ADMIRE+: Curiosity-Exploration-Driven Reinforcement Learning with Dynamic Graph Attention Networks for IP/Optical Cross-layer Routing

Zheng Zhang[(1)], Jiawei Zhang[(1)*], Huangxu Ma[(1)], Ruikun Wang[(1)], Zhiqun Gu[(1)], and Yuefeng Ji[(1)*]

[(1)] State Key Lab of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications (BUPT), Beijing, China, Corresponding author: *{zjw, jyf}@bupt.edu.cn.

Abstract *We propose ADMIRE+, an enhanced collaborative data-driven and model-driven routing engine in IP/optical networks. With the enhanced exploration by intrinsic curiosity and dynamic graph attention networks, 11.5% wavelength saving and 19.6% average latency reduction are demonstrated in a cross-layer network testbed compared to ADMIRE. ©2023 The Author(s)*

Introduction

The IP over WDM architecture is showing a new trend of flat development, enabled by the usage of pluggable coherent transceivers with high-capacity and long-distance [1-3]. With this technique, WDM optical signals can be directly emitted by an IP router and then sent to the Reconfigurable optical add/drop multiplexers (ROADMs) without intermediate Optical-Electrical-Optical (OEO) conversions. It would become a cost-effective solution to build the future metro core networks to support B5G/6G [4] and datacenter interconnection [5].

With this new trend and emerging applications in a metro area, routing in IP over WDM networks will become more complicated and dynamic. The IP/optical cross-layer routing, which essentially is a traffic grooming [6] problem, have been well investigated in a series of literatures [7,8]. Recently, due to the remarkable decision-making ability, deep reinforcement learning (DRL) was applied to traffic grooming [9-12]. In our previous work, we proposed a collaborative data-driven and model-driven routing engine (ADMIRE) [9-10], which uses convolutional neural network (CNN)-enabled DRL to dynamically modify the edge weights of auxiliary graph (AGEW). However, its performance is limited by the insufficient ability to explore the network environment. For one thing, the CNN-based agent in ADMIRE shows poor performance on precepting network states with Euclidean structure [13]. For another, the reward set as the negative number of newly activated wavelengths mostly is zero (i.e., sparse reward), which leads to ineffective action-value feedback. The above issues increase the difficulty of exploration and eventually leads to a local optimal solution.

In this paper, we propose ADMIRE+, an enhanced version of ADMIRE, which realizes an enhanced exploration to the network environment through intrinsic curiosity module (ICM) and dynamic graph attention networks enabled DRL (DGAT-DRL). By analysing state and action to generate curiosity towards environmental novelty, ICM adds an intrinsic reward to encourage agent to preferably explore (i.e., **internal exploration**). Meanwhile, DGAT is responsible for efficiently extracting the features of dynamic networks (i.e., **external exploration**). Through joint exploration to the environment, ADMIRE+ makes more intelligent RWA decisions for the network. We evaluate ADMIRE+ in a nine-node testbed, which achieves 11.5% wavelength saving and 19.6% average latency reduction than ADMIRE. Simulation results in a large topology also verify the superiority of ADMIRE+.

ADMIRE+ with Enhanced Exploration Mechanism

ADMIRE+ collaborates the exploration to the internal/external network environment and determines suitable AGEW to make optimized RWA decisions. Fig. 1 presents the procedure of ADMIRE+ which consists of three modules.

Module 1: DGAT-enabled external exploration. As shown in Fig. 1(a), we integrate DGAT [14] to the agent for efficiently analysing the state s_t from environment. Initially, the agent extracts the correlation between nodes through analysing the network state s_t^n, which is the remaining wavelength capacity of all links. The message that indicates the importance of neighbour node j to node i is called attention coefficient e_{ij}, which is respectively calculated and normalized in Eqs. (1-2), where $\vec{h}_i$ is the feature of node i, W is a linear transformation. Specifically, by executing the attention mechanism A at the end, the attention becomes dynamic and makes the agent explore the external environment better. Finally, as shown in Eq. (3), the updated feature $\vec{h'}_i$ of node i is generated by averaging the aggregated features based on K independent attention mechanisms. Fig. 1(a) illustrates the multi-head attention (with K = 3 heads) by node 1 on its neighbours.

