DEEP REINFORCEMENT LEARNING-BASED DYNAMIC LIGHTPATH PROVISIONING FOR ELASTIC OPTICAL NETWORKS

Hai-Chau Le and Van Duc Khuat

Posts and Telecommunications Institute of Technology, Hanoi, Vietnam

Abstract: In this paper, we develop a deep reinforcement learning-based routing, modulation format and spectrum assignment (RMSA) algorithm for elastic optical networks capable of provisioning dynamically lightpath services. In order to enhance the network performance, the developed RMSA exploits deep reinforcement learning (DRL) mechanism for selecting efficient route and spectral resource by learning experiences of dynamic lightpath provisioning. Numerical simulations have been utilized to estimate the performance of the elastic optical network applied the proposed DRL-based RMSA solution. The obtained results demonstrate that our proposed network solution outperforms the conventional shortest path algorithm-based one significantly and offers a notable performance enhancement in terms of blocking probability and accepted traffic volume.

Keywords: Deep reinforcement learning, elastic optical network, routing and spectrum assignment, network control algorithm.

I. INTRODUCTION

To cope with the Internet traffic explosion and the popular adoption of new networking paradigms, development of cost-effective, dynamic and heterogeneous bandwidth-abundant flexible optical backbone networks is essential [1, 2]. Recently, elastic optical networks (EONs) have been emerged and realized as one of the most promising networking technologies for the next-generation backbone networks [3]. Compared to traditional fixed-grid (e.g., 50 GHz) wavelength-division multiplexing (WDM) network, EON enables flexibly setting-up bandwidth-variable superchannels by grooming series of finer-granularity subcarriers and adapting the modulation formats according to the QoT of lightpaths [4, 5]. Thanks to that, elastic optical networks are capable of provisioning dynamic bandwidth-flexible end-to-end connection and offer service providers the flexibility to customize their infrastructure dynamically according to application requirements [6, 7]. In EONs, routing and spectrum allocation includes three sub-problems, namely routing, modulation and spectrum assignment (RMSA/RSA), that is known as an NP hard. Up to now, many studies have intensively investigated the routing, modulation and spectrum assignment (RMSA) problem in order to fully exploit the benefits of EONs. However, many technological issues and challenges must be dealt with to realize and commercialize elastic optical networks due to the requirements of more complicated network planning and more sophisticated optical-path provisioning schemes [4-6].

Furthermore, different from static RMSA problems for which explicit optimization models can be formulated, dynamic RMSA problems that target the optimization of provisioning lightpaths dynamically and flexibly are more challenging [8-10]. Due to the time-varied arrivals and departures of requested lightpaths as well as the uncertainty of future traffic demands, EON state becomes dramatically destabilized and hence, the efficiency of the optimizations based on the current state is deteriorated [6, 8]. Unfortunately, the existing works only employ pre-determined (fixed) RMSA policies regardless of the time-varying EON states or depend on simple empirical policies based on manually extracted features, i.e., lack of comprehensive perceptions of the holistic EON states, and consequently are unable to achieve real adaptive service provisioning in EONs.

In the meantime, recent advances in deep reinforcement learning (DRL) have demonstrated beyond human-level performance in handling large-scale online control tasks [11, 12]. The application of DRL in the optical communication and networking domain has received intensive research interests and opens a new approach. DRL parameterizes action policies with deep neural networks (DNNs) [13, 14] that can perceive complex system states from high-dimensional input data, such as, images, and traffic matrices. By accumulating action experiences from repeated interactions with the target systems and by reinforcing actions leading to higher rewards, DRL is able to learn successful policies (i.e., correct configurations of the DNNs) progressively. Several applications of emerging deep reinforcement learning (DRL) techniques in EONs for enabling an autonomic

Contact author: Hai Chau Le
Email: chaulh@ptit.edu.vn
Manuscript received: 25/8/2022, revised: 24/10/2022, accepted: 12/11/2022.

(self-driving) and cognitive networking framework have been introduced and show a lot of potential [15-18]. This new approach enables self-learning-based service provisioning capabilities by employing DRL agents to learn policies from dynamic network operations to enhance the overall network performance.

