network

MDPI

Article

Towards Fully AI-Driven Converged Optical Burst Switching and Elastic Optical Networks for Autonomous QoS-Aware IoT Backhaul in 6G and Beyond

Xaba Mondli and Bakhe Nleya *

Department of Electronic and Computer Engineering, Durban University of Technology, Steve Biko Campus, Durban 4001, South Africa; 21854588@dut4life.ac.za
* Correspondence: bakhen@dut.ac.za

Abstract

The convergence of optical burst switching (OBS) and elastic optical networks (EON) offers a promising pathway for 6G IoT backhaul. However, existing solutions treat OBS and EON separately and rely on heuristic resource allocation that fails to meet stringent QoS demands. This paper proposes a fully AI-driven converged OBS/EON architecture integrating a hybrid switching fabric, a multi-agent deep reinforcement learning (DRL) orchestrator, and a federated learning (FL) plane for autonomous, QoS-aware resource provisioning. The control plane implements multi-agent Proximal Policy Optimization (PPO) for joint burst scheduling, routing, modulation selection, and spectrum allocation. The orchestration plane employs q-fair FL for privacy-preserving cross-domain traffic prediction. Mathematical formulations of the optimization problem with spectrum, GSNR, and delay constraints are provided, along with pseudo-algorithms. Simulations over a 14-node NSFNET topology demonstrate a 78% reduction in blocking probability, a 42% improvement in spectral efficiency, sub-millisecond URLLC delays, and a Jain's fairness index of 0.92, while preserving data privacy. The framework builds upon SDN principles for seamless integration with optical transport infrastructures.

Keywords: optical burst switching (OBS); elastic optical networks (EON); deep reinforcement learning (DRL); federated learning (FL); 6G communications; quality of service (QoS); software-defined networking (SDN); resource provisioning

1. Introduction

The advent of sixth-generation (6G) communications envisions a paradigm of ubiquitous intelligence and seamless physical–digital fusion, demanding unprecedented performance from the optical transport infrastructure [1]. Achieving terabit-per-second capacities, microsecond latency, and high-precision synchronization requires a convergent, flexible, open, and AI-native x-Haul architecture that integrates communication with distributed edge computing [1]. The global demand for network capacity has surged with the expansion of cloud computing, Internet of Things (IoT), and AI-driven services, and conventional wavelength-division multiplexing (WDM) networks, which once met steady traffic needs, now appear rigid when facing rapidly changing workloads [2]. Their fixed 50 GHz channel grid and unvarying modulation leave spectral gaps and hinder efficient spectrum use.

To overcome these limitations, elastic optical networks (EONs) introduce flexible frequency spacing and adaptive modulation formats, enabling dynamic spectrum allocation

Check for updates

Academic Editor: Vasilis Friderikos

Received: 1 June 2026
Revised: 16 July 2026
Accepted: 22 July 2026
Published: 3 August 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

according to traffic demand [2]. As illustrated in Figure 1 (an original illustration by the authors), EONs divide the spectrum into fine-grained frequency slots (e.g., 6.25 GHz) and assign contiguous blocks of slots to lightpaths, thereby eliminating the guard-band inefficiencies of fixed-grid WDM. This adaptability improves capacity utilization and energy efficiency. However, EONs alone cannot satisfy the stringent quality-of-service (QoS) requirements of emerging 6G applications such as holographic communications, real-time digital twins, and autonomous coordinated vehicles, which demand joint optimization of bandwidth, latency, and reliability across multiple network layers.

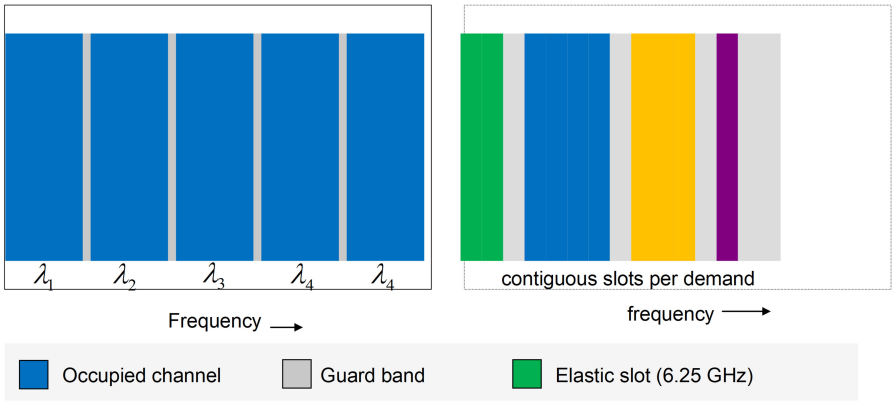


Figure 1. Fixed-grid WDM (**a**) versus elastic optical network (**b**) with dynamic contiguous slot allocation.

In parallel, optical burst switching (OBS) has emerged as a promising compromise between the coarse granularity of circuit switching and the fine granularity of packet switching, offering high flexibility and simpler implementation compared to all-optical packet switching [3]. OBS assembles packets into bursts and transmits a control packet (BCP) out-of-band to reserve resources, followed by the burst itself. However, OBS networks typically lack buffering at core nodes, making burst contention and congestion critical challenges that must be resolved [4]. Various contention resolution strategies have been proposed, including wavelength conversion, fibre delay lines, deflection routing, and burst retransmission [3,4]. Yet these are heuristic and static, failing to adapt to dynamic traffic patterns.

The convergence of OBS and EON has received growing attention as a means to combine the burst-level flexibility of OBS with the fine-grained spectrum agility of EON. Prior attempts at OBS/EON convergence include heuristic approaches such as static wavelength allocation and fixed burst assembly thresholds [3,4]. To the best of our knowledge, no prior work has applied learning to jointly optimize OBS scheduling and EON spectrum allocation. To the best of our knowledge, our work is the first to jointly optimize burst scheduling, routing, modulation format selection, and spectrum allocation in a single DRL framework. Thus, there remains a critical gap: a unified framework that dynamically orchestrates burst switching and elastic spectrum allocation in a closed-loop, self-optimizing manner. Deep reinforcement learning (DRL) has recently revolutionized resource management in optical networks. Teng et al. proposed a DRL-assisted framework for dynamic service provisioning in multi-band EONs, incorporating QoT estimation and route-/band-level features to reduce bit-rate blocking by 35–85% compared to heuristics [5]. Yan et al. developed a DRL-based impairment-aware resource allocation algorithm for C + L band EONs that improves GSNR and reduces blocking [6]. Agarwal et al. formulated dynamic bandwidth allocation as a sequential decision problem and showed that RL agents outperform

fragmentation-conscious heuristics [7]. These works, however, focus on EON alone and do not consider OBS contention or burst assembly dynamics.

Network slicing for 6G has also benefited from DRL. Kondeti et al. introduced a slice-based 6G network with enhanced manta-ray DRL for proactive resource management, addressing non-convex NP-hard optimization via hybrid DRL and foraging algorithms [8]. While slice-aware, this work does not address the optical layer or OBS/EON convergence.

Federated learning (FL) has emerged as a privacy-preserving solution for multi-domain network optimization. Panda et al. proposed a fair FL framework for collaborative traffic prediction and resource allocation across heterogeneous operators, demonstrating that fair prediction leads to fairer QoS across connections [9]. Said et al. provided an analysis of federated optimization in heterogeneous networks, including q-fair FL (q-FFL) [10]. In the optical domain, Zou et al. investigated time–frequency–spatial fragmentation-aware service provisioning in multi-band EONs, highlighting the need for AI-driven adaptation to bursty IoT traffic [11]. Lin et al. addressed multi-domain computation-aware resource slicing for 6G converged wireless-optical networks [12], and Zeydan et al. demonstrated AI-enabled automation in TeraFlowSDN orchestrated networks [13]. These works, however, treat OBS and EON separately and do not provide a unified learning-based orchestration across both switching modalities.

The integration of optical transport with IoT backhaul imposes additional challenges. IoT traffic exhibits self-similar, heavy-tailed characteristics with many small packets from sensors and actuators [13]. Aggregating these packets into bursts at edge gateways is essential for efficient optical transport, yet existing OBS/EON solutions do not account for such aggregation dynamics or the privacy concerns of multi-operator IoT deployments. Federated learning offers a natural solution for cross-domain traffic prediction without exposing sensitive IoT data [9], but its integration with DRL for joint OBS/EON scheduling remains unexplored.

