International Journal of Advanced Smart Convergence Vol.13 No.2 16-24 (2024)
http://dx.doi.org/10.7236/IJASC.2024.13.2.16

IJASC 24-2-3

A reinforcement learning-based network path planning scheme for SDN in multi-access edge computing

MinJung Kim and Ducsun Lim*

Master Degree, Department of Computer Software, Hanyang University, Korea
**Post-Doc, Department of Computer Software, Hanyang University, Korea*
*rlaalswjd405@hanyang.ac.rk, *imcoms@hanyang.ac.kr*

Abstract

With an increase in the relevance of next-generation integrated networking environments, the need to effectively utilize advanced networking techniques also increases. Specifically, integrating Software-Defined Networking (SDN) with Multi-access Edge Computing (MEC) is critical for enhancing network flexibility and addressing challenges such as security vulnerabilities and complex network management. SDN enhances operational flexibility by separating the control and data planes, introducing management complexities. This paper proposes a reinforcement learning-based network path optimization strategy within SDN environments to maximize performance, minimize latency, and optimize resource usage in MEC settings. The proposed Enhanced Proximal Policy Optimization (PPO)-based scheme effectively selects optimal routing paths in dynamic conditions, reducing average delay times to about 60 ms and lowering energy consumption. As the proposed method outperforms conventional schemes, it poses significant practical applications.

Keywords: *Multi-access edge computing, software-defined networking, 6G, reinforcement learning*

1. Introduction

Modern society is in an era of accelerated data-driven digital innovation, which translates into new opportunities through technological progress across various industries. With the rapid increase in Internet speed, the importance of data generation, collection, and processing is increasing, emphasizing the need for efficient and intelligent network management system [1-2]. In this context, SDN, MEC, and 6G, the next-generation communication technologies, are providing an important turning point in networking technology by revolutionizing the design and management of network structures.

SDN increases the flexibility of network operations by separating the control and data planes, offering the advantages of increased network efficiency and reduced operational costs [3]. However, SDN is accompanied by security vulnerabilities and difficulties in complex network configurations and management. MEC processes data at the network edge by deploying computing resources, thereby shortening response times and

Manuscript Received: April. 5, 2024 / Revised: April. 11, 2024 / Accepted: April. 16, 2024
Corresponding Author: imcoms@hanyang.ac.kr
Tel: +82-2-2220-1088
Post-Doc, Department of Computer Software, Hanyang University, Korea

reducing data traffic loads, which significantly benefits latency-sensitive mobile applications and IoT devices [4-5].

6G will elevate these technologies one step further, enabling hyperconnectivity and high-speed communication, and will provide an essential foundation for future technologies, such as extensive IoT networks, ultra-realistic media, and autonomous vehicles. This enables unprecedented levels of connectivity and speed in communication technologies and poses new technological challenges, including spectrum allocation, signal interference, and energy consumption [6].

In this study, we applied Proximal Policy Optimization (PPO) to SDN, MEC, and 6G environments to optimize network path determination, resource allocation, and service placement. In this study, we developed an intelligent network management solution that maximizes efficiency and adapts to real-time changes. It assessed the use of reinforcement learning in environments that integrate SDN, MEC, and 6G [7]. The proposed enhanced PPO scheme swiftly selects optimal routing paths under dynamic conditions, thereby enhancing efficiency. This method reduced the average delay times to approximately 60 ms and decreased energy consumption, proving that the enhanced PPO outperformed the existing scheme.

The remainder of this paper is organized as follows. Chapter 2 reviews the existing research on SDN, MEC, and 6G, and the application of Reinforcement Learning, particularly PPO. Section 3 discusses the system model and explains the integration of SDN, MEC, and 6G. Chapter 4 details the proposed PPO-based network optimization technique. Section 5 presents the simulation results for various scenarios. Finally, Section 6 summarizes the findings and outlines future research directions.

2. Related Work

In this chapter, we review the research on SDN, MEC, and 6G, and analyze the use of reinforcement learning technologies such as PPO in network systems. Additionally, we discuss the interrelations between these technologies and how reinforcement learning can optimize them collectively.

2.1 Software-Defined Networking (SDN)

Existing research on SDN focuses on network flexibility and centralized management, particularly in dynamic traffic management and automation of routing policies [8]. However, these studies do not adequately address issues of scalability and performance in complex network topologies. Reference [9] highlights the need for improved routing algorithms and resource allocation strategies. Our research aims to develop adaptive algorithms that dynamically respond to network changes, maintaining high performance and scalability.

