IEEE INFOCOM 2023 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS) | 978-1-6654-9427-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/INFOCOMWKSHPS57453.2023.10225865

Towards AI-driven On-demand Routing in 6G Wide-Area Networks

Bin Dai†, Wenrui Huang†, Xinbin Shi†, Mengda Lv†, and Yijun Mo‡
†School of Electronic Information and Communications, Huazhong University of Science and Technology, China
‡ School of Computer Science and Technology, Huazhong University of Science and Technology, China
§ {daibin, wenrh, sugeladi, lvmengda, moyj}@hust.edu.cn

***Abstract*—In upcoming sixth generation (6G) networks, it is a critical challenge to support a plethora of innovative services across wide-area networks. To realize the dedicated QoS provisioning and meet the diverse quality of service (QoS) requirements of services in terms of criteria like bandwidth, delay, jitter and loss ratio, we proposed an AI-driven on-demand routing framework to support the diverse end-to-end QoS Provisioning in large-scale wide-area networks. Specifically, we make further efforts on solving the instability and non-convergence issues of the AI-driven routing algorithm and enhance it with the assistance of expert knowledge on traffic engineering and the latest advance on reinforcement learning. Furthermore, the simulation results show that our algorithm outperforms other benchmark routing algorithms with efficient learning and a significant reduction in delay, jitter and loss ratio by the traffic data sets of the real-world wide-area networks.**

***Index Terms*—Wide-Area Networks, Software Defined Networks, On-demand Routing Optimization, Deep Reinforcement Learning**

I. Introduction

Nowadays, the fifth generation (5G) network is facing an increasing number of mobile vertical applications, such as autonomous vehicles, augmented reality (AR) /virtual reality (VR), and the industrial Internet of Things (IoT). The vertical applications require different service level agreements (SLAs). To guarantee strict SLAs for vertical applications, ultra-reliable low-latency communication (URLLC) has been introduced in the 5G new radio network and the core network [1]. However, URLLC capabilities in 5G standards have been laid out for radio access network (RAN) and 5G core networks to guarantee certain SLAs of the vertical applications, like time-sensitive network (TSN) technology, without consideration for the diverse requirements of the various vertical applications carried over wide-area networks. In oncoming beyond 5G and 6G networks, it is critical to support massive access in emerging terrestrial and satellite networks and various vertical applications in wide-area networks [2]. Therefore, it is essential to design a flexible network routing framework to meet the diverse quality of service (QoS) requirements of the various vertical applications in 6G wide-area networks.

Network routing optimization has long been a hot spot in networking. However, facing the growth of complexity of network architecture and the popularity of new network applications, traditional routing algorithms may not be able to make optimal decisions on network routing without considering the actual network states. Fortunately, artificial intelligence (AI) has achieved great success in numerous fields recently. Therefore, AI-driven network routing algorithms have emerged as a new trend in networking. In particular, by taking full advantage of iterative learning ability, deep reinforcement learning (DRL) has attracted considerable attention to network routing optimization. Meantime, a highly intelligent and fully autonomous human-oriented system is the primary goal of the future 6G network [3]. Driven by the above-mentioned issues and prospects, a novel AI-driven network routing framework for 6G wide-area networks has arisen naturally.

There has been some pioneering research toward AI-driven network routing in recent years. The article [4] applies deep Q-learning networks (DQN) to verify whether machine learning is beneficial to network routing. DRL techniques have also been applied in the case of QoS-aware routing [5]. The article [6] developed a DQN-based DeepRoute algorithm with the upper-confidence bound calculation to learn the optimal path for network utility maximization. The article [7] leveraged DRL for network routing in the network with heavy traffic, aiming at reducing the probability of congestion and finding the optimal path of packet forwarding. The above-mentioned proposals originated from the Q-learning algorithm, it does not work well in the case of exploration in continuous action space. Therefore, the policy gradient [8] is proposed to solve the problem of exploration in continuous action space. The article [9] designed a deep deterministic policy gradient (DDPG) agent to adapt the network state automatically, optimize routing configuration, and minimize network delay. The article [10] proposed traffic engineering (TE) aware exploration and actor-critic-based prioritized experience replay, to optimize the DDPG agent. Simulation results showed that TE-aware DRL significantly improved network utility and reduced end-to-end latency.

