Intent based AI model in packet-optical networks towards 6G

Paola Iovanna(1), Marzio Puleri(1), Giulio Bottari(1), Fabio Cavaliere(1)

(1) Ericsson Research, Pisa, Italy, paola.iovanna@ericsson.com

Abstract *This paper presents opportunities and challenges to apply an intent-based AI model to a packet-optical transport in the path to 6G. A method is reported with related assessment.*

1 - Introduction

As the demand for network and service capacity rapidly increases with the advent of 5G and moving towards 6G, new challenges arise in managing a multitude of vertical services with guaranteed QoS. To reduce infrastructure costs, networks are evolving into intent-driven networks [1], introducing a higher level of automation and optimization. The network should dynamically adapt to new situations and changing needs autonomously and automatically, limiting human intervention to the definition of high-level system needs that the network will translate into requirements and actions to ensure a good match over time. By continuous monitoring, the network identifies intent violations to be repaired with a set of actions. This approach poses further challenges to the transport network, which plays an increasingly important role in the evolution towards 6G. In [2], the complexity to expose a suitable abstraction of the transport is reported, while in [3][4] a smart E2E platform targeting network and computing self-optimization to provide the committed Quality of Experience (QoE) to intelligent applications is proposed. Some work has been done and is on-going in defining interfaces for transport solution in an intent-based system [5]. Anyway, there are further critical aspects such as dynamic re-configurability and monitoring to be considered when intent-based approach is applied to a transport based on optical technology that can play a relevant role thanks to its capability to ensure low latency, high bandwidth, and low power consumption. In this context, this paper proposes an intent-based method to guarantee QoS for vertical services spanning a geographical area, considering opportunity and challenges when the method is applied to optical technology. The paper discusses the mapping of high-level services into transport requirements for vertical services, relevant optical technology for meeting 5G to 6G requirements in the access area, and a model for intent-based applied to packet-optical networks. Simulation results are presented to assess the performance of the proposed method. Session 2 describes examples of the high-level services mapping in transport requirements for vertical services. Session 3 provides a summary of the optical technology relevant to meet 5GB to 6G requirements in the access area; Session 4 describe a model for Intent based applied to a packet-optical networks that has been assessed by simulation and a Proof of Concept. Session 5 reports some simulation results to assess performances of the method. Sessions 6 provides the conclusions.

2 – Reference scenario

According to [4], a service must be available in a specific deployment area, such as local or geographical. For each service, multiple transport connections must be automatically established and managed to ensure that the service requirements are met. Each service is composed of several traffic flows characterized by different sources-destination pairs in the transport domain, which can also include mobility. A service can be an aggregate of final UEs, such as sensors on a factory production line or sensors on patients in an ambulance. In most cases, the granularity of each flow is lower than the granularity of a wavelength. Therefore, a suitable interworking of optical with packet technology is a viable solution, at least at the edge point of the transport network. The packet-optical approach combines the flexibility of packet technology with the high bandwidth and low latency of optical technology.

Figure 1 presents an exemplary packet-optical network based on packet switches (PS) coupled with optical switches (OS).

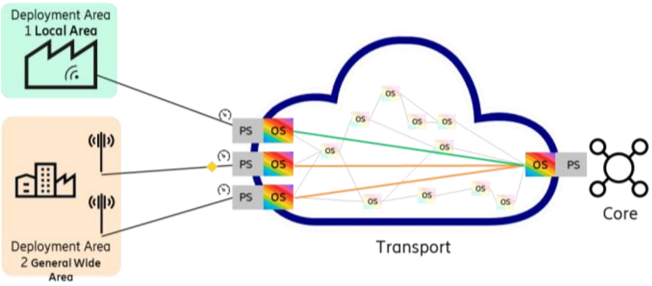

Figure 1 – Exemplary packet-optical network

Two services are deployed on the network and mapped mapped on the corresponding transport resources interconnecting RAN/CN functions:

- A service dedicated to smart manufacturing applications in a local area (indoor or campus)
- A service dedicated to covering applications in an urban scenario (general wide area)

The first service requires a single transport connection (shown in green in the figure) to connect the RAN and CN functions, while the second service, being deployed across a wider geographic area, needs two transport connections between the RAN and CN. If the transport network uses a combination of packet and optical technologies, as in this example, E2E optical paths are used to offload packet traffic to the optical layer and guarantee low and deterministic latency. Additionally, in a cloud-based scenario, transport connections should be dynamically adapted to the processing load on the servers where the network functions run. Optical switching is needed to move aggregated packet traffic from one server location to another, which presents a challenge in reconciling the QoS requirements (such as bandwidth and latency) with the reconfiguration times of the optical network nodes.

