Autonomous Intent-driven Optimization of PONs: A Vendor-agnostic DRL Demonstration

L. Inglés[1,2,3,*], L. Anet Neto[1,3], C. Rattaro[2], M. Morvan[1,3], A. Castro[2], L. Nuaymi[1]

[1] *IMT Atlantique, 655 Av. du Technople, 29280 Plouzané, France*
[2] *Universidad de la República, 565 Av. Julio Herrera y Reissig, 11300 Montevideo, Uruguay*
[3] *Lab-STICC CNRS UMR 6285, Technopôle Brest-Iroise - CS 83818, 29238 Brest Cedex 3, France*

[*] *lucas.ingles-loggia@imt-atlantique.fr*

Abstract: We demonstrate a Deep Reinforcement Learning application for autonomous T-CONT configuration in passive optical networks. Using DDQN, we achieve vendor-agnostic dynamic upstream-latency optimization in a fixed–mobile convergence use case.

1. Overview

Passive Optical Networks (PONs) have become the dominant solution for last-mile broadband access due to their cost-efficiency. This is mainly thanks to its point-to-multipoint (PtMP) architecture, which allows operators to mutualize infrastructure deployment costs between several clients. Standardization efforts have also been a key to its massive deployment by promoting multi-vendor interoperability between operator and client-side equipment [1].

However, PtMP topologies introduce an inherent challenge to network operators: multiple end-users must share the same upstream transmission medium (feeder fiber) and channel (wavelength). Access to such shared resource is managed by the Optical Line Terminal (OLT), which allocates transmission time windows to the users or Optical Network Units (ONU) through a Dynamic Bandwidth Allocation (DBA) mechanism. Network performances indicators such as latency, bit-rate and packet jitter are thus extremely dependent on the nature of user traffic within a PON tree (statistically multiplexed in its essence) but also on the performances of DBA algorithms. Also, even if interoperability is enforced by the standards, the actual details of the DBA algorithms are vendor-specific, forcing operators to often rely on blunt, static configurations of PON transmission containers (T-CONTs) for each service. In practice, this means that despite the benefits of DBA over fixed bandwidth allocation, a degree of network over-provisioning persists, driven by simplistic, "set-and-forget" T-CONT engineering rules.

In this context, the rapid evolution of artificial intelligence (AI) and machine learning (ML) has extended to optical access networks, motivating the exploration of learning-based techniques for more efficient bandwidth allocation solutions. Among these, Deep Reinforcement Learning (DRL) has emerged as a powerful framework for enhancing network performance and autonomy. By enabling agents to learn optimal control policies through continuous interaction with their environment, DRL facilitates automated configuration and proactive optimization in complex networking scenarios. But even if its use in optical access networks represents a significant step toward intelligent, self-optimizing infrastructures capable of meeting diverse and dynamic service demands, it has been mainly addressed from a vendor perspective, by means of changes in the DBA algorithms themselves [2–4].

Finally, a central challenge in applying DRL to operational networks, which is often neglected in recent publications, is the convergence time and the risk of service degradation during the agent's training phase. Different DRL algorithms address such exploration-exploitation trade-off in various ways, with learning stability being a key selection criterion for production environments. Algorithms like the Double Deep Q-Network (DDQN) [5], for instance, mitigate value overestimation by decoupling action selection from evaluation, leading to more stable and reliable convergence, a particularly desirable property in network control applications.

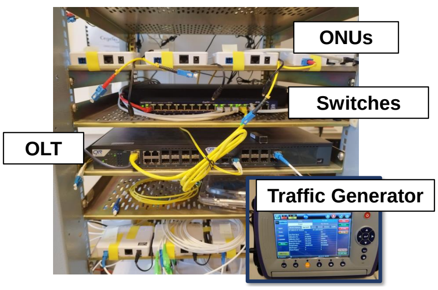

Fig. 1: Commercial GPON testbed with OLT, 16 ONUs, optical splitters, the traffic generator and switches for control and data planes.

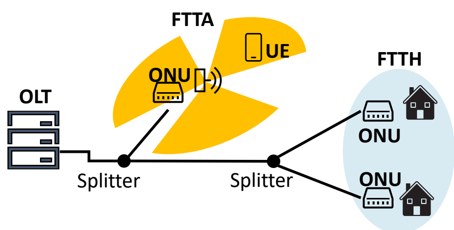

Fig. 2: Fixed-Mobile Convergence over PON. Point-to-multipoint architecture sharing fiber infrastructure between fixed and mobile users.

