Graph Sequence Attention Network-Enabled Reinforcement Learning for Time-Aware Robust Routing in OSU-Based OTN

Huangxu Ma, Jiawei Zhang*, Yuefeng Ji*
State Key Lab of Information Photonics and Optical Communications, Beijing Univ. of Posts and Telecommunications (BUPT), Beijing, China
*Corresponding author emails: *{zjw, jyf}@bupt.edu.cn*

Abstract: We propose a time-aware robust routing in OSU-based OTN through the newly designed graph sequence attention network-enabled reinforcement learning. Simulation results show > 28% OSU frame loss reduction compared to the baselines. © 2022 The Author(s)

1. Introduction

Benefit from the requirements of Fixed 5G (F5G) applications on stable and secure networking with high service level agreement (SLA), optical transport network (OTN) has been sunk from backbone to metro-aggregation enabled by optical service unit (OSU) [1]. For one thing, with flexible resource scheduling of OSU-OTN, network operators can allocate resources on-demand to accommodate future time-varying service requirements with high resource utilization. For another, long-term fixed routes are expected to ensure SLA, as the adjustment of established pipelines in OSU-OTN will bring about service degradation. Such an anticipatory resource allocation within fixed routes, named time-aware routing, should be driven by a suitable predictor. Nevertheless, traffic prediction can only predict general trends of traffic. Those traffic growth beyond expectations, such as sudden social events, will cause a serious impact as it may insufficient to be carried by the allocated route, and then loss of OSU frame (LOF) occur. Therefore, it's significant to enhance the robustness of time-aware routing, that is, to reduce LOF caused by unexpected traffic growth.

The robustness of time-aware routing can be effectively enhanced by extending load balancing to the time dimension, i.e. balancing the load of each link at each moment over a period. Recently, using reinforcement learning (RL) for network load balancing has shown significant advantages over heuristics in online optimal decision-making [2, 3]. However, previous RL-based studies only achieve balance by dynamically adjusting routing at each moment, ignoring the balancing requirements of long-term deployments in OSU-OTN. Moreover, the existing neural networks applied to RL agent can hardly meet the need for efficient perception of topologies with time-domain states. As one of the most famous variants of GNN which had been used for routing optimization combined with RL [4], graph attention network (GAT) introduces the attention mechanism to achieve adaptive matching of weights to neighbors [5], enabling efficient extraction of graph information and powerful generalization across topologies. Despite the advantages of GAT over GNN and even traditional convolutional neural network (CNN) for single-moment routing with RL, the optimization in reducing LOF is still limited when making decisions for time-aware robust routing, as no time-series correlation of bandwidth-occupied states in the topology are considered.

In this paper, we first propose a time-aware robust routing strategy in OSU-OTN, where a newly designed graph sequence attention network-enabled RL (GSAT-RL) is adopted. In the RL agent, a gated recurrent unit (GRU) is introduced in GSAT to enhance the perception of the time-series. GSAT-RL tends to balance bandwidth requirements of time-series in the network. We compare GAST-RL with baselines including classical heuristic and RL-based algorithms under real traffic. Results show that the proposed strategy can achieve much lower LOF caused by traffic growth beyond expectations and exhibits a strong generalization ability.

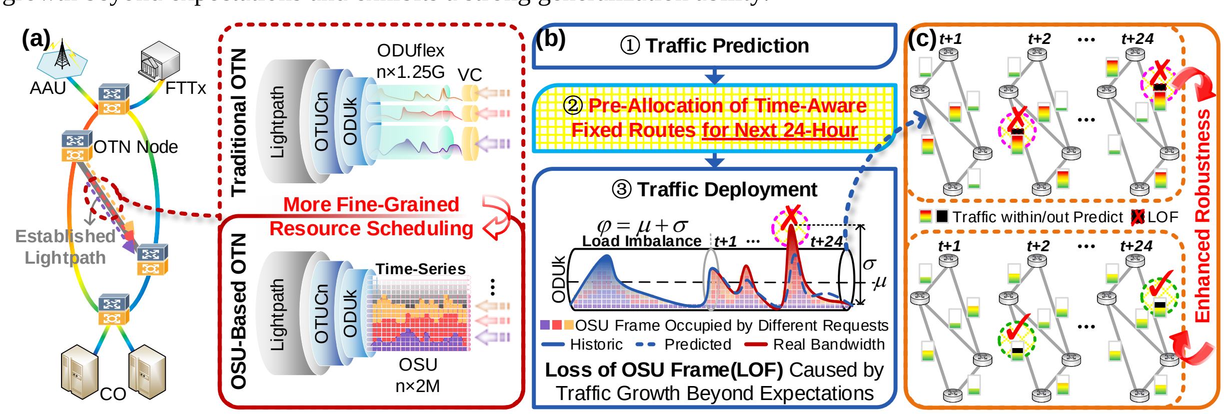

Fig. 1. (a) Architecture of OSU-OTN; (b) Overall process for time-aware routing; (c) Enhanced robustness of the network.

