Spectrum Adaptive Awareness Routing and Spectrum Allocation Based on Reinforcement Learning

Zechuan Guan
Beijing Institute of Technology
School of Information and Electronics
Beijing, China
3120210722@bit.edu.cn

Fu Wang
Beijing University of Posts and Telecommunications
School of Electronic Engineering
Beijing, China
wangfu6672@gmail.com

Ze Dong
Beijing Institute of Technology
School of Information and Electronics
Beijing, China
zdong9@bit.edu.cn

Zhipei Li
Beijing Institute of Technology
School of Information and Electronics
Beijing, China
lizhipei@bit.edu.cn

Huan Chang
Beijing Institute of Technology
School of Information and Electronics
Beijing, China
changhuan@bit.edu.cn

Ran Gao
Beijing Institute of Technology
School of Information and Electronics
Beijing, China
gaoran@bit.edu.cn

***Abstract*—We propose a spectrum adaptive awareness routing and spectrum allocation (SAA-RSA) scheme based on reinforcement learning in elastic optical networks. This scheme combines spectrum fragment awareness and reinforcement learning to realize adaptive spectrum allocation. Based on the proposed scheme, the action space of RL is optimized, thus improving the performance of network and reducing traffic blocking probability. The simulation results show that when the network load is 300 Erlang, compared with traditional KSP-FF method and reinforcement learning method without considering spectrum fragment, the scheme can reduce blocking probability by 27.56% and 12.15% respectively.**

Keywords—RSA, Reinforcement learning, Deep Q-network, Elastic optical networks, Spectrum Slice Degree

I. Introduction

Elastic optical network (EON) can effectively utilize spectrum resources and allocate required frequency slots on demand for different service requirements. EON is considered as one of the key technologies of next-generation optical networks with great potential due to its flexibility. However, Due to the variability of the optical path spectrum, the EON spectrum is severely fragmented and the utilization of the spectrum is low. In order to improve the spectrum utilization, we need to fully consider the network state and select the appropriate spectrum when the optical path is established.

The key problem of EON is routing and spectrum allocation (RSA), which is a NP problem mathematically, and the best solution cannot be found through traditional methods. Traditional methods use heuristic algorithms with high time complexity. In recent years, reinforcement learning has been widely studied in optical networks [1]. The performance of RL in resource allocation problems is excellent. Chen et al. extracts EON state characteristics through DNN to achieve RL strategy [2]. In order to comprehensively extract EON features such as topology information and path information, Xu et al. further introduces FNN and RNN to optimize RL strategies [3]. Based on probabilistic failure model, Jiao et al. applies Q learning algorithm to RL strategy and establishes node reliability analysis model [4]. Zhang et al.propose a scheme to ensure EON's ability to cope with failures under high network traffic load [5]. Zhang et al. introduces DQN into DQN algorithm and modeling through Markov decision process (MDP), the long-term blocking probability in online RSA can be effectively reduced [6]. However, the algorithm used for spectrum allocation is not good enough, which makes the alternative spectrum actions in the RL strategy action space numerous and poor quality. The existing RL strategies tend to generate high spectral fragmentation rate and greatly increase the complexity of the algorithm.

In this paper, we propose a spectrum adaptive awareness RSA (SAA-RSA) scheme based on reinforcement learning. We define a spectrum slicing degree to reflect the condition in which the service allocation segments (SSD) the spectrum in the link. RSA model is established in the context of reinforcement learning. Then, numerical simulation is carried out in NSFNET network topology to verify the performance of the proposed algorithm. The simulation results show that when the network load is 300 Erlang, compared with traditional KSP-FF method and reinforcement learning method without considering spectrum slice degree, the scheme can reduce blocking probability by 27.56% and 12.15% respectively.

II. PROPOSED SSD-DDQN ALGORITHM

In an elastic optical network, as dynamic arriving services are constantly established and released, spectrum fragments on links may cause that incoming services cannot find proper connections, causing service congestion, and seriously affecting network performance. In this section, we propose a spectrum slice degree to reduce the spectrum fragments generated by spectrum allocation. Then, we introduce the Double DQN model in reinforcement learning and combine it with the spectrum slice degree.