In this paper, we deal with routing, modulation format and spectrum assignment problem by applying deep reinforcement learning technique in elastic optical networks capable of provisioning dynamically lightpath services. By learning experiences of dynamic lightpath provisioning, our developed DRL-based RMSA takes advantage of deep reinforcement learning (DRL) mechanism to figure out and assign efficient path and spectral resource to the lightpaths. Numerical experiments are performed on a typical network topology and corresponding traditional elastic optical network with the shortest path algorithm under the same fiber configurations and network conditions is used for benchmarking. The attained results show that the network with the DRL-based RMSA solution outperforms that of the conventional shortest path algorithm significantly and offers a notable performance improvement in terms of blocking probability and accepted traffic volume.

II. DYNAMIC RMSA SCHEME USING DEEP REINFORCEMENT LEARNING

A. Dynamic Routing, Modulation format and Spectrum Assignment Problem

In this work, we consider a single fiber elastic optical network that can accommodate optical paths (also called lightpaths) dynamically and flexibly. The network adopts the distance adaptive approach [15] to determine the modulation format, m($m \in$ [1, 2, 3, 4]), according to the physical distance of lightpaths with four typical modulation formats of BPSK (m=1), QPSK (m=2), 8-QAM (m=3) and 16-QAM (m=4). We also assume that there is no spectrum conversion capability equipped.

Let $G(V, E, F)$ denote the elastic optical network topology and state, where V and E represent the sets of nodes and fiber links, $F=\{F_{e,f} \mid e \in E, f \in [1, W]\}$ contains the state of each frequency slot (FS) f on each fiber link e. A lightpath request between a source-destination node pair (s, d) is modelled as $R_t(s, d, b, \tau)$ with $s, d \in E$, b Gbps and τ denoting the required bandwidth and service duration.

To accommodate the lightpath R_t, an end-to-end routing path between s and d need to be computed and a proper modulation format m must be determined while a suitable number of spectrally contiguous FS's (i.e., the spectrum contiguous constraint) have to allocated on each link along the selected path from s to d according to b and m. Hence, the spectrum allocated on different fibers to R_t must be aligned.

Unlike the static RMSA problem where traffic requests are known in prior and the objective is to minimize the total spectrum usage, in the dynamic RMSA problem (i.e., online or dynamic lightpath provisioning), traffic requests may arrive and release in real time and need to be immediately served upon their arrivals. The dynamic RMSA problem, hence, targets minimizing the long-term request blocking probability, which is defined as the ratio of the number of blocked requests to the total number of requests over a period [6].

B. DRL-based RMSA Design

The DRL-based RMSA successively learns the optimal RMSA policy based on its perception of network states (e.g., topology, spectrum utilization and in-service lightpaths) and the feedback from the environment (i.e., network operations) using deep reinforcement learning. Figure 1 illustrates the operation principles of the developed DRL-based RMSA.

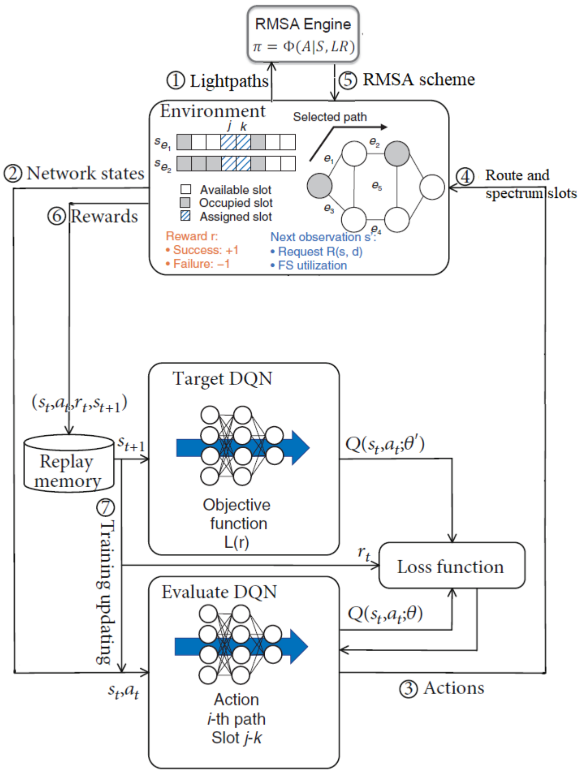

Fig. 1. DRL-based RMSA scheme with Double DQN.