Although the recent DRL-based networking works have been able to tackle the issue of continuous action space exploration, and achieved a good result in network utility maximization. They still have some shortcomings, especially in wide-area networks.

(1) They ignore the impact of the diverse QoS requirements for various vertical applications. A few of them attempted to apply the DRL agent to implement the QoS-aware network routing. In [5], A QoS-aware adaptive

Corresponding author: Yijun Mo (email: moyj@hust.edu.cn).

routing algorithm was proposed in the designed multi-layer software-defined networks (SDN). The article [11] proposed a novel network representation method based on feature engineering and verified its performance in the use case of QoS-aware routing. But none of them mentioned how to meet the diverse QoS requirements for various vertical applications based on the DRL agent in detail.

(2) The DRL-based network routing algorithms (i.e. Q-learning, DDPG) may suffer from instability and non-convergence issues [12]. With the increase of the number of nodes and links in wide-area networks, the action space of the DRL agent is exponential growth, resulting in "dimension disaster" and non-convergence problems for the DRL network [13]. That is to say, facing the high-dimensional space exploration for network routing decisions in large-scale wide-area networks, the existing DRL-based network routing algorithms are insufficient.

(3) They ignore the impact of the distinction between the real-world network environment and the evaluated network environment. Although some of the works trained and evaluated the DRL agent with the real-world network typologies. They generated the traffic matrices by the mathematical traffic models (i.e. Poisson distribution, gravity model) to fit the actual network traffic. However, the rare mathematical traffic model can well match the actual wide-area network traffic at the temporal and spatial dimensions simultaneously.

To address the above-mentioned shortcomings, in this article we propose a novel AI-driven On-demand Routing (OdR) framework for wide-area networks. To the best of our knowledge, the AI-driven network routing framework for diverse QoS provisioning has barely been treated in the literature. To adapt differential QoS requirements of various vertical applications, we span the link weights to a one-dimensional array and design the QoS-sensitive reward function, that learns the on-demand routing policies to meet the diverse QoS requirements. We make further efforts on solving the instability and non-convergence issue of DRL-based network routing with the assistance of prior knowledge from traffic engineering, taking the advantage of expert knowledge on traffic engineering to explore the action space of routing policies. In addition, we also attempt to optimize the DDPG-based on-demand routing algorithm by learning from the latest research advance on continuous action space exploration. The main contributions of this article are briefly summarized as follows.

(1) We propose a novel AI-driven on-demand routing framework with diverse end-to-end QoS provisioning for wide-area networks, in which a comprehensive reward function has been designed to meet the differential QoS demand for the various vertical applications. In consequence, the DRL agent can make the optimal network routing decision according to the QoS requirements of the vertical applications encapsulated in the packet.

(2) To solve the instability and non-convergence issue, we enhance the AI-driven on-demand routing algorithm with the assistance of expert knowledge on traffic engineering and the latest research advance on continuous action space exploration. We propose an AI-driven on-demand routing with link congestion inference (OdR-DDPG-CI) algorithm, in which an end-to-end link congestion inference algorithm is put forward to estimate the link congestion probability. The OdR-DDPG-CI algorithm uses the congestion probability of each link to guide the direction of action space exploration of link weights. In addition, we made an endeavor to complement the AI-driven on-demand routing algorithm by taking the advantage of twin delayed deep deterministic policy gradient (TD3) and soft actor-critic (SAC) from the recent advance on DRL, and propose the OdR-TD3 and OdR-SAC algorithms respectively.

(3) To the best of our knowledge, this is the first work evaluating the performance of AI-driven on-demand routing algorithms by the traffic data sets of real-world wide-area networks. Traffic matrices in past research work are generated by mathematical traffic models. In this article, we have revealed the great difference between artificial data sets and real-world data sets, especially in the temporal pattern.

The remainder of this article is organized as follows. In section 2, we propose an AI-driven on-demand routing framework to support the diverse end-to-end QoS provisioning. In section 3, we extend the AI-driven on-demand routing algorithm on behalf of expert knowledge on traffic engineering and TD3, SAC learning algorithms. To evaluate the performance of the above-mentioned algorithms, a set of experiments have been conducted in section 4. Finally, we conclude the paper and discuss open challenges of AI-driven routing optimization in wide-area networks and research trends in section 5.

