Deep Reinforcement Learning for Infrastructure as a Service over Flexible Optical Networks

Michael Doherty, Alejandra Beghelli

Optical Networks Group, University College London (UCL), United Kingdom
michael.doherty.21@ucl.ac.uk

Abstract *We apply a single deep reinforcement learning agent for dynamic virtual network provisioning. Benchmarked against state of the art heuristics, our approach achieves an order of magnitude lower blocking probability. Interpretability analysis provides insight to the agent's use of spectrum resources.*

Introduction

Infrastructure as a Service (IaaS) is a cloud computing paradigm where storage, computing and networking resources are leased to customers by means of virtualisation. In its most complex form, customers can request an entire network. In that case, its topology and node/link capacities must be specified. Such a specification is known as a virtual network request, as shown in Fig. 1.

Given the ever-increasing Internet traffic, flexible-grid elastic optical networks (EON)[1] are a promising approach to increase spectrum usage efficiency. In EONs, spectrum is divided into fine-grained frequency slot units (FSU). This allows greater flexibility and spectral efficiency but judicious selection of FSUs is required to avoid fragmenting the spectrum into unusable blocks.

For IaaS provision over EON, resources (e.g. compute, storage) must be allocated at nodes and spectrum on connecting links, subject to the constraints of continuous and contiguous FSU blocks. This is known as the virtual optical network embedding (VONE) problem. For efficient resource usage and revenue generation, low-blocking VONE strategies must be in place.

Previous Work

Heuristic algorithms[2][3][4][5][6] and a distributed protocol[7] have been proposed to solve the VONE problem and constitute the state of the art.

However, hand-crafted heuristics do not guarantee optimal resource allocation. Optimal solutions can be found using exact solution methods (e.g. integer linear programming), but they are not feasible for dynamic environments or large networks. Consequently, the application of deep reinforcement learning (DRL) to the VONE problem has been studied recently[8].

DRL presents an opportunity to effectively search the solution space and learn a superior allocation policy. In the literature, multi-agent DRL approaches have been applied to VONE because the large combinatorial space of node- and path-selections has been considered too large for a single agent[9]. However, multi-agent DRL is more complex and may lead to less-optimised solutions due to the difficulty of agent cooperation. Separate agents to allocate nodes and links[8] or nodes, links and backup resources[9] have been proposed and an arbitrary selection of heuristics used for performance comparison. The latter is a result of a lack of heuristic benchmarks in this area.

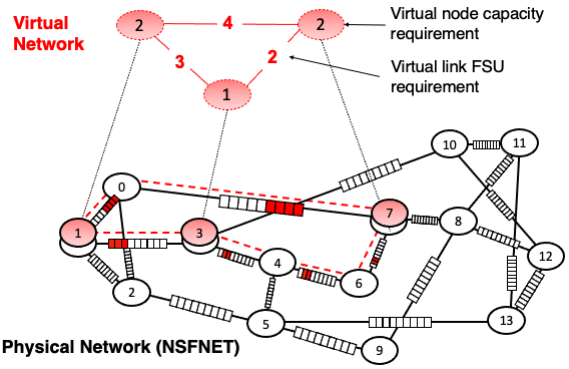


Fig. 1: The embedding of a virtual network to an elastic optical network, with spectrum divided into frequency slot units (FSU)

Contribution

In this paper, a continuation of our work in[10], we present the first benchmarking of VONE heuristics and compare the performance of a single DRL agent against the best-performing heuristic. By using a single agent (as opposed to separate agents solving sub-problems sequentially as in[8]) for the VONE problem, a better quality solution can be obtained. The DRL agent reported here extends the best-performing agent in[10] by en-

abling it to work with an increased action space, a key aspect for realistic application. We also present preliminary results interpreting the actions selected by the agent and comparing with the heuristic decisions. We expect these results to help develop more efficient strategies for IaaS provision.

Network and Traffic Model

The substrate physical network is an EON modeled as an undirected graph with N_s nodes and L_s bidirectional links, each with capacity and FSUs. Virtual networks are undirected graphs consisting of N_v nodes and L_v links, each with specific capacity and bandwidth (FSUs) requirements.

Traffic is dynamic, with virtual network requests arriving and departing stochastically. Requests arrive following a Poisson process and depart according to an exponential distribution with the inverse of the holding time.

Deep Reinforcement Learning Algorithm

The deep reinforcement learning paradigm separates the learning process into the environment and the agent components. The environment models the substrate EON and handles the generation and allocation of virtual network requests, interpretation of the agent's actions, and the reward function. The agent's interaction with the environment involves the network state observation, the agent's action, and the resulting reward.