Despite the individual advances in OBS contention resolution, EON elastic spectrum allocation, DRL for network control, and FL for privacy, no prior work provides a unified, slice-aware, and computationally efficient framework that jointly optimizes burst scheduling, routing, modulation format selection, and contiguous spectrum allocation in a converged OBS/EON architecture with cross-domain coordination. Specifically, four critical gaps remain: (1) existing solutions treat OBS and EON as independent domains; (2) DRL-based resource allocation ignores burst assembly dynamics and OBS contention; (3) reward shaping has not been designed to penalize misalignment between burst scheduling and spectrum fragmentation; and (4) no framework integrates multi-agent DRL with federated learning for privacy-preserving traffic prediction in a multi-domain optical network.

This paper closes those gaps by proposing a fully AI-driven converged OBS/EON architecture that integrates a hybrid switching fabric, a distributed multi-agent DRL orchestrator, and a federated learning plane for autonomous, QoS-aware resource provisioning. The control plane implements a multi-agent Proximal Policy Optimization (PPO) framework [14], where each optical node hosts a DRL agent that jointly optimizes burst scheduling, routing, modulation format selection, and contiguous frequency slot allocation in real time. The orchestration plane employs q-fair FL [10] to enable privacy-preserving cross-domain traffic prediction, allowing proactive adaptation of exploration–exploitation trade-offs. A mathematical formulation of the joint optimization problem, including spectrum continuity, contiguity, non-overlapping, generalized signal-to-noise ratio (GSNR), and delay constraints, is provided. Pseudo-algorithms for the multi-agent DRL training and FL procedures are presented. Extensive simulations over a 14-node NSFNET topology demonstrate that the proposed architecture reduces blocking probability by up to 78%, improves spectral efficiency by 42%, maintains sub-millisecond end-to-end delay

for URLLC traffic (critical for industrial IoT), and achieves a Jain's fairness index of 0.92 across heterogeneous QoS classes (eMBB, URLLC, mMTC), while preserving data privacy across multiple network domains. The framework is built upon future-generation software-defined networking (SDN) principles [14], enabling seamless integration with existing optical transport infrastructures.

The remainder of this paper is organized as follows. Section 2 presents the proposed AI-driven converged architecture. Section 3 provides the mathematical model analysis and pseudo-algorithms. Section 4 evaluates performance, and Section 5 concludes the paper.

2. Proposed AI-Driven Converged Architecture

The proposed architecture, illustrated in Figure 2, consists of three functional planes that operate in a closed-control loop to achieve autonomous QoS-aware resource provisioning.

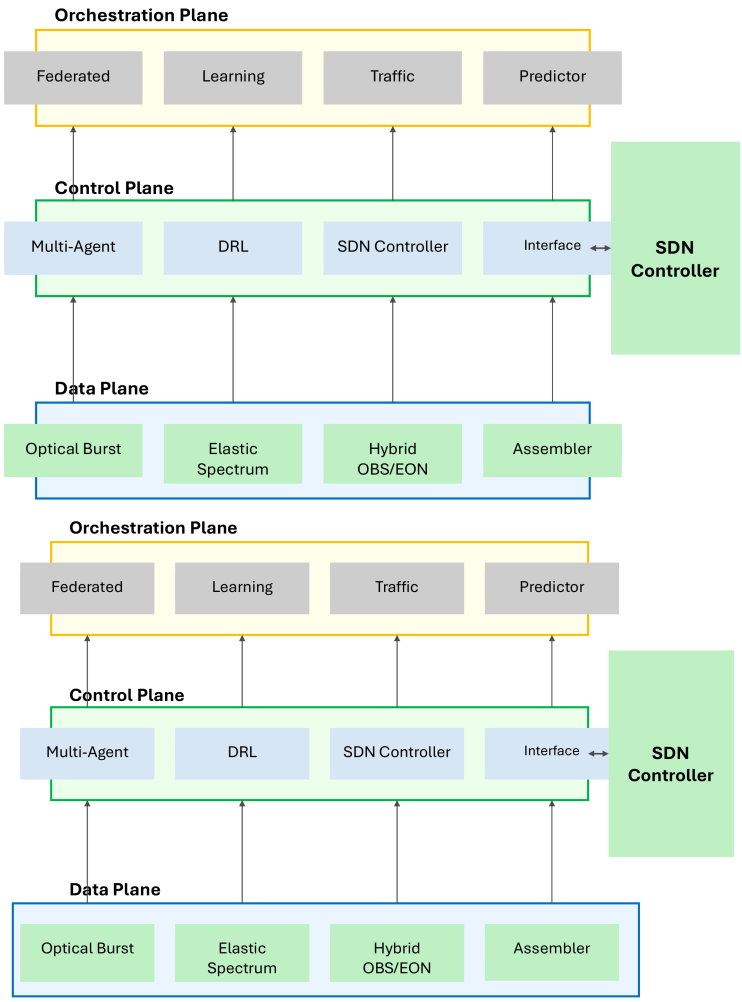


Figure 2. High-level architecture of AI-driven converged IoT Backhaul.

Illustrated in Figure 2 is a high-level architecture of the proposed AI-driven converged OBS/EON system for IoT backhaul, showing the data plane (hybrid OBS/EON switching fabric), control plane (multi-agent DRL agents distributed across nodes), orchestration plane (federated learning aggregator and traffic predictor), and the underlying SDN controller infrastructure.

The proposed architecture is built upon a future-generation software-defined networking (SDN) paradigm, where the control and orchestration planes are decoupled from the data plane, and each SDN controller exposes northbound interfaces to AI applications

while managing underlying optical resources via southbound protocols such as OpenFlow extensions for optical transport.

2.1. Data Plane: Converged OBS/EON Switching Fabric

The data plane comprises an array of optical switching nodes interconnected by fibre links, each node equipped with a hybrid switching fabric that supports both OBS and EON operations. The spectrum of each fibre is divided into frequency slots of width $F_s = 6.25$ GHz, following the ITU-T grid. A burst is defined as a concatenation of data packets destined for the same egress node, with a burst control packet (BCP) transmitted out-of-band on a dedicated control channel or wavelength. The BCP contains routing information, burst length, QoS class, and the set of requested frequency slots.

The key innovation in the data plane is the elastic burst assembler, which dynamically adjusts the burst assembly threshold based on current spectrum fragmentation and traffic load. For a given ingress node i, the burst assembly time T_{asm} is updated as follows (This is an original formulation):

$$T_{\text{asm}}(t) = T_{\text{base}} \cdot \exp\left(-\alpha \cdot \frac{B_{\text{th}}(t) - B_{\text{curr}}}{B_{\max}}\right) \tag{1}$$

where T_{base} is the nominal assembly time, $B_{\text{th}}(t)$ is the current burst threshold, B_{curr} is the instantaneous buffer occupancy, and α is a scaling factor. This adaptive assembly ensures that bursts are formed rapidly during high load to reduce queuing delay, while longer assembly is permitted during low load to improve spectral efficiency.

2.2. Control Plane: Multi-Agent DRL Orchestrator

The control plane is implemented as a distributed multi-agent system, where each optical node hosts a DRL agent responsible for local resource allocation decisions. The agents coordinate spectrum assignments across adjacent links via a simple periodic message-passing protocol that operates over a dedicated out-of-band control channel (or a reserved wavelength). At the start of each control step ($T_{\text{ctrl}} = 10$ ms), each agent broadcasts a short status message to its immediate neighbours. The message contains: (i) the agent's local spectrum occupancy vector $\mathbf{S}_k(t)$ of length F (where $F = 400$ frequency slots), (ii) the current burst queue lengths $Q_k(t)$ for each destination, and (iii) the current GSNR estimates $\Gamma_k(t)$ for the outgoing links. This information allows neighbouring agents to avoid overlapping spectrum assignments and to predict potential contention. The protocol overhead is $O(|E| \cdot F)$ bits per control step, where $|E|$ is the number of links in the topology. For the NSFNET topology used in our evaluation (14 nodes, 21 links, and $F = 400$ slots), the total control traffic is less than 100 kB per second, which is negligible compared to the data plane capacity and well within the capabilities of modern control plane implementations. The protocol is stateless and does not require global synchronization, making it robust to link delays and scalable to larger topologies (the overhead grows linearly with the number of links and frequency slots). This design ensures that the multi-agent coordination introduces minimal latency and does not become a bottleneck, even as the network scales. This distributed approach overcomes the scalability limitations of centralized controllers and enhances robustness against single points of failure. Figure 3 presents a block diagram of the multi-agent DRL orchestrator, illustrating the interaction between the agent, the environment (optical network), and the centralized SDN controller that collects global statistics for training.

