Q-Learning Based Routing in Optical Networks

Nolen B. Bryant, Kwok K. Chung, Jie Feng, Sommer Harris, Kristine N. Umeh, Michal Aibin *Member, IEEE*
Khoury College of Computer Sciences
Northeastern University, Vancouver, Canada
m.aibin@northeastern.edu

***Abstract*—The rapid increase in bandwidth demand has driven the development of flexible, efficient, and scalable optical networks. One of the technologies that allows for much more flexible resource utilization is Elastic Optical Network. However, there is a need to solve the Routing, Modulation and Spectrum Assignment (RMSA) problem. In this paper, we use reinforcement learning to improve the efficiency of the routing algorithm. More specifically, we implement an off-policy Q-learning and compare it with the state-of-the-art algorithms. The results confirm that Q-learning is highly effective when optimal results need to be found in a large search space.**

***Index Terms*—q-learning, optical networks, routing**

I. Introduction

By 2023, 66% of the global population will be Internet users, an increase of 15% from 2018 [1]. The currently projected growth in traffic demands would cause significant bandwidth bottlenecks within conventional optical network technology [2]. This rapidly increasing demand for bandwidth is pushing the evolution of more flexible, efficient, and scalable optical networks. From the Traditional Wavelength Division Multiplexing (WDM), we now have solutions such as Elastic Optical Networks (EON), which are based on the optical orthogonal frequency-driven multiplexing (O-OFDM) scheme. WDM networks are limited by their fixed wavelength assignment, leading to an underutilized spectrum [3]. Many of the parameters that had to be constants in the WDM networks, such as modulation format and wavelength space between channels, can now be dynamically changed according to the demands of the systems. One solution to this problem and the main candidate for the future of optical transmission technology is EON [4]. The newer O-OFDM technology allows for greater bandwidth efficiency by allocating spectrum into multiple, narrow slices according to the request. However, increasing the elasticity of these networks requires more sophisticated and dynamic algorithms, which can utilize the new flexible spectrum technologies to handle high traffic without violating the constraints of the system: spectrum contiguity, spectrum continuity, and slice opacity; it means that the required spectrum slots must be adjacent, the slots must be the same in all links of the route, and until the allocation is finished, they cannot be reallocated [5].

To meet all of these criteria, researchers are implementing algorithms to solve the Routing, Modulation and Spectrum Assignment (RMSA) problem [6]. In this problem, a route is the path through which the light travels from the source to the destination. When a route is static, the route from source to destination can be set before the light has started traveling through the network, while with a dynamic route, the route may be adjusted in path, depending on the resources available. Modulation affects the way the light wave that carries data through the optical fiber is altered. The data that is coded and sent through the beam is not changed in modulation, but the beam itself is altered. We can think of this as similar to refracting light or changing the amplification of a wavelength. There are six modulation formats that are considered in the scope of Elastic Optical Networks problems: BPSK, QPSK, 8-QAM, 16-QAM, 32-QAM, and 64-QAM. Each modulation has benefits or trade-offs to consider for each signal, depending on the path length, bit-rate, and Optical Signal-to-Noise Ratio (OSNR). Spectrum assignment is how we allocate segments of the signal spectrum to carry client requests, avoiding frequency overlap. There are different types of allocation policies, or 'fits', for assigning connection requests to spectrum segments. This paper uses an implementation of a reinforcement learning algorithm to solve the RMSA problem. Our measure of efficiency when solving this problem is request Blocking Percentage (BP), where we divide the number of rejected requests by the total requests offered to the network.

Reinforcement learning is a type of machine learning that is well suited for sequential decision making due to the action value learning cycle, which incrementally defines 'good' behavior [7]. Reinforcement learning can provide strong, adaptive, and economical solutions to complex large-scale challenges. In our case, we use a very specific approach, Q-learning. Q-learning is an off-policy reinforcement learning algorithm. An off-policy algorithm approximates the optimal action value function, independent of the policy. In particular, Q-learning is about learning a strategy that maximizes overall reward. So, by repeatedly trying all actions in all states, it learns which are best overall, judged by long-term discounted reward.

This paper is divided as follows. In Section I, we introduced and motivated the problem. In the next section, we discuss related work, followed by a problem statement. We conclude our work with the simulation setup and the results.

