Building Autonomic Elastic Optical Networks with Deep Reinforcement Learning

Xiaoliang Chen, Roberto Proietti, and S. J. Ben Yoo

Conventional schemes for service provisioning in next-generation elastic optical networks (EONs) rely on rule-based policies that suffer from scalability issues and can lead to poor resource utilization efficiency due to the lack of knowledge about the essential characteristics of EONs. The authors discuss the application of emerging deep reinforcement learning (DRL) techniques in EONs for enabling an autonomic (self-driving) and cognitive networking framework.

Abstract

Conventional schemes for service provisioning in next-generation elastic optical networks (EONs) rely on rule-based policies that suffer from scalability issues and can lead to poor resource utilization efficiency due to the lack of knowledge about the essential characteristics of EONs (e.g., traffic profiles, physical-layer impairments). This article discusses the application of emerging deep reinforcement learning (DRL) techniques in EONs for enabling an autonomic (self-driving) and cognitive networking framework. This new framework achieves self-learning-based service provisioning capabilities by employing DRL agents to learn policies from dynamic network operations. Such capabilities can remarkably reduce the amount of human effort invested in developing effective service provisioning policies for emerging applications, and thus, can facilitate fast network evolutions. Based on the framework, we first present DeepRMSA, a DRL-based routing, modulation, and spectrum assignment (RMSA) agent for EONs. Then, as today's networks are often composed of multiple autonomous systems, we extend the autonomic networking framework to multi-domain EONs by applying multi-agent DRL (where multiple autonomous DRL agents learn through jointly interacting with their environments). Comparisons of the results from numerical simulations show significant advantages of the proposed framework over the existing rule-based heuristic designs.

Introduction

Elastic optical networking (EON) has emerged as one of the most appealing solutions for meeting the challenges of the next-generation networks [1]. By virtue of its flexible spectrum allocation mechanisms, EON is able to support not only the ever-increasing volume of Internet traffic but also user-customized dynamic service provisioning at the optical layer [2]. To fully exploit the benefit of EON, effective network control and management (NC&M) is necessary. While software-defined networking (SDN) has enabled a centralized and programmable NC&M paradigm for elastic optical networks (EONs) [3], existing works mostly make use of hard-coded rule-based policy designs for service provisioning in EONs. These designs are usually built on domain-specific knowledge, such as the data plane operation principle and mathematical optimization theories, which entails significant efforts of networking experts and operators. More importantly, as network conditions (e.g., topology, traffic profile) may change, these designs might need to be revisited periodically, causing scalability issues and thus hindering the fast evolution of the networks.

The past few years have witnessed dramatic advances in machine learning (ML) since the breakthroughs in deep learning algorithms. ML allows learning complex system functions from big data while obviating the need for domain-specific knowledge. The application of ML in optical networking has attracted intensive research interest [4–6]. In [4], the authors proposed an ML-aided fault management system for detecting soft failure patterns in optical networks. We introduced ML to the NC&M of multi-domain optical networks in [5] and therein demonstrated a cognitive (tendency-aware) inter-domain traffic engineering design enabled by a deep neural network (DNN)-based traffic estimator. An ML approach for modeling the quality of transmission of optical connections was also developed. The authors of [6] studied the problem of ML as a service in inter-data-center optical networks, where tenants manage their virtual network slices with the assistance of commercial ML models trained by third-party entities, and discussed the related vulnerability issues in such scenarios. All of the above existing works adopt supervised ML techniques, which require large volumes of data for training. In addition to the fact that such data are difficult to obtain, these works still rely on artificially defined policies that utilize the ML models.

Deep reinforcement learning (DRL) has recently shown compelling potential of learning successful policies for large-scale online control problems [7]. In contrast to supervised or unsupervised ML approaches demanding large amounts of data, DRL parameterizes policies with DNNs (analogous to human brains, which can sense complex system states directly from high-dimensional data, e.g., images and traffic matrices) and progressively approximates the optimal policies by training the DNNs with experiences from online operations. Thus, DRL enables self-learning capabilities that allow learning agents to learn and to adapt to systems autonomically without human intervention. Such self-learning is especially crucial for intelligent network and service management. Previous works have reported a few applications of DRL in the areas of communication and networking [8, 9]. The authors of [8] proposed a DRL algorithm

Digital Object Identifier:
10.1109/MCOM.001.1900151

The authors are with the University of California.

that enables experience-driven traffic engineering by jointly learning a network environment and its dynamics. In [9], the authors presented a DRL-based data center network management framework and demonstrated a DRL agent that can learn the optimal topology configurations regarding different applications. Nevertheless, the application of DRL in EONs remains underexploited.

In this article, we leverage DRL to propose an autonomic networking framework for EONs. We first elaborate on the network architecture, modular NC&M system design, and the operation principle of the proposed framework. Based on the framework, we present DeepRMSA, a DRL-based routing, modulation, and spectrum assignment (RMSA) agent for EONs. We demonstrate the effectiveness of DeepRMSA through numerical simulations. Next, we extend the autonomic networking framework to a multi-domain EON scenario by applying multi-agent DRL (MADRL). Finally, we summarize the article.