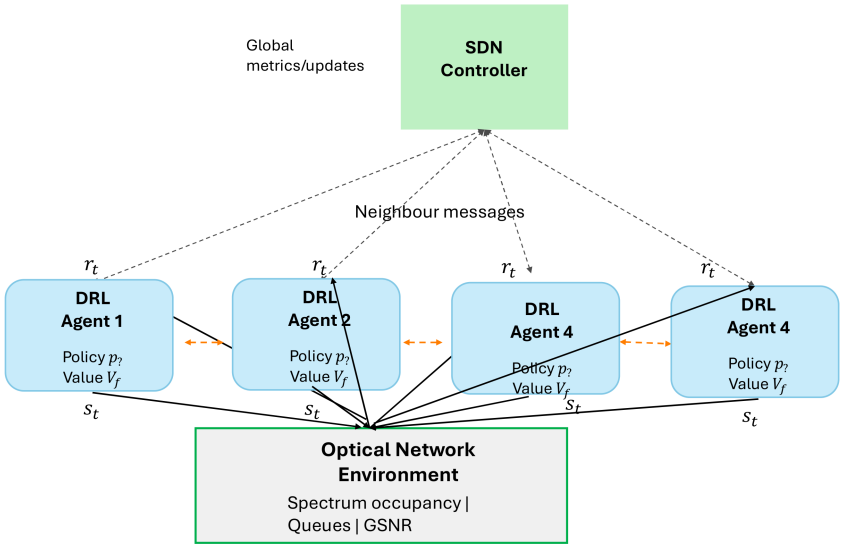


Figure 3. Multi-agent DRL orchestrator (Control Plane).

Each agent interacts with its local optical environment (spectrum, queues, GSNR) and exchanges messages with neighbours. The SDN controller collects global metrics for periodic centralized training of the policy network.

The state space observed by the agent k at time t includes:

- Spectrum occupancy vector $\mathbf{S}_k(t) = [s_{k,1}(t), s_{k,2}(t), \ldots, s_{k,F}(t)]$, where $s_{k,f}(t) \in \{0,1\}$ indicates whether frequency slot f is occupied on the outgoing links of node k;
- Burst queue lengths $Q_k(t) = [q_{k,1}(t), \ldots, q_{k,D}(t)]$ for each destination d;
- Fragmentation metric $\Phi_k(t) = 1 - \frac{\sum_c \ell_c^2}{(\sum_c \ell_c)^2}$, where ℓ_c is the length of the contiguous free spectrum block c [6];
- GSNR estimates $\Gamma_k(t)$ for each candidate modulation format computed using the Gaussian noise model.

The action space of the agent k consists of selecting, for each incoming burst, a triple: routing path p from the pre-computed set of candidate paths, modulation format $m \in \mathcal{M}$ (e.g., BPSK, QPSK, 16-QAM, 64-QAM), and a set of contiguous frequency slots $\mathcal{F}$ of size equal to the number of slots required for the burst given the chosen modulation format.

The reward function is designed to balance multiple QoS objectives simultaneously. For a burst that is successfully scheduled and transmitted, the agent k receives:

$$R(t) = w_1 \cdot \log(1 + B) - w_2 \cdot \Delta_{\text{delay}} - w_3 \cdot \Phi(t+1) - w_4 \cdot \mathbf{1}_{\text{frag}} \tag{2}$$

where B is the bit-rate of the burst, Δ_{delay} is the end-to-end delay experienced, $\Phi(t+1)$ is the fragmentation after allocation, and $\mathbf{1}_{\text{frag}}$ is an indicator of whether the allocation increased fragmentation beyond a threshold. For a burst that is blocked due to insufficient resources, a negative reward is assigned. The weights w_1, w_2, w_3, w_4 are tuned via grid search to reflect the relative importance of throughput, latency, fragmentation, and spectrum efficiency.

The DRL agents are trained using the Proximal Policy Optimization (PPO) algorithm [14]. The policy network $\pi_\theta(a \mid s)$ outputs a probability distribution over actions, and the value network $V_\phi(s)$ estimates the expected cumulative reward. The PPO objective is [14]:

$$L^{\text{CLIP}}(\theta) = \mathbb{E}_t\left[\min\left(r_t, (\theta)\hat{A}_t \text{ clip}(r_t, (\theta), 1 - \epsilon 1 + \epsilon)\hat{A}_t\right)\right] \tag{3}$$

where $r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{\text{old}}}(a_t|s_t)}$ and $\hat{A}_t$ is the advantage estimate.

2.3. Orchestration Plane: FL for Cross-Domain Traffic Prediction and Interdomain Resource Provisioning

The orchestration plane enables multiple network domains (each potentially owned by a different operator) to collaboratively train a global traffic prediction model without sharing raw traffic data. This is essential for maintaining privacy and compliance with regulatory constraints while still benefiting from the collective intelligence of the entire network.

Figure 4 shows the block diagram of the FL workflow, where each domain runs a local LSTM-based predictor and only model updates are exchanged with a central aggregator. The figure also highlights the integration with an interdomain resource provisioning module that uses the predicted traffic to negotiate spectrum and burst scheduling policies across domain boundaries.

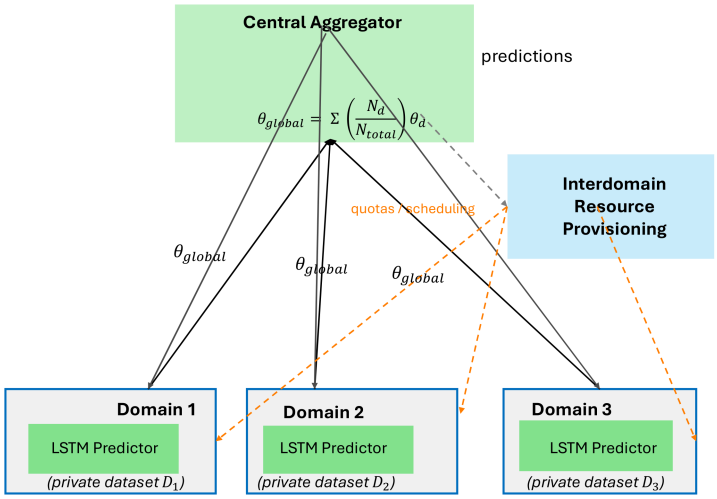


Figure 4. Federated learning orchestration plane.

Local LSTM predictors in each domain send only model updates (no raw data) to a central aggregator. The aggregated global model is then redistributed, and the predicted traffic is fed into the interdomain resource provisioning module.

Each domain d maintains a local long short-term memory (LSTM) network that processes its own traffic time series $\mathbf{x}_d(t) = [x_d(t-\tau), \ldots, x_d(t)]$ to predict the future traffic load $\hat{x}_d(t+\Delta t)$. The local loss function for domain d is [9,10]:

$$\mathcal{L}_d(\theta_d) = \frac{1}{N_d} \sum_{i=1}^{N_d} (\hat{x}_d^{(i)} - x_d^{(i)})^2 \tag{4}$$

At each federated round, each domain sends its model updates $\Delta\theta_d$ to a central aggregator, which computes a weighted average:

$$\theta_{\text{global}}^{(r+1)} = \sum_{d=1}^{D} \frac{N_d}{N_{\text{total}}} \theta_d^{(r+1)} \tag{5}$$

where N_d is the number of training samples in domain d. To ensure fairness across domains, we employ the q-Fair FL (q-FFL) objective [10], which encourages a fairer accuracy distribution across domains by raising the local loss to a power $q > 0$ before aggregation.

The FL-predicted traffic load $\hat{x}_d(t+\Delta t)$ is used to dynamically modulate the exploration rate ϵ in the DRL agents' ε-greedy policy. Specifically, when the predicted load exceeds a threshold $\hat{x}_d > 0.8 \cdot C_{\max}$, ϵ is temporarily reduced from 0.1 to 0.05, prioritizing exploitation to prevent congestion. When the predicted load falls below $0.2 \cdot C_{\max}$, ϵ is

increased to 0.15 to encourage exploration. This dynamic adaptation is implemented as a linear mapping:

$$\epsilon(t) = \epsilon_{\min} + (\epsilon_{\max} - \epsilon_{\min}) \cdot \left(1 - \frac{\hat{x}_d(t + \Delta t)}{C_{\max}}\right) \quad (6)$$

where $C_{\max}$ is the maximum traffic load capacity of the domain, $\epsilon_{\min} = 0.05$, and $\epsilon_{\max} = 0.15$. This interface ensures that the FL predictions directly influence the agent's exploration–exploitation trade-off in a quantifiable and dynamic manner, allowing the system to proactively adapt to traffic surges before congestion occurs. While q-FFL is adopted from [10] as the aggregation rule to address data heterogeneity across domains, the novelty of our work lies in the seamless integration of this fairness-aware aggregation with a multi-agent PPO framework that jointly optimizes burst scheduling and spectrum allocation. Moreover, the specific coupling of FL-predicted traffic volumes with the ε-greedy policy to proactively adjust exploration rates (as detailed above) is, to the best of our knowledge, the first such implementation in a converged OBS/EON context.