II. AI-driven On-demand Routing Framework and Algorithms

A. AI-driven On-demand Routing Framework for Wide-area Networks

The AI-driven on-demand routing framework requires a flexible architecture to support the diverse end-to-end QoS demands of vertical applications. As shown in Fig. 1, if a certain flow request of a vertical application with a certain QoS demand enters the wide-area network, the aggregation router in the data plane will parse and report its demand to the control plane. Subsection II.B depicts how to represent the QoS demand of a vertical application by meta-data block technique [15]. The on-demand routing agent in the control plane makes routing decisions with the current state of the network and the demand of flow requests. As distinguished from the currently developed AI-driven routing agent, a congestion inference module is designed in this article to calculate the congestion probability of links, guiding the direction of action space exploration.

B. QoS Demands Representation of Vertical Applications

We use a programmable data plane to encapsulate the QoS requirement of the vertical applications in the packet header as its meta-data block. To simplify the representation of QoS

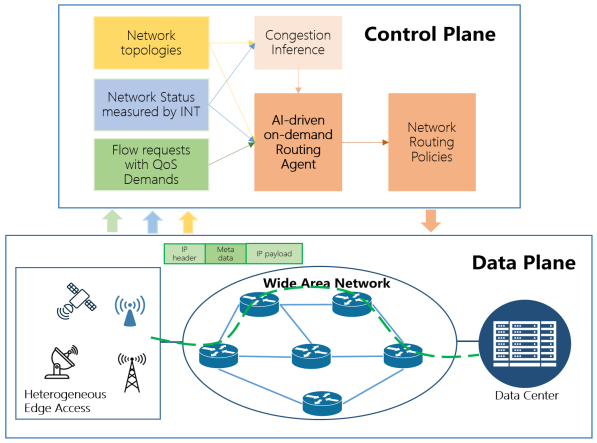


Fig. 1. The AI-driven on-demand routing framework for wide-area networks.

requirements via meta-data block, we classify the QoS demand of the vertical applications into K categories by certain criteria such as delay-sensitive and loss-sensitive flows.

C. AI-driven On-demand Routing Agent

It is assumed that the scale-free network topology is a directed graph $G = (V, E)$, where V is the node set of the network, edge set E is the link connecting the nodes, and the number of nodes and edges are $n_v = |V|$ and $n_e = |E|$, respectively. $\mathcal{T}_{i,j}^k$ is a two-dimensional array that refers to the flow request with the QoS requirement k between node i and node j. $\mathcal{W}_t^k$ is a one-dimensional array that quantifies the link weights for the k-th QoS provisioning in the t-th time interval. Specifically, $\mathcal{W}_t^k = \{W_{l_t}^k | l \in E\}$.

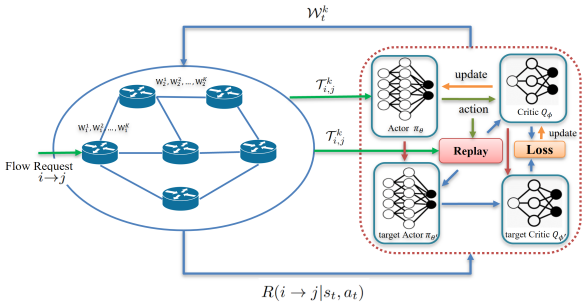


Fig. 2. The DDPG-based on-demand routing agent.

As shown in Fig. 2, a DDPG-based on-demand routing (OdR-DDPG) agent is designed to seek the optimal routing policy for flows with diverse QoS provisioning. The OdR-DDPG agent consists of five core components: the actor, the critic, the target actor, the target critic, and the replay buffer. The actor is a deterministic policy network that decides the current action based on the input state. The critic is a Q-value network that evaluates the quality of the current action. The current action is executed to obtain the next state and reward R, which are stored in the replay buffer. The state-action pair and reward stored in the replay buffer are used to train the actor and critic networks.

D. AI-driven On-demand Routing Algorithm

In this article, we present an OdR-DDPG algorithm, to learn the optimal packet routing policy that can achieve a maximum cumulative reward by exploring the state space (i.e., the network state, traffic matrices) and action space (i.e., the link weights) iteratively.