DRL-Based Autonomic Networking Framework

Figure 1 shows the block diagram of the proposed DRL-based autonomic networking framework. The framework is built on the basis of the SDN architecture, with decoupled data and NC&M planes. The data plane adopts EON technologies to provision dynamic and flex-grid (e.g., at a granularity of 6.25 GHz) optical connections for clients from metro networks, data centers, and research facilities. Optical performance monitoring functionalities (e.g., monitoring of optical signal-to-noise ratio) are also employed for sensing the states of data plane operations. The NC&M plane employs a remote and centralized SDN controller for service provisioning management. The SDN controller utilizes advanced network modeling languages and SDN protocols to communicate with SDN agents (locally attached to data plane equipment) for collecting service requests, distributing service schemes, and inquiring device conditions and monitoring data on demand.

We design the service provisioning mechanism based on the principle of DRL. Specifically, upon an event (e.g., reception of a service request) that triggers a specific DRL application (e.g., DRL-based RMSA or failure restoration), the SDN controller makes the feature engineering module generate an EON state representation for the corresponding DRL agent. The feature engineering module retrieves various network state data (e.g., pending requests, in-service connections, and resource utilization) from the traffic engineering database and tailors the data to meet the demand of the DRL agent. The DNNs of the DRL agent take as input the state data and output a service provisioning policy to the SDN controller. Here, a service provisioning policy can be a probability distribution over a set of available service schemes. The SDN controller in turn determines a service scheme with the policy. Based on the service provisioning outcome, corresponding feedback is sent to the reward system. The reward system translates the feedback into an immediate reward for the DRL agent. The reward enables the DRL agent to quantitatively measure the quality of the action taken (i.e., the service scheme selected). For example, an agent gets a reward of 1 if a request is successfully serviced and 0 otherwise. The service provisioning sample (i.e., the state, action, and reward tuple) is stored in the experience buffer, which afterward produces training signals to update the DNNs. In particular, the DRL agent tunes the DNNs to reinforce actions (i.e., increase the corresponding probabilities) leading to higher long-term cumulative rewards. This way, through repeated service provisioning practice, the DRL agent can progressively learn effective policies. Meanwhile, as the DRL agent performs training constantly upon new observations, it is able to adapt to gradual network evolutions. Different DRL agents can also work in collaboration through knowledge transfers for faster convergence and improved network-wide performance. Eventually, the DRL-based service provisioning design enables a fully autonomic EON system with self-learning and self-adapting capabilities. Note that, with slight modifications, the proposed framework is also applicable for networks using different data plane technologies (e.g., packet networks).

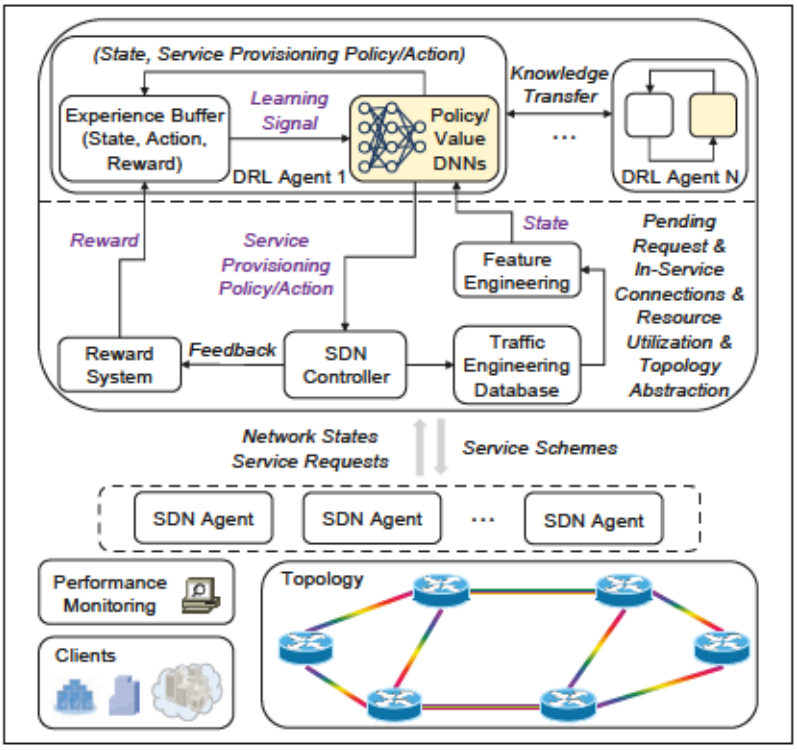

Figure 1. DRL-based autonomic networking framework. SDN: software-defined networking; DNN: deep neural network; DRL: deep reinforcement learning.