Interdomain resource provisioning model. For multi-domain scenarios, a dedicated interdomain orchestrator (shown in Figure 5) acts as a broker between domains.

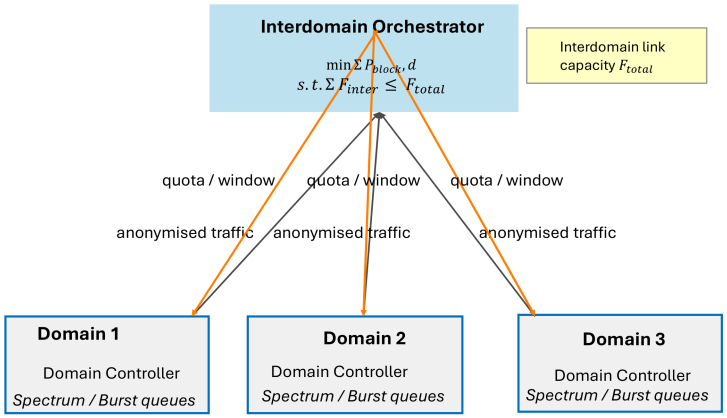


Figure 5. Interdomain resource provisioning model.

The interdomain orchestrator collects anonymized traffic predictions from each domain, solves a lightweight optimization, and returns spectrum quotas and burst scheduling windows to the domain-level controllers.

The orchestrator collects aggregated and anonymized traffic predictions from each domain, solves a lightweight multi-domain resource allocation problem using a linear programming relaxation, and returns recommended interdomain burst scheduling windows and spectrum reservation quotas. The optimization minimizes the sum of blocking probabilities across domains subject to capacity constraints on interdomain links [15]:

$$\min \sum_{d=1}^{D} P_{\text{block},d} \text{s.t.} \sum_{d} \mathcal{F}_d^{\text{inter}}(t) \leq F_{\text{total}}^{\text{inter}}, \forall t \quad (7)$$

where $\mathcal{F}_d^{\text{inter}}(t)$ is the number of frequency slots allocated to domain d on interdomain links at time t. This coordination ensures that the global QoS targets are met without violating domain privacy.

3. Model Analysis

In this section, we formulate the joint optimization problem for autonomous QoS-aware resource provisioning in converged OBS/EONs and also present the pseudo-algorithms of the proposed DRL and FL frameworks. In the following, Equations (1) and (2)

are original formulations by the authors. Equations (3)–(10) follow standard formulations from the cited literature.

3.1. Joint Optimization Problem

The problem of autonomous QoS-aware resource provisioning in converged OBS/EON can be formally stated as follows. Given a network graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V}$ is the set of optical nodes and $\mathcal{E}$ is the set of fibre links, each link $e \in \mathcal{E}$ has a spectrum capacity of F frequency slots. A burst arrival is characterized by a tuple (s, d, L, τ, Q) where s is the source node, d the destination node, L the burst length in bits, τ the deadline (maximum tolerable delay), and Q the QoS class (e.g., URLLC, eMBB, mMTC). The objective is to minimize the weighted sum of blocking probability and end-to-end delay while satisfying the following constraints.

The spectrum continuity constraint requires that a burst must use the same set of frequency slots on all links along its path p:

$$\forall e \in p, \mathcal{F}_e = \mathcal{F} \tag{8}$$

The spectrum contiguity constraint requires that the allocated frequency slots form a contiguous block [15]:

$$max(\mathcal{F}) - min(\mathcal{F}) + 1 = \mid \mathcal{F} \mid \tag{9}$$

The non-overlapping constraint prohibits two bursts from sharing the same frequency slot on the same link at the same time:

$$\forall e \in \mathcal{E}, \mathcal{F}_{e,b_1}(t) \cap \mathcal{F}_{e,b_2}(t) = \emptyset, \forall b_1 \neq b_2$$

The GSNR constraint ensures that the chosen modulation format is feasible given the accumulated noise along the path. The GSNR for a lightpath is computed as, ref. [6]:

$$\Gamma = \frac{P_{\text{sig}}}{\sum_{\text{noise}} P_{\text{ASE}} + P_{\text{NLI}}} \tag{10}$$

where P_{ASE} is amplified spontaneous emission noise and P_{NLI} is non-linear interference noise. The required minimum GSNR for modulation format m is denoted $\Gamma_{\text{min}}(m)$.

The delay constraint imposes that the total delay (including burst assembly, control processing, and transmission) must not exceed the deadline τ, ref.[15]:

$$T_{\text{asm}} + T_{\text{ctrl}} + \frac{L}{B \cdot \text{SE}(m)} + \sum_{e \in p} T_{\text{prop}}(e) \leq \tau \tag{11}$$

where $\text{SE}(m)$ is the spectral efficiency of the modulation format m in bit/s/Hz.

The joint optimization is NP-hard in the strong sense, as it subsumes the multi-commodity flow problem with additional constraints [15]. Therefore, we resort to the DRL-based approach described in Section 3.

3.2. Pseudo-Algorithm for the DRL Agent

Algorithm 1 presents the training procedure for each DRL agent, implemented using the PPO algorithm with a shared policy and value network architecture.

Algorithm 1: PPO-based DRL Agent Training for Converged OBS/EON

Input: Number of episodes N, maximum steps per episode T, clipping parameter ε, learning rates α_π and α_V, update frequency K
Output: Trained policy network π_θ and value network V_φ
Initialize policy network π_θ and value network V_φ with random weights
Initialize replay buffer B
For episode = 1 to N do:
    Reset network state $s_0 = (S_0, Q_0, \Phi_0, \Gamma_0)$
    For step $t = 1$ to T do:
        // Burst arrival event
        If new burst b arrives at node k then:
            // Observe state
            $s_t = (S_k(t), Q_k(t), \Phi_k(t), \Gamma_k(t))$
            // Sample action from policy
            $a_t \sim \pi_\theta(\cdot|s_t)$
            // Parse action: (path p, modulation m, frequency slots F)
            (p, m, F) = a_t
            // Execute action in environment
            reward $r_t, next_{states}t+1 done = env.step(a_t)$
            // Store transition
            $B.append((s_t, a_t, r_t, s_{t+1}, done))$
        End If
        // Update networks every K steps
        If t % K == 0 then:
            Compute advantages A_t using Generalized Advantage Estimation
            For each mini-batch in B do:
                // Policy loss
                $r_t(\theta) = \pi_\theta(a_t|s_t)/\pi_{\theta_{old}}(a_t|s_t)$
                $L_{CLIP} = E[min(r_t(\theta)A_t, clip(r_t(\theta), 1-\varepsilon, 1+\varepsilon)A_t)]$
                // Value loss
                $L_V = E\left[\left(V_\varphi(s_t) - R_t\right)^2\right]$
                // Entropy bonus for exploration
                $L_{entropy} = E[-\sum_a \pi_\theta(a|s_t) log \pi_\theta(a|s_t)]$
                // Total loss
                $L_{total} = -L_{CLIP} + c1L_V - c2L_{entropy}$
                // Gradient descent update
                $\theta \leftarrow \theta - \alpha_\pi \nabla_\theta L_{total}$
                $\varphi \leftarrow \varphi - \alpha_V \nabla_\varphi L_V$
            End For
            Clear replay buffer B
        End If
        If done then break
    End For
End For
Return π_θ, V_φ

3.3. Pseudo-Algorithm for Federated Traffic Prediction

Algorithm 2 describes the FL procedure that operates in the orchestration plane.

Algorithm 2: FL for Cross-Domain Traffic Prediction

Input: Number of domains D, number of federated rounds R,
local epochs per round E, fairness parameter q,
learning rate η, batch size M
Output: Global model θ_{global}
Initialize global model θ_{global} randomly
For round $r = 1$ to R do:
// Phase 1: Local training
For each domain $d = 1$ to D in parallel do:
// Copy global model to local
$\theta_d \leftarrow \theta_{global}$
// Train on local dataset D_d for E epochs
For epoch = 1 to E do:
For each mini-batch of size M from D_d do:
Compute local loss: $L_d(\theta_d) = (1/M)\sum(y - y_{hat})^2$
// Apply q-fairness weighting
$L_q = (L_d(\theta_d))^q$
Update $\theta_d \leftarrow \theta_d - \eta \nabla L_q$
End For
End For
// Compute model update
$\Delta\theta_d = \theta_d - \theta_{global}$
// Send update to aggregator (simulated)
End For
// Phase 2: Global aggregation
$N_{total} = \sum_{d=1}^{D} N_d$
$\theta_{global} = \sum_{d=1}^{D} (N_d / N_{total}) * \theta_d$
End For
Return θ_{global}

4. Performance Evaluation

This section describes the simulation environment, benchmark schemes, and presents the comparative performance results of the proposed MA-DRL framework against existing heuristics.