$$e_{ij} = A^T LeakyReLU((W\vec{h}_i, W\vec{h}_j)) \quad (1)$$

$$\alpha_{ij} = softmax_j(e_{ij}) \quad (2)$$

$$\vec{h'}_i = \sigma(\frac{1}{K}\sum_{k=1}^{K}\sum_{j\in N_i}\alpha_{ij}^k W^k \vec{h}_j) \quad (3)$$

The network state s_t^n updated by DGAT is concatenated by the request state s_t^r, which includes one-hot encoded source/destination node and 24-hour traffic bandwidth demand. Then they are fed to a fully connected neural networks (FCNN) to output customized action a_t. By relational reasoning of graph message, DRL agent with DGAT is excelled at exploring the complex dynamic network.

Module 2: Cross-Layer routing with optimal AGEW. AG is a classical model for traffic grooming, which simplifies the cross-layer routing problem by finding the least cost path in a single layer. To dynamically optimize AG for each service flow, a_t are designed as eight kinds of AGEW, which are selected from discrete action spaces designed as the cartesian product of a set of empirical values [7]. As shown in Fig. 1(b), we assign the AGEW to tune the strategy of traffic grooming and make RWA decision for the current network state s_t. Accordingly, the network environment converts to the state s_{t+1}.

Module 3: Internal exploration motivated by ICM. After conducting the action a_t, DRL agent receives feedback from the environment and learns how to maximize the reward for a specific task. To avoid the adverse effects of sparse reward on exploration efficiency, we propose a dense feedback mechanism by using double rewards. In addition to the extrinsic reward r_t^e provided by the environment, we add an intrinsic reward r_t^i by importing ICM [15], which is shown in Eq. (4).

The generation process of the dense reward r_t is shown in Fig. 1(c). The agent forms a cognition and interest to the environment through predicting what will happen next [16]. Prediction error is used to measure how familiar the agent is with the next state s_{t+1}. Generally, higher familiarity with the next state means less curiosity. Firstly, we encode the states s_t, s_{t+1} to extract the state features $\emptyset(s_t)$, $\emptyset(s_{t+1})$. In the forward model F, we predict the next state feature $\hat{\emptyset}(s_{t+1})$ by the action a_t and the current feature $\emptyset(s_t)$. Specially, the intrinsic reward r_t^i is calculated as Eq. (5), where $\eta > 0$ is a scaling factor. To make the agent ignore irrelevant features, we introduce inverse model G with self-supervision to modify the encoded model.

$$r_t = r_t^e + r_t^i \quad (4)$$

$$r_t^i = \frac{\eta}{2}\left\|\hat{\emptyset}(s_{t+1}) - \emptyset(s_{t+1})\right\|_2^2 \quad (5)$$

$$\min_{\theta_I} L_F(\hat{\emptyset}(s_{t+1}), \emptyset(s_{t+1})) \quad (6)$$

$$\min_{\theta_I} L_I(\hat{a}_t, a_t) \quad (7)$$

By reward shaping, the optimization target of DRL changes to maximize the sum of reward r_t and minimize loss function of model F and G as shown in Eqs. (6-7), where θ_F and θ_I are parameters to optimize the prediction error between predicted values and actual values. The intrinsic reward from ICM motivates the agent to pursue novelty and surprising states for improving the internal exploration competence to the environment.