DeepRMSA

While EON introduces unprecedented flexibility to optical-layer spectrum management, the design of RMSA algorithms is not trivial. Figure 2 shows an example of RMSA in a five-node topology, where two lightpath requests $\mathcal{R}_1$ (from node 1 to node 4) and $\mathcal{R}_2$ (from node 2 to node 5) arrive sequentially, each demanding bandwidth of 2 or 4 frequency slots (FSs). For the sake of clarity, we reduce the optimization dimension of the RMSA problem by fixing the routing paths as 1-2-4 and 2-4-5, respectively, omitting the modulation format assignment procedure. Based on the spectrum utilization state on each link, two FS-blocks

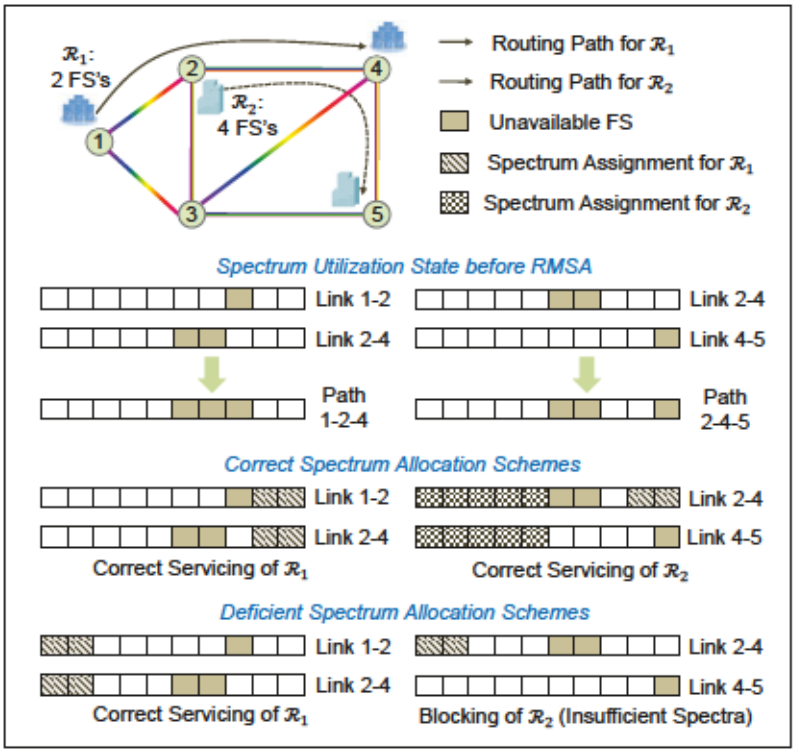

Figure 2. An illustrative example of RMSA in a five-node EON.

(i.e., [1,5] and [9,10]) are available on path 1-2-4. However, the only correct policy is allocating FS-block [9,10] to $\mathcal{R}_1$ (where both of the requests are successfully serviced), since otherwise, $\mathcal{R}_2$ will be blocked due to the lack of spare spectra on link 2-4. Note that more practical RMSA problems involving realistic-scale topologies and larger link capacities while allowing flexible routing and modulation format choices would be much more complicated than that given by the above example. In this context, previous works have proposed a number of optimization models and heuristic designs for RMSA problems [10, 11]. While the optimization models suffer from high complexity and can hardly be applied to problems with realistic scales, the heuristic designs all apply fixed policies based on artificially defined rules, resulting in suboptimal performance. In this section, based on the proposed autonomic networking framework, we present a DRL-based cognitive RMSA agent, namely, DeepRMSA, for EONs.

Design

Given an EON, the target of DeepRMSA is to learn the optimal RMSA policy for each lightpath request so that the long-term network throughput is maximized. Next, we describe the key components of DeepRMSA, including the designs of *State* (what network state DeepRMSA sees), *Action* (how DeepRMSA decides the RMSA schemes), *Reward* (numerical incentives characterizing the action performance of DeepRMSA), and *Training* (how DeepRMSA learns).

State: Effective state representations enable DeepRMSA to sense critical information required for RMSA. For each lightpath request, we make DeepRMSA read a state representation containing the information of the request's source and destination nodes and the current spectrum utilization state of the EON. To convey the spectrum utilization state, we calculate K candidate routing paths for the request and obtain for each of the paths:

- The size and the starting position of the first available FS-block (i.e., the available FS-block with the lowest starting index)
- The number of required FSs based on the applicable modulation format
- The average size of available FS-blocks
- The total number of available FSs

Note that instead of extracting key features on different paths, one can directly feed DeepRMSA with the original link-by-link spectrum utilization matrix to avoid any information loss. However, this would cause scalability issues and add significant difficulty to the successful training of DeepRMSA. As a future research task, we will study more effective state representation methods for DeepRMSA.

Action: DeepRMSA selects one from the K candidate paths for each lightpath request. Then a modulation format can be determined based on the routing path selected and the distance-adaptive resource allocation model discussed in [12]. After evaluating the performance of DeepRMSA with different degrees of flexibility in spectrum assignment (by including in the action space different numbers of candidate FS-blocks on each path), we make DeepRMSA use a fixed first-fit spectrum assignment scheme (always allocating the first available FS-blocks).

Reward: DeepRMSA gets a reward of 1 if a request is successfully accommodated and −1 otherwise.