In the OdR-DDPG algorithm, the reward function is critical to meeting the differential QoS demand for various vertical applications. The reward function R in the t-th time interval is formulated as:

$$R(i \to j|s_t, a_t) = -(\theta_1^k \hat{\mathcal{D}}_{i\to j} + \theta_2^k \hat{\mathcal{J}}_{i\to j} + \theta_3^k \hat{\mathcal{L}}_{i\to j}) \quad (1)$$

where $\mathcal{D}_{i\to j}$, $\mathcal{J}_{i\to j}$ and $\mathcal{L}_{i\to j}$ are the packet delay, jitter and loss ratio from node i to node j, $\hat{\mathcal{D}}_{i\to j}$, $\hat{\mathcal{J}}_{i\to j}$ and $\hat{\mathcal{L}}_{i\to j}$ are the normalized quantities of $\mathcal{D}_{i\to j}$, $\mathcal{J}_{i\to j}$ and $\mathcal{L}_{i\to j}$, and θ_1^k, θ_2^k, $\theta_3^k \in (0, 1)$ are the tunable weights that determine the importance of QoS metrics for applications with the k-th class QoS demand.

In our past work [15], we found that the OdR-DDPG algorithm has the issue of convergence difficulties and falling into the local optima prematurely. The main reasons are summarized as follows: (1) DDPG algorithm is susceptible to hyper-parameters, and any irrational setting can lead to unstable learning. (2) DDPG algorithm overestimates the Q values of the critic network. The accumulated estimation errors will lead to the agent falling into local optima. (3) The OdR-DDPG algorithm spans the link weight to the one-dimensional array, which significantly increases the complexity of the action space exploration, especially in large-scale wide-area networks.

III. Extensions of AI-driven On-demand Routing

To address these above-mentioned challenges, we conduct some exploratory research and extend the OdR-DDPG algorithm with the assistance of expert knowledge of traffic engineering and the latest breakthrough in DRL.

A. OdR-DDPG-CI Algorithm

To improve the exploration performance of the high-dimensional action space in the OdR-DDPG algorithm, we propose a directional action space exploration method with link congestion inference [16]. According to the expert knowledge of traffic engineering, the congested links with the higher congestion probability should be selected by the candidate routing paths with a lower probability, and vice versa. Therefore, we adopt a link weight optimization method based on link congestion inference, which infers the link congestion state by the end-to-end network measurement, to guide the direction of action space exploration.

If the delay or packet loss of a path is greater than the threshold, the path is called a congested path, and at least one link in the path is assumed as a congested link. We use the method of Boolean algebra to establish the congestion relationship between the paths and links. The weight of links can be adjusted by the congestion probability of all links, which is formulated as follows:

$$d\hat{W}_{b_t}^k = p_b \rho + (1 - p_b) dW_{b_t}^k \quad (2)$$

where ρ is the weight factor, which is defined by real-time link status.

It can be seen from the above formula that the congestion probability can guide the probability weighting of the weight of the l-th link for the k-th QoS demand to generate the final optimized link weight. The link with little congestion probability has a greater link weight and is more likely to be selected.

B. OdR-TD3 Algorithm

To solve the problem of overestimating the Q value of DDPG, we extend the OdR-DDPG algorithm based on TD3, named OdR-TD3. In OdR-TD3, the following three approaches are adopted to suppress Q-value overestimation.

(1) Double Q-learning. That is, two sets of critic networks represent different Q values, and then select the smallest Q value as the target Q value to resolve the problem. Compared with Fig.1, Q_{ϕ_1}, Q_{ϕ_2} is two Critic networks, and $Q_{\phi_1'}, Q_{\phi_2'}$ is two Critic target networks. There is also an experience replay pool.

(2) Delayed update. The actor-network delays the update to make its training more stable. The update frequency of the actor-network is slower than that of the critic network. The critic network is updated every step, while the actor-network is updated at every d step ($d \geq 2$), which can make the learning of the Q-value more stable.