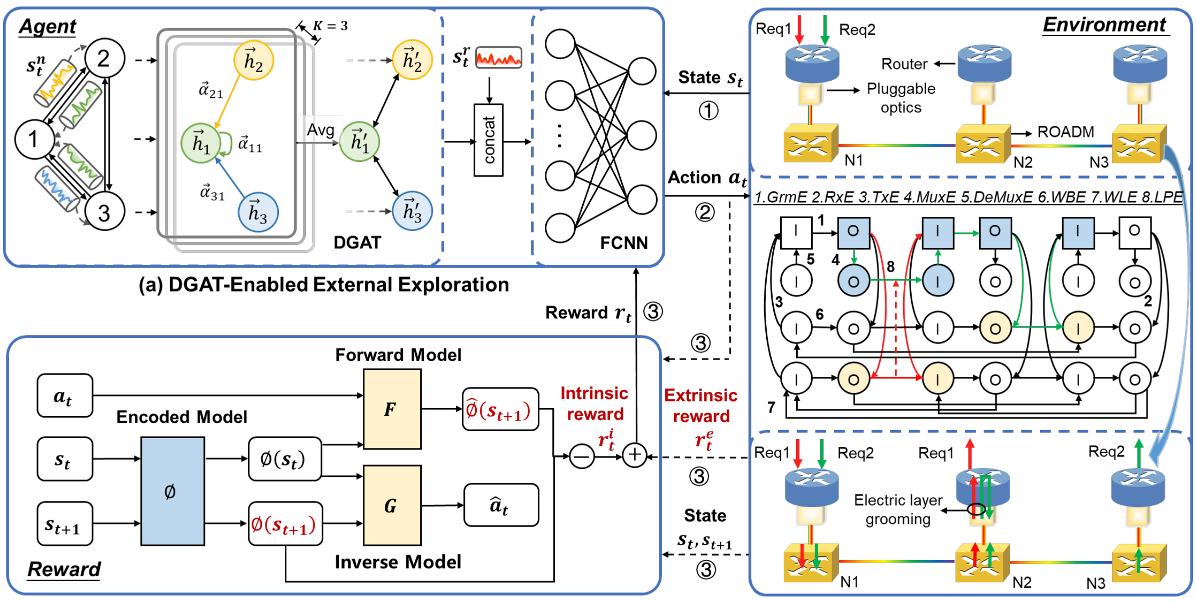

Fig. 1: Procedure of ADMIRE+. (a) DGAT-enabled external exploration, (b) Cross-layer routing with optimal AGEW, (c) ICM-motivated internal exploration.

Experimental Setup and Results

We demonstrate the performance and feasibility of ADMIRE+ in a cross-layer network testbed as shown in Fig. 2(a), which consists of nine hybrid optical-electrical switching nodes and twelve bidirectional links. Each node is set with a ROADM and an E-switch. Each fiber includes three wavelengths with total capacity of 3×10 Gbps. And the fiber length between nodes is set to 20 km. The traffic dataset with 100 bidirectional flows was collected from real base stations in Shenzhen, China. The 24 h variable traffic load is tidal over a day ranging from 0 to 1.2 Gbps and mostly peaks in the afternoon. Fig. 2(b) shows the verification process of ADMIRE+. Firstly, we load the dataset into the traffic generator and analyse (TGA). Meanwhile, connection requests are sent to the network management system (NMS) and the SDN controller. The SDN controller is used to execute ADMIRE+ and set up connections automatically. Then service flows generated by the TGA are routed in the testbed. The network statuses from ROADMs and E-switches are reported to the NMS and displayed in the graphical interface. At last, service flows will be looped back to the TGA for analysis.

We train the DRL agent of ADMIRE+ with Deep Q-Networks (DQN) [17]. The model adopts 3 attention heads and 3 DGAT layers with 32 hidden neurons. The following is a FCNN layer with 256 hidden neurons, applying a ReLU activation. In ICM, the encoded model consists of 2 FCNN layers with 512×256 hidden neurons, the forward and inverse model both have 3 FCNN layers with 256×64×32 hidden neurons.

We compare ADMIRE+ with DGAT-DRL, ADMIRE and the traditional grooming policy to minimize the number of wavelength-links (minWL) [8]. The above four algorithms are applied to determine the routing of 200 unidirectional flows under the objective of minimizing wavelength usage and average latency respectively. By enhancing exploration to the complex and sparse environment, ADMIRE+ achieves the best optimization performance. In Fig. 2(c), ADMIRE+ clearly outperforms DGAT-DRL, ADMIRE and minWL with wavelength utilization increased by 6.1%, 11.5% and 17.9% respectively. As shown in Fig. 2(d), ADMIRE+ reduces the end-to-end average latency of 200 flows with 12%, 19.6% and 28.8% than others. We also validated ADMIRE+ in GEANT [18] with 23 nodes and 37 bidirectional links, and we set the length of each fiber between nodes to 20 km. As shown in Fig. 2(e), ADMIRE+ realizes 3.7%, 7.1% and 10.3% wavelength saving than others. In Fig. 2(f), it reduces 4.5%, 10.8% and 27.1% average latency than others.