Training: We design the training of DeepRMSA by applying and modifying the Asynchronous Advantage Actor-Critic (A3C) algorithm [7]. DeepRMSA employs multiple parallel actor-learners (can be seen as incarnations of a DRL agent), each interacting with its own copy of the system environment, to obtain more abundant and diversified training samples. Every actor-learner employs a policy DNN to generate action policies (i.e., probability distributions of taking actions) and a value DNN for estimating the long-term cumulative reward of each system state. The actor-learners maintain and update global policy and value DNNs asynchronously. In DeepRMSA, we adopt DNNs with fully connected architectures [5] that have the same design of input and hidden layers. The output layer of each policy DNN consists of K neurons, outputting the probabilities of selecting the corresponding paths, while each value DNN has only one output neuron.

The procedures of an actor-learner's thread in DeepRMSA are as follows. First, the actor-learner synchronizes the local DNNs with the global parameters and initializes an EON state (step 1). Then, upon receiving each lightpath request, the actor-learner invokes the local DNNs to generate an RMSA policy and an estimation of the cumulative reward taking as input the current EON state (step 2). An RMSA scheme is determined according to the generated policy (step 3). Afterward, the actor-learner receives an immediate (and real) reward, which, together with the EON state, the action taken, and the reward estimation, are stored in an experience buffer (step 4). The actor-learner performs training (i.e., tuning the

parameters of the DNNs) every time the experience buffer contains $2L - 1$ samples (step 5). In particular, for each of the first L samples, the actor-learner calculates the cumulative reward within a window of length L (i.e., the cumulative reward of servicing the current plus the next $L - 1$ requests). With these L samples, the actor-learner then updates the parameters of the global DNNs by a small step toward the direction of minimizing the policy and value losses. The policy loss comprises a term that enables to reinforce actions with larger advantages and an entropy term to encourage exploration (avoid being trapped in local optima). Here, advantages are defined as the differences between the real and the estimated cumulative rewards, indicating how much actions turn out to be better than expected. The value loss is straightforward as the average error of reward estimation. Finally, the actor-learner removes the L samples from the buffer. The rationale of applying such a window-based training mechanism is that by making the cumulative reward for every sample involve L requests, we smooth out the oscillations of the optimization targets (i.e., the cumulative rewards) due to random request arrivals and thus stabilize the training of DeepRMSA. Note that although a large value of L facilitates more stabilized training, on the other hand, it hinders the training signals from new observations being quickly applied (thus hindering the quick response of DeepRMSA to the changing network conditions). Therefore, we envision a moderate value (e.g., 50) to be a proper setup for L. Steps 2 to 5 are repeated until the EON stops operating.

Evaluations

We assessed the performance of DeepRMSA with numerical simulations using the 14-node NSFNET topology. We assumed each link could accommodate 100 FSs. The dynamic lightpath requests were generated according to the Poisson process, with the average arrival rate and service duration being 12 and 14, respectively, and the bandwidth requirements evenly distributed within [25,100] Gb/s. We calculated $K = 5$ candidate paths for each request. We implemented DeepRMSA with 16 actor-learners, each employing DNNs of five hidden layers (128 neurons per layer). For the training of DeepRMSA, we set L and the learning rate as 50 and 10^{-5}, respectively. The running of the simulations consumed 2.33 CPUs of 2.6 GHz and 0.64 GB memory on a standard Linux server, with a duration of ~120 hours. We compared the performance of DeepRMSA with those of two baselines, namely, shortest path routing and first-fit spectrum assignment (SP-FF), and k-shortest path routing and first-fit spectrum assignment (KSP-FF) which has been shown to achieve state-of-the-art performance [11]. Figure 3 plots the results of request blocking probability. It can be seen that DeepRMSA successfully beats both of the baselines after training of around 30,000 epochs and eventually can achieve a blocking reduction of 45.9 percent compared to KSF-FF. The average spectrum utilization ratios from DeepRMSA (after training of 200,000 epochs), KSP-FF, and SP-FF are 32.6, 30.4, and 27.2 percent, respectively. Since DeepRMSA enables accommodating more requests, it utilizes the largest amount of spectrum resources. To verify the robustness of DeepRMSA, we also conducted simulations using the 11-node COST239 topology. The simulation setup remained unchanged, except that the average request arrival rate and service duration were set as 16 and 25, respectively. Results show that DeepRMSA still outperforms KSP-FF with a blocking reduction of 40.7 percent (8.3×10^{-3} vs. 1.4×10^{-2}). The average spectrum utilization ratios from DeepRMSA and KSP-FF are 41.0 and 40.0 percent, respectively.