When a lightpath request, $R_t(s, d, b, \tau)$, arrives (s and d are the source and destination nodes, b is the demanded data rate and τ is service duration), the RMSA engine fetches the current network state, s_t, and calls the Q-network to compute the estimated action value (i.e., Q) for each RMSA solution of lightpath requests. The neural network input is a given state s_t and the output is the value of each function. The action values can be represented by $Q(s_t, a_t; \theta)$, where θ denotes the parameters of the neural network, action $a_t \in$ A in the given state. The network receives the RMSA policies related to the previous operations as feedback and produces an immediate reward r_t for the agent; then, the network moves to the next state s_{t+1}. Then, r_t, s_t, a_t, and s_{t+1} are stored in a replay memory, from which the algorithm derives training data for updating the DRL agent. Here, there are two independent and identical deep neural networks (DQNs), a target DQN ($Q(s_t, a_t; \theta')$) and an evaluate DQN ($Q(s_t, a_t; \theta)$). The evaluate DQN is utilized to compute the Q value for each action, while the target DQN produces the Q values to train the parameters of the evaluate DQN. After that, the action with the maximum Q value is selected. Both the evaluate DQN and the target DQN employ the same neural network structure as the basic module, which uses a simple fully connected neural network, including one hidden layer. The neural network starts in state s_t and follows the value of each action. It attempts to minimize the loss function, *L(r)*.

At each time step, the agent takes an action and the action space is pre-determined at the DRL-based formulation phase. Then, the agent receives an observation and a reward from the environment. Here, observation is defined as the status of the current environment which consists of a request and a appropriate status of FS utilization. On the other hand, reward is a function representing how good the action is. The agent takes actions on the basis of the observation, and parameters of the agent's action-decision function, i.e., a deep neural network (DNN) is updated to maximize the total number of reward.

One of the important problem is how to define the action space; assignable FSs can be varied and rely on routing paths. Actually, our DRL-based RMSA adopts the mechanism introduced in [16]. RMSA is mapped to discrete Action space *A*. The DRL-based RMSA applies *K*-shortest path algorithm to select one from the *K* found shortest paths, and assigns the spectrum by choose the first index of used FSs. Hence, Action space is defined as $A=\{1, 2, \ldots, W\times K, W\times K+1\}$, where *W* and *K* are the numbers of FSs and paths, respectively. When assignable resources do not exist in all *K* shortest paths, the Action option will be *Doing Nothing*. The Action of *Doing Nothing* leads to block the lightpath request. A feature vector is a vector of characteristics of the current state that an Agent uses as a reference when deciding on an Action. Since the states of the whole network would be more helpful in improving performance, the DRL-based RMSA implements the Action-decision function to utilize both entire FS utilization tensor and convolutional neural network (CNN). The DNN determines which candidate should be used to accommodate as many future connection requests as possible. To cope with dynamic changes of the assignable candidate number, the masking approach is employed. Moreover, in order to provide useful information for training efficient RSA algorithms, the developed DRL-based RMSA employs a reward function as following:

$$r_t = \begin{cases} +1 & if\ assignable \\ -1 & otherwise. \end{cases}$$

III. NUMERICAL SIMULATIONS

In this section, we have evaluated the performance of the dynamic lightpath provisioning elastic optical network using the developed deep reinforcement learning-based RMSA algorithm. Numerical experiments are performed on a typical physical network topology named US national science foundation network (NSFNET) that consists of 14 nodes and 22 links (as illustrated in Figure 2).

In our simulations, the network is assumed to be a single fiber optical network in which each link includes only one optical fiber. The optical fiber capacity is *W* spectrum slots and spectrum slot capacity is *BW* (*W* is set at 64 and *BW* is 12.5 GHz); it means that a fiber is able to carry maximally 64 spectrum slots and, a slot bandwidth is supposed to be 12.5 GHz. The network can set up and release optical paths flexibly and dynamically. Modulation format of each lightpath is distance-adaptively assigned. In the network, optical paths can be modulated by one of four typical formats that are BPSK, QPSK, 8-QAM and 16-QAM. The slot bandwidth and the appropriate optical reach of BPSK, QPSK, 8-QAM, and 16-QAM optical signals are given in Table I. In our tested network, spectrum conversion resource will not be deployed.

