European Conference on Optical Communication (ECOC) 2022 © Optica Publishing Group 2022

Reinforcement-Learning-based Multilayer Path Planning Framework that Designs Grooming, Route, Spectrum, and Operational Mode

Takafumi Tanaka(1), Katsuaki Higashimori(1),

(1) NTT Network Innovation Laboratories, Hikarinooka, Yokosuka, Kanagawa, 239-0847 Japan takafumi.tanaka.mg@hco.ntt.co.jp

Abstract *We propose a reinforcement-learning-based multilayer path planning framework that designs grooming and optical path parameters. Simulation results show that the proposed method can improve blocking probability by 20 % compared to conventional heuristic methods. ©2022 The Author(s)*

Introduction

To accommodate the continuously growing internet traffic, the field of optical networking has long addressed the Routing and Spectrum Assignment (RSA) problem that efficiently assigns route and frequency resources to optical path demands. In recent years, there have been many attempts to apply machine learning to various tasks in optical networks. In the RSA problem, research is underway on using reinforcement learning (RL) to autonomously learn the optimal optical path planning through trial and error. In the Wavelength Division Multiplexing (WDM) layer, several methods for selecting optical paths and frequency slots using RL have been proposed and reported to achieve lower blocking probabilities than the First-Fit, the simple and basic RSA algorithm[1],[2].

This paper proposes a framework for designing multilayer paths (MLP) using RL in multilayer networks where logical paths such as the IP layer and Optical Data Unit (ODU) layer are further accommodated in the optical paths in the WDM layer. There have already been reported on multilayer path design using RL[3], which can appropriately select the establishment of new optical paths and use grooming paths. However, it assumes fixed grid in the WDM layer; therefore, there is room for further improvement in terms of frequency utilization efficiency. In contrast, we propose a multilayer path design framework that, in addition to selecting grooming paths, allows the selection of the many operational modes (combinations of modulation format and symbol rate) supported by recent digital coherent transceivers and frequency slots on the flexible grid. Simulations demonstrated that the RL-based MLP planning method is superior to conventional heuristic methods. In addition, the Reward, which is the target parameter for optimization by RL, can be tailored to the desired requirements, such as blocking probability and the number of optical paths.

MLP Planning using Auxiliary Graph

This section provides an overview of the common MLP design method[4] using the auxiliary graph. First, an auxiliary graph is constructed upon the arrival of MLP request. This auxiliary graph is a composite of the Establishment Candidate Graph (ECG), a graph whose edges are candidates for newly established optical paths, and the Grooming Capable Graph (GCG), a graph whose edges are existing optical paths with enough timeslot resources to accommodate the MLP request. The optimal solution is the path connecting the source and destination nodes on the auxiliary graph with the shortest route. We can determine the location of new and existing optical paths arbitrarily by setting weights for each edge of the auxiliary graph. This policy for setting the edge weights is called the Edge Weight Policy. In addition, to determine the parameters required for the newly established optical path, the Spectrum Policy, which determines the frequency slot, and the Operational Mode Policy, which determines the transmission mode consisting of a combination of modulation format and symbol rate, are used.

Here is an example of a heuristic MLP design method. In the Edge Weight Policy, H is the number of hops traversed on the physical topology, $W(e) = W^{GCG} + H \times W_h^{GCG}, e \in E(GCG)$, $W(e') = W^{ECG} + H \times W_h^{ECG}, e' \in E(GCG)$. For example, by setting $W(e)$ to be large relative to $W(e')$, we can prioritize the use of grooming by reducing the number of new paths as much as possible. In this paper, we refer to this policy as MaxGrooming. The Spectrum Policy is First