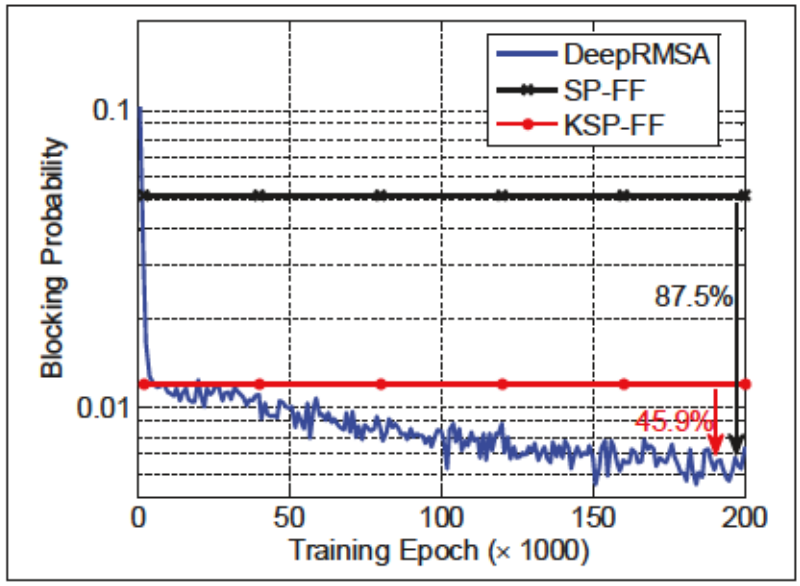

Figure 3. Comparison of request blocking probability between DeepRMSA and the baselines.

Autonomic Multi-Domain Networking with MADRL

Today's Internet is essentially a multi-domain network, where multiple autonomous systems work cooperatively to provide global interconnectivity. With the explosion of emerging applications (e.g., distributed science computing) and the rapid evolution of data center networks, there is an increasing demand for dynamic and high-capacity end-to-end services across multiple domains. In this context, multi-domain EON becomes one of the most promising solutions for the next-generation Internet backbone [3]. The challenges of realizing efficient multi-domain networking lie in designing powerful inter-domain service provisioning paradigms to enable well-coordinated resource allocations across multiple autonomous EONs with very limited intra-domain information. Recent studies have investigated both distributed and semi-centralized architectures for multi-domain EONs [3, 5]. While the distributed solutions lead to poor resource efficiency, the semi-centralized solutions suffer from scalability and survivability issues and may violate the domain autonomy by having a multi-domain orchestrator to dictate the operations of domains. Therefore, we envision a multi-broker-based multi-domain NC&M architecture, where multiple incentive-driven broker agents work with domain managers (DMs) through service level agreements (SLAs) and can cooperate or compete freely, to facilitate more scalable, diversified, and efficient inter-domain

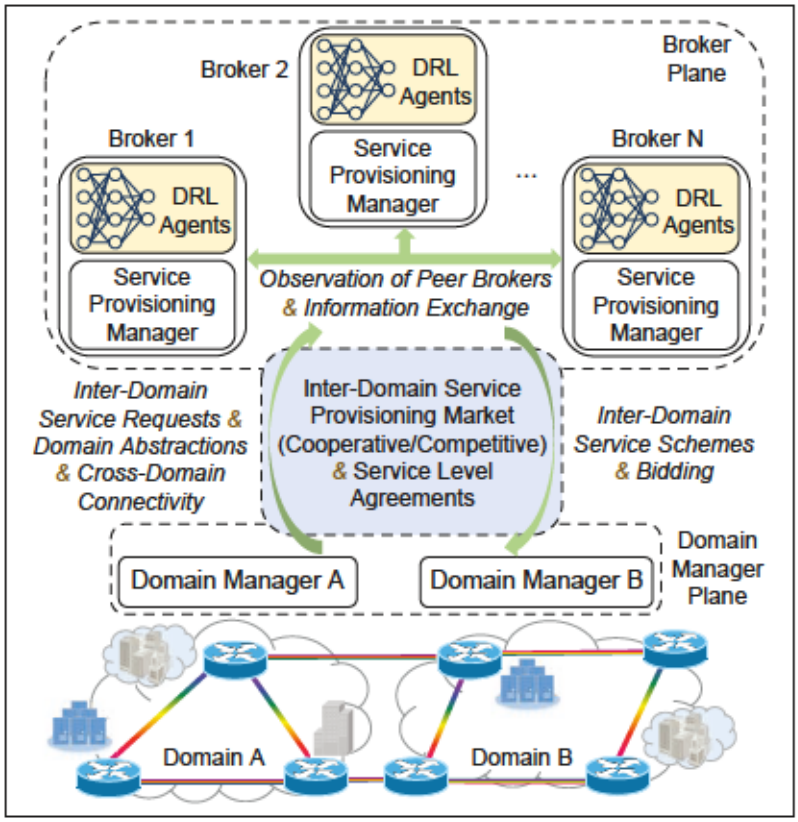

Figure 4. MADRL-based autonomic multi-domain networking framework.

service provisioning [13]. In particular, multi-broker-based NC&M enables each DM to participate in inter-domain service provisioning by subscribing to the services from one or multiple brokers. DMs submit inter-domain service requests and report different levels of domain abstractions (according to mutual confidence and SLAs) to brokers, while brokers return service schemes and biddings. These activities can be seen as forming an inter-domain service provisioning market constrained by SLAs.