The observation space includes the current virtual network request, the state of FSUs on each link, and the remaining node resource capacities.

The action space is divided into node and path sections. The node action space dimensionality depends on the number of virtual and substrate nodes, given by $(1 \times N_s^{N_v})$. For a request comprising L_v virtual links, the path action space dimensionsionality becomes $(L_v \times k * N_f)$, where N_f is the number of FSUs per link. The reward function is kept simple, providing a signal to optimize the agent's policy without guiding its behavior. The environment determines success or failure based on the availability of node resources and bandwidth. Success is rewarded with a value of 0, while failure yields a value of -10.

The DRL algorithm is an implementation of Proximal Policy Optimisation[11], modified to allow multi-step invalid action masking.

Multi-step Invalid Action Masking. The technique of invalid action masking limits the available choice of actions by the agent based on knowledge of the environment e.g. available resources, and has been successfully applied to optimisation problems in optical networks[12][13][14]. The technique was extended in previous work to allow context-dependent multi-dimensional actions, e.g. node action followed by path action, to be masked[10], and is referred to as multi-step invalid action masking. This technique is further developed in this work to recursively mask actions that correspond to already-allocated virtual links in the same virtual network.

Training. Each training episode starts with an unoccupied substrate network. A training episode comprises 10^4 timesteps, with one request per timestep, during which experiences are collected in a rollout buffer size n_{steps}. The policy optimisation step occurs when the rollout buffer is full. The agent was trained for 100 episodes. The traffic load for training was set to 60 Erlangs. Empirically, 60 Erlangs training load with n_{steps} =50, discount factor γ = 0.8 and generalized advantage estimation λ-factor=0.9 was found to result in efficient training for this network topology and traffic distribution.

Training was performed on the NSFNET topology (14 nodes, 21 bidirectional links) as the substrate network, equipped with 100 FSUs per link and 30 compute units per node. The virtual link requests were randomly selected from$\{$2,3,4$\}$ FSUs. The virtual node requests were randomly selected from $\{$1,2$\}$ compute units. These capacity assumptions meant that link capacity was the limiting factor of the number of virtual requests that could be accepted.

The virtual network topology was restricted to a 3-node ring topology to facilitate comparisons between heuristic and agent performance. The mean service holding time is set to 10 time units, with the traffic model as described in the previous section. Random number generation for the traffic model was seeded to ensure diverse patterns across training episodes and reproducible patterns during evaluation.

Results

The evaluation was performed on the same environment model as used in training. In order to benchmark the agent performance, 3 state of the art heuristics were evaluated across traffic loads. Each heuristic comprises a different node-mapping and path-mapping component. Each combination of these components was evaluated, resulting in 9 distinct heuristics. The heuristics are Consecutiveness-Aware Local Resource Ca-

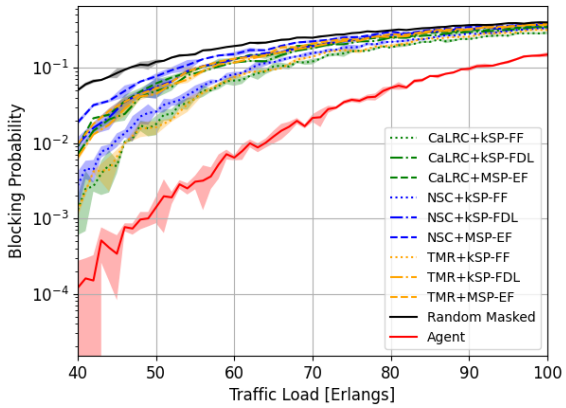


Fig. 2: The mean and standard deviation of blocking probability for the heuristics and DRL agent, across traffic loads.

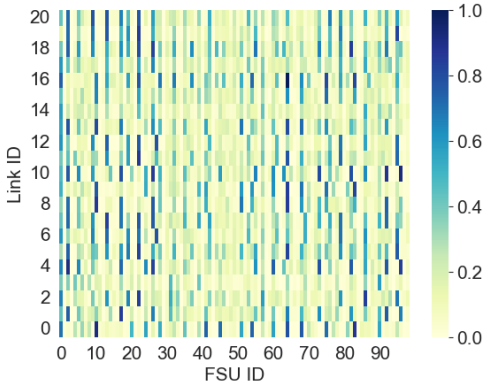
pacity k-Shortest Path First Fit (CaLRC+kSP-FF), Node Switching Capacity kSP Fragmentation Degree Loss (NSC+kSP-FDL), and Topology and Multi Resources Modified Shortest Path Exact Fit (TMR+MSP-EF). The "+" denotes the separation between the node-mapping and the path-mapping components. The reader is directed to[2][4][6] for details of these algorithms. Some heuristics were omitted because they have already been shown to be inferior (as the "LPM" heuristic[3] is shown in[8]) or are designed for a different variant of the problem[5].