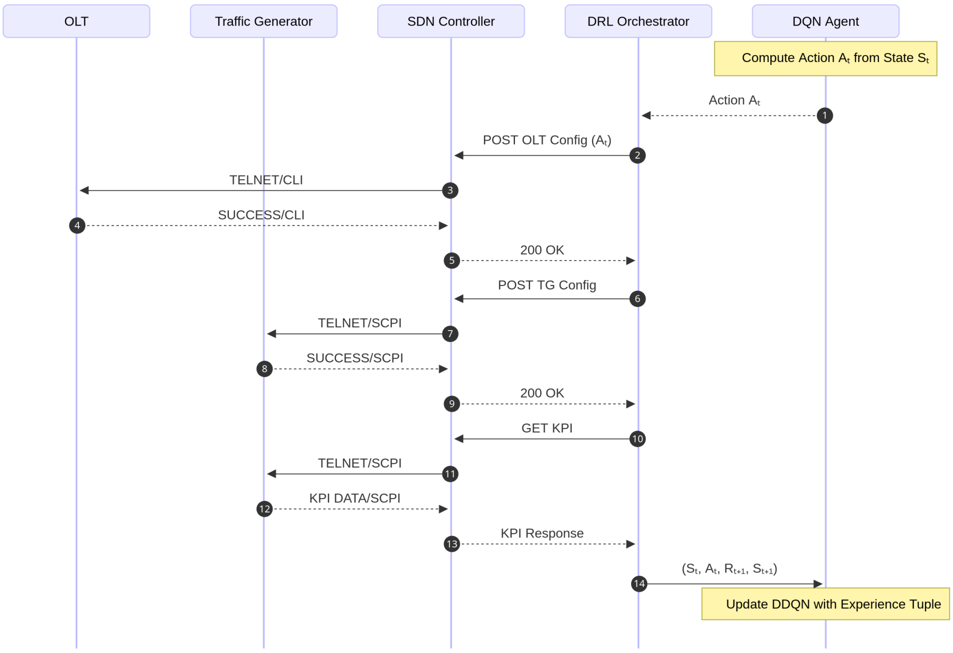

Fig. 3: Agent-to-network interaction flow for real-time PON reconfiguration.

2. Innovation

The opacity of vendor-specific DBA forces operators into static, conservative T-CONT configurations that cannot adapt to dynamic traffic, leading to over-provisioning and sub-optimal resource utilization.

We demonstrate a vendor-agnostic approach that optimizes T-CONT configurations without modifying proprietary DBA logic. Using a DDQN agent interfaced with a commercial G-PON (see Fig. 1) via SDN/RESTCONF, we enable autonomous, intent-driven reconfigurations. We assume a fixed-mobile convergence scenario as depicted in Fig. 2, with one service per type of ONU. One ONU emulates a mobile site (FTTA) with T-CONT type 1, while the remaining fifteen emulate residential users (FTTH) with T-CONT type 3. Different traffic profiles emulate realistic fixed and mobile demand patterns [6], including diurnal variations and distinct service requirements. The agent observes the network state (including normalized FIR/PIR values, traffic demand, congestion flags, and time-of-day features) and selects actions to adjust FIR (T-CONT 1) and PIR (T-CONT 3) parameters. Learning is driven by a reward function that heavily penalizes QoS violations (congestion events for both FTTA and FTTH services) while encouraging resource efficiency through moderate penalties on allocated bandwidth. This design prioritizes service quality while discouraging over-provisioning.

Fig. 3 illustrates the complete control loop of our system. At each optimization step, the DDQN agent computes an action $A_t \in \{\text{increase, decrease, maintain}\}$ for FIR and PIR parameters based on the current network state S_t. The DRL Orchestrator translates this action into vendor-specific commands and applies them to the OLT via the SDN Controller using RESTCONF/YANG. Simultaneously, the orchestrator configures the traffic generator to emulate the corresponding traffic patterns. Once configured, the system uses the SDN controller to gather Key Performance Indicators (KPIs) from all traffic flows, including latency, throughput, and congestion metrics. These measurements, combined with the applied action, form an experience tuple $(S_t, A_t, R_{t+1}, S_{t+1})$ that is fed back to the DDQN agent for continuous learning and policy refinement. This closed-loop operation enables autonomous, real-time adaptation to varying traffic demands without human intervention.

3. OFC Relevance

This work demonstrates a vendor-agnostic method to optimize Passive Optical Networks (PONs) using a commercial SDN-enabled OLT and realistic traffic. We showcase adaptive T-CONT configuration via Deep Reinforcement Learning (DRL) in a live network, offering a path toward autonomous, intent-driven operation. The approach is applicable to any TDM/TDMA-PON management plane with minimal performance impact.

4. Demo Content & Implementation

Our live demonstration is structured into two main phases designed to showcase the limitations of static configurations and the capabilities of our DRL-based solution, while explaining the complete optimization pipeline.

The first phase demonstrates that no static T-CONT configuration can consistently meet QoS requirements under the dynamic traffic patterns generated. Through interactive experiments, attendees will be able to act as network operators, manually adjusting T-CONT parameters and directly observing the impact of their decisions on key performance indicators. This phase intentionally presents a simplified version of a real operational scenario; in practical deployments, the consequences of relying on static configurations (such as resource misallocation, service degradation, or QoS violations) are even more pronounced than those shown here.

