Routing Based On Deep Reinforcement Learning In Optical Transport Networks

Conference Paper · January 2019

DOI: 10.1364/OFC.2019.M2A.6

CITATIONS
23

READS
852

7 authors, including:

Jose Suarez-Varela
Telefónica I+D
44 PUBLICATIONS **438** CITATIONS
SEE PROFILE

Li Kuang
Huawei Technologies
18 PUBLICATIONS **177** CITATIONS
SEE PROFILE

Pere Barlet-Ros
Universitat Politècnica de Catalunya
124 PUBLICATIONS **1,673** CITATIONS
SEE PROFILE

Albert Cabello
Universitat Politècnica de Catalunya
190 PUBLICATIONS **3,279** CITATIONS
SEE PROFILE

Routing Based On Deep Reinforcement Learning In Optical Transport Networks

José Suárez-Varela[1], Albert Mestres[1], Junlin Yu[2], Li Kuang[2], Haoyu Feng[3], Pere Barlet-Ros[1], Albert Cabellos-Aparicio[1]

1. Universitat Politècnica de Catalunya (UPC), Barcelona, Spain
Email: {jsuarezv, amestres, acabello, pbarlet}@ac.upc.edu
2. Network Research Department, Huawei Technologies Co., LTD., Shenzhen, China
Email: {yujunlin2, kuangli}@huawei.com
3. Ottawa Optical System Competency Centre, Huawei Technologies Co., LTD., Ottawa, Canada
Email: fenghaoyu@huawei.com

Abstract: This paper addresses the use of Deep Reinforcement Learning for automatic routing in Optical Transport Networks at the electrical-layer level. We propose a DRL-based solution that achieves both high performance and fast learning.

OCIS codes: (060.4251) Networks, assignment and routing algorithms; (060.4256) Networks, network optimization.

1. Introduction

Recent advances in Deep Reinforcement Learning (DRL) are showing a significant improvement in decision-making and automated control problems [1]. In this context, there is a growing interest in the research community to apply DRL-based solutions for optimization problems in networks. However, even the most advanced DRL algorithms still fail to achieve the ability to generalize when applied in general to network-related scenarios. This hinders the DRL agent from making good decisions when facing network states not explored during training.

In this paper, we address the application of DRL techniques specifically for routing in Optical Transport Networks (OTN). We show that a good representation of the state and action spaces is crucial to facilitate the learning process to the agent. This enables to reduce the level of knowledge abstraction required to the agent and, consequently, contributes to potentially achieve better performance. In our context, the use of representations that can better capture the singularities of network topologies can considerably simplify the learning process to the DRL agent. To this end, our approach is to design state and action representations that convert the challenging problem of routing traffic over complex OTN topologies to a problem easier to resolve. All this, without losing generality. We propose a DRL-based solution specifically designed to route traffic in OTNs that makes it easier for the agent to learn the overall utilization and the dependencies among the end-to-end paths in the network topology. As a result, this also facilitates the detection of possible singularities in the network, such as potential bottlenecks.

We evaluate the proposed DRL-based solution in an OTN scenario to route traffic demands at the level of the electrical domain. Moreover, we compare these evaluation results with those obtained using other DRL-based proposals in previous works addressing also routing problems. All this evaluation includes the use of specific traffic models that simulate real-world network scenarios.

2. Routing scenario and proposed DRL-based solution

Fig. 1 shows a scheme of the routing scenario. Our DRL agent makes decisions at the electrical domain. As a consequence, it considers an OTN with some predefined lightpaths and operates over a logical topology, which is only composed by the Reconfigurable Optical Add-Drop Multiplexer (ROADM) nodes and the lightpaths that connect them. Then, the role of the DRL agent is to route new traffic demands through particular sequences of lightpaths (i.e., end-to-end paths). Note that this is a challenging task for the agent since it should learn during training some singularities of the OTN topology such as potential bottlenecks as well as understand the inter-dependencies among end-to-end paths. All this learning process, it is additionally hampered by the uncertainty in the generation of future service requests. Since the agent works at the electrical domain, traffic demands are considered requests of Optical Data Units (ODUk) signals defined in the ITU-T Recommendation G.709 [2]. For the sake of simplicity, we consider 5 different types of traffic demands (ODU0 to ODU4) whose bandwidth requirements are expressed in terms of multiples of the minimum