2.2 Multi-access Edge Computing (MEC)

MEC aims to reduce response times and alleviate backhaul network loads by optimizing data processing at the network edge [10]. Research has focused on resource allocation and scheduling [11], but often overlooks integration challenges with existing infrastructures and the impact of user mobility. Our research will design robust resource management strategies that consider the dynamic nature of mobile edge environments.

2.3 6G Communication

6G technology aims to enable hyperconnectivity and high-speed communication [12]. It focuses on signal processing, frequency spectrum management, and energy-efficient networks, addressing interference and security challenges [13]. However, comprehensive solutions for 6G's diverse challenges, such as ultra-low latency and massive connectivity, are lacking. Our research will develop holistic approaches integrating advanced signal processing, efficient spectrum utilization, and robust security protocols.

2.4 Reinforcement Learning

Reinforcement learning, especially Proximal Policy Optimization (PPO), is increasingly applied to network optimization for its ability to make decisions in dynamic environments [14]. Current research lacks focus on its applicability and efficiency in real-world scenarios, failing to address scalability and adaptability in large-scale networks. Our research aims to develop scalable reinforcement learning algorithms to handle the complexities of real-world network environments.

2.5 Integration of SDN, MEC, and 6G through Reinforcement Learning

Integrating SDN, MEC, and 6G presents opportunities for further optimization. SDN can enhance MEC deployment and efficiency through better resource allocation and traffic management. MEC supports 6G’s low latency and high-speed requirements. Reinforcement learning, through algorithms like PPO, offers a unifying approach to optimize these technologies collectively. It allows for dynamic adjustment to network changes, optimizing resource allocation and improving overall performance. Our research will explore adaptive, scalable solutions leveraging SDN, MEC, and 6G's strengths to meet modern network demands.

This paper aims to bridge these gaps by introducing innovative solutions enhancing network scalability, integration, and performance through reinforcement learning technologies like PPO.

3. System Model

This model serves as a foundation for applying the PPO to address challenges related to workload, latency, energy consumption, and route optimization. It is specifically designed to support resource allocation and route optimization decisions by dynamically reflecting network changes in real-time. Figure 1 illustrates the structure of the system model.

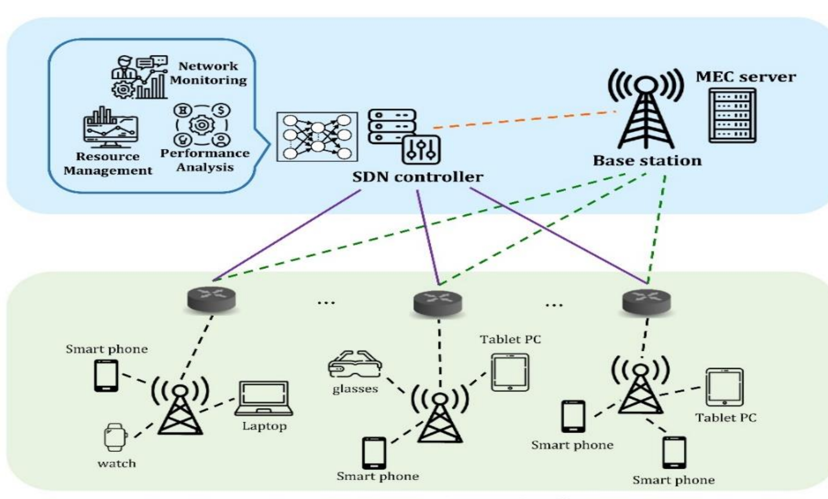

Figure 1. Architecture of the system model.

The SDN controller centrally manages the entire network and controls the switches and routers using protocols such as OpenFlow. It ensures efficient traffic management and optimal routing by dynamically adjusting to network conditions. The MEC servers, typically located in cell towers or ISP data centers, help alleviate traffic loads and minimize service latency by deploying computing resources at the network edge for local data processing. These MEC servers communicate directly with local devices, such as smartphones and IoT devices, to process data closer to the source, thereby reducing latency and improving response times [15].