II. Related Works

Different methods have been investigated to optimize only the spectrum assignment portion of RMSA. Some of the most common methods include Shortest Path First, which provides lower computational complexity and is usually used as a baseline solution [8]. Other articles specifically focus on modulation selection [9], [10] to further optimize routing decisions.

Recently, many researchers who propose RMSA solutions have done it with machine learning [11]–[20]. Although there are optical network papers specifically using Q-learning algorithms, all of the ones we found have either been using Deep Q-learning [21] or are applying the Q-learning algorithm to different areas of optical networks such as edge scheduling [22] and policy determination [23].

Therefore, we believe that the application of a Q-learning algorithm to the RMSA problem with the optimization goal of lowering the overall Request Blocking Percentage is novel. We implement our Q-learning algorithm using the CEONS simulator [24].

III. Problem Statement

Our goal is to find the most efficient candidate path (as shown in Fig. 1) from the source node to the destination node for each request so that the overall request Blocking Percentage (BP) is minimized. As traffic requests and content demand increase globally, networks will require more intelligent routing algorithms, and Q-learning is a potential fit for this environment.

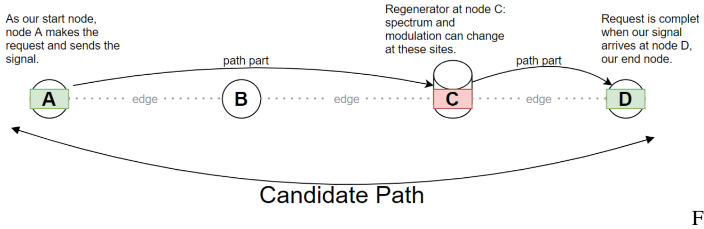

As our start node, node A makes the request and sends the signal.
Regenerator at node C: spectrum and modulation can change at these sites.
Request is complet when our signal arrives at node D, our end node.
path part
path part
A
edge
B
edge
C
edge
D
Candidate Path

Fig. 1. Path Structure and Definitions: A candidate path is shown here. Note that the path parts will be combined to complete the request from source to end nodes (green). Path parts begin at source node or at a node with a regenerator (red) and end at end node or node with a regenerator. Paths may use different modulations.

A. Network Model Notation

We will use the same network notation as in [8]. The optical network is represented as a graph $G(V, E, B, L)$ where V is the set of nodes, E is the set of fiber links (directed edges), B the maximum number of frequency slices that each fiber link can accommodate, and L are the lengths of the fiber links for each $e \in E$. There are six modulation formats that we will consider for each network: $BPSK$, $QPSK$, $8-QAM$, $16-QAM$, $32-QAM$, and $64-QAM$. The set of modulation formats is denoted by M. For each $m \in M$, we have the maximum distance supported by the modulation given as $dist(m)$.

During simulations, a set D of requests is created with each point in time indicated by $t \in T$. The bit rate of each request $c(d)$ is used to calculate the number of slices needed $n(c(d), m)$ for a modulation m. We will exclude the first 50k requests from our results calculations, as we have chosen that to be our train stage. The blocked requests included in the set Dbl will be used to calculate the final Blocking Percentage of the simulation run.