Conclusions

We proposed ADMIRE+ for routing optimization in IP/optical networks. Experimental results in a nine-node testbed indicated that ADMIRE+ can further optimize the RWA decision with enhanced exploration by ICM and DGAT. ADMIRE+ achieves 11.5% wavelength saving and 19.6% average latency reduction compared to our previous work ADMIRE. The simulated results in large topology also have a satisfactory performance, with 7.1% wavelength saving and 10.8% latency reduction.

Acknowledgements

This work is supported by the National Nature Science Foundation project (62021005, 61971055) and Xiaomi Young Talents Program.

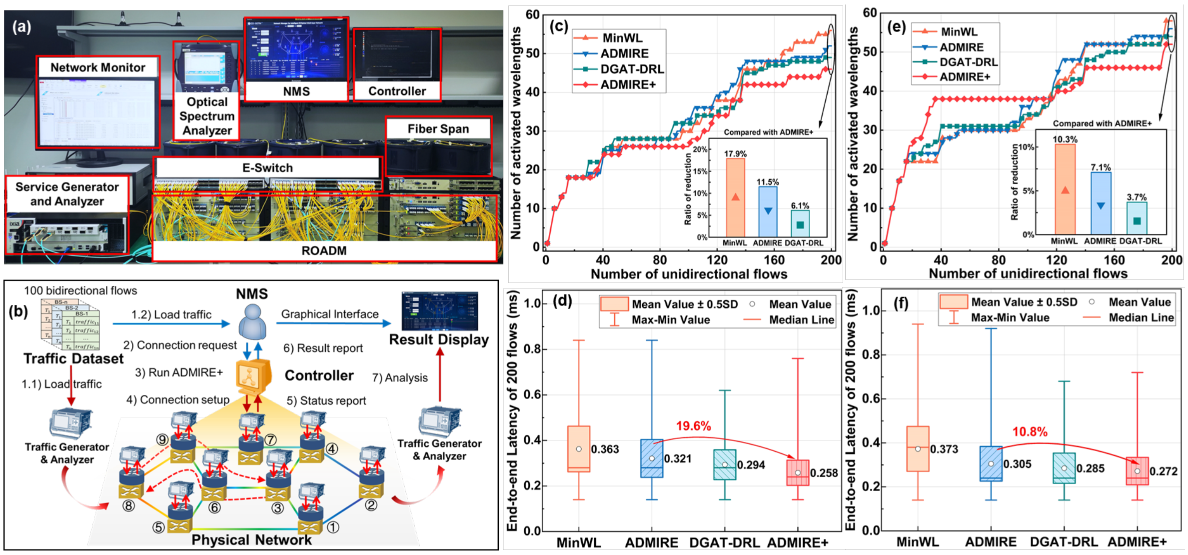

Fig. 2: (a) ADMIRE+ testbed, (b) Verification process, (c) No. of wavelength vs. No. of unidirectional flows in a 9-node topology, (d) End-to-end latency vs. 4 algorithms in a 9-node topology, (e) No. of wavelength vs. No. of unidirectional flows in GEANT, (f) End-to-end latency vs. 4 algorithms in GEANT.

References

[1] A. Gumaste, J. Pedro, P. Momtahan and H. Bock, "Optimal Line-Rates for IP-over-DWDM in Metro and Core Networks: Comparison of ZR+ and Xponder Architectures," in 2023 Optical Fiber Communications Conference and Exhibition (OFC), pp. 1-3, 2023, DOI: https://doi.org/10.1364/OFC.2023.Tu2D.1

[2] S. Melle, T.Zami, N.Rossi and B.Lavigne, "Optimal transponder technology for transporting 800 GbE services in IP-over-WDM backbone networks," in 2023 Optical Fiber Communications Conference and Exhibition (OFC), pp. 1-3, 2023, DOI: https://doi.org/10.1364/OFC.2023.Tu2D.2