A. Spectrum Slice Degree

During the dynamic operation of optical networks, the arrival of new services will cause the continuous spectrum resource segmentation on the link where the services reside. SSD is used to reflect the degree to which the spectrum on a link is segmented by a new connection. The expression is as follows:

$$SSD = \sum_{i=1}^{N_k} (n_{l_i} + n_{r_i}) \Big/ N_k \qquad (1)$$

Corresponding author: Fu Wang (wangfu6672@gmail.com). This work is supported by National Key Research and Development Program of China (2022YFB2902500); National Natural Science Foundation of China (62206017, 61835002, 61935005); Open Fund of IPOC (BUPT) (IPOC2020A006).

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

Where, N_k is the length of the link. n_{l_i} and n_{r_i} represent whether there is idle spectrum in the left and right adjacent parts of the spectrum allocated on the i link. If there is, it is denoted as 1; otherwise, it is 0. An SSD value higher than 1 indicates that the new service generates more segments on the link than the original network, while a value lower than 1 indicates that the allocated spectrum segments are less than the original network.

Assume that there are 6 links in the network, with 8 frequency gaps on each link, as shown in Fig 1. The small blue squares on the link represent occupied gaps, and the small white squares represent unoccupied gaps. Assume that new services are allocated on the network, occupying links 2, 3, and 4, and the service request bandwidth is 2. There are two alternatives, *a* and *b*, that can meet the business requirements. In the traditional RSA algorithm, two schemes are randomly selected. However, the SSD value of scheme *a* and Scheme *b* is 1.33 and 0.33 respectively when spectrum slice degree is calculated respectively. The algorithm based on SSD will choose scheme b, which generates less spectrum segmentation.

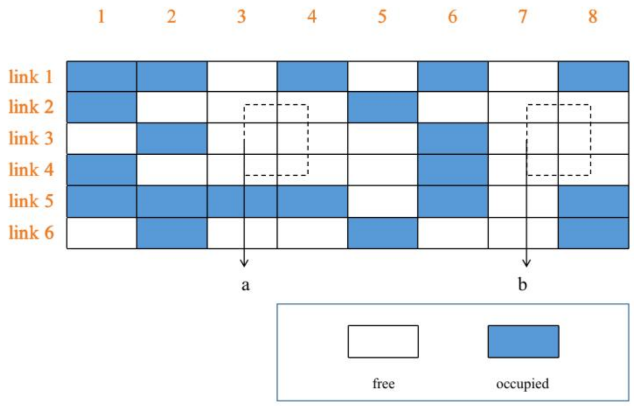

Fig. 1. Example of spectrum slice degree

When the optical network is faced with differentiated arrival service requirements, SSD can provide better spectrum selection and improve the coherence of the spectrum in the network. To improve spectrum utilization and reduce bandwidth blocking probability.

B. *Reinforcement Learning Model*

We define the state space representation, action space and reward function of RSA. The reinforcement learning model is shown in Fig 2.