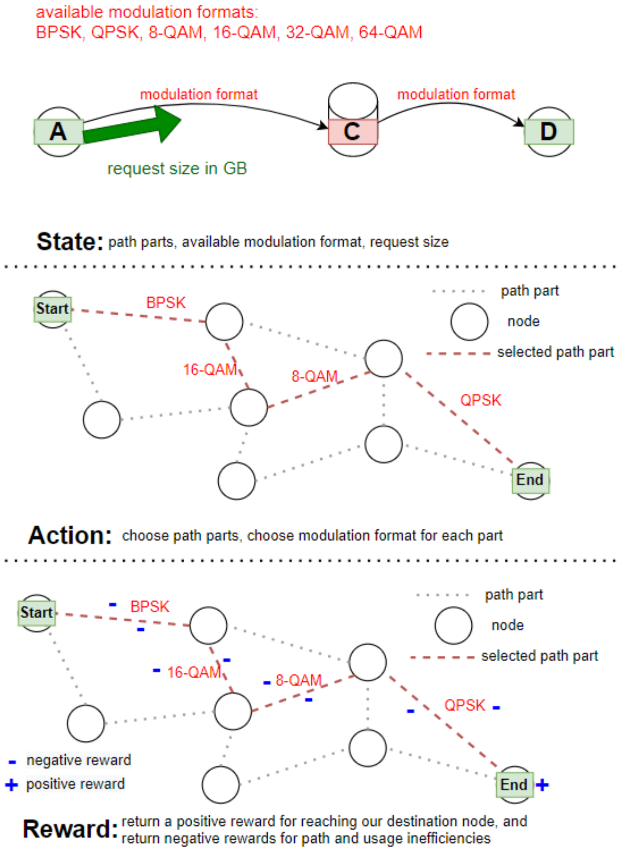

available modulation formats:
BPSK, QPSK, 8-QAM, 16-QAM, 32-QAM, 64-QAM
modulation format
modulation format
A
C
D
request size in GB
State: path parts, available modulation format, request size
Start
BPSK
16-QAM
8-QAM
QPSK
End
path part
node
selected path part
Action: choose path parts, choose modulation format for each part
Start
- BPSK
- 16-QAM
- 8-QAM
QPSK -
End +
path part
node
selected path part
- negative reward
+ positive reward
Reward: return a positive reward for reaching our destination node, and return negative rewards for path and usage inefficiencies

Fig. 2. State, Action, and Reward: Our state is defined as path parts with available modulation formats, and the request size. Our action is defined as choosing the path parts for the request, from source to terminal node, and choosing modulation formats for each of those parts. Our reward will be calculated by taking into account factors such as distance, slices occupied, regenerators used, spectrum used up, regenerators used up, request blocked and destination reached. Our algorithm will give a negative reward for each additional path part, and when modulations are chosen, there are negative rewards for modulation inefficiencies.

B. Optimization Problem

Fundamentally, we hope to optimize the traffic handling in the optical network by solving the RMSA problem to find the most efficient candidate path with the goal of reducing the total BP, as seen in the equation below. Requests can be blocked for various reasons, all leading to a lack of resources for allocation (such as not enough spectrum, or modulations that overuse regenerators in the network nodes).

$$\begin{aligned} BP = (&SpectrumBlockedVolume \\ &+ RegeneratorBlockedVolume)/TotalVolume \end{aligned}$$

Our Q-learning algorithm will train its Q-table with the first 50k requests from D by exploring and rewarding possible actions included in the set A from the states included in the set S within the requested paths (as seen in Figure 2). The

Q-table contains the "quality" score of the action a from the state s. The score $Q(s,a)$ is the maximum expected future reward that is expected to be obtained from taking that action in that state.

During training, the Q-table is updated iteratively through exploration and calculated using the Bellman equation:

$$\begin{aligned} \Delta Q(s,a) &= Q(s,a) + \alpha * TD \\ TD &= r + \gamma(maxQ(s,a') - Q(s,a)) \end{aligned} \tag{1}$$

In the above equation, α is the learning rate ($0 < \alpha \leq 1$), the factor that determines exploration versus exploitation by weighting the importance of newly acquired scores. r is the reward calculated from the reward function R. γ is the discount factor ($0 < \gamma \leq 1$), the factor that weights the importance of future rewards. T is each point in time while D is the set of requests. The right half of this equation multiplies the discount factor by the difference between the maximum possible reward in the next state and the reward in the current state.

The reward R will be calculated with the reward scoring details per factor summarized in Table I.

TABLE I
STATE AND REWARDS

Factor	**Impact**
Distance	Shorter distance → higher reward
Spectrum	Less spectrum used → higher reward
Regenerators	Less regenerators used → higher reward
Link utilization	If fully utilized → penalize

If a candidate path is allocated successfully:

$$R = 100 * (1 - MaxPathOccupiedSlices\%),$$

else:

$$R = -1800$$

then, for each part of the path:

$$\begin{aligned} RPerPart &= (R(\text{as shown above}) * PartLength)/ \\ &\quad SupportedModulationLengthForThatVolume \end{aligned}$$

After adapting the Q-learning algorithm to the RMSA problem (as shown in Figure 3), we can assess the potential of Q-learning by performing simulations and comparing the results with those of other algorithms. In our simulations, we will compare our Q-learning results with the results of a Shortest Path First (SPF) algorithm and an Adaptive Modulation and Regenerator-Aware (AMRA) dynamic routing algorithm [9].