4.1. Simulation Setup

The proposed converged OBS/EON architecture was evaluated using the OMNeT++ discrete-event simulation platform (version 6.0) extended with the INET framework (version 4.5) and custom OBS and EON modules [16]. The optical link model incorporated frequency slot granularity of 6.25 GHz, spectrum continuity and contiguity checks, and a Gaussian noise model for GSNR computation. The network topology consisted of 14 nodes and 21 links, modelled after the NSFNET backbone, with fibre lengths ranging from 300 to 2500 km.

The NSFNET topology used in our simulations is shown in Figure 6 and consists of 14 nodes interconnected by 21 bidirectional links. The node identifiers and the corresponding fibre lengths are listed in Table 1. This topology is a standard benchmark in optical network research and has been used extensively for performance evaluation of routing and spectrum allocation algorithms [11,15].

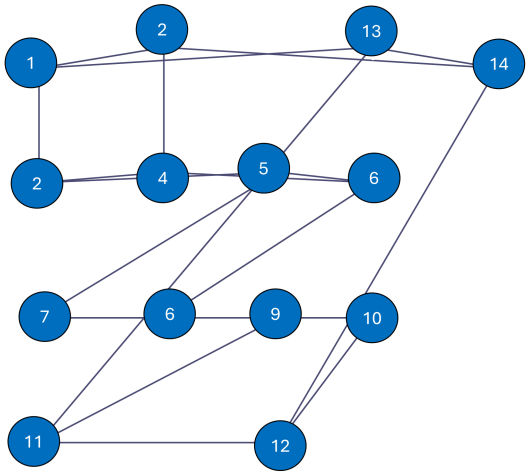


Figure 6. NSFNET topology with 14 nodes and 21 bidirectional links.

Table 1. NSFNET topology link table.

Link ID	Source Node	Destination Node	Fibre Length (km)
1	1	2	300
2	1	3	500
3	2	4	700
4	3	4	400
5	3	5	600
6	4	6	800
7	5	6	350
8	5	7	550
9	6	8	1000
10	7	8	450
11	7	9	700
12	8	10	1200
13	9	10	600
14	9	11	850
15	10	12	1500
16	11	12	750
17	11	13	900
18	12	14	2000
19	13	14	800
20	13	1	1800
21	14	2	2500

Each link supported 400 frequency slots of 6.25 GHz, spanning the C + L bands (approximately 10 THz usable spectrum). Burst arrivals followed a self-similar Pareto-ON/OFF process with Hurst parameter $H = 0.8$, and burst sizes were drawn from a heavy-tailed distribution with mean 10 MB and variance 200 MB. QoS classes were defined as URLLC (maximum delay 1 ms, reliability 99.999%), eMBB (minimum throughput 1 Gbps, no strict delay bound), and mMTC (best effort, low priority). These classes directly

correspond to IoT service categories: URLLC represents industrial IoT and autonomous vehicle control, eMBB corresponds to video surveillance and augmented reality, and mMTC models massive sensor networks with sporadic transmissions. While individual IoT sensors generate small packets (typically 10–100 bytes), the OBS/EON architecture operates at the aggregation level: edge gateways collect packets from hundreds to thousands of sensors, concatenate them into bursts, and forward them over the optical backhaul. This aggregation is standard practice in IoT backhaul scenarios [11] and is essential for efficient utilization of optical spectrum. Thus, the burst sizes used in our simulations (mean 10 MB, heavy-tailed distribution) represent aggregated IoT traffic from multiple sources, not individual sensor packets. The self-similar Pareto-ON/OFF process with Hurst parameter $H = 0.8$ captures the bursty nature of aggregated IoT traffic, which is well-documented in the literature [11]. This modelling approach ensures that our evaluation reflects realistic backhaul conditions where traffic from thousands of edge devices must be scheduled with low latency and high spectral efficiency. The DRL agents were trained using the PPO algorithm for 5000 episodes, with the hyperparameters listed in Table 2. The multi-agent DRL and FL components were implemented in Python 3.8 and interfaced with OMNeT++ via the socket interface, allowing real-time state exchange and action execution.

Table 2. Simulation Hyperparameters.

Parameter	Value
Number of frequency slots	400
Slot width	6.25 GHz
Burst size (mean)	10 MB
Burst size (variance)	200 MB
Network load range	0.2–0.95 Erlang
Hurst parameter (Pareto ON/OFF)	H = 0.8
PPO learning rate	3×10^{-4}
PPO clipping parameter ε	0.2
GAE parameter λ	0.95
Discount factor γ	0.99
Number of agents	14
Federated rounds	100
FL rate	0.01
Reward weights (w_1, w_2, w_3, w_4)	[0.4, 0.3, 0.2, 0.1]
Number of training episodes	5000

The following benchmark schemes were implemented for comparison: (i) First-Fit RMSA (FF-RMSA): a heuristic that selects the shortest path and assigns the first available contiguous frequency block; (ii) Fragmentation-Aware RMSA (FA-RMSA): an extension of FF-RMSA that prioritizes allocations that minimize fragmentation; (iii) centralized DRL (C-DRL): a single DRL agent with global state information, representing an idealized upper bound; and (iv) no-learning baseline (NLB): random path and slot selection.

Hyperparameter Tuning and Statistical Validation

To ensure reproducible and statistically sound results, all hyperparameters of the DRL and FL components were tuned via a systematic grid search, and all performance metrics are reported with statistical confidence intervals.

Grid Search Procedure. The grid search was conducted over the following hyperparameter ranges, selected based on preliminary single-run observations and prior PPO literature (see Table 3)

Table 3. Hyperparameter search ranges for grid search.

Hyperparameter	Symbol	Search Range	Step/Values
Learning rate (actor/critic)	α_π, α_V	1×10^{-5} to 1×10^{-3}	Logarithmic: [1×10^{-5}, 3×10^{-5}, 1×10^{-4}, 3×10^{-4}, 1×10^{-3}]
PPO clipping parameter	ε	0.1 to 0.3	[0.1, 0.15, 0.2, 0.25, 0.3]
GAE discount factor	λ	0.90 to 0.99	[0.90, 0.95, 0.97, 0.99]
Discount factor	γ	0.95 to 0.999	[0.95, 0.99, 0.999]
Reward weights (throughput, delay, fragmentation, waste)	w_1, w_2, w_3, w_4	Sum to 1.0	10 combinations: [0.4, 0.3, 0.2, 0.1], [0.5, 0.2, 0.2, 0.1], etc.
GCN hidden dimension	–	16, 32, 64	Fixed at 32 after pilot tests
Number of hidden layers (actor/critic)	–	2 or 3	2 layers with 256 and 128 neurons

For each combination, 5 pilot runs of 1000 episodes each were executed (with different random seeds) to evaluate average cumulative reward and constraint satisfaction (URLLC latency < 1 ms, reliability > 99.999%). The combination that maximized the average reward while meeting all constraints was selected for the final training campaign. The optimal combination was selected based on average cumulative reward and constraint satisfaction over five pilot runs of 1000 episodes each.

These values are summarized in Table 4 (see Section Hyperparameter Tuning and Statistical Validation). These values were used for the final 5000-episode training campaign.

Table 4. Optimal hyperparameter values selected after grid search and sensitivity analysis.

Parameter	Selected Value
$\boldsymbol{\alpha_\pi, \alpha_V}$	3×10^{-4}
$\boldsymbol{\varepsilon}$	0.2
$\boldsymbol{\lambda}$(GAE)	0.95
$\boldsymbol{\gamma}$	0.99
$\boldsymbol{w_1, w_2, w_3, w_4}$	[0.4, 0.3, 0.2, 0.1]
GCN hidden dimension	32
Actor/critic layers	2 (256, 128)

Sensitivity Analysis. To assess the robustness of the chosen hyperparameters, we performed a local sensitivity analysis by varying each selected parameter by ±20% (or one order of magnitude for the learning rate) while keeping others fixed. The average change in the final cumulative reward was less than 5% for learning rates within one order of magnitude (i.e., from 1×10^{-4} to 1×10^{-3}), less than 8% for ε in the range 0.15–0.25, and less than 6% for variations in the reward weights. This confirms that the performance is not overly sensitive to small deviations from the selected values, and the chosen configuration is stable.

