2022 IEEE the 14th International Conference on Wireless Communications and Signal Processing

Service-Oriented Topology Reconfiguration of UAV Networks with Deep Reinforcement Learning

Ziyan Chen*, Nan Cheng*, Zhisheng Yin†, Jingchao He* and Ning Lu‡
*School of Telecommunications Engineering, Xidian University, Xi'an, China
†School of Cyber Engineering, Xidian University, Xi'an, China
‡Department of Electrical and Computer Engineering, Queen's University, Kingston K7L 3N6, Ontario, Canada
{zychen_3, jchhe}@stu.xidian.edu.cn, dr.nan.cheng@ieee.org, zsyin@xidian.edu.cn, ning.lu@queensu.ca

2022 14th International Conference on Wireless Communications and Signal Processing (WCSP) | 978-1-6654-5085-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/WCSP55476.2022.10039281

***Abstract*—The high mobility of UAVs makes it flexible to provide on-demand service function chains (SFCs) for users in a large geographic area where the terrestrial network is usually not available. Considering sequential and sparsely geographically distributed service requests, the UAV network topology should be dynamically adjusted to provide guaranteed quality of services. This is especially challenging since the UAV movement, data transmission, and virtual function deployment and computing are highly coupled. In this paper, we investigate the topology reconfiguring of UAV networks to construct SFCs by jointly programming multi-UAV trajectories intelligently. Specifically, the dynamic UAV-SFC construction problem is formulated to maximize the net benefit of constructing the SFC by optimizing the UAV trajectory. The net benefit is defined as the delayed benefit deducting energy consumption costs. Then, we propose a deep Q-network (DQN)-based algorithm for real-time decision-making of multi-UAV actions to program multi-UAV trajectories jointly. Simulation results show that our proposed approach of topology reconfiguration can significantly reduce the delay in completing services and save the energy consumption of UAVs.**

***Index Terms*—Topology reconfiguration, UAV networks, SFC, Deep reinforcement learning**

I. Introduction

With the development of next-G wireless communication networks, many new services are emerging, such as the Internet of Vehicles (IoV), smart cities, autonomous driving, etc., the provisioning of which has attracted great attention in academia and industry [1]. Compared with conventional services, such new services usually have significantly diverse requirements in terms of very strict requirements (such as microsecond delay) and emerging requirement categories (such as coverage, security, and intelligence). However, such multi-dimensional requirements have not been well addressed by existing best-effort service provision strategies, which motivates the exploration of on-demand service provision schemes toward 6G.

Service function chain (SFC) has been widely investigated to support virtual network functions (VNFs) to flexibly schedule multi-domain system resources and customized index requirements [2]. In [3], a reconfigurable service provision framework based on SFC is investigated, and a heuristic algorithm for the SFC programming is proposed, which balances the resource consumption of both air and ground nodes and the aggregation ratio (AR) metric is proposed to evaluate the trade-off between communication and computation. To reduce the propagation and CPU processing delay, a low-complexity heuristic algorithm based on mixed integer nonlinear programming is proposed to deploy the SFC in NFV networks [4]. Considering the heterogeneity of multi-tier nodes in space-air-ground integrated networks (SAGIN), models of the migration cost and the delay incurred by VNF hot migration and instantiating are established in [5], where the dynamic VNF mapping and scheduling are jointly realized. Appreciating previous works, the deployment of SFC relies heavily on the capacity and cost of servers, but the impact of the networking topology on constructing an SFC has not been considered in previous works.

UAV networks have been widely used for network services such as edge computing due to their low cost, rapid deployment, and flexible schedule [6], [7]. Particularly, UAVs can dynamically modify their position and transmission mode according to network requirements, resulting in the high dynamic topology structure of the UAV network [8]. Therefore, by carefully programming the UAV trajectory, the network performance and service quality can be improved [9]. Considering the limited power of the UAV, the speed and trajectory of the UAV are jointly optimized to maximize its energy efficiency while ensuring a constrained throughput [10]. Besides, a multi-agent reinforcement learning (MARL) based joint trajectory optimization of UAVs is proposed to improve the energy efficiency of data collection [11]. However, few related works have addressed constructing the SFC in UAV networks with challenging considerations, e.g., dynamic topology, limited energy, limited computation, etc.