(3) Target strategy smoothing. The OdR-TD3 not only adds noise to the action output but also adds noise to the target action, to calculate the target action-value function. By smoothing the changes of the Q function along with different actions, it can make the Q estimated value more accurate and the algorithm more robust. The target policy smoothing can be formulated as follows:

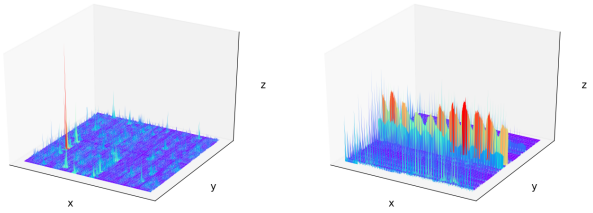
$$a' = \pi_{\theta'}(s') + \epsilon, \ \epsilon \sim clip(N(0, \widetilde{\sigma}, -c, c) \tag{3}$$

where ϵ is the random noises, following normal distribution. $clip(N(0, \widetilde{\sigma}, -c, c)$ indicates that each element of $N(0, \widetilde{\sigma})$ is in the range of $[-c, c](c > 0)$.

C. OdR-SAC Algorithm

We propose an on-demand routing algorithm based on the SAC algorithm, named OdR-SAC. SAC is a deep reinforcement learning algorithm of actor-critic networks. It is also an off-policy algorithm like DDPG, including one actor-network, two critic networks, and two critic target networks. The training phase of SAC learns a policy network and two Q networks. SAC introduced the entropy into the learning phase of the Q-value function, aiming at maximizing both Q-value and entropy. The entropy in OdR-SAC is applied to adjust the randomness of network routing policies. To prevent premature fall into the local optima, the OdR-SAC algorithm introduces the maximum entropy via the flexible Bellman formula as follows:

$$\begin{aligned} &H(\pi(\cdot|s') = -E_a log \pi(a'|s') \\ &\pi = arg \max_{\pi} E[\sum_{t} R(s_t, a_t) + aH(\pi(\cdot|s_t))] \end{aligned} \tag{4}$$



(a) traffic data sets generated by the gravity model

(b) traffic data sets from the real-world wide-area networks

Fig. 3. The comparison of the artificial and real-world traffic data sets. (The traffic volume (z) with respect to different originate-destination pair (y) at certain interval (x)).

where $\pi(a'|s')$ is the probability of selecting action a' in the state s'.

To curb overestimation, the Q value takes the minimum of the two critic target networks. The loss function value of the critic network is calculated with the difference between the Q value of the critic target networks. The loss calculation and update of the actor-network follow the KL divergence principle.

IV. Performance Evaluation

A. Traffic Data Sets from the Real-world Wide-area Networks

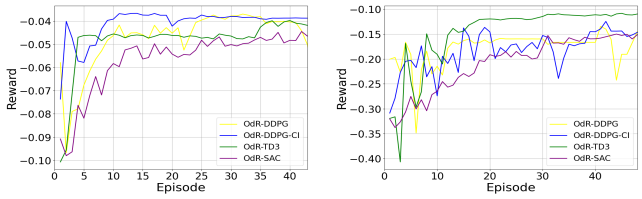
In contrast to past research work on AI-driven routing algorithms, in this article, we are the first time to use traffic data sets[1] [2] from real-world wide-area networks to train and evaluate the network model. In past work, the training traffic data sets were generated by the traffic generation model, but the traffic generation model can not well fit the feature of network traffic of the real-world wide-area networks. In detail, there are some drawbacks to the data sets from the mathematical model: (1) If the originate-destination pair is nearby, the traffic volume in traffic matrices generated by the traffic generation model will be overestimated. (2) In real-world wide-area networks, the traffic at the temporal dimension is self-correlation which has been ignored by the past traffic generation model. The extreme nature of the network environment leads to the unsatisfactory performance of algorithm models trained with data sets generated by the mathematical models. To verify the theoretical analysis, we compare the data sets generated by the gravity model and the data sets from real-world wide-area networks as shown in Fig. 3. It has been shown that the real-world data sets represent significant self-correlation. The real-world data sets originated from the GÉANT network and the Abilene network, that is the public network traffic data sets from the tracking of the Internet traffic.

B. Experiment Setting

We conduct an extensive emulation environment to evaluate the performance of the proposed AI-driven on-demand routing algorithms. We implement the proposed AI-driven on-demand routing framework based on SDN architecture and set up the emulation environment by the Mininet platform that includes

[1]http://www.cs.utexas.edu/yzhang/research/AbileneTM/,2004.