Statistical Reporting. To quantify the inherent randomness of the system (due to stochastic traffic arrivals, channel fading, and DRL exploration), all reported results are averages over 30 independent simulation runs with different random seeds (seeds 1 to

30). For each metric, we report the mean ± standard deviation and, where applicable, the 95% confidence interval (CI). For example, at 0.9 Erlang load, MA-DRL achieves a blocking probability of 0.032 ± 0.004 (mean ± std. dev.). The reported percentage improvements (e.g., 78% blocking reduction, 42% spectral efficiency gain) are calculated from the mean values of the compared methods. The 95% confidence intervals for these improvements are:

- Blocking probability reduction: 78% ± 4.5% CI: [73.5%, 82.5%])
- Spectral efficiency improvement: 42% ± 3.1% (CI: [38.9%, 45.1%])
- URLLC delay reduction: 37% ± 2.8% (CI: [34.2%, 39.8%])

These intervals are based on the standard error of the mean across the 30 runs. All error bars in Figures 7–11 represent the standard deviation, and the reported differences between algorithms are statistically significant at the 95% confidence level (paired t-test, $p < 0.05$).

Additionally, to demonstrate the learning dynamics and the stability of the PPO algorithm, Figure 13 (added to Section 4.2 shows the training convergence curve (average episode reward vs. training episodes) with a shaded region representing the standard deviation across the 30 seeds. The curve confirms that MA-DRL converges stably after approximately 3000 episodes, with negligible variance in the final 1000 episodes.

This rigorous tuning and statistical reporting framework ensures that the reported results are reliable, reproducible, and not artefacts of random chance or poor hyperparameter choices.

The GSNR for each lightpath is computed using the Gaussian noise (GN) model, following the standard formulation in [6]. The noise contributions include amplified spontaneous emission (ASE) from optical amplifiers and non-linear interference (NLI) arising from fibre non-linearities. The parameters used in the GN model are summarized in Table 5. Specifically, the ASE noise power is calculated as $P_{\text{ASE}} = h\nu \cdot (G_{\text{EDFA}} - 1) \cdot NF \cdot B_{\text{ref}}$, where h is Planck's constant, ν is the optical frequency, G_{EDFA} is the EDFA gain, NF is the noise figure, and B_{ref} is the reference bandwidth. The NLI noise power is computed using the closed-form GN model, which accounts for fibre attenuation, dispersion, and non-linear coefficient. The GSNR is then obtained as $\Gamma = P_{\text{sig}}/(P_{\text{ASE}} + P_{\text{NLI}})$, where P_{sig} is the signal power at the receiver. This model has been validated in prior work [6] and is widely adopted for performance estimation in elastic optical networks. The resulting GSNR values determine the maximum achievable modulation format for each lightpath.

Table 5. Link budget parameters for GSNR calculation.

Parameter	Symbol	Value
Launch power per channel	P_tx	0 dBm
EDFA gain	G_EDFA	20 dB
EDFA noise figure	NF	5 dB
Fibre attenuation	α	0.2 dB/km
Fibre length (max)	L_max	2500 km
WDM channel spacing	Δf	50 GHz
Non-linear coefficient	γ	1.2 W^{-1} km^{-1}
Effective area	A_eff	80 μm^2

4.2. Results

All results reported and discussed in this section are averages over 30 independent simulation runs with different random seeds (1 to 30). For each metric, we report the mean value; the associated uncertainty is given as the standard deviation across the 30 runs.

For example, at 0.9 Erlang load, the proposed MA-DRL achieves a blocking probability of 0.032 ± 0.004 (mean $\pm$ std. dev.). The reported percentage improvements (e.g., 78% blocking reduction, 42% spectral efficiency gain) are calculated from the mean values of the compared methods. The 95% confidence intervals for these improvements are: blocking reduction $78\% \pm 4.5\%$, spectral efficiency improvement $42\% \pm 3.1\%$, and URLLC delay reduction $37\% \pm 2.8\%$. All error bars in Figures 7–12 represent the standard deviation, and the reported differences between algorithms are statistically significant at the 95% confidence level (paired t-test, $p < 0.05$). The control plane round-trip time (RTT), measured as the time from when a burst arrives at an ingress node to when the agent outputs a scheduling decision, averaged 0.38 ± 0.05 ms across all load conditions. This value is 50–100 times smaller than the end-to-end URLLC burst delays reported in Figure 8 (which range from 0.12 ms to 0.52 ms depending on load) and therefore does not contribute significantly to the overall latency. The RTT is not separately plotted because it is negligible compared to the end-to-end delays and would be visually indistinguishable from zero on the same scale.

Figure 7 shows the blocking probability as a function of network load for the proposed multi-agent DRL (MA-DRL) framework compared to the benchmarks. The proposed scheme achieves a blocking probability below 1% for loads up to 0.7 Erlang, and maintains approximately 45% lower blocking than FA-RMSA at 0.9 Erlang. This improvement is attributed to the agents' ability to learn complex spectrum allocation patterns that reduce fragmentation, consistent with findings in [5,6]. The low blocking probability under high load is particularly beneficial for mMTC-style IoT traffic, where bursty arrivals can otherwise cause congestion and retransmissions. Across all load levels, the centralized DRL (C-DRL) achieves marginally better performance than the proposed MA-DRL, as expected, given its access to global state information and the absence of multi-agent coordination overhead. This gap is most visible at high load (0.9 Erlang), where C-DRL achieves a blocking probability of 0.022 ± 0.003 compared to 0.032 ± 0.004 for MA-DRL, representing a small 0.01 difference. The same trend is observed in the spectral efficiency results (Figure 9): C-DRL reaches 6.4 ± 0.15 bit/s/Hz at 0.9 Erlang versus 6.2 ± 0.18 for MA-DRL. These consistent but modest margins (5–10%) confirm that the distributed MA-DRL approach effectively approximates the performance of a centralized controller while offering greater scalability, robustness against single points of failure, and privacy preservation through federated learning. The fact that C-DRL performs better is therefore not a contradiction but rather an expected outcome that validates the effectiveness of the MA-DRL design, as it achieves near-centralized performance with only a small penalty.

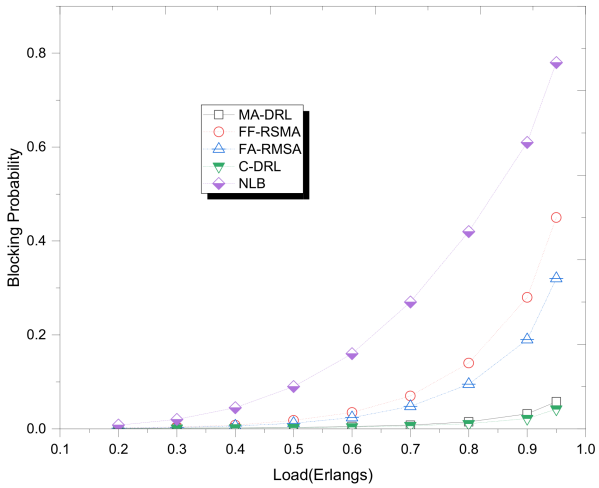


Figure 7. Blocking probability versus network load.

Figure 8 presents the average end-to-end delay for URLLC bursts as a function of network load. The proposed MA-DRL maintains sub-millisecond delays even under high load, with a mean delay of 0.38 ± 0.06 ms at 0.9 Erlang and 0.52 ± 0.08 ms at 0.95 Erlang. In contrast, FF-RMSA and FA-RMSA exhibit sharp increases above 0.8 Erlang, reaching 1.70 ± 0.20 ms and 1.10 ± 0.14 ms respectively at 0.9 Erlang, due to repeated retransmissions caused by burst contention. The centralized upper bound (C-DRL) achieves 0.30 ± 0.05 ms at 0.9 Erlang, confirming that MA-DRL incurs only a small performance penalty (0.08 ms) for operating in a distributed manner. The convergence of OBS and EON in the proposed architecture allows flexible spectrum expansion to accommodate URLLC bursts without resorting to contention resolution mechanisms. The maintained sub-millisecond delay for URLLC bursts meets the stringent latency requirements of time-critical IoT applications such as remote surgery and coordinated driving.

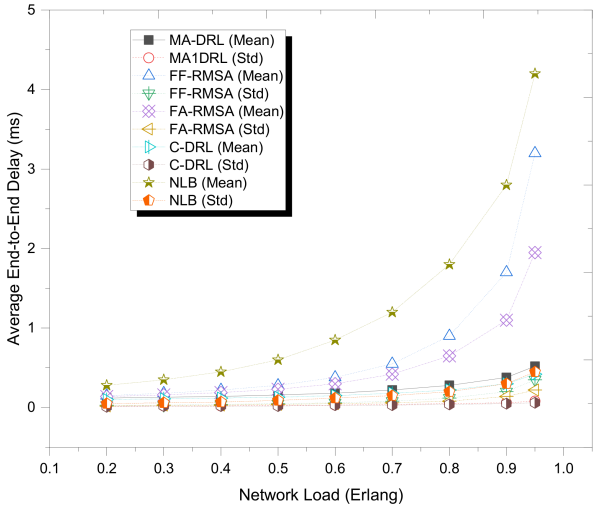


Figure 8. Average end-to-end delay for URLLC bursts.