[3] T. Zami, N. Rossi and B. Lavigne, "Optimal Channel Spacing for Next-Gen WDM Networking with 800ZR+ Elastic Optical Transponders," in 2023 Optical Fiber Communications Conference and Exhibition (OFC), pp. 1-3, 2023, DOI: https://doi.org/10.1364/OFC.2023.Tu2D.3

[4] W. Jiang, B. Han, M. A. Habibi and H. D. Schotten, "The Road Towards 6G: A Comprehensive Survey," in IEEE Open Journal of the Communications Society, vol. 2, pp. 334-366, 2021, DOI: https://doi.org/10.1109/OJCOMS.2021.3057679

[5] W. Fang, M. Lu, X. Liu, L. Gong and Z. Zhu, "Joint defragmentation of optical spectrum and IT resources in elastic optical datacenter interconnections," in Journal of Optical Communications and Networking, vol. 7, no. 4, pp. 314-324, April 2015, DOI: 10.1364/JOCN.7.000314.

[6] E. Modiano, "Traffic grooming in WDM networks," in IEEE Communications Magazine, vol. 39, no. 7, pp. 124-129, July 2001, DOI: 10.1109/35.933446

[7] W. Yao, G. Sahin, M. Li and Byrav Ramamurthy, "Analysis of multi-hop traffic grooming in WDM mesh networks," 2nd International Conference on Broadband Networks, Boston, MA, USA, 2005, pp. 165-174, vol. 1, DOI: 10.1109/ICBN.2005.1589612

[8] H. Zhu, H. Zang, K. Zhu, and B. Mukherjee, "A novel generic graph model for traffic grooming in heterogeneous WDM mesh networks," IEEE/ACM Transactions on Networking, vol. 11, no. 2, pp. 285-299, 2003, DOI: https://doi.org/10.1109/TNET.2003.810310

[9] 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), pp. 1-3, 2022, DOI: https://doi.org/10.1364/OFC.2022.M3F.4

[10] J. Zhang, Z. Chen, B. Zhang, R. Wang, H. Ma and Y. Ji, "ADMIRE: collaborative data-driven and model-driven intelligent routing engine for traffic grooming in multi-layer X-Haul networks," in Journal of Optical Communications and Networking, vol. 15, no. 2, pp. A63-A73, 2023, DOI: 10.1364/JOCN.470690

[11] M. Shimoda and T. Tanaka, "Mask RSA: End-To-End Reinforcement Learning-based Routing and Spectrum Assignment in Elastic Optical Networks," 2021 European Conference on Optical Communication (ECOC), Bordeaux, France, 2021, pp. 1-4, DOI: https://doi.org/10.1109/ECOC.52684.2021.9606169

[12] T. Tanaka, "Adaptive Traffic Grooming Using Reinforcement Learning in Multilayer Elastic Optical Networks," in 2023 Optical Fiber Communications Conference and Exhibition (OFC), pp. 1-3, 2023, DOI: https://doi.org/10.1364/OFC.2023.Tu2D.6

[13] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang and P. S. Yu, "A Comprehensive Survey on Graph Neural Networks," in IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 1, pp. 4-24, Jan. 2021, DOI: 10.1109/TNNLS.2020.2978386

[14] S. Brody, U. Alon, and E Yahav, "How attentive are graph attention networks?". arXiv preprint arXiv:2105.14491, 2021, DOI: https://doi.org/10.48550/arXiv.2105.14491

[15] D. Pathak, P. Agrawal, A. A. Efros, and T. Darrell, "Curiosity-Driven Exploration by Self-Supervised Prediction," 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 2017, pp. 488-489, DOI: https://doi.org/10.1109/CVPRW.2017.70

[16] H. Hu, S. Song and G. Huang, "Self-Attention-Based Temporary Curiosity in Reinforcement Learning Exploration," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 51, no. 9, pp. 5773-5784, Sept. 2021, DOI: 10.1109/TSMC.2019.2957051

[17] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," arXiv preprint arXiv:1312.5602, 2013, DOI: https://doi.org/10.48550/arXiv.1312.5602

[18] S. Uhlig, B. Quoitin, J. Lepropre, and S. Balon, "Providing public intradomain traffic matrices to the research community," ACM SIGCOMM Computer. Communication Review. 36(1), 2006, pp. 83-86, DOI: https://doi.org/10.1145/1111322.1111341