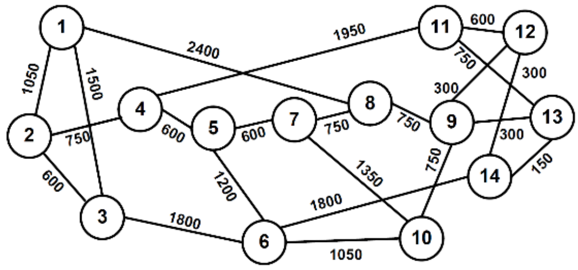

Fig. 2. Experimental network topology – NSFNET.

Moreover, we also applied the following parameters for the numerical simulation. Traffic requests are supposed to arrive sequentially and follow Poisson distribution with the average arrival rate of λ (requests per time unit). Distribution of lightpath holding time (mean hold time - MHT) is assumed to be a negative exponential one with the mean hold time of $1/\mu$ (time units) (μ is fixed at 10^{-2}). Consequently, the given network traffic load in Erlangs is λ/μ. On the other hand, the capacity of each requested lightpath between node pairs is also randomly assigned between 25 and 100 Gbps following a uniform distribution. Table I summarizes major simulation parameters of the experimental network.

TABLE I. SIMULATION PARAMETERS

Parameter			Value
Network topology			NSFNET
Number of nodes			14
Number of links			22
Spectrum slot number per link			64
Slot bandwidth			12.5 GHz
Modulation format	BPSK	*Slot capacity*	12.5 Gbps
		Optical reach	>2500 km
	QPSK	*Slot capacity*	25 Gbps
		Optical reach	2500 km
	8-QAM	*Slot capacity*	37.5 Gbps
		Optical reach	1250 km
	16-QAM	*Slot capacity*	50 Gbps
		Optical reach	625 km
Capacity of requested lightpaths			25-100 Gbps
Mean hold time			100 time units

In our investigation, the deep Q-network includes 2 Convo. 2 layers (each with 16 convolution kernels), 3 Conv. 3 layers (each with 1 convolution kernel) and 2 fully connected layers ([128,50]).We calculate K = 5 candidate paths for each s-d pair, i.e., the number of nodes in the output layer is 5, γ and ε are set to be 0.99 and 0.1 respectively.

Performance of the developed DRL-based RMSA adopted elastic optical network with the capability of provisioning dynamic lightpath services is tested and measured in terms of connection blocking probability and accepted traffic volume. Here, the blocking probability is calculated as the ratio of the blocked connection number to the total number of lightpaths requested. The relative accepted traffic volume is determined as the ratio of the traffic volume obtained by the developed network to that of corresponding conventional elastic optical network utilizing a popular RMSA algorithm, that is the shortest path algorithm [4], under the same network configuration. Hereafter, the results obtained by using the proposed DRL-based algorithm and the conventional one will be denoted as *DRL* and *Conventional* respectively.

A. *Blocking Probability*

In fact, for dynamic lightpath provisioning elastic optical network, blocking probability is a key indicator to determine the network ability of dynamically and effectively providing lightpath services. The less blocking probability is, the better network performance is achieved. Figure 3 shows the obtained blocking probability of the DRL-based RMSA algorithm in comparing to that of the conventional shortest path algorithm when the traffic arrival rate ranges from 0.1 to 1.0. The attained graphs describe that our developed solution outperforms the conventional one over all the tested traffic load. It implies that the performance of both comparable solutions is degraded rapidly as the traffic load is increased however, thanks to the use of DDQN for learning and selecting route, modulation and spectrum resources, our solution offers dramatically less blocking probability, especially in small traffic load area. When the traffic load becomes larger, due to the spectrum collision, more connection requests are refused, in this case the reward function is critically decreased (see Figure 4), and this leads to an increase of blocking probability.