Random selection of actions from amongst the available masked choices was also evaluated ("Random Masked" in figure 2). Figure 2 shows the blocking probability of each heuristic, evaluated for 5 episodes of 10^4 timesteps from 40 to 100 Erlangs. The traffic range is selected for the values at which blocking is first observed for the agent and before blocking exceeds 15%.

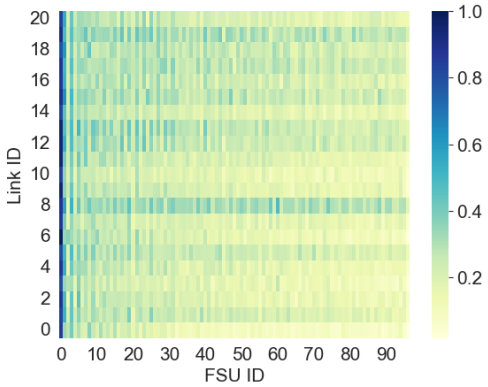
Notably, the kSP-FF path mapping heuristic results in the best performance for all node heuristics, followed by kSP-FDL then MSP-EF. The heuristic with the lowest blocking probability across the traffic range is CaLRC+kSP-FF.

CaLRC+kSP-FF is compared with the agent performance, evaluated across the same traffic distributions, in Figure 2. The agent achieves an order of magnitude lower mean blocking probability at 40 Erlangs and 15% lower at 100 Erlangs. This significant improvement enables network operators to provide an acceptable level of service at higher traffic loads than heuristics are capable of.

To understand how the agent achieves this superior performance, the utilisation of FSUs across the network was recorded for 1 evaluation episode at 100 Erlang traffic, for both the agent and CaLRC+kSP-FF heuristic. Figure 3 shows the resulting heatmaps, normalised in each case by the peak number of requests in which a link-FSU was utilised. Comparison of Figures 3(a) and (b) show the more complex distribution of utilised resources for the agent compared to the heuristic, which utilises the first available FSU. The darker horizontal bands of Figure 3(b) show the heuristic strongly favours certain links, e.g. 8, 19, compared to the more balanced distribution of the agent.



(a) DRL Agent



(b) CaLRC+kSP-FF heuristic

Fig. 3: Normalised heatmaps of link and FSU utilisation

Conclusions

We have identified CaLRC+KSP-FF as the best-performing heuristic for the VONE problem. Our single DRL agent exhibits blocking one order of magnitude lower than this heuristic. Analysis of the agent's use of spectrum resources suggests that a more balanced link and FSU utilisation leads to this significantly better performance.

Acknowledgements

Financial support from EPSRC Centre for Doctoral Training in Connected Electronic and Photonic Systems (CEPS CDT) and EPSRC Programme Grant TRANSNET (EP/R035342/1) is gratefully acknowledged.

References

[1] O. Gerstel, M. Jinno, A. Lord, and S. B. Yoo, "Elastic optical networking: A new dawn for the optical layer?", *IEEE Communications Magazine*, vol. 50, no. 2, s12–s20, Feb. 2012, ISSN: 1558-1896. DOI: 10.1109/MCOM.2012.6146481.

[2] L. Gong and Z. Zhu, "Virtual Optical Network Embedding (VONE) Over Elastic Optical Networks", en, *Journal of Lightwave Technology*, vol. 32, no. 3, pp. 450–460, Feb. 2014, ISSN: 0733-8724, 1558-2213. DOI: 10.1109/JLT.2013.2294389. [Online]. Available: http://ieeexplore.ieee.org/document/6679238/ (visited on 01/06/2023).

[3] B. Chen, Y. Zhao, and J. Zhang, "Energy-efficient virtual optical network mapping approaches over converged flexible bandwidth optical networks and data centers", *Opt. Express*, vol. 23, no. 19, pp. 24 860–24 872, Sep. 2015. DOI: 10.1364/OE.23.024860. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-23-19-24860.