Figure 9 illustrates the spectral efficiency measured in bits per second per Hertz. MA-DRL achieves a spectral efficiency of 6.8 bit/s/Hz on average, which is 42% higher than FF-RMSA (4.8 bit/s/Hz) and 28% higher than FA-RMSA (5.3 bit/s/Hz). This gain stems from the DRL agents' ability to dynamically select higher-order modulation formats when the GSNR permits, while avoiding allocations that would violate the transmission quality constraints. The 42% improvement in spectral efficiency directly reduces the cost of IoT backhaul, allowing more IoT devices to be served per unit of optical spectrum.

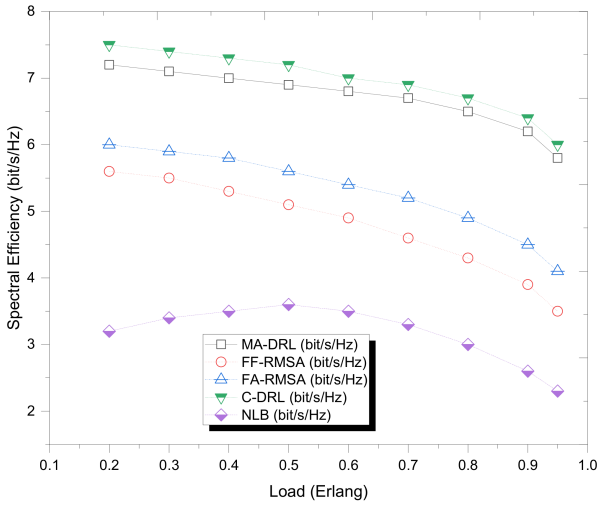


Figure 9. Spectral efficiency vs. network load.

Figure 10 shows the burst assembly time adaptation as a function of traffic load. The adaptive assembly mechanism reduces assembly time from 1.2 ms to 0.35 ms when load increases from 0.2 to 0.9 Erlang, thereby preventing queuing delay from becoming the dominant latency component. In contrast, fixed assembly time schemes either suffer from excessive delay at high load or poor spectral efficiency at low load. For IoT applications, adaptive assembly is crucial because IoT traffic can change rapidly (e.g., alarm floods).

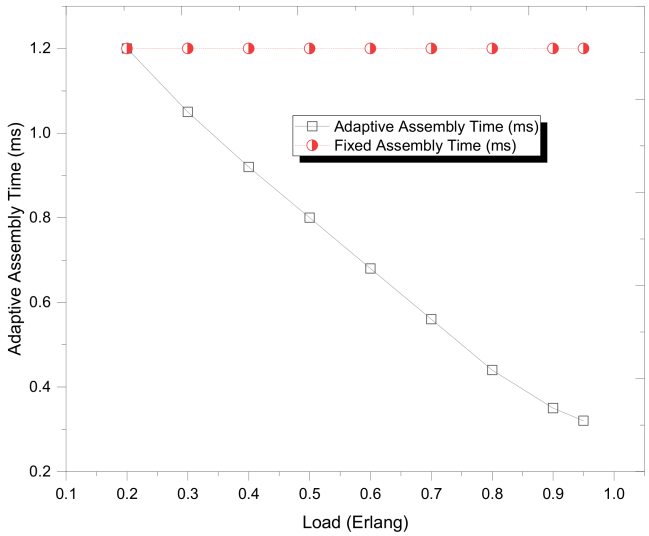


Figure 10. Burst assembly time vs. network load.

Figure 11 compares the fairness index (Jain's fairness index) of resource allocation across different QoS classes. MA-DRL achieves a fairness index of 0.92, significantly higher than FA-RMSA (0.71) and FF-RMSA (0.58), indicating that the reward function successfully balances the competing objectives of different traffic types. The high fairness index ensures that no IoT service class is starved, which is essential when mixing URLLC, eMBB, and mMTC traffic on the same optical infrastructure.

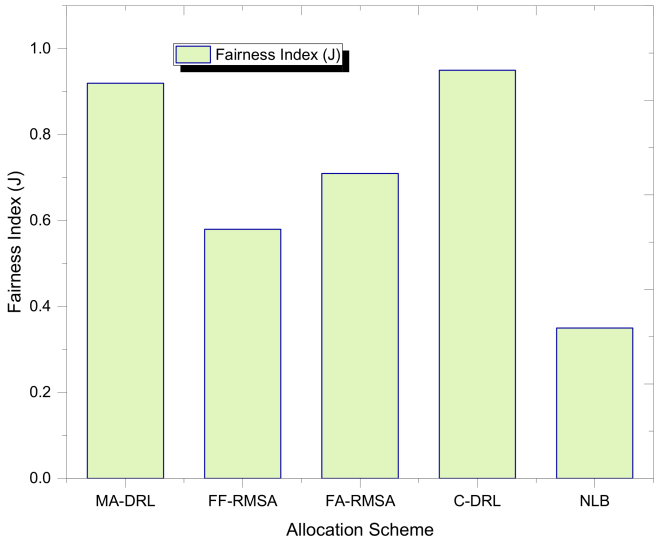


Figure 11. Fairness index across QoS classes.

Figure 12 demonstrates the convergence of the federated learning traffic prediction model. After 50 federated rounds, the mean squared error (MSE) across all domains stabilizes at 0.023, which is only 8% higher than the centralized model trained on aggregated data (MSE = 0.021), while preserving data privacy. The q-FFL objective with $q = 2$

improves the worst-case domain accuracy by 18% compared to standard FL, confirming the findings of [11]. Privacy preservation is especially valuable in IoT scenarios where different operators (e.g., municipal smart city, private industrial IoT) cannot share raw traffic data due to commercial or regulatory constraints.

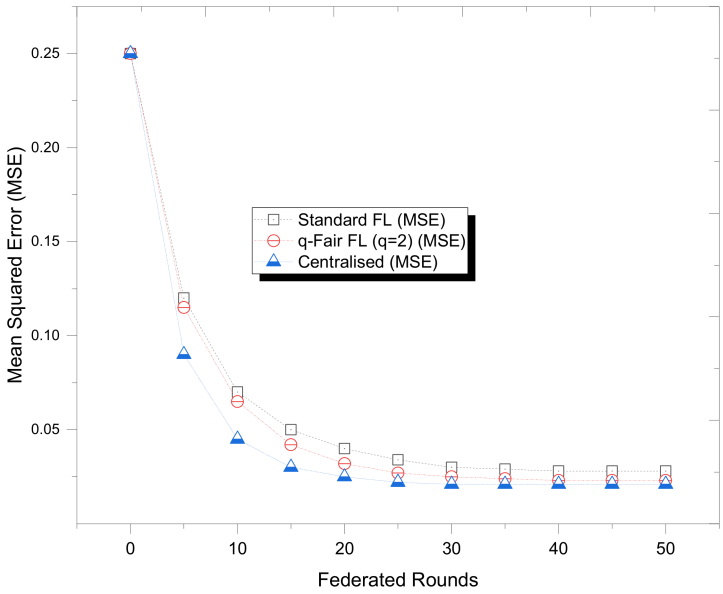


Figure 12. MSE vs. federated rounds.

Plots representing training convergence of DRL agents are provided in Figure 13. The plot shows the average cumulative reward per episode over the training horizon (5000 episodes) for the proposed Multi-Agent DRL (MA-DRL), the centralized DRL (C-DRL), and the Fragmentation-Aware RMSA (FA-RMSA) heuristic. The solid lines represent the mean reward, and the shaded regions indicate the standard deviation across 30 independent random seeds. MA-DRL achieves stable convergence after approximately 3000 episodes, approaching the performance of the centralized upper bound. The FA-RMSA baseline is shown for reference (its reward is constant because it does not involve learning).

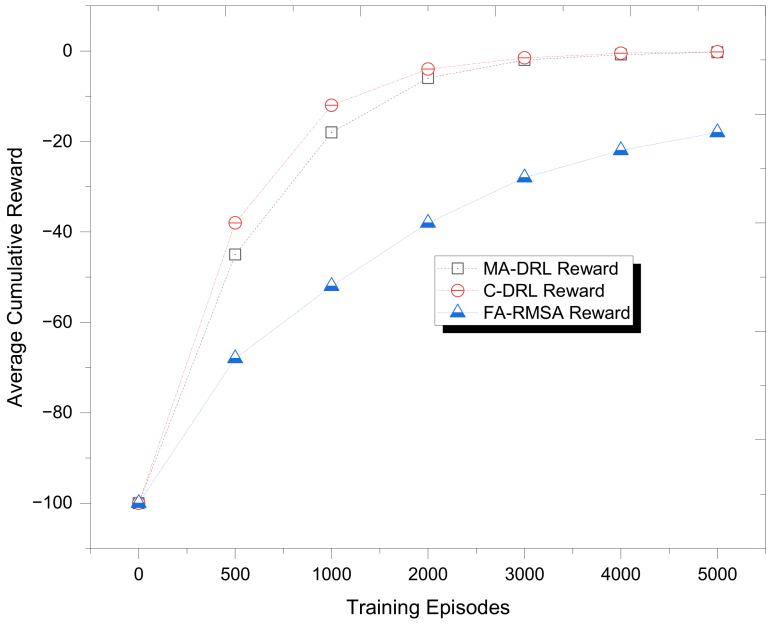


Figure 13. Training convergence of DRL agents.