In this paper, we investigate the SFC construction of on-demand services by reconfiguring UAV network topology, where the delay of completing service and the energy consumption of UAVs are jointly optimized. Considering the dynamic networking of multiple UAVs and the random service requests with diverse user requirements, we propose a framework to construct the on-demand SFC by optimizing the UAV network topology. To realize the service-oriented topology reconfiguration, we formulate an optimization problem with an RL method to solve it. Finally, a UAV trajectory planning approach is obtained with intelligent real-time decision-making in stepping flight action. The main contributions of this work are summarized as follows

- We propose a framework for constructing the SFC of a

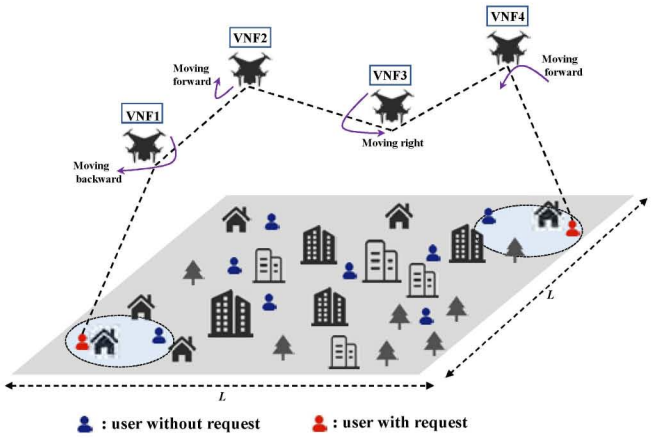


Fig. 1. Illustration of the SFC in UAV network

UAV network to provide end-to-end on-demand services, where the delay model contains the transmission and computing delay and the energy consumption model contains the flight, communication, and computing energy consumption. Particularly, we define the net benefit as the delayed benefit minus the energy cost to evaluate the topology reconfiguration for optimizing the SFC.

- Based on the proposed framework, we formulate an optimization problem to maximize the net benefit of constructing the SFC of an end-to-end communication service by programming the UAV trajectory for reconfiguring the topology of the UAV network. The total energy consumption of UAVs is constrained to ensure the completion of service. A predefined received power is also constrained, which guarantees reliable transmission.
- To solve the trajectory planning problem, the primal optimization problem is transformed into an RL problem. We propose a Deep Q-Network (DQN) based algorithm for the real-time decision-making of multi-UAV actions. Results show that our proposed algorithm can obtain lower delay and lower energy consumption than benchmarks, and the gains increases as the task data.

The remainder of this paper is organized as follows. Section II introduces the system model, including the channel model, the delay model, and the energy consumption model. In Section III, the problem is formulated, and a basic idea of using the DQN method to solve this problem is proposed. Section IV presents simulation results to evaluate the performance of the proposed scheme. We conclude this paper and guide our future work in Section V.

II. System Model

The work of this paper is to construct the on-demand SFC by optimizing the UAV network topology. This paper assumes that only one end-to-end request is generated at the moment, but the method and model of this paper can also be extended to the situation of multiple requests.

A. Network and Communication Model

We consider M users distributed in an $L \times L$ square area covered by a UAV network consisting of N UAVs shown in Fig. 1. we denote $M = \{1, 2, \cdots, m\}$ as the set of M users and $N = \{1, 2, \cdots, n\}$ as the set of N UAVs. For constructing an SFC for end-to-end networking service, N UAVs are equipped with different functions to complete a workflow in this work. Generally, discrete-time sampling is adopted in data forwarding systems. We consider a discrete-time setting $K = \{1, 2, \cdots, k\}$ with equal slot duration τ.

The ITU-R proposed the sigmoid function-based line-of-sight (LoS) probability of the air-to-ground (A2G) channel can be expressed as [12]

$$P_{ij}^{LoS}(\theta_{ij}) = \frac{1}{1 + a \cdot \exp(-[\theta_{ij} - a] \cdot b)}, \tag{1}$$

$$\theta_{ij} = \frac{180^\circ}{\pi} \cdot \sin^{-1} \frac{h_i}{d_{ij}}. \tag{2}$$

where θ_{ij} and d_{ij} are the elevation angle and distance between the j^{th} user and the i^{th} UAV, respectively, h_i is the flying height of the i^{th} UAV, and a, b are parameters representing characteristics of environment. Thus, the probability of NLoS channel is