[4] H. Wang, X. Xin, J. Zhang, Y. Sun, and Y. Ji, "Dynamic virtual optical network mapping based on switching capability and spectrum fragmentation in elastic optical networks", in *2016 21st OptoElectronics and Communications Conference (OECC) held jointly with 2016 International Conference on Photonics in Switching (PS)*, Jul. 2016, pp. 1–3.

[5] M. Zhu, S. Zhang, Q. Sun, G. Li, B. Chen, and J. Gu, "Fragmentation-Aware VONE in Elastic Optical Networks", en, *Journal of Optical Communications and Networking*, vol. 10, no. 9, p. 809, Sep. 2018, ISSN: 1943-0620, 1943-0639. DOI: 10.1364/JOCN.10.000809. [Online]. Available: https://opg.optica.org/abstract.cfm?URI=jocn-10-9-809 (visited on 01/06/2023).

[6] W. Wei, H. Gu, K. Wang, X. Yu, and X. Liu, "Improving Cloud-Based IoT Services Through Virtual Network Embedding in Elastic Optical Inter-DC Networks", *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 986–996, Feb. 2019, ISSN: 2327-4662, 2372-2541. DOI: 10.1109/JIOT.2018.2866504. [Online]. Available: https://ieeexplore.ieee.org/document/8449303/ (visited on 05/04/2023).

[7] D. Bórquez-Paredes, A. Beghelli, A. Leiva, *et al.*, "Agent-based distributed protocol for resource discovery and allocation of virtual networks over elastic optical networks", en, *Journal of Optical Communications and Networking*, vol. 14, no. 8, p. 667, Aug. 2022, ISSN: 1943-0620, 1943-0639. DOI: 10.1364/JOCN.450314. [Online]. Available: https://opg.optica.org/abstract.cfm?URI=jocn-14-8-667 (visited on 01/06/2023).

[8] G. Li, C. Xi, and R. Zhu, "Multi-Agent Deep Reinforced Virtual Network Embedding in Elastic Optical Networks", en, in *2022 20th International Conference on Optical Communications and Networks (ICOCN)*, Shenzhen, China: IEEE, Aug. 2022, pp. 1–3, ISBN: 978-1-66545-898-6. DOI: 10.1109/ICOCN55511.2022.9900943. [Online]. Available: https://ieeexplore.ieee.org/document/9900943/ (visited on 01/06/2023).

[9] F. He and E. Oki, "Shared Protection-Based Virtual Network Embedding Over Elastic Optical Networks", en, *IEEE Transactions on Network and Service Management*, vol. 19, no. 3, pp. 2869–2884, Sep. 2022, ISSN: 1932-4537, 2373-7379. DOI: 10.1109/TNSM.2022.3178350. [Online]. Available: https://ieeexplore.ieee.org/document/9782680/ (visited on 01/06/2023).

[10] M. D. Doherty, Y. Zhang, and A. Beghelli, "Masked deep reinforcement learning for virtual network embedding on elastic optical networks", in *2023 International Conference on Optical Network Design and Modeling (ONDM)*, 2023.

[11] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, *Proximal Policy Optimization Algorithms*, arXiv:1707.06347 [cs], Aug. 2017. DOI: 10.48550/arXiv.1707.06347. [Online]. Available: http://arxiv.org/abs/1707.06347 (visited on 01/20/2023).

[12] M. Shimoda and T. Tanaka, "Mask RSA: End-To-End Reinforcement Learning-based Routing and Spectrum Assignment in Elastic Optical Networks", en, in *2021 European Conference on Optical Communication (ECOC)*, Bordeaux, France: IEEE, Sep. 2021, pp. 1–4, ISBN: 978-1-66543-868-1. DOI: 10.1109/ECOC52684.2021.9606169. [Online]. Available: https://ieeexplore.ieee.org/document/9606169/ (visited on 01/09/2023).

[13] Z. Shabka and G. Zervas, *Resource Allocation in Disaggregated Data Centre Systems with Reinforcement Learning*, arXiv:2106.02412 [cs], Nov. 2021. DOI: 10.48550/arXiv.2106.02412. [Online]. Available: http://arxiv.org/abs/2106.02412 (visited on 01/18/2023).

[14] J. W. Nevin, S. Nallaperuma, N. A. Shevchenko, Z. Shabka, G. Zervas, and S. J. Savory, "Techniques for applying reinforcement learning to routing and wavelength assignment problems in optical fiber communication networks", *Journal of Optical Communications and Networking*, vol. 14, no. 9, pp. 733–748, Sep. 2022, ISSN: 1943-0639. DOI: 10.1364/JOCN.460629.