2. Time-Aware Robust Routing in OSU-OTN

Fig. 1 (a) shows a metro-aggregation network supported by OSU-OTN. Requests are accessed by active antenna units (AAU) or fiber access terminals, and transmitted to one of the candidate central offices (CO). OSU-OTN enables higher resource utilization than traditional OTN, as massive requests uniformly encapsulated by the fine-grained OSU frames can be mapped into one optical data unit (ODUk) with the guarantee of hard isolation [1]. We perform routing in the logical topology where each edge is an established lightpath with one optical channel transport unit (OTUCn), modelled as a directed graph. Multiple ODUk in a lightpath are represented as parallel edges of the graph.

For improving utilization of bandwidth and providing long-term fixed routes, traffic prediction is used to predict bandwidth requirements for all services over the next 24-hour, which are used for pre-allocating time-aware routes as shown in Fig. 1 (b). The remaining bandwidth for each ODUk is expressed as a 24-hour time-series, and services are reserved in selected routes for bandwidth throughout the next 24-hour. However, sudden traffic growth due to unexpected factors in regional network load is unpredictable and will likely result in exceeding the remaining bandwidth of ODUk after deployment. In which case, the services deployed in these ODUk will lose OSU frames beyond capacity. To mitigate the impact of sudden traffic growth, the robustness of time-aware routing should be enhanced, i.e. the traffic load should be balanced across the network in the time dimension as shown in Fig. 1 (c). The network after balancing shows a more positive effect in carrying sudden traffic growth with lower LOF. We define $\varphi_l = \mu_l + \sigma_l$ as the degree of time-aware load imbalance of link l, where μ_l and σ_l denote the mean and standard deviation of the time-series. φ is a more comprehensive representation of time-aware load imbalance than a direct mean value or peak value, as it reflects both average bandwidth usage and multiple bandwidth peaks. Furthermore, the maximum value of φ in the network is Φ, defined as the degree of imbalance of the network. We minimize Φ while routing to enhance the robustness of the network. We assume that the global routing optimization is triggered when the load is at the lowest point every day.

3. GSAT-RL Algorithm

Through rounds of training between the network environment and the agent, RL achieves a reward-driven search for the optimal action under a certain state. The overall structure of GSAT-RL specifically applied for time-aware robust routing is shown in Fig. 2, where the states, actions, and rewards are defined as follows. 1) **State**: The network state includes the edge state and the node state, which respectively represent the remaining capacity of each link in the next 24-hour and the betweenness centrality of each node. The request state consists of its one-hot encoded source node, destination node, and predicted bandwidth demand of 24-hour. 2) **Action**: The action is chosen from the action space which is set to all routes between all possible source and destination nodes by the RL agent. 3) **Reward**: When a request is deployed on the selected route, a reward value, set to $\max(\varphi_{l_1}, \varphi_{l_2}, \dots, \varphi_{l_n})$, is calculated promptly to evaluate the quality of the chosen action for the current state, driving the agent to select routes with a more balanced strategy. Note that $l_1, l_2, \dots, l_n$ are the edges where the bandwidth usage changes after action is performed.

