalopoulou and G. Ellinas, "Survey on Machine Learning for Traffic-Driven Service Provisioning in Optical Networks," *IEEE Commun. Surv. Tutorials*, vol. 25, no. 2, pp. 1412-1443, 2023.
[5] B. Shariati, *et. al.*, "Traffic Monitoring and Analytics Framework for Optical Access Networks," *Proc. CSNDSP*, pp. 852-856, 2022.
[6] M. Balanici, G. Bergk, P. Safari, B. Shariati, J. K. Fischer and R. Freund, "Demonstration of a Real-Time ML Pipeline for Traffic Forecasting in AI-Assisted F5G Optical Access Networks," *European Conf. on Optical Commun.*, Tu2.5, pp. 1–4, 2022.
[7] B. Shariati, *et. al.*, "Telemetry Framework with Data Sovereignty Features," *Opt. Fiber Commun. Conf. (OFC)*, M3G.2, pp. 1-3, 2023.
[8] B. Lim, *et. al.*, "Temporal Fusion Transformers for interpretable multi-horizon time series forecasting," *Int. J. Forecasting*, vol. 37, 2021.