The relationship between the SDN controller, MEC servers, and local devices is crucial for the system's overall efficiency. The SDN controller oversees the entire network, making high-level routing and resource allocation decisions. MEC servers handle localized data processing, reducing the burden on the central network and enhancing performance for end-users. Local devices interact with MEC servers for quick data processing and rely on the SDN controller for broader network connectivity and resource management. By integrating these components, the system ensures real-time adaptability and optimized resource usage, leveraging PPO to continually refine decision-making processes based on current network conditions.

3.1 Local Computing

In local computing, tasks are processed directly within a smart device, and the associated delay time and energy consumption are calculated as follows (1) and (2).

$$L_{local} = \frac{T_i}{R_{local}} \tag{1}$$

$$E_{local} = P_{local} \times \frac{T_i}{R_{local}} \tag{2}$$

The delay time, L_{local} calculated from local processing, represents the time required for the data to be processed within the smart device. This was calculated by dividing the number of bytes of data, T_i by the device's processing speed, R_{local}. Concurrently, the total energy consumption, E_{local} during local processing was determined by multiplying the power consumed, P_{local}, by the processing time.

3.2 Remote Computing

In remote computing, data are processed on a nearby MEC server at the network edge rather than on a local device. This reduces communication delays by avoiding data transmission to the central data center. The latency and energy consumption of this process are expressed in Equations (3) and (4), respectively.

$$L_{remote} = \frac{T_i}{R_{remote}} + D_{network} \tag{3}$$

$$E_{remote} = P_{remote} \times \left(\frac{T_i}{R_{remote}}\right) \tag{4}$$

where L_{remote} represents the delay associated with remote processing, T_i denotes the number of bytes of transmitted data, and R_{remote} represents the processing speed of the MEC server. In addition, $D_{network}$ indicates the network delay. E_{remote} represents energy consumption during remote processing, whereas P_{remote} represents the average power consumption of the server. These variables are crucial for evaluating the efficiency of remote computing and play essential roles in optimizing network performance and energy management.

3.3 Optimization of path planning

In SDN, path optimization minimizes link costs by assigning weights to each link and calculating path costs based on traffic volume, as defined quantitatively in Equation (5).

$$C = \sum_{i=1}^{n} \alpha_i \cdot x_i \tag{5}$$

In this model, C represents the total path cost and each α_i signifies the weight assigned to the i-th link, representing its cost. The weight α_i can be determined based on network performance indicators such as bandwidth usage, latency, cost, or other relevant metrics, which vary depending on the performance optimization criteria set by the network designer and operator. The variable x_i denotes the amount of traffic passing through the i-th link and calculates its impact on the overall network cost.

The scope of optimization varies depending on specific network scenarios or topologies. The proposed PPO-based optimization is particularly effective in dynamic and complex network environments where traffic patterns are highly variable and real-time adaptability is crucial. Examples include large-scale data centers, ISP networks with fluctuating traffic loads, and mobile networks with high user mobility. In these scenarios, PPO can significantly improve performance and resource utilization by continually refining decision-making processes based on current network conditions.

4. Proposed Scheme

The proposed method enhances network performance by integrating SDN flexibility, MEC local computing, and 6G high-speed communication using the PPO reinforcement learning algorithm to optimize routing by considering traffic, delay, and energy consumption.

4.1 State, Action, and Reward -

The state consists of variables that quantitatively represent the overall condition of the network, defined as follows in Equation (6).

$$S = (C, L, R) \tag{6}$$

The C represents the connection status of each node and link outlining the network's structure. L indicates the traffic load or data currently processed in each link. R reflects the available resources such as the computing power or storage space of each node. The actions, detailed in Equation (7), involve changes, such as adjusting routing paths and reallocating resources.

$$A = (\Delta R, \Delta P) \tag{7}$$

The ΔR represents changes to routing paths that direct specific traffic through designated links, thereby enhancing network efficiency and minimizing latency. ΔP involves adjusting the allocation of resources to specific nodes or links, aiming to optimize the use of network resources and improve overall performance. A Reward Function was used to evaluate the network performance resulting from the selected actions. It incorporates various performance metrics, including throughput, latency, and energy usage, as defined in Equation (8).

$$R(S,A) = \omega_1 \cdot T - \omega_2 \cdot D - \omega_3 \cdot E \tag{8}$$

T represents the throughput, which is the amount of data processed per unit time. D denotes the average latency, which is the average time required for the data to travel from the source to its destination. E represents energy usage, which reflects the total amount of energy required to operate a network. The weights $\omega_1, \omega_2, \omega_3$ represent the importance of throughput, latency, and energy usage respectively.

