AI Learns G-PON: Toward Adaptive T-CONT Configuration for Fixed-mobile Convergence

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 the use of a commercial G-PON as an AI-enhanced self-optimizing substrate. We refine the T-CONT configuration with deep reinforcement learning to optimize transmission latency and ensure stable and adaptive performances for fixed-mobile convergent scenarios.

1. Introduction

Passive Optical Networks (PON) form a key component of today's residential broadband, and its extensive infrastructure has been considered to reduce fiber deployment costs in convergent fixed-mobile scenarios [1–3]. This reuse, however, is not straightforward. PON's point-to-multipoint topology and time-division multiplexing, well-suited for residential traffic, are poorly aligned with the stringent latency, jitter, and bit-rate requirements of some mobile services. In addition, the policies adopted by operators for dynamic bandwidth allocation (DBA) configuration, often implemented in a set-and-forget fashion, lack flexibility and could impose a trade-off between over-provisioning and degraded quality of service (QoS) if fixed and mobile traffic coexist in the same PON tree. In other words, fixed–mobile convergence can unlock new value for operators, but it also exposes the drawbacks of a cost-effective architecture and the rigidity of traditional network engineering practices. Addressing this tension requires more adaptive and responsive control over resource allocation than what is adopted today in PONs, with artificial intelligence (AI) expected to play a pivotal role [4].

Several studies have investigated the application of AI for bandwidth allocation in PONs. However, most have concentrated on optimizing DBA procedures [5, 6], often from a vendor-centric perspective. In contrast, our work adopts the network operator's perspective by enabling AI-driven dynamic configuration of transmission containers (T-CONTs). This approach seeks to reconcile the stringent performance requirements of mobile services with the shared nature of PONs, while abstracting away the vendor-specific details of DBA algorithms. By embedding an additional layer of intelligence directly into optical access management, we assess a self-optimizing infrastructure that can maximize the value of existing fiber assets while avoiding service degradation and minimizing the impact on live network operations. We evaluate three deep reinforcement learning (DRL) approaches in a commercial SDN-enabled G-PON serving 16 clients. We examine the trade-offs between convergence time and the ability to achieve and stabilize target upstream latency below 1 ms for mobile services.

2. Experimental Setup

Our DRL framework is evaluated on a commercial G-PON testbed comprising an OLT and 16 ONUs (Fig. 1). These are configured with mixed upstream traffic profiles derived from the literature [7, 8]: 15 ONUs emulate heterogeneous FTTH user traffic, categorized into three distinct usage profiles (high, medium, and low network demand), while 1 ONU carries the characteristic load of an FTTA site. The resulting aggregated traffic (Fig. 2) exhibits clear diurnal patterns and distinct loads per user category, with shaded zones indicating the traffic variations considered in this work. We assume 1 service per ONU, with T-CONT type 1 for mobile traffic, defined by the fixed information rate (FIR) and T-CONT type 3 for fixed clients, defined by the peak information rate (PIR) and committed information rate (CIR). Under these conditions, we demonstrated in [9] that no static T-CONT configuration can meet the objectives of preventing congestion while maintaining delays below 1ms for FTTA and 5ms for FTTH throughout the day. We also showed that supervised learning can be used for T-CONT optimization. However, this approach required extensive labeled data and was unable to learn autonomously, operating solely offline. By using DRL instead, we can optimize the T-CONT configuration by interacting with the G-PON, continuously improving through trial and error rather than relying solely on static predictions.

Our platform also comprises an SDN controller layer to enable management-plane programmability for all underlying physical network functions (OLT, traffic generator and analyzer, and 10 GbE switch for ONU traffic aggregation). The DRL agent, situated at the orchestration plane, interfaces with the controller through a RESTCONF/YANG-based northbound interface (NBI). This allows the orchestrator to gather comprehensive network telemetry and execute configuration actions across the network, which will later be translated into DRL

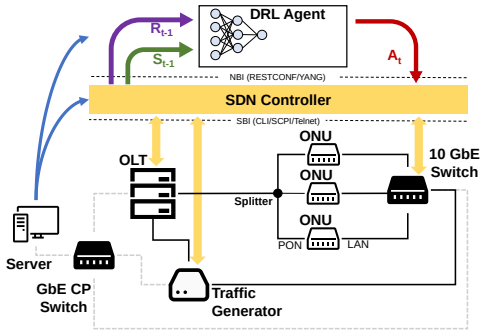

Fig. 1: Architecture of the DRL-optimized PON and abstraction layer.