One key aspect to consider in the intent-based model is the slot time at which node reconfigurations occur. The reconfigurability time in the optical layer should be slower than the traffic variation for scalability and stability reasons. Optimization actions should avoid continuous traffic redirection in alternatives paths requiring continuous re-configuration of the optical layer. In the following section, we present a reference model for AI intent-based model in packet-optical transport.

3 - Optical features in intent-based AI model

According to the intent-based paradigm, an intent is a declarative information object that defines the requirements that an autonomous system and infrastructure are expected to fulfil [6]. Any intent should be automatically translated into transport requirements. To achieve this, it should be possible to dynamically reconfigure the optical nodes to adapt to current traffic or service needs. The combination of packet and optical switching can achieve this goal by providing the necessary dynamism without over-provisioning the packet switches [7]. The Innovative Optical and Wireless Network (IOWN) global forum [8] has analysed the gaps that need to be addressed to move towards an optical transport network that can be dynamically configured end-to-end, even when crossing different optical vendors and technologies. For instance, expanding the use of Wavelength Division Multiplexing (WDM) as an aggregation technology from the inner segments of the optical network to the access ones represents a major cost challenge. Reconfigurable Optical Add Drop Multiplexers (ROADMs) based on Wavelength Selective Switches (WSS) are an essential ingredient of today optical communication networks [9] but their cost is not affordable for an access network. An example of cost-effective small-size ROADM (referred as Mini ROADM in the rest of this paper), suitable for an access network covering distances up a few tens kilometres, is reported in [10][11][12]. It is based on silicon photonics micro rings that can be integrated with III-V Semiconductor Optical Amplifiers (SOA) to realize lossless devices [13]. Silicon micro-rings can also be used in tuneable optical filters at the receiver of pluggable transceivers [12], in radio units, digital units or packet switches and routers. Used in combination with tuneable transmitters, like the ones specified in the Recommendation ITU-T G.698.4 [14], these filters enable full tuneable transceivers that, with proper protocols for negotiating the transmitting and receiving wavelengths, are self-configurable. This enables the programmability of the optical layer in centralized RAN architectures based on power splitters, like the ones that reuse the optical infrastructure of Passive Optical Networks (PON) [15] or other variants of broadcast-and-select architectures [16]. Current G.698.4 specifies 25 Gbit/s tuneable transceivers that can be implemented in a small form-factor. Increasing the capacity to 100 Gbit/s, and beyond, coherent transceivers become an option to consider. Coherent PON [15] is under study at ITU-T Study Group 15 and the compatibility of the Mini ROADM with 100 Gbit/s coherent transceivers has been demonstrated in [11]. But no programmability would be possible without monitoring protocols that can provide continuous feedback to offer immediate notifications to prevent failures or performance degradation. Monitoring techniques at the optical layers are specified in Recommendation ITU-T G.697 [17] and optical transceivers enable remote monitoring capabilities are reported in [18].

4 – Example of intent-based AI model in packet-optical network

Considering the transport characteristics and the key principles of an intent-based architecture, an AI intent-based model system for transport has been defined. An intent serves as the input of the system, representing a service with related geographical area and a list of QoS requirements (peak rate, average rate, latency, burstiness) that must be guaranteed with resource optimization. Next, the system defines a set of concurrent traffic flows in the transport network associated with the intent, which should be managed considering changes in the aggregated bandwidth of each flow over time. Finally, the system acts on the transport connections in accordance with the model depicted in Figure 2.

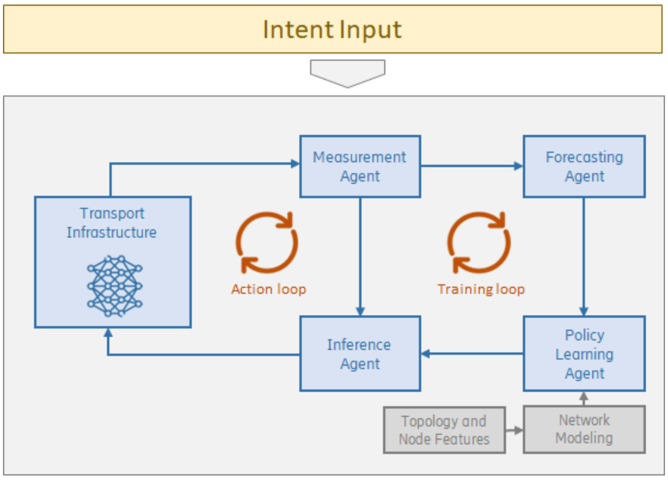

Figure 2 – Scheme of the AI Intent-based model

Although the model is general, this section refers to a scenario where each node is composed of a packet and an optical node, such as a ROADM. At the edge of the network, the packet switches aggregate and tag the radio traffic entering the transport network, representing the traffic flows with source-destination addresses. The use of packet switches at the edge enables traffic management for flows and maintains flow identifiers in the mapping on the optical connections throughout the service lifecycle. The network model represents the ROADM with its switching capability and DWDM transmissions, including the number of ROADM ways and the links to the packet nodes where they are present. Latency parameters are associated with the node to facilitate traffic off-loading from packet to optical. The quality of transmission (QoT) parameters (e.g., transmission power, OSNR, Bit error rate) are evaluated statically and included in the modelling.