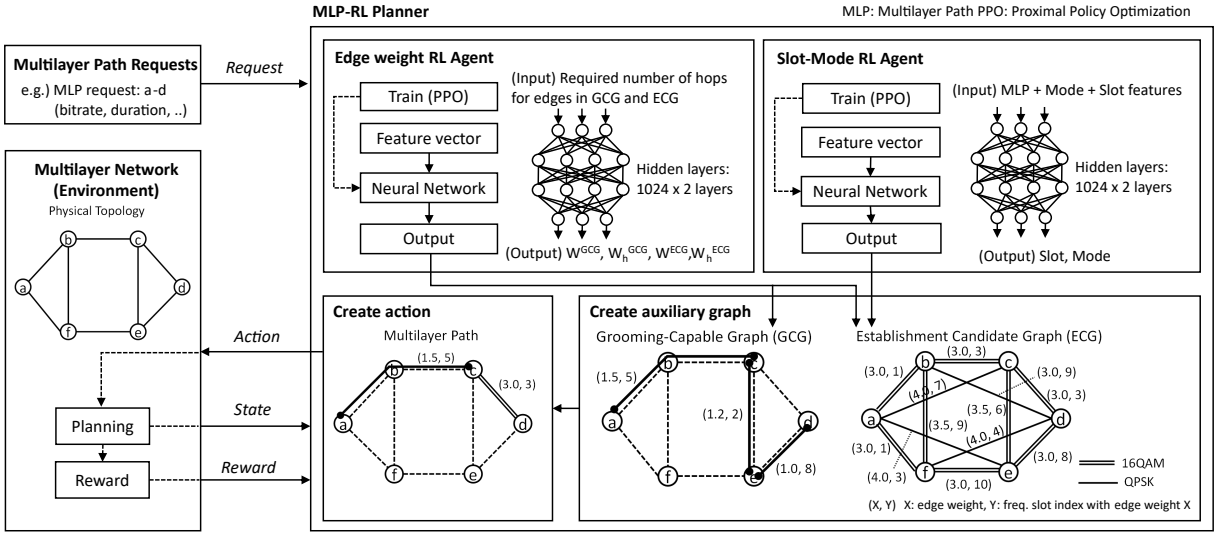

Fig. 1: Proposed MLP planning framework

Fit, the standard RSA algorithm, and the Operational Mode Policy selects the mode with the highest capacity among the modes that can reach the shortest path between the source and destination nodes. In this paper, we refer to this policy as MaxCapacity.

Deep Reinforcement Learning Framework for MLP Planning

RL is a type of machine learning that repeatedly interacts with the Agent, which acts as the brain to design the MLP and optimizes the Agent's policies to maximize the Reward, an indicator of whether the design is good or bad. Figure 1 shows the multi-layer path design framework using RL proposed in this paper: the MLP-RL Planner, which designs MLPs, has two RL Agents that learn the parameters needed to generate the auxiliary graph.

The first is the Edge Weight RL Agent, which plays the role of the Edge Weight Policy that optimizes the edge weights for selecting new optical paths and existing optical paths on the auxiliary graph. The feature vector input to the neural network representing the Agent's policies has a size $2 \times N \times N \times M$, where N is the number of nodes in the physical topology and M is the number of transmission modes. For each GCG and ECG, the number of hops that the optical path corresponding to the edge connecting the two points goes through on the physical topology is expressed in a connection matrix. This feature vector also has a transmission mode dimension, with a value of 0 for edges where the possible transmission distance for each transmission mode on the ECG is less than the path of the optical path candidate. For this feature vector input, the output of four parameters is W^{GCG}, W_h^{GCG}, W^{ECG}, and W_h^{ECG}.

The second is the Slot-Mode RL Agent, which corresponds to the Operational Mode Policy and Spectrum Policy that optimize the transmission mode and frequency slot for each edge of the ECG. The feature vector input to the neural network includes information about MLP (source and destination nodes, hop count, the average number of degrees on the route), operational mode (difference between transmission distance of the MLP and the operational mode, difference between the bit rate of the MLP request and the capacity of the operational mode, number of optical paths required to accommodate the MLP request), and frequency slots (frequency slot usage on the shortest path considering spectrum continuity constraint). And the output from the neural network includes information about the transmission mode and frequency slot to be applied for the MLP request. Here, if the bit rate of the MLP request is greater than the capacity of the transmission mode, multiple optical paths are used to accommodate it.

We assume that two neural network architectures for each Agent are identical for simplicity. This neural network architecture and feature vector parameters were determined through trial and error, but there is room for further improvement in these hyperparameter designs.

After setting the parameters output from these two neural networks to each edge of GCG and ECG, MLP-RL Planner determines the Actions (MLP including information on existing and new optical paths) by searching for the shortest route on the auxiliary graph. The Environment reflects this MLP in the multilayer network and returns the

Tab. 1: MLP planning schemes

	Heuristic	RL-Edge	RL-SlotMode	RL-All
Edge weight	MaxGrooming[4]	RL	MaxGrooming	RL
Operational mode	MaxCapacity[4]	MaxCapacity	RL	RL
Frequency slot	First Fit	First Fit	RL	RL

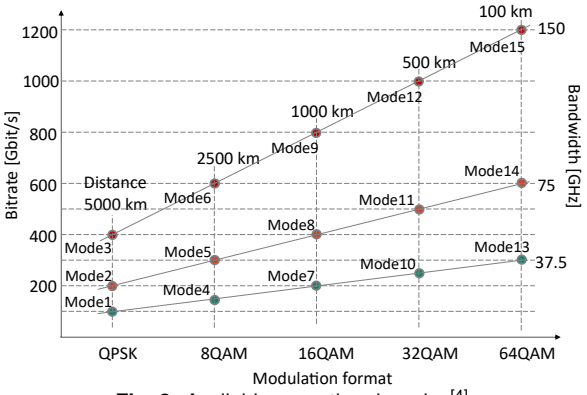

Fig. 2: Available operational modes[4]

