

Deep-reinforced impairment-aware dynamic resource allocation in nonlinear elastic optical networks

Chenglin Shi^{1,3}, Min Zhu^{2,3}, Jiahua Gu^{2,3}, Tianyu Shen^{2,3}, Xueqi Ren^{2,3}

1. School of Integrated Circuits, Southeast University, Nanjing, 210096, China

2. National Mobile Communication Research Laboratory, Southeast University, Nanjing, 210096, China

3. Purple Mountain Laboratories, Nanjing, 211111, China, minzhu@seu.edu.cn

Abstract: This paper proposed a dynamic resource allocation scheme in nonlinear elastic optical networks based on deep reinforcement learning, which achieves significant blocking probability reductions of more than 44.1% compared with baseline algorithms.

1. Introduction

Elastic optical networks (EONs) have emerged as a promising technology to accommodate dynamic and diverse demands of next-generation applications by allocating resources adaptively [1]. Most existing related works on resource allocation in EONs are based on predefined transmission-reach (TR) limits, i.e., each modulation format has a corresponding TR limit, which requires a large margin for the most connections to guarantee quality of transmission (QoT) in the worst case through resource overprovisioning and hence causes inefficient resource utilization [2]. Hence, some studies have been proposed to significantly improve spectral efficiency by accurately accounting for physical-layer nonlinear impairments (NLI) using the Gaussian Noise (GN) model in EONs, compared with existing TR-based methods [3, 4]. The impairment-aware dynamic resource allocation in EONs includes two main challenges: 1) the routing and spectrum allocation (RSA), where the spectrum continuity and contiguity constraints are considered [1]; and 2) the modulation format assignment accounting for the NLI [3]. This is called the routing, modulation, and spectrum assignment (RMSA) problem designed for nonlinear EONs. The authors in [3] calculate the NLI interference from other connections and jointly optimize physical layer resources of each connection in static EONs by using a mixed integer programming (MILP) and a decomposition heuristic. In [4], the authors proposed a hybrid NLI estimation technique along with a sophisticated K-least congested path routing strategy to solve the NLI-aware resource allocation problem in dynamic EONs. However, the above these works only rely on fixed heuristic policies based on simple empirical design regardless of the time-varying EON states, and therefore cannot achieve dynamic adaptive resource allocation in EONs.

Recently, Deep Reinforcement Learning (DRL) has demonstrated beyond human-level performance for resource allocation problems. DRL can learn successful policies progressively without any prior knowledge of the target system's behavior, by accumulating action experiences from repeated interactions with the target systems and by reinforcing actions leading to higher rewards [5]. Moreover, DRL can support a variety of optimization objectives just by setting different reinforcement rewards. The authors of [6] present a DRL-based online RMSA framework for dynamic traffic demands in EONs, which just adopts the TR limits to determine the modulation format and does not consider the actual network state. This paper, to the best of our knowledge, is the first attempt to address an impairment-aware dynamic RMSA issue in nonlinear EONs based on DRL technology. We structure a deep neural network (DNN) to perceive the complex EON state, and a self-learning intelligent agent to achieve autonomous and optimal nonlinear RMSA policy. Simulation results in the 14-node NSFNET topology show that the proposed DRL-based nonlinear RMSA outperforms the baseline nonlinear RMSA heuristics.

2. Nonlinear RMSA Problem Formulation

We model the EON as a directed graph $G(V, E, F)$, where V and E are sets of nodes and fiber links, F denotes the frequency slot (FS) usage on each link $e \in E$. A dynamic lightpath request (LR) with bandwidth b and service duration τ from source node o to destination node d can be modeled as $R(o, d, b, \tau)$. To serve R , we need to compute an end-to-end routing path $p_{o,d}$, determine a QoT-guaranteed modulation format m , and allocate a set of spectrally continuous and contiguous FS's according to b and m on each fiber link along $p_{o,d}$. Under the assumption of Nyquist spectral shaping [7], the number n_i of FS's allocated to R_i can be computed as $n_i = \lceil b / (m_i \cdot B_{PM-BPSK}) \rceil$, where $B_{PM-BPSK}$ is the data rate a FS can provide with modulation format PM-BPSK, and $m_i \in M = \{1, 2, 3, 4\}$ denotes the spectral efficiencies of PM-BPSK, PM-QPSK, PM-8QAM and PM-16QAM respectively. We can also use m_i to denote the corresponding modulation format assigned to connection R_i . For each available m_i , its required minimum signal-to-noise-ratio (SNR) threshold SNR_{th}^m for $m_i \in M$ under a certain pre-forward error correction (FEC) bit-error rate (BER) (4×10^{-3} in this paper) is given in Table I [7]. The connection QoT can be estimated based on the GN model [7], which is an analytical model to calculate NLIs in dispersion-uncompensated links. By combing various PLIs