The model is based on two loops: the “Training loop” and the “Action loop”.

Training Loop

The training loop is used to learn the policy for deciding where to route traffic in an optimal way in the future. The policy is designed to minimize the transition between optical and packet switches, the load on the ports of the packet switches, and the number of re-routings over time. The policy is periodically updated to follow the evolution of traffic behaviour. Historical data captured by the measurement agent is used by the forecasting agent to estimate, using statistical approaches, the possible evolution of the traffic. The measurement agent captures data for each traffic flow before it enters the transport network and evaluates its peak rate, average rate, and burstiness. Typical AI-based intent systems rely on historical data, but in this proposed approach, traffic forecasting is considered to avoid acquisition of very large data sets for learning. This choice also helps in reducing the representation space used for learning the optimal strategy and speeds up the learning phase, so that a new learning phase can occur during the usage of the policy learnt in the previous step. The policy learning agent is an Adaptive Markov Decision Process (AMDP) [19][20] that operates in a fully observable environment, where all data flows and queuing in the switches are monitored at runtime, providing a complete image of the state of the network. Given the network modelling that exposes the wavelengths and node features, the AMDP aims at maximizing the continuity of the wavelength between the source and destination nodes and, reducing the transition over packet nodes, to maximize offload from packet to optical. The reward rule used by the policy learning agent also tries to minimize the load of the packet nodes and the number of re-routing over time, while ensuring latency for each service and keeping the node load level under control. The learnt policy is passed cyclically to the action loop.

Action Loop

The action loop encompasses the transport infrastructure, measurement agent, and inference agent. In this case, the measurement agent provides the current traffic parameters for each ingress flow to the inference agent, that applies the learnt policy at runtime to decide on the optimal route configuration. As the policy learning agent is time-varying, so is the inference agent. This means that the action taken at two different instants having the same input traffic condition will differ to account for the evolution of the network status. Such behaviour helps in limiting node reconfiguration due to a reaction to a particular situation, while promoting system stability.

5 – Assessment results

The behaviour of the two loops was tested through simulation and implementation of a proof-of-concept (PoC) that includes commercial packet switches and two-way ROADM realized on integrated silicon photonics. The simulations address several scenarios, including mesh networks. The stimuli were generated modelling real measured traffic traces related to mobile broadband traffic (MBB) and industrial traffic (IndT). The MBB traffic exhibits a circadian behaviour, while the IndT is usually a constant bit rate (CBR). Since MBB traffic is more challenging for the network, we primarily focused on it. The simulation was configured to allow a load level up to 85% of packet nodes with a forecasting period of 72 hours. The simulated flows were 120.

Figure 3 shows the achieved load levels per packet node and output port, as well as the category over time, using the policy learnt during the forecasted period. Packet nodes load level was usually below 85% as desired. The learnt policy, offloading packet nodes, also minimized the transition between optical and packets nodes, maximizing the usage of optical nodes. Additionally, re-routing was minimized, with typically only a single re-routing occurring at the beginning, and a second re-routing occurring in the middle of the simulation only in very overloaded conditions.

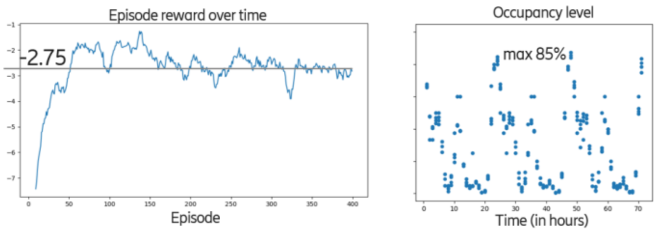

Figure 3 - Example of training

6 – Conclusions

This paper introduces an intent-based AI model to manage services with guaranteed QoS in a packet-optical transport system in the path to 6G. New transmission and switching technologies, are emerging to enable automation mechanisms such as intent-based approaches. The model's feasibility and high performance have been demonstrated in a packet-optical exemplary network. Future study should address open points especially in optical technology, such as monitoring, which should be granular enough to provide continuous feedback through suitable protocols. This also involves the dynamic translation of QoS parameters into QoT.