$$P_{ij}^{NLoS}(\theta_{ij}) = 1 - P_{ij}^{LoS}(\theta_{ij}). \tag{3}$$

The LoS and NLoS path loss between UAV i and terrestrial user j are given by [12]

$$L_{ij}^{LoS} = 20 \log(\frac{4\pi f_c d_{ij}}{c}) + \xi^{LoS}, \tag{4}$$

$$L_{ij}^{NLoS} = 20 \log(\frac{4\pi f_c d_{ij}}{c}) + \xi^{NLoS}, \tag{5}$$

where f_c is the carrier frequency, c is the speed of light, and ξ^{LoS} and ξ^{NLoS} represent the additional path loss of LoS and NLoS channels in different environments [12]. By using (4) and (5), the average path loss of the A2G channel in this network can be calculated as [12]

$$L_{ij}^{avg} = P_{ij}^{LoS} \cdot L_{ij}^{LoS} + P_{ij}^{NLoS} \cdot L_{ij}^{NLoS}. \tag{6}$$

The received power of the current node can be calculated from the transmit power of the previous node, which is given by

$$P_{ij}^{r} = P_{ij}^{t} - L_{ij}^{avg}, \tag{7}$$

where P_{ij}^{t} is the transmission power of the UAV.

Besides, we consider communications between UAVs to be frequency division, i.e., Orthogonal Frequency Division Multiplexing (OFDM) [13], and the channel between UAVs is dominantly determined by the LoS. Thus the average path loss of the air-to-air (A2A) channel is assumed as $L_{ij}^{avg} = L_{ij}^{LoS}$.

B. Delay Model

The end-to-end delay of the service is determined by the number of hops transmitted and the single-hop delay, which is calculated by

$$D_{e2e} = \sum_{i=1}^{N_{hop}} D_{hop}^{i}, \tag{8}$$

where N_{hop} is the number of experienced hops for serving the request and D_{hop} is the single-hop delay, which consists of propagation delay, transmission delay, processing delay, and queuing delay, and can be expressed as

$$D_{hop} = D_{tran} + D_{prop} + D_{proc} + D_{que}. \tag{9}$$

The service defined in this model is an end-to-end service, and the UAV network only accepts a single service within a period, so the queuing delay is not considered here; In addition, the scope of services considered in this model is limited, and the propagation delay is extremely low and can be ignored here. Therefore, the single-hop delay is mainly composed of transmission and processing delays.

1) Transmission delay: The signal-to-noise ratio received from both A2G and A2A links can be respectively obtained as

$$\Gamma_{AGij} = \frac{P^r_{AGij}}{\sigma_j^2}, \Gamma_{AAij} = \frac{P^r_{AAij}}{\sigma_j^2}, \tag{10}$$

where P^r_{AGij} is the received power of node j over the A2G channel between nodes i and j, P^r_{AAij} is the received power of node j over the A2A channel between nodes i and j, σ_j^2 is the additive Gaussian white noise power at node j. Using (10), the capacity of A2G and A2A channels can be respectively obtained as

$$T_{AGij} = \frac{B}{2} \cdot \log_2(1 + \Gamma_{AGij}), \tag{11}$$

$$T_{AAij} = \frac{B}{2} \cdot \log_2(1 + \Gamma_{AAij}), \tag{12}$$

where B represents the available bandwidth allocated to each UAV.

In this paper, data is transmitted in a stream, and the actual maximum available rate of this end-to-end link is evaluated as the minimum of the set of per-hop rates. Therefore, for the t^{th} time slot, the available rate of the link is $T_{e2e} = \min\{T_{AG}, T_{AA}\}$, where T_{AG} and T_{AA} are the sets of the rate of A2G and A2A channels, respectively. The amount of data that the user requires to transmit is denoted as D, and the time required to transmit this data can be calculated as $\frac{D}{T_{e2e}}$ when the transmit power and the location of the UAV remain unchanged.

2) Processing delay: Each intermediate node in the SFC undertakes a certain processing task. We assign the power of i^{th} UAV for computing is P_{proc}. The processing delay is proportional to the amount of data D and inversely proportional to processing power P_{proc}. For easy analysis, we mainly consider the trajectory optimization of UAVs and assume the processing delay as a constant, denoted by D_{proc}.