including amplified spontaneous emission (ASE) noise, self-channel interference (SCI) and cross-channel interference (XCI), we can calculate the SNR for each connection R_i as,

$$SNR_i = \frac{G_i}{G_i^{ASE} + G_i^{SCI} + G_i^{XCI}}, \quad (1) \quad G_i^{ASE} = (e^{\alpha L} - 1) h \nu n_{sp} N_i^{span}, \quad (2)$$

$$G_i^{SCI} = \frac{3\gamma^2 G^3}{2\pi\alpha|\beta_2|} \ln\left(\frac{\pi^2|\beta_2|}{\alpha} \Delta f_i^2\right) N_i^{span}, \quad (3) \quad G_i^{XCI} = \frac{3\gamma^2 G^3}{2\pi\alpha|\beta_2|} \sum_{j \neq i} \ln\left|\frac{\Delta f_{ij} + \Delta f_j/2}{\Delta f_{ij} - \Delta f_j/2}\right| N_{ij}^{span}, \quad (4)$$

where G_i , G_i^{ASE} , G_i^{SCI} and G_i^{XCI} denotes power spectral density (PSD), ASE noise, SCI, XCI of R_i respectively. α , γ , β_2 , h , n_{sp} , ν , L denotes the power attenuation, the fiber nonlinear coefficient, the fiber dispersion, Planck's constant, the spontaneous emission factor, the optical carrier frequency and the length of each span respectively. Δf_i , Δf_j and Δf_{ij} denotes the bandwidth of R_i , the bandwidth of R_j and the center frequency difference between R_i and R_j respectively. The number of spans propagated by R_i along the route is denoted by N_i^{span} , and N_{ij}^{span} denotes the number of spans shared by the routes of R_i and R_j . Note that, in dynamic RMSA, LR's arrive and expire on-the-fly and need to be served upon their arrivals, and hence R_i incurs XCI from both existing and future LR's. But the XCI effect from the future LR's on R_i is not available at the moment of the provisioning R_i . Therefore, we assume that: 1) the number of future LR's is estimated by multiplying the average arrival rate and the service duration of current R_i , 2) the bandwidth of future LR is identical to the average bandwidth of all LR's, 3) the number of FS's used by a future LR is calculated by applying the lowest modulation format, and 4) the allocated FS's for future LR's are placed next to the highest FS of current R_i in the fiber links. R_i is served successfully only if its SNR satisfies the SNR threshold of the used m_i , i.e., $SNR_i \geq SNR_{th}^m$. The objective of the dynamic nonlinear RMSA is to minimize the request blocking probability, which is defined as the ratio of the number of blocked LR's to the total number of LR's over a period.

3. DRL-based Nonlinear RMSA Design

Fig. 1(a) illustrates the operation principle of the proposed DRL-based Nonlinear RMSA. The intelligent agent is presented as a DNN, referred to as a policy network. Upon R_i arrives at t time, the policy network takes the state information s_t including the R_i and the current network state as the input (*step 1*), and outputs the probability distribution $\pi_t(A|s_t, \theta)$ over all possible action space A and θ represents the policy parameters of the DNN (*step 2*). Based on the π_t , the DRL-based nonlinear RMSA agent takes an action $a_t \in A$ and attempts to establish the lightpath for R_i (*step 3*). A reward r_t related to the RMSA operation is feedback (*step 4*). The r_t , together with s_t and a_t , can be used to train the nonlinear RMSA agent. The objective of the agent is to maximize the long-term cumulative reward, which is defined as $\Gamma_t = \sum_{t' \in [t, \infty)} \gamma^{t'-t} \cdot r_{t'}$, [6]. The $\gamma \in (0, 1]$ is a discount factor that decays future rewards.