For the feature extraction of the graph with time-series states in the RL agent, we propose to introduce GRU into GAT to improve the attention mechanism. GSAT enhances the ability to analyze the correlation of time-series between neighbor nodes and enables more effective graph information transfer for RL agent which is shown in Fig. 2 by assigning different attention weights to neighbors. 1) **Message Initialization**: Convolve the node state v of each node with the edge state e of all its adjacent incoming edges. After which they are summed up as the initial message $\vec{m}$ of each node. 2) **Message Update**: For each node whose feature is time-series, GRU is used for feature expansion, replacing the matrix W in GAT. The output of GRU is given as Eq. (1-4), where $W_z, W_r, W_c, U_z, U_r, U_c$ are trainable parameters, $\vec{m}_i^t$ is the current input value of node i [6]. Note that h_i^t implies the information transmitted at the previous time. We keep $H_i = [h_i^0, \dots, h_i^T]$ as the expanded message of node i and pass it to calculate the attention weight between neighbor nodes. For node i and its neighbor node j, the attention weight $\vec{\alpha}_{ji}$ is calculated as Eq. (5), where a represents the weight of a single-layer feedforward network used to learn the relative importance between adjacent nodes. Furthermore, multi-head attention can be introduced for a better performance by expanding the length of H_i from GRU and splitting it to k rows. Then the updated node message $\vec{m}_i'$ is given in Eq. (6). 3) **Q-Value Calculation**:

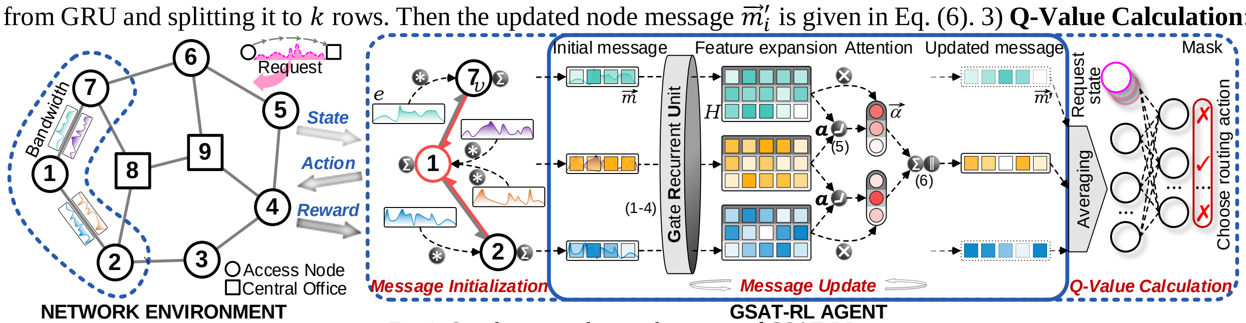

Fig. 2. Simulation topology and structure of GSAT-RL.

The nodes' messages after multiple rounds of updates are averaged and fed into multiple hidden layers along with the request state. The outputs are the Q-value of 132 actions. We mask actions for which source node, destination node, remaining bandwidth of 24-hour, and delay ($\leqslant$100μs) do not meet the current requirements by adding $-inf$ to the corresponding Q-value while training. After selecting the route with the maximum Q-value, the ODUk with the lowest φ is selected by each segment.

In our simulation, we trained the RL agent with Rainbow DQN. The agent is designed as a GSAT model, which adopts 3 attention heads and 3 GSAT layers with 16 hidden features. The following are 3 hidden layers with (256, 256, 132) hidden neurons correspondingly. Hyperparameters are shared among the RL-based algorithms.

$$z^t = \sigma(W_z \vec{m}_i^t + U_z h_i^{t-1}) \quad (1) \qquad r^t = \sigma(W_r \vec{m}_i^t + U_r h_i^{t-1}) \quad (2) \qquad \tilde{h}_i^t = \tanh(W_c \vec{m}_i^t + U_c(r^t \odot h_i^{t-1})) \quad (3)$$

$$h_i^t = (1 - z^t) \odot h_i^{t-1} + z^t \odot \tilde{h}_i^t \quad (4) \qquad \vec{\alpha}_{ji} = \mathrm{softmax}_j(\mathrm{ReLU}(\boldsymbol{a}[H_i || H_j])) \quad (5) \qquad \vec{m}_i' = ||_{k=1}^{K} \sigma\left(\textstyle\sum_{j \in \mathcal{N}_i} \vec{\alpha}_{ij}^k H_j^k\right) \quad (6)$$

4. Simulation Results and Analysis

We consider a topology as shown in Fig. 2, with 7 access nodes (*Node 1-7*), 2 CO nodes (*Node 8-9*), and 12 established lightpaths, each of which uses a pair of 200Gbit/s optical modules and contains two ODU4 (100Gbit/s per ODU4). The distances of lightpaths are randomly generated from 5 to 30km. The normalized bandwidth requirements are extracted from the real base stations in Shenzhen, China.