Service provisioning in multi-broker-based multi-domain EONs can be modeled as incomplete-information repeated games in nature, where multiple players (i.e., brokers and DMs) compete or cooperate for profit (e.g., throughput and revenue gain) maximization. In addition to the fact that the number of brokers and DMs can be large, multi-domain EON systems are typically very complex (heterogeneous and private topologies, traffic patterns, and policies) and dynamic. Therefore, conventional game-theoretic approaches (e.g., analyzing the Nash equilibria) can hardly be used for optimizing the service provisioning strategies in such a scenario. On the other hand, MADRL has shown appealing prospects in solving multi-agent cooperation and competition tasks [14]. By equipping each agent with the DRL functionality and implementing proper rewarding mechanisms, MADRL enables multiple agents to learn how to adapt to each other and eventually converge to the equilibria. In this section, we introduce MADRL to the service provisioning design of multi-broker-based multi-domain EONs, and extend the autonomic networking framework in Fig. 1 to present an autonomic multi-domain networking framework.

Figure 4 shows the schematic of MADRL-based multi-domain networking. Each EON domain operates autonomically according to the principle depicted in Fig. 1 and receives inter-domain services from the brokers. The brokers learn service provisioning policies with MADRL. Specifically, each broker employs a number of DRL agents to model the behaviors of DMs and its peer brokers and learn the policies for different services. The modular function design and the operation principle of the DRL agents resemble those presented in Fig. 1, except that a service provisioning manager, instead of an SDN controller, handles the communication and service scheduling tasks. The DRL agents take as input state representations including both traffic-engineering-related data (e.g., in-service requests and multi-domain abstractions) and features from the peer brokers (e.g., observations of historical actions and cooperative information exchanges). The brokers train the DRL agents asynchronously through dynamic inter-domain service provisioning experiences, thus enabling an autonomic multi-domain networking system with multiple intelligent agents learning to reach joint optimization situations.

MADRL-Based Inter-Domain RMSA

We present an MADRL-based inter-domain RMSA design [15] as a proof-of-concept demonstration for the proposed autonomic multi-domain networking framework. We assume that brokers abstract each domain as consisting of a few domain border nodes interconnected by a virtual node (see the example in Fig. 5) and have full visibility of the inter-domain connectivity. With the multi-domain abstraction, each broker performs inter-domain RMSA operations following procedures similar to those of DeepRMSA. Specifically, upon receiving an inter-domain lightpath request, a broker invokes its DRL agent to recommend a routing path within the abstraction. The routing path essentially suggests a domain sequence for the request by specifying the border nodes to be traversed. Then the involved DMs set up the corresponding lightpath domain by domain, calculating intra-domain segments between border nodes and allocating spectrum resources on them (by applying either DRL-based or heuristic policies [3]). The modeling and training of the DRL agents also resemble those of DeepRMSA but with two modifications. First, the features for each candidate path are derived from only the spectrum utilization of inter-domain links and do not include the number of required FSs (the physical length of each path and therefore the modulation format cannot be determined). For the sake of simplicity, we do not consider communications among brokers so that the DRL agents perform independent learning (i.e., treating the behaviors of peer brokers as part of the holistic system environment). Second, the DRL agents are trained with the Advantage Actor-Critic (A2C) algorithm reduced from A3C (with only one actor-learner).

We evaluated the performance of the MADRL-based inter-domain RMSA design with the 4-domain topology in Fig. 5 and two brokers. We assumed the link capacity to be 100 FSs and that each domain border node

was equipped with 10 optical-electrical-optical converters. The dynamic lightpath requests were generated following an arrival rate of 15 and an average service duration of 15. We assumed a fixed market partitioning scheme where inter-domain requests originating from nodes [3, 4, 5, 7, 8, 11, 12, 15, 17, 18, 23, 24] are handled by Broker 1, while the others are forwarded to Broker 2. The implementation and training of the DRL agents used similar parameter setups to those of the evaluations for DeepRMSA. We also implemented the Least-Hop and Balanced-Load routing algorithms as the baselines. For all the algorithms, the DMs applied the KSP-FF algorithm for provisioning intra-domain requests. Figure 6 shows the evolution of request blocking probability from Brokers 1 and 2 during the training. We can see that MADRL enables the brokers to improve their policies steadily from dynamic service provisioning. MADRL surpasses both of the baselines after training of 160,000 epochs, and eventually facilitates 23.9 and 23.1 percent blocking reductions for Brokers 1 and 2, respectively, compared to the best performance of the baselines. Meanwhile, it is apparent that the MADRL-based design facilitates higher overall multi-domain throughput, indicating a situation where interests of DMs and brokers are jointly optimized.

Conclusion

This article presents a DRL-based autonomic networking framework for EONs. The proposed framework enables self-learning-based service provisioning by employing DRL agents to learn policies through dynamic EON operations. Based on the framework, we elaborate on the design of DeepRMSA, a DRL agent for RMSA in EONs. Further, we extend the autonomic networking framework to a multi-domain EON scenario by introducing a multi-broker-based NC&M architecture and applying MADRL. Numerical results show notable advantages of the proposed framework compared to the existing heuristic-based designs.