[2]https://knowledgedefinednetworking.org/data/datasets_v0/geant2.tar.gz



(a) GÉANT network (b) Abilene network

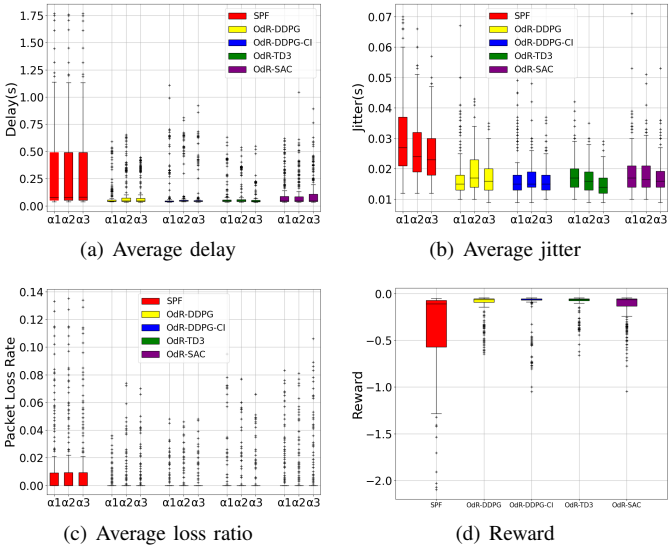
Fig. 4. The training curve under the GÉANT network and Abilene network.

a set of virtual hosts and switches. We use Ryu as the SDN controller of the control plane and DDPG model in the PyTorch library to implement an AI-driven on-demand routing agent.

C. Experiment results

For the traffic data sets of the GÉANT network, we conduct experiments to compare the performance of the proposed AI-driven on-demand routing algorithms with the widely used baseline algorithm. We used the end-to-end average packet delay, loss ratio, and jitter as the performance metrics.

As shown in Fig. 4, we can see that in two well-known network topologies, the GÉANT network and the Abilene network, all the proposed AI-driven on-demand routing algorithms can achieve a good convergence. Compared to the final reward value, the OdR-DDPG-CI algorithm performed better than the OdR-DDPG algorithm. This also proves that it is valuable to apply the expert knowledge of traffic engineering to guide the action exploration of the routing agent. The training speed of the OdR-SAC algorithm is the slowest, but the reward function value has been steadily increasing. This is because the OdR-SAC algorithm is more robust and the learning effect is more stable. However, the OdR-SAC algorithm is more complex and hard to train well.

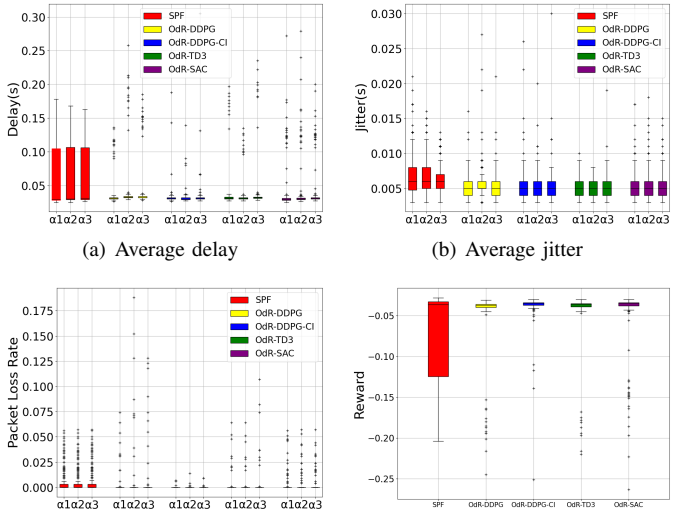


(a) Average delay (b) Average jitter

(c) Average loss ratio (d) Reward

Fig. 5. The end-to-end average delay, jitter, loss ratio, and the average training reward under the GÉANT network.