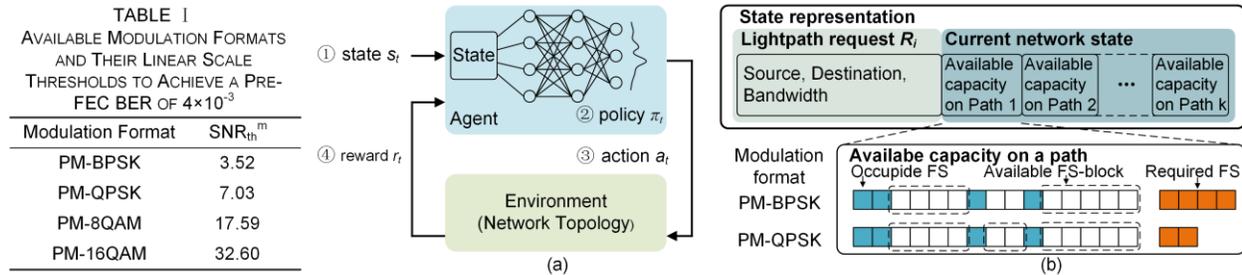


Fig. 1. (a) Operation principle of the DRL with DNN and (b) state representation.

Fig. 1(b) shows the state representation s for the DRL-based Nonlinear RMSA, which includes the information of R_i and spectrum utilization on the K -shortest candidate paths with J available FS-blocks based on M different modulation formats. For each k of the K candidate paths, we use M different modulation formats to calculate the number n_i of required FSs and try to find appropriate FS-blocks to accommodate the required FSs. The size and the starting index of each available FS-block, the average size of the available FS-blocks and the total size of the available FS-blocks are also viewed as part of state. In Fig. 1(b), if the agent selects the PM-QPSK instead of the PM-BPSK, the number n_i of required FSs decreases (i.e., $4 \rightarrow 2$) and the number J of the available FS-blocks increases accordingly (i.e., $2 \rightarrow 3$). Although a higher level of modulation format can save spectrum resources, its SNR requirement gets stricter. Hence, we need to verify whether the connection that consists of the k -th path, m -th modulation format and j -th FS-block hypothetically-assigned to R_i meet the corresponding SNR constraint. When the following three cases occur: 1) the number of available candidate paths between o and d is smaller than K , 2) the number of available FS-blocks is smaller than J , 3) the pre-computed connection fails to satisfy the SNR constraint,

we would assign an array of -1 for corresponding part to keep the format of state representation consistent. Thus, there are $K \cdot M \cdot J$ actions in the action space. If R_i is served, the agent receives a reward r_t of 1; otherwise, r_t is -1.

The training of the agent adopts the framework of policy gradient reinforcement algorithm in an iteration way. The agent initiates $\Delta\theta = 0$ at the beginning of each iteration. In each iteration, we simulate N episodes for one LR set to explore the probabilistic space of possible actions using the current policy and use the resulting data to improve the policy. The LR set includes all LRs arriving within a fixed number T of time unit. Specifically, in each time unit, the resources occupied by the expired requests are released. Then the agent obtains s_t based on the state model mentioned above, and decides a nonlinear RMSA scheme (i.e., action a_t) to serve R_i and receives a r_t accordingly. Thus, the state s_t , action a_t and reward r_t are recorded for all time units of each episode. The values of (s_t, a_t, r_t) are used to compute Γ_t at each time unit of each episode. To reduce the variance originated from the gradient estimates, a baseline value b_t is obtained by averaging Γ_t , where the average is taken at the same time unit across all episodes with the same LR set (i.e., $b_t = (1/N) \sum_{i=1}^N \Gamma_t^i$). The $\Delta\theta$ is calculated to optimize policy network parameter $\theta \leftarrow \theta + \Delta\theta$ via gradient descent equation $\Delta\theta = \mu \cdot \sum_{t=1}^L \sum_{i=1}^N \nabla \log \pi_{\theta}(s_t^i, a_t^i) (\Gamma_t^i - b_t^i)$, where μ is the learning rate.