Open issues for building autonomic EONs include:

1. How to effectively train DRL agents for large-scale topologies with more complex system state representations
2. How to improve the robustness of the DRL agents against fast network evolutions and sudden changes (e.g., topology changes due to network failures/anomalies)

In addition to developing more advanced learning algorithms, a potential solution for improving the scalability of the proposed framework could be applying a distributed learning mechanism that employs multiple cooperative DRL agents exploring different subsets of the network in parallel. Meanwhile, although it is feasible for DRL agents to learn policies only through online operations, a more practical and time-efficient approach is to build accurate service provisioning simulators that can faithfully simulate real-world network operations and to train DRL agents offline with the simulators before activating them for online use. Building such accurate simulators requires full knowledge about the network operation principles and traffic profiles, and effective modeling techniques, which demands further research efforts. Lastly, the recent advances in ML have enabled the capabilities of learning temporal and spatial information from time-series and topological inputs, which can potentially be leveraged to improve the adaptability (to traffic evolutions or topology changes) of the DRL agents.

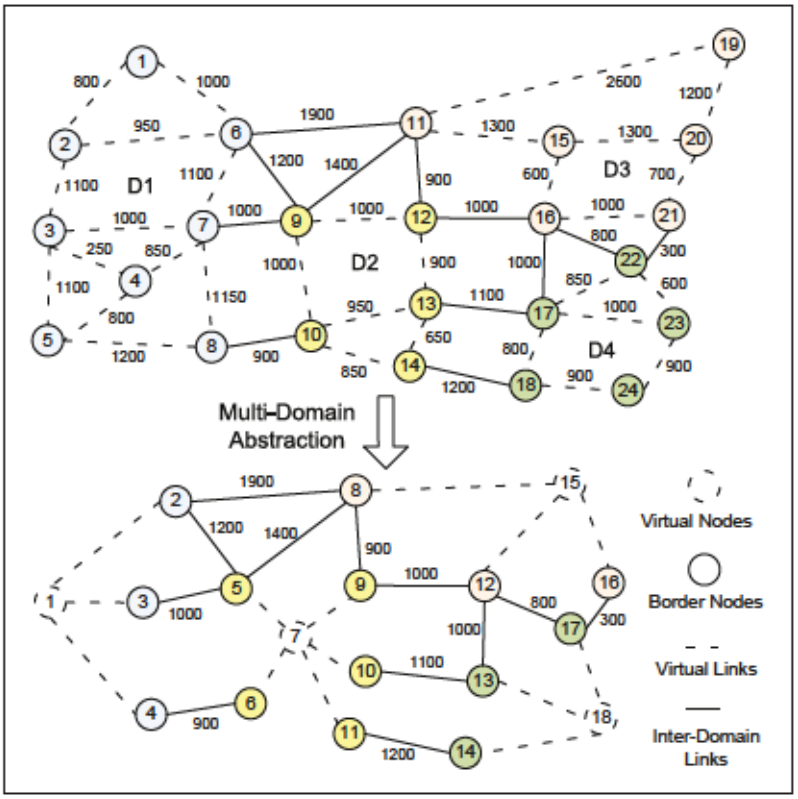

Figure 5. Multi-domain abstraction of a 4-domain physical topology by brokers.

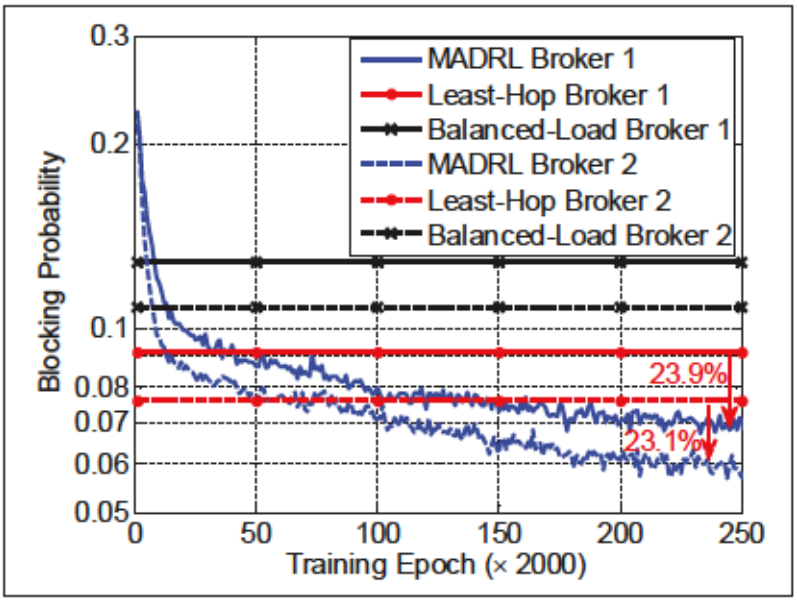

Figure 6. Comparison of inter-domain request blocking probability between the MADRL-based approach and the baselines.

Acknowledgments

This work was supported in part by DOE DE-SC0016700 and NSF ICE-T:RC 1836921.