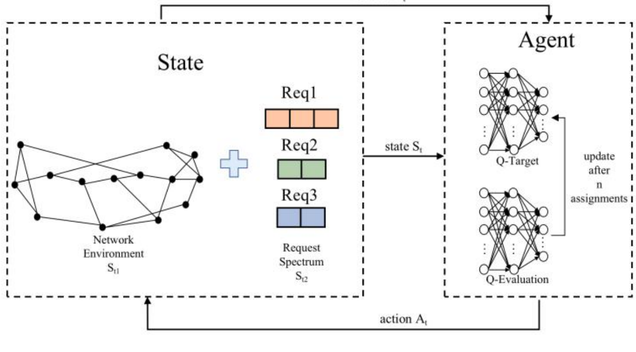

Fig. 2. The reinforcement learning model

- State space: The state space contains two types of characteristics to fully reflect the business request information and optical network status. One is the service request information, including the source node, destination node, spectrum width of the request, and duration of the service. The other is the network environment information including the candidate link and its internal spectrum usage. Service request information includes source node *o*, destination node *d*, service duration *t*, and spectrum width *w* required by services.The network environment information includes the initial index w_K^{J1} and width w_K^{J2} of the *J* candidate spectrum block in the *K* link, the total width w_K^{J3} and average width w_K^{J4} of the available spectrum block in the *K* candidate link, and the spectrum fragment rate w_K^{J5} on the link. The above information forms a $1 \times [2N+2+K(2J+3)]$ state space matrix, where *N* represents the number of nodes in the network.
- Action space: In the application scenario of RSA, the frequency gap in the whole spectrum is constantly occupied and released, and the request width of each service is uncertain. Therefore, the action space changes with the network status. According to the spectrum slicing degree proposed above, every time a routing path is determined, its corresponding optimal spectrum allocation can also be determined, which means that we can get a smaller but better motion space. Therefore, we first find *K* alternative paths through the KSP algorithm, and then find out the optimal spectrum allocation corresponding to each alternative path according to SSD. If there are less than *K* available alternative links, the missing link is set as the first available link. If the number of available links is 0, service congestion occurs. The size of the action space is *K*.
- Reward function: The RSA scheme based on reinforcement learning proposed by us aims to reduce blocking probability of business requests, so we set feedback reward as follows: when the Agent tries to make connection allocation, the environment will give back a reward. Reward=1 if the selected action satisfies the service needs. Reward=-1 if the action causes service congestion.

C. *Training*

DQN [7] has two neural networks with the same structure but different parameters. The target network Q_{tar} is used to store historical information, has historical data, and periodically updated; The evaluation network Q_{eva} has the latest parameters and calculates the output action for the incoming business request.The formula for calculating the target Q value is as follows:

$$Y_t = R_{t+1} + \gamma \max_a Q(S_{t+1}, a; \theta_t') \quad (2)$$

However, the calculation method of the target Q value in DQN may lead to the problem of overestimation. Therefore, we introduce the Double DQN algorithm [8]. This algorithm changes the calculation method of target Q value, separates selection from evaluation, and avoids the problem of overestimation.The Q value of Double DQN is calculated by the following formula:

$$Y_t = R_{t+1} + \gamma Q(S_{t+1}, \arg\max_a Q(S_{t+1}, a; \theta_t^{'})) \quad (3)$$

TABLE I shows the training process of RSA problem based on SSD-DDQN algorithm.

TABLE I. TRAINING BASED ON SSD-DDQN SCHEME

Algorithm 1:Training based on SSD-DDQN

Initialize evaluation Q-network Qeva;
Initialize target Q-network Qtar;
for** episode = 1, 2, 3, ⋯ **do
    Initializing the optical network;
    Observe the state space St ;
    Obtain the action space At through SSD;
    for** i from 1 to k **do
        Select the action at by Qeva;
        Get the reward value rt and state st+1;
        Store the data in Qeva;
    end
    Update the data from Qeva to Qtar;
end

III. NUMERICAL SIMULATIONS

Simulation experiments are carried out in NSFNET topology with 14 nodes and 21 bidirectional links. The number of available spectrum blocks on each link is 256. The number of spectrum blocks required by network requests appears randomly in 2, 4, 8, and 16 with equal probability. Service arrival follows the Poisson distribution, and the duration of each service follows the negative exponential distribution. The traffic load is the product of service arrival rate and service duration.

In DDQN, the attenuation value γ for future reward is set to 0.99, the learning rate α for error is set to 0.0001, and the random exploration rate ε is set to 0.1.

Fig. 3 shows the network blocking probability of SSD-DDQN algorithm with different K values. Blocking probability decreases with the increase of K value when K value is 1, 2 and 3 respectively. This is because the larger the K value is, the larger the action space in action selection, and the greater the possibility to select the path with the best return, therefore, the better business blocking probability performance presented by SSD-DDQN algorithm.

Fig 4 compares the traditional KSP-FF algorithm [9], the DQN algorithm without considering spectrum slice, and the SSD-DDQN algorithm proposed in this paper. Since SSD-DDQN fully considers the impact of the segmentation degree of the link spectrum on the network, the blocking probability is the lowest. The experimental results show that when the network load is 300 Erlang, the SSD-DDQN can reduce the blocking probability by 27.56% and 12.15% respectively, compared with the KSP+FF method and DDQN without considering the spectrum slice degree.

IV. CONCLUSION

In this paper, an RSA scheme based on spectrum slice degree combined with reinforcement learning is proposed. The simulation results show that when the network load is 300 erlang, the proposed scheme can reduce the blocking probability by 27.56% and 12.15% respectively, compared with the KSP+FF method and reinforcement learning method without considering the spectrum slice degree. It can make better decisions and efficiently find more reliable paths for requests.