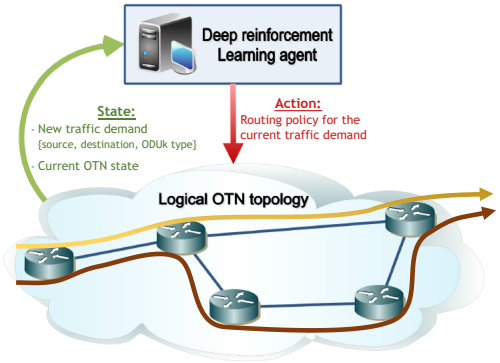


Fig. 1: Routing scenario

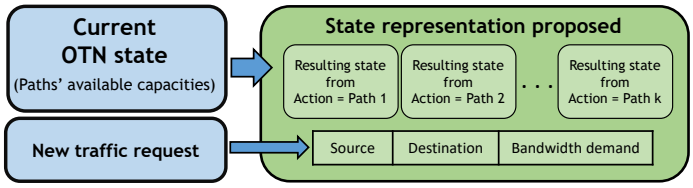


Fig. 2: State representation proposed.

ODU0 bandwidth unit (hereafter "ODU0 BU")[1]. Accordingly, the capacity of the network lightpaths is also defined as the number of ODU0 BUs they can carry.

2.1. *State and action representations*

We propose a solution based on DRL to specifically tackle routing optimization in OTN scenarios. The design of this solution involves both how to define the network state (i.e., the *observation space*) and the set of actions that the agent can apply (i.e., the *action space*).

For the action space, we propose discrete sets of actions. We consider that the DRL agent has a list of "k" candidate paths for each source-destination pair in the OTN. For instance, this list may include the "k" shortest paths for every source-destination pair. Then, each epoch the agent receives a new traffic request and must select one of the "k" candidate paths that connect the source and the destination of such request.

For the state representation, instead of using link-level statistics as in previous works in the literature, we propose using statistics at the *path-level*. This way, the agent does not need to abstract knowledge from the link-level to the path-level. We assume that the DRL agent has access to monitoring data of the OTN. In particular, it should have access to the current available capacity in the "k" candidate end-to-end paths for each source-destination pair in the network[2]. Fig. 2 shows an scheme of our representation. Here, our approach is to provide directly the agent with information about the resulting available capacities after routing the current traffic demand through each of the 'k" candidate paths (Fig. 2 right). Intuitively, this simplifies the problem given that the representation proposed shows the agent a comprehensive picture that describes the consequences of applying every possible action.

Additionally, the state representation must include a description of the current traffic request to be routed through the network. To represent it, we use a tuple $\{source, destination, bandwidth\ demand\}$, where "source" and "destination" are vectors with one-hot encoding (as they do in Deep-RMSA [3]), and the "bandwidth demand" is an integer that represents the bandwidth requirement (in ODU0 BUs).

3. Evaluation and discussion

In our experiments we trained a DRL agent using an implementation of the Trust Region Policy Optimization (TRPO) algorithm [4] provided by ChainerRL (v0.3.0) [5]. The agent is modeled with a neural network with two hidden layers, each one with 64 neurons. The number of neurons in the input depends on the state representation used and the neurons of the output layers are 4. We selected a value of 0.995 for the *discount factor* (γ) and 0.97 for the *exploration parameter* (λ), which obtained the best results.

In our evaluation scenario, we assume that service requests do not expire. Thus, an episode begins with an empty network and it ends when a new service request cannot be routed because there is not enough available capacity in the path selected by the agent. Hence, the final objective of the agent is to maximize the total bandwidth of service requests that can be routed in the network. That is to postpone as much as possible the event that a request cannot be routed. As a result, the immediate reward of the agent is the bandwidth of the current service request (in ODU0 BUs)