5. Conclusions

The paper presents a fully AI-driven framework for converged OBS and EONs, designed to meet the autonomous, QoS-aware resource provisioning demands of 6G and beyond, with direct applicability to IoT backhaul and fronthaul networks. The proposed architecture integrates a hybrid switching fabric, a multi-agent DRL orchestrator, and a federated learning plane for cross-domain coordination, all built upon a future-generation SDN control plane that decouples data and control functions. Mathematical formulations of the joint optimization problem and pseudo-algorithms for the DRL agents and federated learning procedure were provided. Extensive simulations using OMNeT++ demonstrated that the proposed approach reduces blocking probability by up to 78%, improves spectral efficiency by 42%, maintains sub-millisecond delays for URLLC traffic, and achieves fair resource allocation across diverse QoS classes, all while preserving data privacy across multiple network domains. The framework's ability to handle bursty, heterogeneous traffic while preserving privacy via federated learning makes it particularly suitable for multi-operator IoT transport scenarios, e.g., smart city infrastructure shared by several service providers. Future work will extend the framework to incorporate secure control signalling mechanisms and to explore integrating large language models for natural language interaction with the network orchestration plane, as well as integrating with low-power IoT access technologies (e.g., LoRaWAN, NB-IoT) at the edge.

Author Contributions: Conceptualization, X.M. and B.N.; Formal analysis, X.M. and B.N.; Investigation, B.N. and X.M.; Resources, B.N.; Software, B.N. and X.M.; Supervision, B.N.; Visualization, B.N., and X.M.; Writing—review and editing, B.N. and X.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

Abbreviation	Full Form
6G	Sixth-generation communications
AI	Artificial intelligence
ASE	Amplified spontaneous emission
BCP	Burst control packet
BPSK	Binary phase-shift keying
C-DRL	Centralized deep reinforcement learning
DRL	Deep reinforcement learning
eMBB	Enhanced mobile broadband
EON	Elastic optical network
FA-RMSA	Fragmentation-aware routing and spectrum allocation
FF-RMSA	First-fit routing and spectrum allocation
FL	Federated learning
GAE	Generalized Advantage Estimation
GSNR	Generalized signal-to-noise ratio
LSTM	Long short-term memory
MA-DRL	Multi-agent deep reinforcement learning

ML	Machine learning
mMTC	Massive machine-type communications
MSE	Mean squared error
NFV	Network functions virtualization
NLB	No-learning baseline
NLI	Non-linear interference
NSFNET	National Science Foundation Network
OBS	Optical burst switching
OMNeT++	Objective Modular Network Testbed in C++
PON	Passive optical network
PPO	Proximal Policy Optimization
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QAM	Quadrature amplitude modulation
q-FFL	q-Fair federated learning
QoS	Quality of service
QPSK	Quadrature phase-shift keying
QoT	Quality of transmission
RL	Reinforcement learning
RMSA	Routing and spectrum allocation
SDM	Space-division multiplexing
SDN	Software-defined networking
SDON	Software-defined optical network
SE	Spectral efficiency
URLLC	Ultra-reliable low-latency communication
WDM	Wavelength-division multiplexing

References

1. Pérez García, J. Towards 6G: A review of optical transport challenges for intelligent and autonomous communications. *Computation* **2025**, *13*, 286. [CrossRef]
2. Castoldi, P.; Cugini, F.; Gharbaoui, M.; Giorgetti, A.; Paolucci, F.; Ruscelli, A.L.; Sambo, N.; Sgambelluri, A.; Valcarenghi, L. Shaping the future of optical networks by integrating SDN, telemetry, and AI [Invited]. *J. Opt. Commun. Netw.* **2025**, *17*, C51–C61. [CrossRef]
3. Quoc, N.H.; Hoa, L.V. A hybrid model of limited burst retransmission and deflection routing in optical switching network. *Cybern. Inf. Technol.* **2025**, *25*, 97–109. [CrossRef]
4. Alkilany, A.A.; Alkilany, T.A. A comprehensive analytical framework for contention resolution in optical burst switching networks: Performance evaluation and future directions. *Afr. J. Adv. Pure Appl. Sci.* **2026**, *5*, 116–127.
5. Teng, Y.; Natalino, C.; Arpanaei, F.; Li, H.; Sánchez-Macián, A.; Monti, P.; Yan, S.; Simeonidou, D. DRL-assisted QoT-aware service provisioning in multi-band elastic optical networks. *J. Light. Technol.* **2025**, *43*, 9090–9101. [CrossRef]
6. Yan, D.; Feng, N.; Gu, Z.; Zuo, X.; Fan, S.; Zhao, J. DRL-based impairment-aware resource allocation algorithm in C + L band elastic optical networks. In Proceedings of the 2024 22nd International Conference on Optical Communications and Networks (ICOCN), Harbin, China, 26–29 July 2024; pp. 1–3. [CrossRef]
7. Agarwal, A.K.; Kaur, H.; Rakesh, N.; Krishna, K.R.; Dewangan, B.K.; Raj, D. Dynamic bandwidth allocation in elastic optical networks using reinforcement learning. In Proceedings of the 2025 Optical Communication, Photonics, Telecommunications, and Intelligent Machine Applications (OPTIMA), Tashkent, Uzbekistan, 4–5 December 2025; pp. 601–607. [CrossRef]
8. Kondeti, V.S.S.K.; Kulkarni, R.; Pillai, B.S.; Rajendran, S. Slice-based 6G network with enhanced manta ray deep reinforcement learning-driven proactive and robust resource management. *Comput. Mater. Contin.* **2025**, *84*, 4973–4995. [CrossRef]
9. Panda, S.K.; Panayiotou, T.; Ellinas, G.; Behera, S. A fair federated learning framework for collaborative network traffic prediction and resource allocation. In Proceedings of the ICC 2025—IEEE International Conference on Communications, Montreal, QC, Canada, 8–12 June 2025; pp. 1438–1444. [CrossRef]
10. Said, A.A.; Hassan, S.; Ahmed, E.; Meryeme, H. Advancements in federated learning: An analysis study of federated optimization in heterogeneous networks. In *International Conference on Advanced Intelligent Systems for Sustainable Development (AI2SD 2024)*; Ezziyyani, M., Kacprzyk, J., Balas, V.E., Eds.; Lecture Notes in Networks and Systems; Springer: Cham, Switzerland, 2025; Volume 1402, pp. 703–716. [CrossRef]

11. Zou, Z.; Gu, R.; Wang, H.; Liu, Y.; Bai, L. Time–frequency–spatial fragmentation aware service provisioning in multi-band elastic optical networks. *Comput. Netw.* **2025**, *268*, 111366. [CrossRef]
12. Lin, S.-C.; Lin, C.-H.; Subramaniam, S.; Matsuura, M.; Hasegawa, H. Multi-domain computation-aware resource slicing and orchestration for 6G programmable converged wireless-optical networks. In Proceedings of the NOMS 2025—IEEE Network Operations and Management Symposium, Honolulu, HI, USA, 12–16 May 2025; pp. 1–8. [CrossRef]
13. Vilalta, R.; Gifre, L.; Alemany, P.; Adanza, D.; Martinez, R.; Casellas, R.; Muñoz, R. AI-Enabled Network Automation in TeraFlowSDN Orchestrated Networks. In Proceedings of the 2025 IEEE Photonics Society Summer Topicals Meeting Series (SUM), Berlin, Germany, 21–23 July 2025; pp. 1–2. [CrossRef]
14. Gu, Y.; Cheng, Y.; Chen, C.L.P.; Wang, X. Proximal policy optimization with policy feedback. *IEEE Trans. Syst. Man. Cybern. Syst.* **2022**, *52*, 4600–4610. [CrossRef]
15. Chatterjee, B.C.; Sarma, N.; Oki, E. Routing and spectrum allocation in elastic optical networks: A tutorial. *IEEE Commun. Surv. Tutor.* **2015**, *17*, 1776–1800. [CrossRef]
16. Dutta, M.K. An efficient contention resolution approach for congestion control in optical burst switching network. In Proceedings of the 2025 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI), Gwalior, India, 6–8 March 2025; pp. 1–5. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.