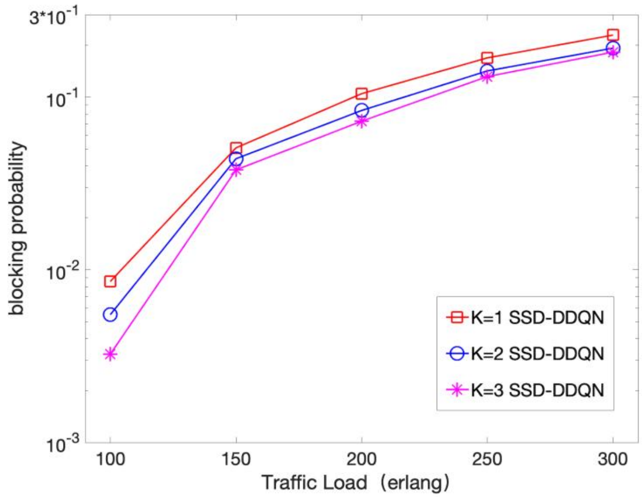

Fig. 3. Blocking probability of SSD-DDQN with different K values

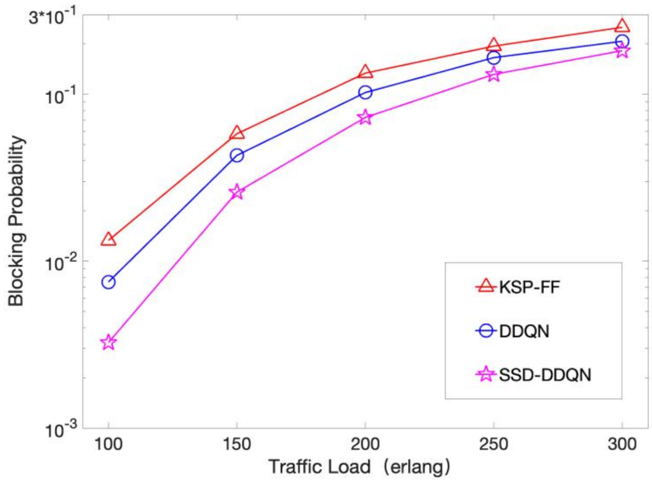

Fig. 4. Comparison between KSP-FF, DDQN, and SSD-DDQN

REFERENCES

[1] Wang, F. , et al. "Dynamic routing and spectrum assignment based on multilayer virtual topology and ant colony optimization in elastic software-defined optical networks." Optical Engineering 56.7 (2017).

[2] Chen, X. , et al. "DeepRMSA: A Deep Reinforcement Learning Framework for Routing, Modulation and Spectrum Assignment in Elastic Optical Networks." Journal of Lightwave Technology pp.99 (2019).

[3] L, Xu. , et al. "Deep Reinforcement Learning-Based Routing and Spectrum Assignment of EONs by Exploiting GCN and RNN for Feature Extraction." Journal of Lightwave Technology vol. 40, no. 15, pp.4945-4955 (2022).

[4] Y. Jiao. , et al. "Reliability-Oriented RSA Combined with Reinforcement Learning in Elastic Optical Networks," International Conference on Optical Communications and Networks (ICOCN) pp.1-3 (2022).

[5] Z. Zhang. , et al., "A Multi-Agents Survivable Deep Reinforced Routing and Spectrum Assignment Algorithm in Elastic Optical Network." International Conference on Electronics Information and Emergency Communication (ICEIEC) pp.1-4 (2022).

[6] Zhang, J. , et al. "Online routing and spectrum allocation in elastic optical networks based on dueling Deep Q-network." Computers & Industrial Engineering 173 108663 (2022).

[7] Mnih, V. , et al. "Human-level control through deep reinforcement learning." Nature 518 529–533 (2015).

[8] Van Hasselt, H. , et al. "Deep Reinforcement Learning with Double Q-Learning." Proceedings of the AAAI Conference on Artificial Intelligence 30(1) (2016).

[9] Yawei,Yin. , et al. "Spectral and Spatial 2D Fragmentation-Aware Routing and Spectrum Assignment Algorithms in Elastic Optical Networks [Invited]." J. Opt. Commun. Netw 5 A100-A106 (2013).