4. Evaluation and Discussion

The performance of the proposed DRL-based nonlinear RMSA is evaluated in 14-node NSFNET. The parameters related to physical impairments are $\alpha = 0.22$ dB/km, $\gamma = 1.3$ (W·km)⁻¹, $\beta_2 = -21.7$ ps²/km, $n_{sp} = 1.58$, $\nu = 193.55$ THz, $L = 100$ km. Each fiber link accommodates 100 FS's of 12.5 GHz each. The dynamic LRs are generated according to Poisson process, with average arrival rate and duration being 10 and 25 time units respectively. The bandwidth of each LR distributed evenly within [80, 320] Gb/s. A uniform PSD $G = 15$ mw/THz is assumed for all connections.

We first evaluated the impact of the DNN scale on the performance of the DRL algorithm. Fig. 3(a) shows the cumulative rewards collected from every 1000 LRs with different scales of DNNs, i.e., 1 hidden layer of 6 neurons (1×6), (1×12), (1×24) and (2×12). We can see that the rewards with three different number of neurons are very close, with 1 hidden layer outperforming 2 hidden layers. This is because (1×12) DNN does not encounter the overfitting issue compared with (2×12) case. Then we study how the (K, J) setups affect the performance of the DRL algorithm. Fig. 3(b) shows the cumulative rewards of $K=5$ is higher than that of $K=1$ or 3 when $J=1$. It indicates that increasing routing diversity can efficiently improve the DRL performance. We can also see that providing more options in spectrum (e.g., $J=3$) does not improve the DRL performance. This is because more flexible spectrum allocation may lead to severe fragmentation, and larger action space is not conducive to learning correct policies.

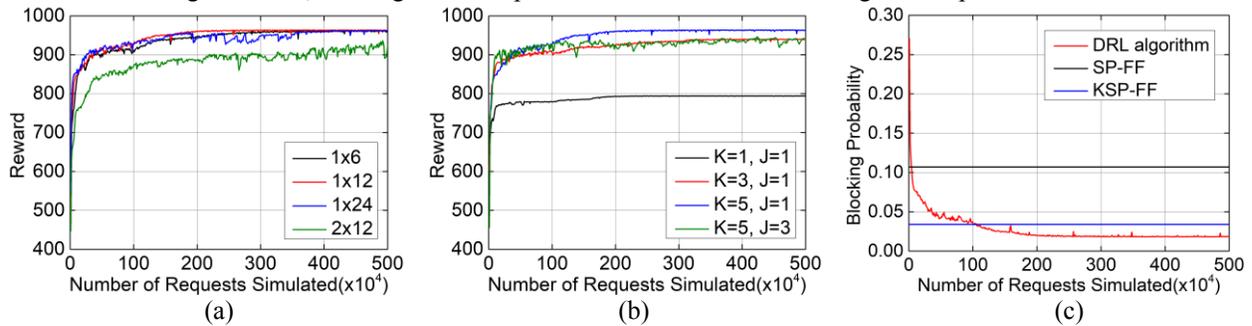


Fig. 3. Cumulative rewards with (a) different DNN scales, (b) different K and J , (c) LR blocking probability

We compare the performance of our DRL algorithm with baseline algorithms, including shortest-path routing and first-fit spectrum allocation (SP-FF) and K-shortest-path routing and first-fit spectrum allocation (KSP-FF). Fig. 3(c) shows our DRL algorithm can achieve a blocking reduction of more than 44% compared with KSP-FF. It is because that DRL-based nonlinear RMSA agent can learn to make better decisions directly from experience interacting with the dynamic EON environment.

5. Conclusion

We propose a DRL-based dynamic RMSA scheme in nonlinear EONs. Simulation results show that the proposed algorithm can significantly reduce the blocking probability of more than 44%, when compared with the baselines.

References

- [1] J. Zhao, et al., IEEE JLT, 33. 22: 4554-4564, 2015.
- [2] M. Jinno, et al., IEEE CM, 48. 8: 138-145, 2010.
- [3] L. Yan, et al., IEEE/OSA JOCN, 7. 11: B101-B108, 2015.
- [4] R. Wang, et al., IEEE ONDM, 2017.
- [5] H. Mao, et al., ACM, 50-56, 2016.
- [6] X. Chen, et al., IEEE JLT, 37. 16: 4155-4163, 2019.
- [7] L. Yan, et al., IEEE JLT, 35. 10: 1766-1774, 2017.