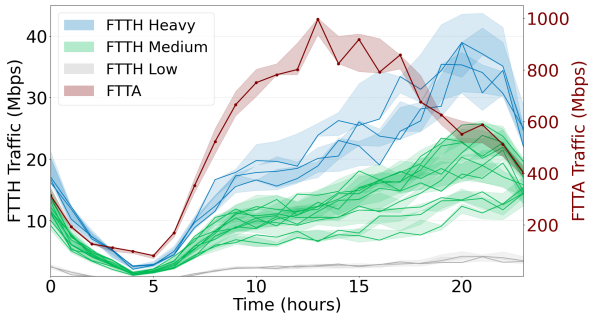

Fig. 2: Assumed traffic patterns for FTTA (right y-axis) and FTTH ONUs (left y-axis).

parameters (actions at an instant t taken based on the states and rewards at an instant $t-1$, cf. Fig. 1).

We evaluate three DRL solutions: Deep Q-Network (DQN) [10], its Double DQN (DDQN) [11] extension, and Proximal Policy Optimization (PPO) [12]. In both DQN and DDQN, a neural network estimates the value (Q-value) of each possible action given a particular state. DDQN improves upon DQN by decoupling action selection from evaluation, thus mitigating overestimation bias and enabling more stable policy learning. In contrast, PPO is a policy-based method that directly learns the policy itself, i.e., the mapping from states to actions. We model the agent–G-PON interaction as a Markov Decision Process (MDP) with the objective of minimizing congestion (delay $<$1 ms for FTTA, $<$5 ms for FTTH) while reducing the total allocated bandwidth. The agent observes network conditions, adjusts T-CONT parameters, and receives feedback through delay, congestion, and resource utilization metrics. Table 1 summarizes the MDP formulation.

Table 1: MDP formulation for G-PON control.

Component	**Description**
State	FIR/PIR, FTTA/FTTH demand & delay, congestion flags, daytime (10 normalized features)
Action	Discrete: increase/decrease FIR, PIR, or do nothing.
Reward	$R = 1 - 10(\text{FTTA}_{\text{cong}} + \text{FTTH}_{\text{cong}}) - 0.5(\text{FIR}_{\text{norm}} + \text{PIR}_{\text{norm}})$
Episode	One full day (24h) of stochastic traffic with 3 configuration updates per hour.

3. Results

We evaluate PPO, DQN, and DDQN algorithms through extensive experimentation on a commercial G-PON testbed. Our analysis focuses on key operational metrics: reward convergence, latency stability, and resource allocation efficiency. Traffic is modeled to reflect realistic daily variations, generated randomly within predefined demand zones. All algorithms employ three-layer neural networks with tuned hyperparameters and are assessed based on convergence speed in episodes (Table 1) and their impact on live network KPIs during training.

Fig. 3 compares the learning curves of PPO, DQN, and DDQN agents using optimized hyperparameters, including an ablated DDQN variant where configuration frequency was reduced from 3 to 1 update per hour. The DDQN agent (green) achieves the highest and most stable reward, demonstrating superior convergence. DQN (orange) shows competitive initial convergence but exhibits greater long-term instability. PPO (blue) displays significant training noise and fails to reach robust performance within 500 episodes. The ablated DDQN (gray), despite using identical hyperparameters, achieves lower and more variable performance due to limited experience diversity resulting from sparse configuration updates.

Fig. 4 illustrates the variations of the FTTA latency per episode for the fine-tuned agents, with shaded areas representing the standard deviation obtained with each algorithm within 1 episode. DDQN achieves the lowest and most stable latency among the three approaches. However, during the initial 20 episodes of training, the convergence transition results in elevated delay levels that exceed the target QoS thresholds. This learning-induced degradation is inherent to the exploration phase but is considerably more severe and prolonged for both PPO and DQN. Figs. 5 and 6 present the agents' exploration of T-CONT parameters under consideration. The curves represent the episode means, while the shaded areas indicate the range between the minimum and maximum values used by the DRL agent's trial-and-error research approach. Notably, each episode corresponds to a full 1-day cycle, which accounts for the substantial variation observed. We highlight that our reward function penalizes congestion more heavily than bandwidth usage. Consequently, the agents initially prioritize FIR-PIR configurations that minimize congestion and subsequently converge toward slightly lower bandwidth allocations.

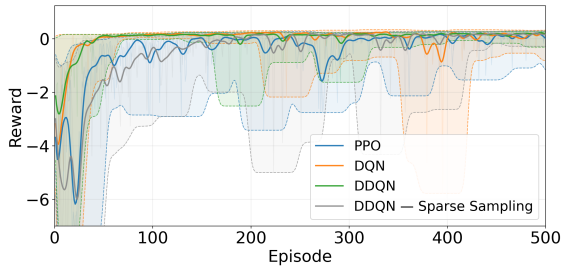

Fig. 3: Reward convergence.