A further comparison between the developed DRL-based RMSA algorithm and the conventional one, in terms of link utilization ratio is described in Figure 5. The proposed algorithm attains higher link utilization ratio than the conventional one. This means that, with the same traffic condition, our algorithm can use the spectrum resource more efficiently to help reduce the spectrum collision and enhance the network performance. Taking advantage of deep reinforcement learning, our solution can wisely select the route and spectrum resource to adapt properly with the traffic condition.

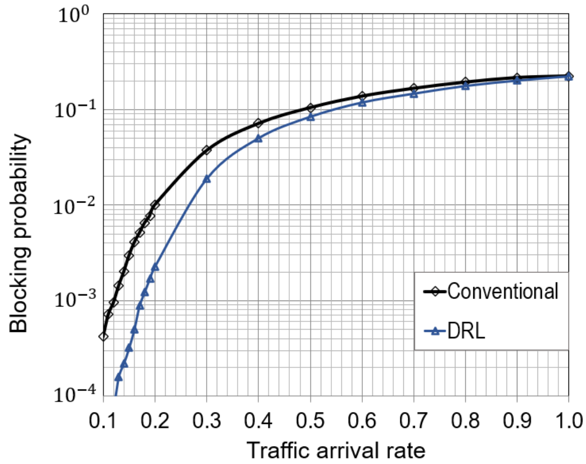

Fig. 3. Blocking probability.

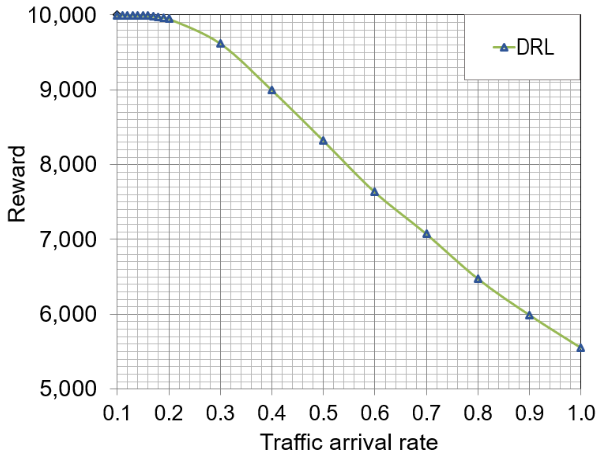

Fig. 4. Reward function of DRL-based RMSA algorithm.

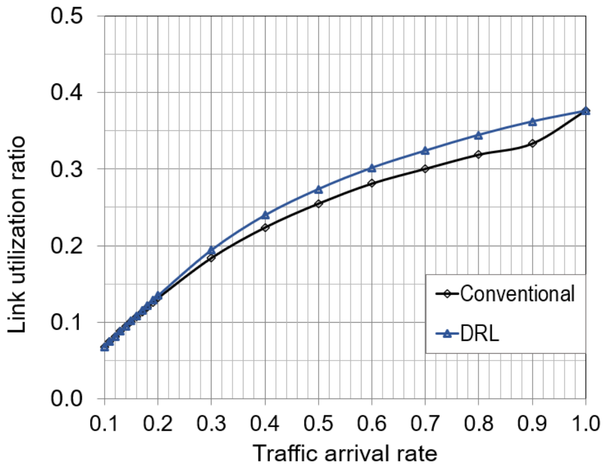

Fig. 5. Comparison of link utilization ratio.

B. *Relative Accepted Traffic Volume*

In order to assess the efficiency of our developed network solution, we have also estimated and compared the traffic volume providing by the DRL-based RMSA and that of conventional algorithm under the same network configuration when the required blocking probability of 10^{-2} and 10^{-3}. Here, the simulation results of the network with the traditional RMSA algorithm are used as the benchmark and, the relative accommodated traffic volumes, the rate between the obtained results to that of the appropriate conventional networks, are plotted in Figure 6. Consequently, the relative traffic volume of the conventional RMSA is a constant (1.0).