4.2 Proposed Scheme

The proposed technique employs PPO to address the network-path optimization challenge. This method aims to enhance real-time network performance, decrease latency, and improve energy efficiency. The pseudocode illustrates the processes involved in making routing decisions and allocating resources (Table 1).

Table 1. Proposed scheme

Step	Description
1	Initialize the policy network π_θ and the value network $V_\emptyset$
2	***for*** each iteration ***do***
3	Collect set of trajectories by running policy π_θ in the environment
4	Compute rewards to go R_t
5	Compute advantage estimates A_t using the value network $V_\emptyset$
6	Update the policy by maximizing the PPO-Clip objective function
7	$L^{CLIP} = \mathbb{E}_t[\min(r_t(\theta)A_t, \widehat{clip(r_t}(\theta), 1-\epsilon, 1+\epsilon)A_t]$
8	Optionally update the value function $V_\emptyset$ by regression
9	***end for***

First, the policy network π_θ and value network $V_\emptyset$ were initialized. These were used to predict the optimal behavior and state value of the network, respectively. During each iteration, data were collected from various states of the network through the implemented policy, rewards R_t and advantage estimate. A_t were subsequently calculated based on these data. This estimate was utilized to update the policy by maximizing the PPO-Clip objective function, which adjusted the policy ratio $r_t(\theta)$ accordingly.

The PPO-based optimization framework integrates seamlessly with SDN, MEC, and 6G infrastructures by dynamically managing resources and optimizing performance based on real-time network data. In SDN, it enhances the controller's decision-making for routing and resource allocation. For MEC, it optimizes task scheduling and reduces latency. In 6G networks, it adjusts parameters for spectrum allocation and interference mitigation. This unified approach ensures cohesive and efficient network operations across all components.

5. Performance Evaluation

We analyzed the simulation results of the PPO-based scheme to evaluate its performance in terms of delay time and energy efficiency. Table 2 summarizes the main hyperparameters used in the experiments.

Table 2. Hyper parameters

Parameter	Description	Value
Learning Rate	The rate of updating the model	0.001
Discount Factor	Determines the present value of future rewards	0.99
Epochs	Total number of training epochs	300
Batch Size	Number of data points processed per batch	64
Epsilon	Clipping parameter, limits the policy update range	0.2

Through simulation, the Enhanced PPO significantly reduced latency compared to the existing technique, particularly in low-latency areas, demonstrating its ability to speed up data transmission and processing under high-traffic conditions. Figure 2 shows the average delay time by epoch, indicating an Enhanced PPO's continued reduction in delay time with each epoch, approximately 60 ms compared with 70 ms for the existing scheme.