Fig. 5 shows the experimental results under the GÉANT network. From Fig. 5, we can see that all proposed AI-driven on-demand routing algorithms significantly reduce the end-to-end average delay, loss ratio, and jitter, compared to Shortest Path First (SPF) algorithm. The performance of the OdR-DDPG-CI algorithm is optimal, reducing the end-to-end delay, loss ratio, and jitter by 20.5%, 18.2%, and 10.9% at least respectively, compared to the other AI-driven on-demand routing algorithms. From Fig. 5(d), the average training reward of all algorithms is revealed. Fig. 6 shows the experimental results under the Abilene network. From Fig. 6, we can see that all proposed AI-driven on-demand routing algorithms significantly reduce the end-to-end average delay, loss ratio, and jitter, compared to the SPF algorithm. The performance of the OdR-DDPG-CI algorithm is also optimal, reducing the end-to-end delay, loss ratio, and jitter by 12.3%, 21.7%, and 9.1% at least respectively, compared to the other AI-driven on-demand routing algorithms. From Fig. 6(d), we can see that the average reward of the OdR-DDPG-CI algorithm is the stablest and highest. Table I shows the satisfaction rate of the QoS demands on packet delay, jitter, and loss ratio under the GÉANT network and Abilene network. The satisfaction rate is the ratio of the flow requests whose certain QoS metrics have been satisfied. We can also observe that all the AI-driven on-demand routing algorithms are superior to the SPF algorithm in the satisfaction rate. It can be seen that the satisfaction rate of the OdR-DDPG-LSE and OdR-TD3 algorithms is in advance of the others. From the experimental results, it can be seen that compared with the other baseline algorithms, the proposed AI-driven on-demand routing algorithms in this article can greatly reduce the average delay, jitter, and loss ratio for the diverse QoS demands.



(a) Average delay (b) Average jitter

(c) Average loss ratio (d) Reward

Fig. 6. The end-to-end average delay, jitter, loss, and the average training reward under the Abilene network.

V. Open Challenges and Research Trends

Although the research on AI-driven network routing algorithms is promising, there still are several open challenges. First, there is no specific AI-driven on-demand routing framework for the diverse QoS requirements of vertical applications

TABLE I
THE SATISFACTION RATE ON PACKET DELAY, LOSS AND JITTER UNDER THE GÉANT NETWORK AND ABILENE NETWORK

	GÉANT Network					Abilene network				
	SPF	OdR-DDPG	OdR-DDPG-CI	OdR-TD3	OdR-SAC	SPF	OdR-DDPG	OdR-DDPG-CI	OdR-TD3	OdR-SAC
delay(%)	95.351	99.819	**99.956**	99.905	99.782	94.328	98.38	98.355	**98.786**	98.326
jitter(%)	99.907	99.987	99.982	**99.99**	99.972	99.08	99.782	99.795	**99.839**	99.756
loss(%)	99.731	99.895	**99.995**	99.93	99.933	98.29	99.692	**99.714**	99.471	99.28

in wide-area networks. Second, all AI-driven routing algorithms face the issue of instability and non-convergence. Third, the data sets of network traffic generated by the traffic model cannot fit the real-world traffic well. To solve those problems, we propose an AI-driven on-demand routing framework for diverse QoS provisioning in wide-area networks. Furthermore, we make efforts to optimize the AI-driven routing algorithms from expert knowledge of traffic engineering. In addition, we use the traffic data sets from real-world wide-area networks to evaluate the performance. Extensive experimental results have shown that our proposed AI-driven on-demand routing algorithms significantly reduce the end-to-end delay, loss ratio, and jitter compared with widely-used baseline algorithms.

In the near future, there are several research directions for AI-driven on-demand routing optimization in wide-area networks.

A. Explainable Artificial Intelligence for Networking

It is an open challenge that AI-driven networking is lack of transparency. It is hard to explain the essential features that influence actions in a DRL agent. In this article, we have made some efforts to explain and optimize the action space exploration of the DRL agent with expert knowledge of traffic engineering. Future research can work on a new explainable artificial intelligence (XAI) model for networking, mapping the inputs, the parameters, and the outputs of the neural networks to the performance metrics of network routing in wide-area networks.

B. Traffic Generation Model for Wide-area Networks

The AI-driven on-demand routing algorithms require large quantities and high-quality labeled data to train the neural network model. However, the costs of massive data labeling in wide-area networks may be unaffordable. Manual data sets labeling is time-consuming and impractical. In this article, we have drawn the conclusion that the data sets of network traffic generated by the existing traffic generation model cannot fit real-world traffic well. In the future, research can be made toward this aspect, putting forward a new traffic generation model for wide-area networks with consideration of temporal and spatial correlations.