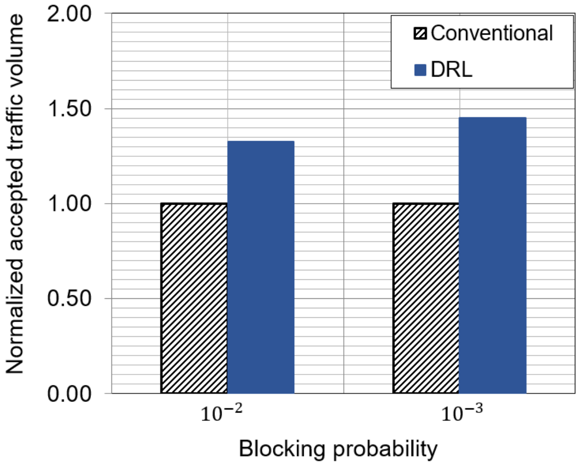

Fig. 6. Accepted traffic volume comparison.

The bar graphs show that, compared to that of the conventional network, our developed solution gains a significant improvement in terms of the accepted traffic volume. With the blocking probability of 10^{-2}, more 32.5% traffic can be accommodated by using our solution. The enhancement even becomes better with less blocking probability, says higher network performance required, up to 45% more traffic volume can be carried, with the blocking probability of 10^{-3}, by adopting our developed algorithm.

IV. CONCLUSION

We have investigated dynamic lightpath provisioning-enable elastic optical networks and applied deep reinforcement learning to deal with the routing, modulation format and spectrum assignment problem. The networks are capable of automatically and dynamically setting-up and releasing bandwidth-flexible lightpaths by using a deep reinforcement learning mechanism. The developed DRL-based RMSA algorithm exploits a Double DQN for learning the optimal dynamic RMSA policies in elastic optical network to improve the network performance. Numerical simulation results demonstrate the efficiency of our developed network solution. Compared to the conventional shortest path algorithm, it can gain more up to 32.5% and 45% traffic volume with the blocking probability of 10^{-2} or 10^{-3}, respectively.

REFERENCES

[1] G. Zhang, M. De Leenheer, A. Morea and B. Mukherjee, "A Survey on OFDM-Based Elastic Core Optical Networking," *IEEE Communications Surveys & Tutorials*, IEEE, vol. 15, no. 1, pp. 65-87, 2013.

[2] A. Lord, P. Wright and A. Mitra, "Core Networks in the Flexgrid Era," *Journal of Lightwave Technology*, IEEE, vol. 33, no. 5, pp.1126-1135, 2015.

[3] H. Furukawa and R. S. Luís, "Petabit-class Optical Networks Based on Spatial-Division Multiplexing Technologies," *2020 International Conference on Optical Network Design and Modeling (ONDM)*, pp. 1-3, 2020.

[4] B. Chatterjee, N. Sarma and E. Oki, "Routing and Spectrum Allocation in Elastic Optical Networks: A Tutorial," *IEEE Communications Surveys & Tutorials*, IEEE, vol. PP, no. 99, pp. 1, 2015.

[5] Chatterjee, N. Sarma, E. Oki, "Routing and Spectrum Allocation in Elastic Optical Networks: A Tutorial," *IEEE Communications Surveys & Tutorials*, vol. 1, pp. 99, 2015.

[6] M. Klinkowski, P. Lechowicz, and K. Walkowiak, "Survey of resource allocation schemes and algorithms in spectrally-spatially flexible optical networking," *Optical Switching and Networking*, vol. 27, pp. 58–78, 2018.

[7] Alberto Paradisi, Rafael Carvalho Figueiredo, Andrea Chiuchiarelli, Eduardo de Souza Rosa, "Optical Communications: Advanced Systems and Devices for Next Generation Networks," *Springer*, 2019.

[8] Devi Chadha, "Optical WDM Networks: From Static to Elastic Networks," *John Wiley & Sons Ltd*, 2019.

[9] Hai-Chau Le, Ngoc T. Dang, "Spectrum Utilization Efficiency of Elastic Optical Networks Utilizing Coarse Granular Routing," *Informatica*, vol. 42, pp. 293–300, 2018.

[10] Y. Li, W. Chen, S. K. Bose and G. Shen, "Efficient Multi-Stage Deployment of Ultra-Low Loss Fibers in Elastic Optical Networks," *Journal of Lightwave Technology*, vol. 38, no. 14, pp. 3542-3552, 15 July, 2020.