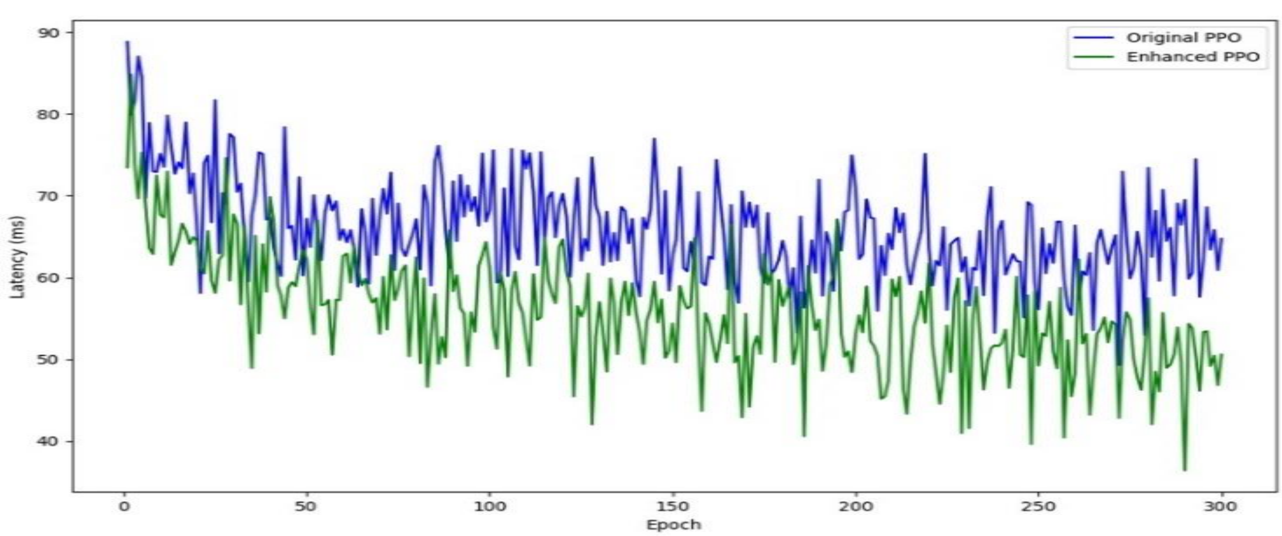

Figure 2. Average Latency over Epochs

Figure 3 shows the energy consumption per epoch. The graph shows that the Enhanced PPO was superior in terms of energy efficiency. In Epoch 300, the Enhanced PPO consumed 30 energy units, whereas the Original PPO consumed approximately 35 units. These data indicate that Enhanced PPO improves energy efficiency. These results demonstrate that the Enhanced PPO delivers superior performance in network path optimization compared with existing schemes, particularly in reducing latency and increasing energy efficiency.

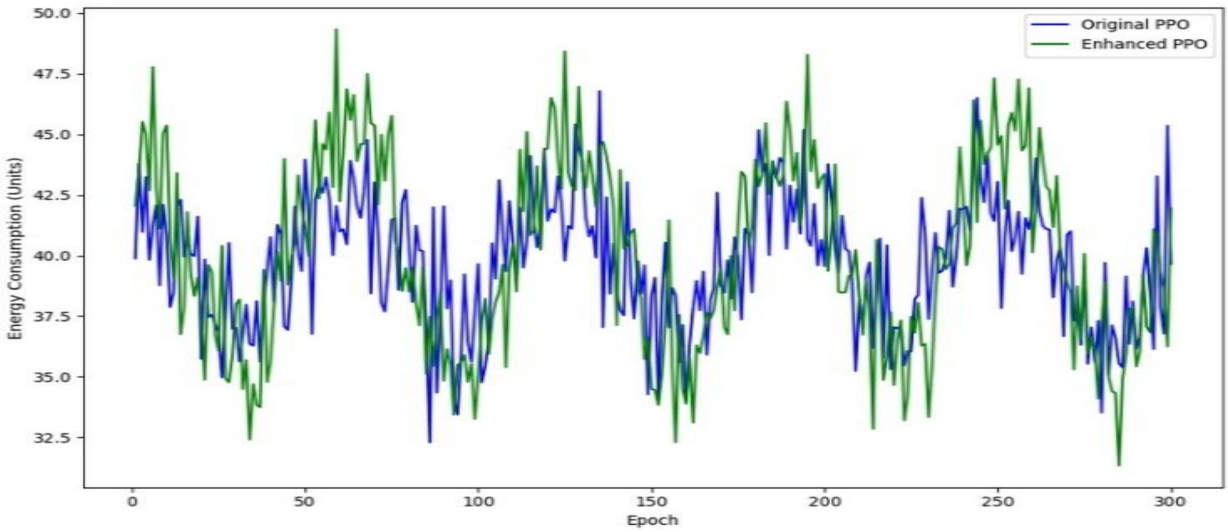

Figure 3. Average Energy Consumption over Epochs

To provide a more detailed analysis of the results, we discuss the potential reasons behind the observed improvements. The Enhanced PPO likely achieves better energy efficiency due to its ability to more accurately predict optimal paths and resource allocation, reducing unnecessary data transmissions and processing. This precise optimization leads to lower energy consumption as network devices operate more efficiently. Additionally, the reduced latency can be attributed to the Enhanced PPO's improved decision-making process, which quickly adapts to network changes, ensuring data packets are routed through the most efficient paths.

Therefore, the Enhanced PPO outperformed existing schemes in terms of performance, making it highly valuable in various fields related to network optimization. This performance improvement serves as an important benchmark for future network designs and operations, demonstrating the potential for significant advancements in energy-efficient and low-latency network solutions.

6. Conclusion

This study shows that the Enhanced PPO outperforms the existing scheme in network path optimization, significantly reducing latency and managing energy more efficiently. Its strong performance under low-latency conditions facilitates effective data processing in high-traffic environments. The enhanced PPO quickly selects optimal paths under dynamic conditions, potentially lowering long-term operational costs. Future research will assess its broader applicability by analyzing its performance across various parameters and network conditions.