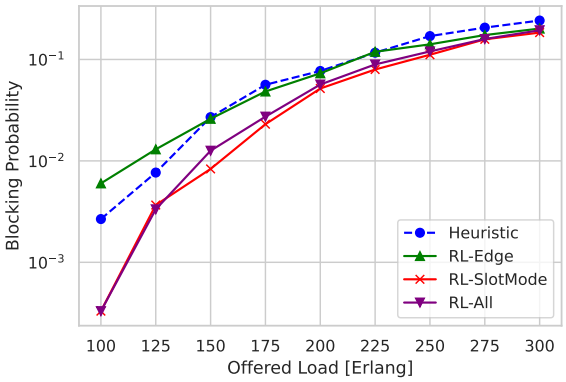

Fig. 3: Blocking probability versus traffic load

State and Reward values to the MLP-RL Planner. We can flexibly configure the Reward according to the requirements. The MLP-RL Planner learns the optimal policy using continuously arriving MLP requests and the multiple States and Rewards obtained by planning based on the requests. In this paper, we adopt the widely proven RL algorithm, Proximal Policy Optimization (PPO)[5].

Experimental Setup and Results

We evaluated the effectiveness of the proposed method by simulation.. The number of MLP requests was set to 3000, with bit rates in 100 Gbit/s increments from 100 Gbit/s to 400 Gbit/s, and the arrival and duration of the MLP requests were assumed to follow Poisson and exponential distributions, respectively. The evaluated network was NSFNET (14 nodes), and the frequency slots available for each link were 12.5 GHz/slot * 160 slots. In addition, 15 operational modes were selectable, as shown in Fig. 2.

The four scenarios were evaluated as shown in Tab. 1. In contrast to the heuristic method, RL-Edge used RL for edge weight optimization, RL-SlotMode used RL for operational mode and frequency slot selection, and RL-All for applied RL to both of them. We assumed the Reward to be +1 for successful and -1 for unsuccessful MLP assignment to minimize the blocking probability.

To demonstrate the effectiveness of the proposed method, a simulation evaluation was performed. The network topology used was NSFNET (14 nodes). The number of MLP requests to be generated was set to 3000, with bit rates in 100 Gbit/s increments from 100 Gbit/s to 400 Gbit/s occurring between randomly selected points in equal proportions, and their occurrence and duration following Poisson and exponential distributions, respectively. The number of optical fibers between nodes was set to 1 and the available frequency slots were 12.5 GHz/slot * 160 slots.

The four scenarios evaluated were as shown in Tab. 1. In contrast to the Heuristic, RL-Edge applies edge weights, RL-SlotMode applies transmission mode and frequency slot selection, and RL-All applies reinforcement learning to both.

Figure 3 shows the blocking probability against offered traffic load. The figure confirms that for Heuristic, RL-Edge, which behaves equivalently to Heuristic in the WDM layer, is equivalent, while RL-SlotMode and RL-All can accommodate about 20 % more MLPs, based on a blocking probability of 10^{-2}, due to optimization in the WDM layer. In addition, the state space is very large, especially for RL-All, so the neural network architectures needs to be further improved to obtain better performance.

Conclusions

We proposed an MLP design framework using RL and confirmed that the accommodative design method based on this framework could accommodate more MLPs by 20 % over the conventional heuristic method. Furthermore, although we set the Reward to minimize the blocking probability in this paper, it is also possible to optimize the Reward according to various requirements, such as setting it to reduce the number of new optical paths.

References

[1] X. Chen, B. Li, R. Proietti, H. Lu, Z. Zhu, and S. J. B. Yoo, "Deeprmsa: A deep reinforcement learning framework for routing, modulation and spectrum assignment in elastic optical networks", *Journal of Lightwave Technology*, vol. 37, no. 16, pp. 4155–4163, 2019. DOI: 10.1109/JLT.2019.2923615.

[2] M. Shimoda and T. Tanaka, "Mask rsa: End-to-end reinforcement learning-based routing and spectrum assignment in elastic optical networks", in *2021 European Conference on Optical Communication (ECOC)*, 2021, pp. 1–4. DOI: 10.1109/ECOC52684.2021.9606169.

[3] Z. Chen, J. Zhang, B. Zhang, R. Wang, H. Ma, and Y. Ji, "Admire: Demonstration of collaborative data-driven and model-driven intelligent routing engine for ip/optical cross-layer optimization in x-haul networks", in *2022 Optical Fiber Communications Conference and Exhibition (OFC)*, 2022, pp. 1–3.

[4] T. Tanaka and M. Shimoda, "Impact of operational mode selection and grooming policies on auxiliary graph-based multi-layer network planning", in *2021 European Conference on Optical Communication (ECOC)*, 2021, pp. 1–4. DOI: 10.1109/ECOC52684.2021.9606005.

[5] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms", *CoRR*, vol. abs/1707.06347, 2017. arXiv: 1707.06347. [Online]. Available: http://arxiv.org/abs/1707.06347.