References

[1] O. Gerstel *et al.*, "Elastic Optical Networking: A New Dawn for the Optical Layer?," *IEEE Commun. Mag.*, vol. 50, no. 4, Apr. 2012, pp. S12–S20.

[2] P. Lu *et al.*, "Highly Efficient Data Migration and Backup for Big Data Applications in Elastic Optical Inter-Data-Center Networks," *IEEE Network*, vol. 29, no. 5, Sept./Oct. 2015, pp. 36–42.

[3] Z. Zhu *et al.*, "OpenFlow-Assisted Online Defragmentation in Single-/Multi-Domain Software-Defined Elastic Optical Networks," *J. Opt. Commun. Net.*, vol. 7, Jan. 2015, pp. A7–A15.

[4] D. Rafique *et al.*, "Cognitive Assurance Architecture for Optical Network Fault Management," *J. Lightwave Tech.*, vol. 36, no. 7, Apr. 2018, pp. 1443–50.

[5] X. Chen *et al.*, "Knowledge-Based Autonomous Service Provisioning in Multi-Domain Elastic Optical Networks," *IEEE Commun. Mag.*, vol. 56, no. 8, Aug. 2018, pp. 152–58.

[6] J. Guo and Z. Zhu, "When Deep Learning Meets Inter-Datacenter Optical Network Management: Advantages and Vulnerabilities," *J. Lightwave Tech.*, vol. 36, no. 20, Oct. 2018, pp. 4761–73.

[7] V. Mnih *et al.*, "Asynchronous Methods for Deep Reinforcement Learning," *Proc. ICML*, 2016, pp. 1928–37.

[8] Z. Xu *et al.*, "Experience-Driven Networking: A Deep Reinforcement Learning Based Approach," *Proc. INFOCOM*, 2018, pp. 1871–79.

[9] S. Salman *et al.*, "DeepConf: Automating Data Center Network Topologies Management with Machine Learning," *Proc. NetAI*, 2018, pp. 8–14.

10] Z. Zhu *et al.*, "Dynamic Service Provisioning in Elastic Optical Networks with Hybrid Single-/Multi-Path Routing," *J. Lightwave Tech.*, vol. 31, no. 1, Jan 2013, pp. 15–22.

[11] Y. Yin *et al.*, "Spectral and Spatial 2D Fragmentation-Aware Routing and Spectrum Assignment Algorithms in Elastic Optical Networks," *J. Opt. Commun. Net.*, vol. 5, no. 10, Oct. 2013, pp. A100–A106.

[12] M. Ju *et al.*, "Leveraging Spectrum Sharing and Defragmentation to P-Cycle Design in Elastic Optical Networks," *IEEE Commun. Lett.*, vol. 21, no. 3, Mar. 2017, pp. 508–11.

[13] X. Chen *et al.*, "Incentive-Driven Bidding Strategy for Brokers to Compete for Service Provisioning Tasks in Multi-Domain SD-EONs," *J. Lightwave Tech.*, vol. 34, no. 16, 2016, pp. 3867–76.

[14] R. Lowe *et al.*, "Multi-agent Actor-Critic for Mixed Cooperative-Competitive Environments," *Proc. NIPS*, 2017, pp. 6379–90.

[15] X. Chen *et al.*, "Multi-Agent Deep Reinforcement Learning in Cognitive Inter-Domain Networking with Multi-Broker Orchestration," *Proc. OFC*, 2019, paper M2A.2.

Biographies

XIAOLIANG CHEN [S'14, M'16] (xlichen@ucdavis.edu) received his Ph.D. degree from the University of Science and Technology of China in 2016. He is currently a research scholar at the University of California (UC) Davis. His research interests include optical networks, software-defined networking, and cognitive networking. He has published more than 50 papers in journals and conferences of IEEE and OSA. He is an Associate Editor of Springer's *Telecommunication Systems Journal*, Wiley's *Emerging Telecommunications Technologies Journal*, and *KSII Transactions on Internet and Information Systems*, and was/is a TPC member of IEEE ICNC 2018 and 2019, and ICC 2018.

ROBERTO PROIETTI (rproietti@ucdavis.edu) received his M.S. degree in telecommunications engineering from the University of Pisa, Italy, in 2004, and his Ph.D. in fiber optical communications from Scuola Superiore Sant'Anna, Pisa, Italy, in 2009. He is a project scientist at UC Davis. His research interests include high-spectrum-efficiency coherent transmission systems and elastic optical networking, access optical networks and radio over fiber, and optical switching technologies and architectures for supercomputing and data center networks.

S. J. BEN YOO [S'82, M'84, SM'97, F'07] (sbyoo@ucdavis.edu) received his B.S., M.S., and Ph.D. degrees from Stanford University, California, in 1984, 1986, and 1991, respectively. He is a Distinguished Professor of electrical engineering at UC Davis. His research at UC Davis includes future computing, photonic communications, cognitive networking, and integrated systems for the future Internet. He is a recipient of the DARPA Award for Sustained Excellence (1997), the Bellcore CEO Award (1998), the Mid-Career Research Faculty Award (2004, UC Davis), and the Senior Research Faculty Award (2011, UC Davis).