Acknowledgement

This work was partially supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) (No.2020-0-01373, Artificial Intelligence Graduate School Program (Hanyang University)) and the research fund of Hanyang University (HY-2024).

References

[1] H. Kim and N. Feamster, Improving Network Management with Software Defined Networking, IEEE Communications Magazine, Vol. 51, No. 2, pp. 114-119, 2013.
DOI: https://doi.org/10.7236/JIIBC.2017.17.2.233

[2] L. Aarikka-Stenroos and P. Ritala, Network Management in the Era of Ecosystems: Systematic Review and Management Framework, Industrial Marketing Management, Vol. 67, pp. 23-36, 2017.
DOI: https://doi.org/10.1016/j.indmarman.2017.08.010

[3] M. Ndiaye, G. P. Hancke, and A. M. Abu-Mahfouz, Software Defined Networking for Improved Wireless Sensor Network Management: A Survey, Sensors, Vol. 17, No. 5, 1031, 2017.
DOI: https://doi.org/10.3390/s17051031

[4] D. Lim et al., DRL-OS: A Deep Reinforcement Learning-based Offloading Scheduler in Mobile Edge Computing, Sensors, Vol. 22, no. 23, pp. 9212, 2022. DOI: https://doi.org/10.3390/s22239212

[5] T. Taleb et al., On Multi-access Edge Computing: A Survey of the Emerging 5G Network Edge Cloud Architecture and Orchestration, IEEE Communications Surveys & Tutorials, Vol. 19, No. 3, pp. 1657-1681, 2017. DOI: https://doi.org/10.1109/COMST.2017.2705720.

[6] L. Yang, and X. Zheng. 6G: A Survey on Technologies, Scenarios, Challenges, and the Related Issues. Journal of Industrial Information Integration 19 (2020): 100158.
DOI: https://doi.org/10.1016/j.jii.2020.100158

[7] Y. Lu and X. Zheng, 6G: A Survey on Technologies, Scenarios, Challenges, and the Related Issues, Journal of Industrial Information Integration, Vol. 19, 100158, 2020.
DOI: 10.1109/JIOT.2022.315767

[8] Y. Zuo et al., Learning-based Network Path Planning for Traffic Engineering, Future Generation Computer Systems, Vol. 92, pp. 59-67, 2019.
DOI: https://doi.org/10.1016/j.future.2018.09.043

[9] O. S. Oubbati et al., SEARCH: An SDN-Enabled Approach for Vehicle Path-Planning, IEEE Transactions on Vehicular Technology, Vol. 69, No. 12, pp. 14523-14536, 2020.
DOI: 10.1109/TVT.2020.3043306

[10] D. Lim and I. Joe, A DRL-Based Task Offloading Scheme for Server Decision-Making in Multi-Access Edge Computing, Electronics, Vol. 12, no. 18, pp. 3882, 2023. DOI: https://doi.org/10.3390/electronics12183882

[11] D. Lim and D. Lim, Deep Reinforcement Learning-Based Task Offloading in Multi-access Edge Computing for Marine IoT, in Proc. Computational Methods in Systems and Software, Cham: Springer International Publishing, pp. 233-244, 2023. DOI: 10.1007/978-3-031-53549-9_23

[12] M. H. Alsharif et al., Sixth Generation (6G) Wireless Networks: Vision, Research Activities, Challenges and Potential Solutions, Symmetry, Vol. 12, No. 4, 676, 2020.
DOI: https://doi.org/10.3390/sym12040676

[13] S. Zhang and D. Zhu, Towards Artificial Intelligence Enabled 6G: State of the Art, Challenges, and Opportunities, Computer Networks, Vol. 183, 107556, 2020.
DOI: https://doi.org/10.1016/j.comnet.2020.107556

[14] G. Dulac-Arnold et al., Challenges of Real-World Reinforcement Learning: Definitions, Benchmarks and Analysis, Machine Learning, Vol. 110, No. 9, pp. 2419-2468, 2021.
DOI: https://doi.org/10.48550/arXiv.1904.12901

[15] D. Yong et al., Joint Optimization of Multi-User Partial Offloading Strategy and Resource Allocation Strategy in D2D-Enabled MEC, Sensors, Vol. 23, No. 5, 2565, 2023.
DOI: https://doi.org/10.3390/s23052565