IV. Simulation

A. Network topologies

This study uses the CEONS simulator and its topologies: US26 and Euro28. Regenerators are used to amplify signals or change the modulation format in the path. All of our simulations ran on topologies that had 250 regenerators per node. We divided the simulations by the number of candidate paths: 2, 3, 5, 10, and 30. Then within each of those categories, we divided the runs up by Erlang values (traffic intensity): 300, 400, 500, 600, 700, 800 and 900. The number of requests was set at 100,000 for all of our simulations. All settings are presented in Fig. 4.

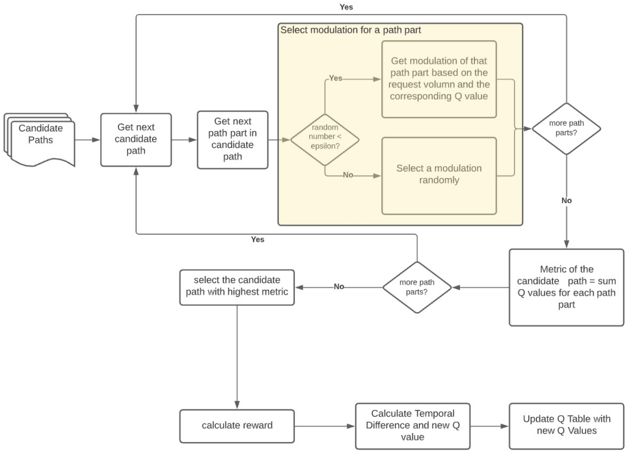
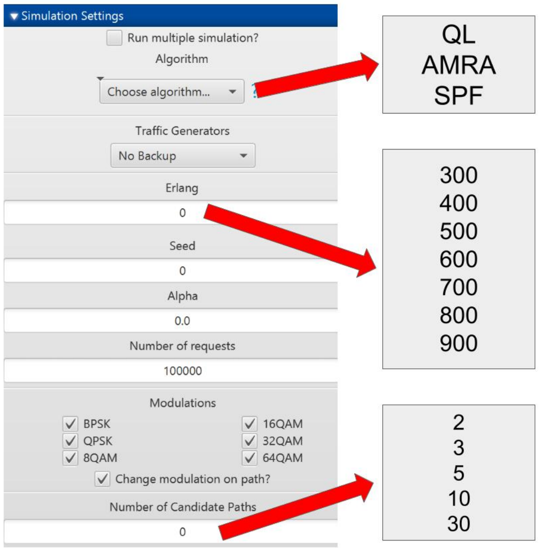

Yes
Select modulation for a path part
Candidate Paths
Get next candidate path
Get next path part in candidate path
random number < epsilon?
Yes
No
Get modulation of that path part based on the request volumn and the corresponding Q value
Select a modulation randomly
more path parts?
No
Yes
Metric of the candidate path = sum Q values for each path part
more path parts?
No
select the candidate path with highest metric
calculate reward
Calculate Temporal Difference and new Q value
Update Q Table with new Q Values

Fig. 3. Q-learning algorithm flowchart



▼ Simulation Settings

□ Run multiple simulation?

Algorithm: Choose algorithm...

Traffic Generators: No Backup

Erlang: 0

Seed: 0

Alpha: 0.0

Number of requests: 100000

Modulations

☑ BPSK ☑ 16QAM
☑ QPSK ☑ 32QAM
☑ 8QAM ☑ 64QAM

☑ Change modulation on path?

Number of Candidate Paths: 0

QL
AMRA
SPF

300
400
500
600
700
800
900

2
3
5
10
30

Fig. 4. Simulation Settings Summary

V. Results

Our main measure of algorithm efficiency is the BP. The BP results of our Q-learning algorithm compared to the AMRA and SPF algorithms on the Euro28 network are presented in Table II, and the BP results of the US26 network are presented in Table III. Both AMRA and Q-learning significantly outperform the SPF algorithm, which verifies our earlier statement that even though SPF is lowering the computational complexity, it should only be used as a baseline solution. Looking into more detail, AMRA is much better with lower ER numbers. Convergence is much faster with lower traffic loads. On the other hand, it is clear that Q-Learning achieves better results with high traffic loads. The reason for that is that there are many more decisions to be made, and Q-Learning awards only directions that do work, excluding ones that can lead to resource overutilization. As a final note, we can see that increasing the number of candidate paths improves results for all algorithms, but it is significantly improving the efficiency of Q-Learning; a much bigger search space allows Q-Learning to train, learn, and adapt to achieve higher efficiency in making the routing decisions.