[11] V. Mnih and K. Kavukcuoglu et al., "Human-level control through deep reinforcement learning," Nature, vol. 518, pp. 529–533, 2015.

[12] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.

[13] N. Luong, D. Hoang, S. Gong, D. Niyato, P. Wang, Y. Liang, and D. Kim, "Applications of deep reinforcement learning in communications and networking: A survey," arXiv preprint arXiv:1810.07862, 2018.

[14] Z. Xu, J. Tang, J. Meng, W. Zhang, Y. Wang, C. Liu, and D. Yang, "Experience-driven networking: A deep reinforcement learning based approach," arXiv preprint arXiv:1801.05757, 2018.

[15] 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," in Journal of Lightwave Technology, vol. 37, no. 16, pp. 4155-4163, 15 Aug.15, 2019.

[16] M. Shimoda and T. Tanaka, "Deep Reinforcement Learning-based Spectrum Assignment with Multi-metric Reward Function and Assignable Boundary Slot Mask," 2021 Opto-Electronics and Communications Conference (OECC), July 2021.

[17] Hado Van Hasselt, Arthur Guez, and David Silver. Deep reinforcement learning with double q-learning. In Thirtieth AAAI Conference on Artificial Intelligence, 2016.

[18] Luong, Nguyen Cong, Dinh Thai Hoang, Shimin Gong, Dusit Niyato, Ping Wang, Ying-Chang Liang, and Dong In Kim. "Applications of deep reinforcement learning in communications and networking: A survey." IEEE Communications Surveys & Tutorials 21, no. 4 (2019): 3133-3174.

GIẢI PHÁP CẤP PHÁT PHỔ TẦN ĐỘNG DỰA VÀO KỸ THUẬT HỌC TĂNG CƯỜNG SÂU CHO MẠNG QUANG LƯỚI BƯỚC SÓNG LINH HOẠT

Tóm tắt: Trong bài báo này, chúng tôi đã nghiên cứu và phát triển một thuật toán định tuyến và gán định dạng điều chế và phổ tần (RMSA) dựa vào kỹ thuật học tăng cường sâu (DRL) cho mạng quang lưới bước sóng linh hoạt hỗ trợ việc cấp phát băng thông động. Nhằm nâng cao hiệu năng của mạng, giải pháp RMSA đề xuất đã khai thác cơ chế học tăng cường sâu để lựa chọn tuyến đường, khuôn dạng điều chế và tài nguyên phổ tần hiệu quả bằng việc học hỏi kinh nghiệm cấp phát phổ tần động và linh hoạt trong mạng. Phương pháp mô phỏng số được áp dụng để đánh giá hiệu năng của giải pháp RMSA dựa trên DRL được đề xuất. Các kết quả thu được cho thấy rằng giải pháp mạng được đề xuất của chúng tôi vượt trội hơn đáng kể so với thuật toán đường đi ngắn nhất thông thường và góp phần cải thiện đáng kể hiệu năng mạng về các thông số xác suất chặn kết nối và khả năng chấp nhận lưu lượng.

Từ khoá: Học tăng cường sâu, mạng quang lưới bước sóng linh hoạt, định tuyến và gán phổ tần, thuật toán điều khiển mạng.

Hai-Chau Le received the B.E. degree in Electronics and Telecommunications Engineering from Posts and Telecommunications Institute of Technology (PTIT) of Vietnam in 2003, and the M.Eng. and D.Eng. degrees in Electrical Engineering and Computer Science from Nagoya University of Japan in 2009 and 2012, respectively. From 2012 to 2015, he was a researcher in Nagoya University of Japan and in University of California, Davis, USA. He is currently a lecturer in Telecommunications Faculty at PTIT. His research interests include optical technologies, network design and optimization and future network technologies. He is an IEEE member.

Van Duc Khuat received the B.E. degree in Information Technology from Posts and Telecommunications Institute of Technology (PTIT) of Vietnam in 2021. Now, he is a lecture assistant in Telecommunications Faculty at PTIT. His research interests include artificial intelligence (AI), bioinformatics and network optimization.