The second phase showcases our complete DDQN-based optimization pipeline performing autonomous real-time resource allocation. We will explain the full control loop: from network state observation and feature extraction to the DDQN agent's decision-making process and subsequent T-CONT reconfiguration. The demonstration will highlight the agent's ability to dynamically adjust FIR and PIR parameters throughout the day, maintaining latency below 1 ms for mobile traffic while optimizing bandwidth utilization without human intervention.

The demonstration will be conducted entirely via a remote connection to the operational testbed in our laboratory. Although no physical equipment will be on site, attendees will interact directly with the system through a dedicated visualization interface. This includes live traffic generation monitoring and a comprehensive Grafana dashboard (Fig. 4) displaying real-time metrics: DRL reward, round-trip delay and traffic load per step; evolution of PIR and FIR values; episode-based reward progression; current FIR, PIR, step, episode, and DRL exploration rate values; plus congestion indicators and action history. This setup allows full observation of system behavior across both demonstration phases. It is important to note that the DDQN agent will be pre-trained prior to the demonstration and will only perform policy refinement in reaction to different traffic profiles during the live demo. This pre-training is necessary given the limited time available for the demo and for interacting with the audience. The training performance of our DRL agent and its comparison with other DRL algorithms have been reported in a regular submission to subcommittee N4.

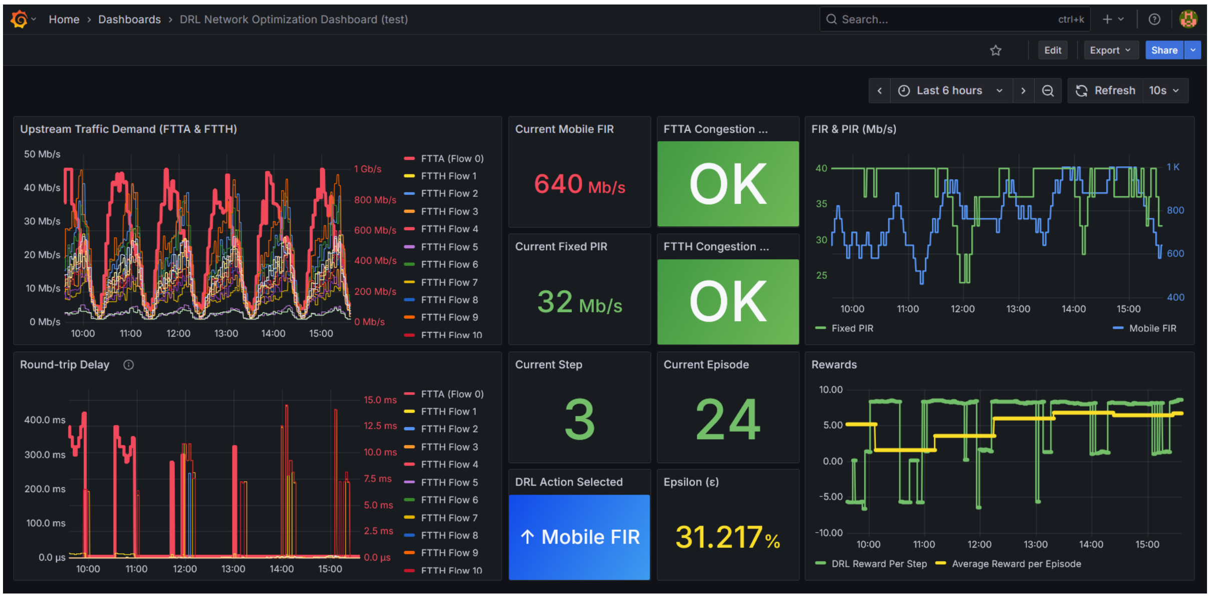

Fig. 4: Graphical user interface of the proposed PON optimization application (during DRL training phase).

Acknowledgements

This work was funded and supported by the DGE IPCEI ME/CT Orange project as well as the Uruguayan CAP (Comisión Académica de Posgrados). It was also conducted in the framework of the Lab'Optic, a joint research initiative between the UMR-6285 Lab-STICC, the UMR 6082 Foton Institute and Orange Innovation.

References

1. International Telecommunication Union, "Recommendation ITU-T G.984.3: Gigabit-capable Passive Optical Networks (GPON): Transmission convergence layer specification," ITU-T Study Group 15, Recommendation G.984.3, Jan. 2014.
2. E. Wong et al., "Machine learning enhanced next-generation optical access networks—Challenges and emerging solutions [Invited Tutorial]," *J. Opt. Commun. Netw.*, 2023.
3. L. Hu et al., "APP-aware MAC scheduling for MEC-Backed TDM-PON mobile fronthaul," *IEEE Commun. Lett.*, 2023.
4. L. Ruan and E. Wong, "Addressing concept drift of dynamic traffic environments through rapid and self-adaptive bandwidth allocation," in *ICCCN*, 2023.
5. H. Van Hasselt et al., "Deep reinforcement learning with double Q-learning," in *AAAI conference on artificial intelligence*, 2016.
6. M. Kihl et al., "Traffic analysis and characterization of Internet user behavior," in *ICUMT*, 2010.