TABLE II
AVERAGE BLOCKING PERCENTAGE (BP) IN THE EURO28 NETWORK

Algorithm	# of candidate paths	300 ER	400 ER	500 ER	600 ER	700 ER	800 ER	900 ER
SPF	2	**0**	**0**	**0**	0.28	1.21	3.9	16.2
	3	**0**	**0**	**0**	0.03	0.98	2.81	16.1
	5	**0**	**0**	**0**	**0**	0.68	2.44	15.9
	10	**0**	**0**	**0**	**0**	0.66	2.43	15.5
	30	**0**	**0**	**0**	**0**	0.6	2.1	14.9
AMRA	2	**0**	**0**	**0**	**0**	**0.55**	**1.21**	**4.2**
	3	**0**	**0**	**0**	**0**	**0.41**	**1.11**	**3.82**
	5	**0**	**0**	**0**	**0**	**0.28**	**1.01**	**3.33**
	10	**0**	**0**	**0**	**0**	0.25	0.98	3.19
	30	**0**	**0**	**0**	**0**	0.19	0.88	2.99
Q-Learning	2	**0**	**0**	**0**	**0**	0.9	1.88	7.4
	3	**0**	**0**	**0**	**0**	0.8	1.5	5.2
	5	**0**	**0**	**0**	**0**	0.34	1.22	3.50
	10	**0**	**0**	**0**	**0**	**0.2**	**0.77**	**2.11**
	30	**0**	**0**	**0**	**0**	**0.11**	**0.67**	**1.99**

TABLE III
AVERAGE BLOCKING PERCENTAGE (BP) IN THE US26 NETWORK

Algorithm	# of candidate paths	300 ER	400 ER	500 ER	600 ER	700 ER	800 ER	900 ER
SPF	2	**0**	**0**	**0**	0.41	1.61	4.22	18
	3	**0**	**0**	**0**	0.23	1.28	3.15	14.5
	5	**0**	**0**	**0**	**0**	0.88	2.01	12.9
	10	**0**	**0**	**0**	**0**	0.74	2.00	12.6
	30	**0**	**0**	**0**	**0**	0.72	1.98	12.5
AMRA	2	**0**	**0**	**0**	**0**	**0.65**	**1.85**	**6.3**
	3	**0**	**0**	**0**	**0**	**0.51**	**1.77**	**4.22**
	5	**0**	**0**	**0**	**0**	**0.38**	**1.25**	**4.03**
	10	**0**	**0**	**0**	**0**	0.35	0.99	3.55
	30	**0**	**0**	**0**	**0**	0.32	0.95	3.02
Q-Learning	2	**0**	**0**	**0**	0.12	1.11	2.01	8.88
	3	**0**	**0**	**0**	**0**	0.99	1.83	6.2
	5	**0**	**0**	**0**	**0**	0.54	1.79	6.0
	10	**0**	**0**	**0**	**0**	**0.25**	**0.57**	**1.88**
	30	**0**	**0**	**0**	**0**	**0.19**	**0.52**	**1.5**

VI. CONCLUSION

We demonstrated the capabilities of the Q-learning algorithm in solving the RMSA problem within Elastic Optical Networks. The best scenarios to apply the Q-learning approach were discussed with clear pros and cons. with Future work could focus on problems related to survivability and different path protection schemes while using Q-learning.

REFERENCES

[1] Cisco, "Global Cloud Index: Forecast and Methodology, 2016–2021 (white paper)," CISCO, Tech. Rep., 2018.

[2] M. Aibin and K. Walkowiak, "Resource requirements in fixed-grid and flex-grid networks for dynamic provisioning of data center traffic," in *IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, Vancouver, Canada, 2016, pp. 1–4.

[3] B. Mukherjee, "WDM optical communication networks: progress and challenges," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 10, pp. 1810–1824, 2000.

[4] M. Jinno *et al.*, "Concept and Enabling Technologies of Spectrum-Sliced Elastic Optical Path Network (SLICE)," in *Asia Communications and Photonics Conference and Exhibition*, Shanghai, China, 2009.

[5] K. Christodoulopoulos *et al.*, "Elastic Bandwidth Allocation in Flexible OFDM- based Optical Networks," *Journal of Lightwave Technology*, vol. 29, no. 9, pp. 1354 – 1366, 2011.