We have compared our GSAT-RL with widely used first fit (FF), greedy algorithm (GBA), CNN-RL, GNN-RL, and GAT-RL. The link with the highest φ will limit the imbalance of the whole network, shown as Φ in Fig. 3 (a). We use the 5-45 real bandwidth of the requests to be carried on the routes calculated with the predicted ones to verify the ability in reducing LOF, which is shown in Fig. 3 (b). As can be seen from the above two figures, the above algorithms are less effective in enhancing the robustness of routing and reducing the global LOF than GSAT-RL, which achieves a frame loss rate of 1.05% when the number of requests reaches 45, enabling a 28% and 56% reduction in LOF compared to the GAT-RL and GBA respectively. Correspondingly, the higher the degree of imbalance of the network, the poorer the robustness, leading to a larger LOF rate. FF selects the first available route for the request's predicted bandwidth of 24-hour, resulting in poor results. GBA is locally optimal and does not consider the impact on the following deployments. CNN-RL requires inefficient feature extraction in the adjacency matrix of the graph. GNN-RL sums the messages of each neighbor nodes to update directly. GAT-RL incorporates an attention mechanism that measures the relevance between nodes and thus achieves far better results than GNN-RL, yet the attention mechanism is calculated without considering the correlation of the time-series. Fig. 3 (c) shows the convergence process of the above RL-based algorithms. GSAT-RL has better convergence performance as it has a more powerful ability to extract complex information with time-series from topology. Finally, we also verified the generalization ability for topology of the GSAT model in Fig. 3 (d). We remove lightpath 1-2 as a new topology, and update the action space to retrain the RL agent from 0 to 300 episodes. The results show that retaining the GSAT weights trained in the original topology leads to faster convergence and comparable LOF results after convergence compared to retraining the GSAT model.

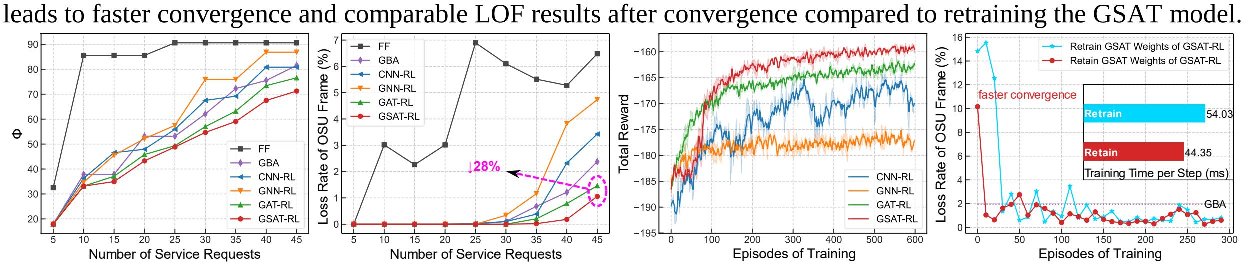

Fig. 3. (a) Φ vs. requests; (b) LOF rate vs. requests; (c) Convergence process; (d)Validation of generalization.

5. Conclusions

We proposed a novel GSAT-RL algorithm for the time-aware robust routing in OSU-OTN to reduce OSU frame loss caused by unexpected traffic growth. Simulation results validated our inspiration that GSAT-RL had a better performance in enhancing robustness and reducing the global LOF than the state-of-art algorithms. The generalization ability for topologies of the GSAT model was also verified.

Acknowledgement

This work is supported by the National Key R&D Program of China (No. 2018YFB1800802), the National Nature Science Foundation of China Project (No. 61871051).

References

[1] ITU-T, "Requirements and Considerations on hitless adjustment of OSU bandwidth," T17-SG15, 2021.
[2] T. Mai, et al., "Multi-Agent Actor-Critic Reinforcement Learning based In-network Load Balance," GLOBECOM, 2020.
[3] P. Sun, et al., "DeepMigration: Flow Migration for NFV with Graph-based Deep Reinforcement Learning," ICC, 2020.
[4] P. Sun, et al., "Combining deep reinforcement learning with graph neural networks for optimal VNF…," IEEE Commun. Lett., 2021.
[5] P. Veličković, et al., "Graph attention networks," ICLR, 2018.
[6] J. Chung, et al., "Empirical evaluation of gated recurrent neural networks on sequence modeling," arXiv preprint, 2014.