C. Inter-domain On-demand Routing Optimization

In upcoming 6G networks, the traffic of vertical applications will travel across multiple network domains, such as heterogeneous edge access networks and wide-area networks. It is a great challenge to guarantee strict end-to-end SLAs across multiple network domains. Therefore, inter-domain or cross-domain on-demand routing is an important research direction in the future.

ACKNOWLEDGEMENTS

This work is supported by the National Key Research and Development Program of China under Grant No. 2020YFB1805203.

REFERENCES

[1] G. Pocovi, H. Shariatmadari, G. Berardinelli, K. Pedersen, J. Steiner and Z. Li, "Achieving ultra-reliable low-latency communications: Challenges and envisioned system enhancements", IEEE Network, vol. 32, no. 2, pp. 8-15, Mar. 2018.

[2] C. Liu, W. Feng, Y. Chen, C. -X. Wang and N. Ge, "Cell-Free Satellite-UAV Networks for 6G Wide-Area Internet of Things," in IEEE Journal on Selected Areas in Communications, vol. 39, no. 4, pp. 1116-1131, April 2021, doi: 10.1109/JSAC.2020.3018837.

[3] W. Saad, M. Bennis and M. Chen, "A Vision of 6G Wireless Systems: Applications Trends Technologies and Open Research Problems", IEEE Network, pp. 1-9, 2019.

[4] Valadarsky A, Schapira M, Shahaf D, et al. Learning to route with deep rl[C]//NIPS Deep Reinforcement Learning Symposium. 2017.

[5] Lin S C , Akyildiz I F , Pu W , et al. QoS-Aware Adaptive Routing in Multi-layer Hierarchical Software Defined Networks: A Reinforcement Learning Approach[C]// IEEE International Conference on Services Computing. IEEE, 2016.

[6] Mohammed B, Kiran M, Krishnaswamy N. DeepRoute on Chameleon: Experimenting with Large-scale Reinforcement Learning and SDN on Chameleon Testbed[C]//2019 IEEE 27th International Conference on Network Protocols (ICNP). IEEE, 2019: 1-2.

[7] Ding R, Xu Y, Gao F, et al. Deep reinforcement learning for router selection in network with heavy traffic[J]. IEEE Access, 2019, 7: 37109-37120.

[8] Lillicrap T P , Hunt J J , Pritzel A , et al. Continuous control with deep reinforcement learning[J]. Computer Science, 2015.

[9] Stampa G , Arias M , Sanchez-Charles D , et al. A Deep-Reinforcement Learning Approach for Software-Defined Networking Routing Optimization[J]. 2017.

[10] Xu Z , Jian T , Meng J , et al. Experience-driven Networking: A Deep Reinforcement Learning based Approach. IEEE INFOCOM, 2018.

[11] J. Suarez-Varela et al., "Feature Engineering for Deep Reinforcement Learning Based Routing," ICC 2019 - 2019 IEEE International Conference on Communications (ICC), Shanghai, China, 2019, pp. 1-6, doi: 10.1109/ICC.2019.8761276

[12] Samuel P. M. Choi and Dit-Yan Yeung. 1995. Predictive Qrouting: A Memory-based Reinforcement Learning Approach to Adaptive Traffic Control. In Proceedings of the 8th International Conference on Neural Information Processing Systems (NIPS).

[13] Lei C . Curse of Dimensionality[M]. Springer US, 2009.

[14] Haipeng Yao, et al. AI Routers Network Mind: A Hybrid Machine Learning Paradigm for Packet Routing. IEEE COMPUTATIONAL INTELLIGENCE MAGAZINE, November, 2019.

[15] Y. Cao, et al. "IQoR: An Intelligent QoS-aware Routing Mechanism with Deep Reinforcement Learning." The 45th IEEE Conference on Local Computer Networks (LCN), November 16-19, 2020, Sydney, Australia.

[16] B. Dai, Y. Cao, Z. Wu and Y. Xu, "IQoR-LSE: An Intelligent QoS On-Demand Routing Algorithm With Link State Estimation," in IEEE Systems Journal, vol. 16, no. 4, pp. 5821-5830, Dec. 2022, doi: 10.1109/JSYST.2022.3149990.