References

[1] J. Niemöller, R. Szabó, A. Zahemszky and D. Roeland, "Creating autonomous networks with intent-based closed loops," in Ericsson Technology Review, vol. 2022, no. 4, pp. 2-11, April 2022, doi: 10.23919/ETR.2022.9904673.

[2] A. Campanella, "Intent Based Network Operations," 2019 Optical Fiber Communications Conference and Exhibition (OFC), San Diego, CA, USA, 2019, pp. 1-3.

[3] L. Velasco et al., "End-to-End Intent-Based Networking," in IEEE Communications Magazine, vol. 59, no. 10, pp. 106-112, October 2021, doi: 10.1109/MCOM.101.2100141.

[4] P. Iovanna et al., "End-to-end network slicing orchestration – A Key Enabler for Industry-Vertical use Cases," in Ericsson Technology Review, vol. 2022, no. 2, pp. 2-10, March 2022, doi: 10.23919/ETR.2022.9904695.

[5] P. Sköldström, S. Junique, A. Ghafoor, A. Marsico and D. Siracusa, "DISMI - an intent interface for application-centric transport network services," 2017 19th International Conference on Transparent Optical Networks (ICTON), Girona, Spain, 2017, pp. 1-4, doi: 10.1109/ICTON.2017.8025180.

[6] TM Forum Autonomous Network Project, IG1253 Intent in Autonomous Networks v1.1.0, January 2022

[7] R. Sabella, F. Testa, P. Iovanna, and G. Bottari, "Flexible packet-optical integration in the cloud age: Challenges and opportunities for network delayering," in IEEE Communications Magazine, vol. 52, no. 1, pp. 35-43, January 2014, doi: 10.1109/MCOM.2014.6710062.

[8] https://iowngf.org/

[9] Dan M. Marom, Paul D. Colbourne, Antonio D'Errico, Nicolas K. Fontaine, Yuichiro Ikuma, Roberto Proietti, Liangjia Zong, José M. Rivas-Moscoso, and Ioannis Tomkos, "Survey of Photonic Switching Architectures and Technologies in Support of Spatially and Spectrally Flexible Optical Networking [Invited]," J. Opt. Commun. Netw. 9, 1-26 (2017)

[10] P. Iovanna et al., "Optical Components for Transport Network Enabling The Path to 6G," in Journal of Lightwave Technology, vol. 40, no. 2, pp. 527-537, 15 Jan.15, 2022, doi: 10.1109/JLT.2021.3117122.

[11] Iovanna P, Bigongiari A, Bianchi A, Menezo S, Romagnoli M, Sorianello V, Cavaliere F. Optical Technology for NFV Converged Networks. Applied Sciences. 2021; 11(4):1522. https://doi.org/10.3390/app11041522

[12] Sorianello, V.;De Angelis, G.; Cassese, T.;Preite, M.V.; Velha, P.; Bianchi, A.; Romagnoli, M;Testa, F. Polarization insensitive silico n photonic ROADM withselectable communication direction for ra dio access networks. Opt. Lett. 2016, 41, 5688–5691

[13] P. Nadimi Goki, M. Imran, F. Fresi, F. Cavaliere and L. Potì, "Lossless ROADM by Exploiting low gain SOAs in fronthaul network," 2019 24th OptoElectronics and Communications Conference (OECC) and 2019 International Conference on Photonics in Switching and Computing (PSC), Fukuoka, Japan, 2019, pp. 1-3, doi: 10.23919/PS.2019.8817763.

[14] Recommendation ITU-T G.698.4, https://www.itu.int/rec/T-REC-G.698.4/en

[15] Junwen Zhang, Guoqiang Li, Sizhe Xing, Nan Chi, Flexible and adaptive coherent PON for next-generation optical access network [Invited], Optical Fiber Technology, Volume 75, 2023, 103190, ISSN 1068-5200,, https://doi.org/10.1016/j.yofte.2022.103190.

[16] J. -K. Rhee, I. Tomkos and Ming-Jun Li, "A broadcast-and-select OADM optical network with dedicated optical-channel protection," in Journal of Lightwave Technology, vol. 21, no. 1, pp. 25-31, Jan. 2003, doi: 10.1109/JLT.2003.808638.

[17] Recommendation ITU-T G.697, https://www.itu.int/rec/T-REC-G.697/en

[18] Mobile Optical Pluggables Alliance (MOPA) Technical paper Version 2.1, March 9, 2023, https://mopa-alliance.org/wp-content/uploads/2023/03/MOPA_Technical_Paper-v2.1-Final.pdf

[19] Sutton R.S., Andrew G. Barto A.G.: "Reinforcement Learning: An Introduction", 2nd Ed., MIT Press, Nov. 2018

[20] Liu L., et al.: "A Solution to Time-Varying Markov Decision Processes", Jan 2018, https://arxiv.org/abs/1605.01018