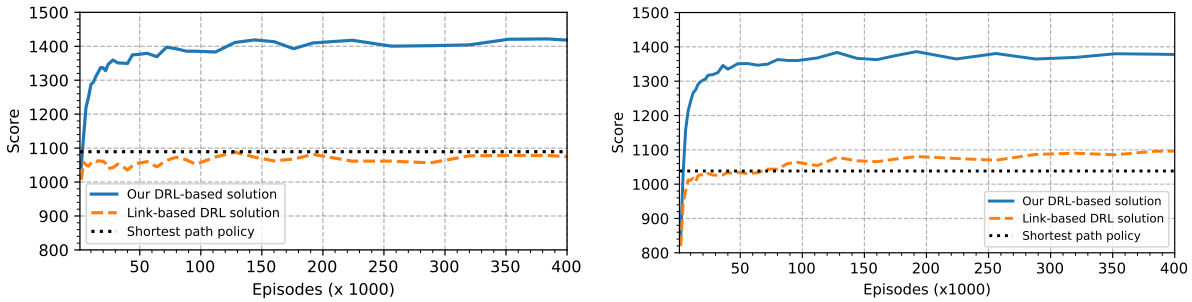
[1] An ODU0 signal can carry 1.244 Gbit/s approximately. Then, we consider the following bandwidth values for the remaining signals: ODU1=2 ODU0 BUs, ODU2=8 ODU0 BUs, ODU3=32 ODU0 BUs and ODU4=64 ODU0 BUs.

[2] Also note that, for other networking problems out of the scope of this paper (e.g., QoS-aware routing) it might be useful to extend our representation with some additional path-level statistics (e.g., estimated end-to-end delays).

in case it was routed properly, otherwise the reward is 0. Note that the objective of DRL-based agents is to learn a policy that maximizes the cumulative expected reward during an entire episode.

We evaluated the agent in the 14-node NSFNET [6], with a custom-built OTN simulator. We consider that all the links have a capacity equal to 200 ODU0 BUs. We compared the results using the DRL solution we propose with those results obtained using DRL agents with a representation that provides only the available capacity of links (hereafter referred to as "Links"). This latter solution is similar to the approach followed by DeepRMSA [3]. In both cases, we use the same action space: the selection of one path among k=4 candidate shortest paths.

We trained the agent in two scenarios with different traffic profiles: (i) with a bimodal synthetic traffic distribution [7], in which 20% of nodes generate 80% of the traffic and elephant-mice distribution [8] for the distribution of the ODUk requests, and (ii) with an uniform distribution for sources, destinations and ODUk types. This latter scenario represents the most challenging case for the DRL agent, given that it cannot exploit particular characteristics of the traffic profile to efficiently direct the exploration during training.



(a) Synthetic traffic distribution

(b) Uniform traffic distribution

Fig. 3: Average score as a function of the training episodes

Fig. 3 shows, for both traffic scenarios, the score achieved by the agent as a function of the training episodes. The score is defined as the average accumulated rewards (i.e., the average bandwidth successfully routed) over 5,000 different episodes. Each subfigure shows the score (higher is better) for both representations mentioned above and the performance achieved by a traditional Shortest Path (SP) routing policy. For both traffic distributions, we observe a similar behavior: the "Links" representation achieves a score comparable to the SP policy, while the agent with the representation we propose is able to route 30% more traffic than using the "Links" representation. Note that in the scenario with synthetic traffic (Fig. 3a), the agent achieves a slightly higher score, since there are more low bandwidth requests, which are easier to be routed properly.

4. Conclusions

We conclude that the DRL-based solution we propose achieves much higher performance than simple solutions using only information about the links' utilization (approx. 30% better).

Acknowledgments: This work has been supported by the Spanish MINECO under contract TEC2017-90034-C2-1-R (ALLIANCE) and the Catalan Institution for Research and Advanced Studies (ICREA).

References

1. V. Mnih et al, "Human-level control through deep reinforcement learning," Nature, 2015.
2. "ITU-T Recommendation G.709/Y.1331: Interface for the optical transport network," http://www.itu.int/rec/T-REC-G.709/, 2016.
3. X. Chen et al, "Deep-RMSA: A deep-reinforcement-learning routing,modulation and spectrum assignment agent for elastic optical networks," OFC, 2018.
4. J. Schulman et al, "Trust region policy optimization," ICML, 2015.
5. "ChainerRL," https://github.com/chainer/chainerrl. Accessed: 2018-03-10.
6. H. Beyranvand et al, "A quality-of-transmission aware dynamic routing and spectrum assignment scheme for future elastic optical networks," Journal of Lightwave Technology, 31(18), 3043-3054, 2013.
7. A. Medina et al, "Traffic matrix estimation: Existing techniques and new directions," SIGCOMM Comput. Comm. Rev., 32, 161-174, 2002.
8. L. Guo et al, "The war between mice and elephants," ICNP 2001.