[6] B. Chatterjee *et al.*, "Routing and Spectrum Allocation in Elastic Optical Networks: A Tutorial," *IEEE Communications Surveys & Tutorials*, no. c, pp. 1–1, 2015.

[7] Y. Bengio, "Learning Deep Architectures for AI," *Foundations and Trends® in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.

[8] M. Aibin, "Dynamic Routing Algorithms for Cloud-Ready Elastic Optical Networks," Ph.D. dissertation, Wroclaw University of Science and Technology, 2017.

[9] M. Aibin and K. Walkowiak, "Adaptive modulation and regenerator-aware dynamic routing algorithm in elastic optical networks," in *2015 IEEE International Conference on Communications (ICC)*. London, UK: IEEE, 6 2015, pp. 5138–5143.

[10] N. Wang and J. P. Jue, "Holding-time-aware routing, modulation, and spectrum assignment for elastic optical networks," in *IEEE Global Communications Conference*, 2014, pp. 2180–2185.

[11] R. Gu *et al.*, "Machine learning for intelligent optical networks: A comprehensive survey," *Journal of Network and Computer Applications*, vol. 157, p. 102576, 5 2020.

[12] A. Chen *et al.*, "A Survey on Traffic Prediction Techniques Using Artificial Intelligence for Communication Networks," *Telecom 2021, Vol. 2, Pages 518-535*, vol. 2, no. 4, pp. 518–535, 12 2021. [Online]. Available: https://www.mdpi.com/2673-4001/2/4/29/htm https://www.mdpi.com/2673-4001/2/4/29

[13] M. Furdek *et al.*, "Machine Learning for Optical Network Security Monitoring: A Practical Perspective," *Journal of Lightwave Technology*, pp. 1–1, 2020.

[14] ——, "An overview of security challenges in communication networks," in *8th International Workshop on Resilient Networks Design and Modeling*, Halmstad, Sweden, 2016.

[15] M. Aibin and K. Walkowiak, "Simulated Annealing algorithm for optimization of elastic optical networks with unicast and anycast traffic," in *International Conference on Transparent Optical Networks*, Graz, Austria, 2014, pp. 2–5.

[16] M. Aibin, "Traffic prediction based on machine learning for elastic optical networks," *Optical Switching and Networking*, vol. 30, pp. 33–39, 11 2018.

[17] I. Khan *et al.*, "QoT Estimation for Light-path Provisioning in Un-Seen Optical Networks using Machine Learning," in *International Conference on Transparent Optical Networks*, no. Ml, 2020, pp. 1–4.

[18] D. Szostak and K. Walkowiak, "Machine Learning Methods for Traffic Prediction in Dynamic Optical Networks with Service Chains," in *International Conference on Transparent Optical Networks*, 2019, pp. 3–6.

[19] M. Aibin and K. Walkowiak, "Monte Carlo Tree Search with Last-Good-Reply Policy for Cognitive Optimization of Cloud-Ready Optical Networks," *Journal of Network and Systems Management*, 2020.

[20] M. Aibin *et al.*, "On Short-and Long-Term Traffic Prediction in Optical Networks Using Machine Learning," *25th International Conference on Optical Network Design and Modelling, ONDM 2021*, 6 2021.

[21] D. C. Abrahão and F. H. T. Vieira, "Resource allocation algorithm for LTE networks using fuzzy based adaptive priority and effective bandwidth estimation," *Wireless Networks*, vol. 24, no. 2, pp. 423–437, 2 2018. [Online]. Available: https://link.springer.com/article/10.1007/s11276-016-1344-6

[22] Q. Zhang *et al.*, "A Double Deep Q-Learning Model for Energy-Efficient Edge Scheduling," *IEEE Transactions on Services Computing*, vol. 12, no. 5, pp. 739–749, 9 2019.

[23] K. Zhan *et al.*, "Intent Defined Optical Network: Toward Artificial Intelligence-Based Optical Network Automation," *Optical Fiber Communication Conference (OFC) 2020 (2020), paper T3J.6*, vol. Part F174-OFC 2020, p. T3J.6, 3 2020. [Online]. Available: https://opg.optica.org/abstract.cfm?uri=OFC-2020-T3J.6

[24] M. Aibin and M. Blazejewski, "Complex Elastic Optical Network Simulator (CEONS)," in *17th International Conference on Transparent Optical Networks (ICTON)*, Budapest, Hungary, 2015, pp. 1–4.