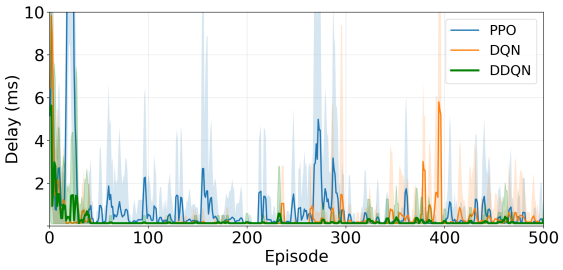

Fig. 4: FTTA delay with standard deviation bounds.

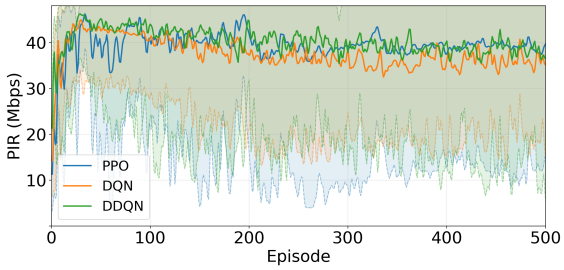

Fig. 5: PIR (T-CONT 3) evolution for FTTH traffic.

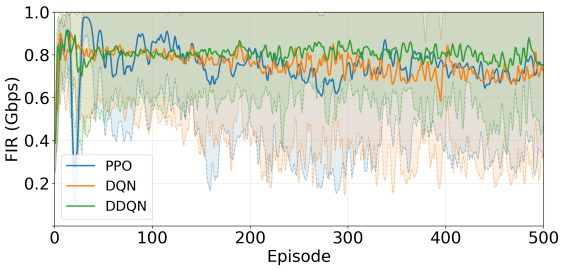

Fig. 6: FIR (T-CONT 1) evolution for FTTA traffic.

4. Conclusions

This work demonstrates the practical viability of DRL-driven T-CONT configuration in commercial G-PON infrastructure for fixed-mobile convergence scenarios. Among the evaluated algorithms, DDQN achieves superior performance with stable convergence and low-latency operation (130 μs on average) after an initial 20-episode learning phase. While this exploration period temporarily degrades SLA compliance, a challenge further amplified in PPO and DQN, the results confirm that tuned action spaces and exploration strategies enable rapid convergence with acceptable operational impact. Increasing measurement frequency also provides the agent with denser training data, which accelerates learning and shortens the period of suboptimal allocations during early adaptation.

Our experimental validation on real equipment, reveals that AI-enhanced PON control is feasible and offers a vendor-agnostic path toward self-optimizing access networks that adapt to variable traffic patterns. Future work could further enhance service continuity by combining autonomous policies with rule-based safeguards and refined MDP formulations, ensuring faster convergence and even more controlled behavior in live environments.

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. We gratefully acknowledge Emilia Foti for her careful review and valuable contributions to the manuscript.

References

1. P. Chanclou et al., "FTTH and 5G Xhaul Synergies for the Present and Future," in *ICTON*, 2019, pp. 1–4.
2. J. Kani et al., "Solutions for Future Mobile Fronthaul and Access-Network Convergence," *J. Light. Technol.*, 2017.
3. L. Anet Neto et al., "Enabling technologies and innovations for 5G-oriented optical access networks," in *Broadband Access Communication Technologies XV*, 2021.
4. E. Wong et al., "Machine learning enhanced next-generation optical access networks—Challenges and emerging solutions [Invited Tutorial]," *J. Opt. Commun. Netw.*, 2023.
5. L. Hu et al., "APP-aware MAC scheduling for MEC-Backed TDM-PON mobile fronthaul," *IEEE Commun. Lett.*, 2023.
6. L. Ruan and E. Wong, "Addressing concept drift of dynamic traffic environments through rapid and self-adaptive bandwidth allocation," in *ICCCN*, 2023.
7. M. Kihl et al., "Traffic analysis and characterization of Internet user behavior," in *ICUMT*, 2010.
8. E. Ericsson, "Mobility report." 2023.
9. L. Inglés et al., "Experimental Demonstration of Demand-Driven PON Configuration for Fixed-Mobile Convergence," in *ECOC*, 2025.
10. V. Mnih et al., "Playing Atari with deep reinforcement learning," *arXiv preprint arXiv:1312.5602*, 2013.
11. H. Van Hasselt et al., "Deep reinforcement learning with double Q-learning," in *AAAI conference on artificial intelligence*, 2016.
12. J. Schulman et al., "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.