C. Energy consumption model

We consider the energy consumption model of the UAV network from the following three components.

1) Energy consumption for data transmission: Each UAV receives data from its previous node and forwards it to the next node after processing. We assume that in the t^{th} time slot, the transmit power of the i^{th} UAV remains unchanged, denoted as P_i^t, therefore, for the UAV network, the energy consumption of data transmission can be expressed as

$$E_T = \sum_{i=1}^{N} \sum_{t=1}^{K} P_i^t \cdot \Delta T_t, \tag{13}$$

where ΔT_t is the length of each slot.

2) Energy consumption for VNF deployment and SFC calculation: This energy consumption consists of two parts:

a. Constant calculation of energy consumption of VNF installation and maintenance: the energy consumption of this part is a fixed value, which can be recorded as ϕ_{iA};

b. The energy consumption of SFC computing: In SFC, each UAV needs to process the data sent by users, and the energy consumption of processing is proportional to the amount of data that needs to be processed. Here we consider the following equation to calculate the computational energy consumption of the i^{th} UAV

$$\phi_{iB} = E_{iB} \cdot D, \tag{14}$$

where E_{iB} is the energy consumption generated by the i^{th} UAV processing the unit data correspondingly, and D is the data that the i^{th} UAV needs to process. From the above part, *a* and part *b*, the VNF deployment and SFC computing energy consumption of the i^{th} UAV can be obtained as

$$\phi_i = \phi_{iA} + \phi_{iB}. \tag{15}$$

From (15), the energy consumption of VNF deployment and SFC computing can be expressed as

$$E_C = \sum_{i=1}^{N} \phi_i. \tag{16}$$

3) Energy consumption for UAV flying: We consider a UAV flying in a plane with a fixed height h, assuming that the UAV network needs K time slots to complete the user's service, and in the t^{th} time slot, the speed of the i^{th} UAV remains unchanged, so the flight energy consumption of each UAV increases linearly with its flight distance. For the i^{th} UAV, its flight energy consumption can be expressed as

$$E_f^i = d_i^t \cdot P_i^f, \tag{17}$$

where d_i^t is the distance that the i^{th} UAV flew in the t^{th} slot. Since the time slot is small enough, it can be considered that the flight direction of the UAV in this time slot does not change, so d_i^t can be expressed as

$$d_i^t = \sqrt{\left(x_i^t - x_i^{t-1}\right)^2 + \left(y_i^t - y_i^{t-1}\right)^2}, \tag{18}$$

where (x_i^t, y_i^t) is the location of the i^{th} UAV in the t^{th} slot, after calculating the flight energy consumption of the i^{th} UAV,

the total flight energy consumption of the overall UAV network in a time slot can be calculated from this formula

$$E_F = \sum_{i=1}^{N} E_f^i. \tag{19}$$

III. Problem Formulation and Solutions

A. Problem Formulation

To construct the on-demand SFC, we reconfigure the topology of the UAV network. In this paper, we mainly consider the two indicators, including delay and energy consumption. To minimize delay and energy consumption, we define the net benefit as the optimization goal, which can be calculated by the delayed benefit minus the energy consumption cost. To maximize the net benefit, we formulate an optimization problem based on the proposed framework. The specific content is as follows.

1) Optimization goal: To minimize delay and energy consumption, the net benefit P is defined as the delay benefit R minus the energy consumption cost C, and it is necessary to maximize the net benefit P, which can be expressed as

$$P = R - C, \tag{20}$$

R can be calculated as

$$R = r_s \cdot \frac{\alpha_{\max} - \alpha_s}{\alpha_{\max}}, \tag{21}$$

where r_s denotes the benefits weight of delay violation degree, $\alpha_{\max}$ denotes the max delay violation degree that can be tolerated by a service request, α_s represents the delay violation degree of service s, and the completion time of service s is D_{e2e} given by [5]

$$\alpha_s = \frac{D_{e2e} - D_s}{D_s}, \tag{22}$$

where D_s is the maximum tolerable delay.

The total cost is calculated from three aspects, including data transmission cost, UAV flight cost, and VNF deployment and calculation cost. The specific calculation formula is as follows

$$C = b_1 \cdot E_T + b_2 \cdot E_C + b_3 \cdot \sum_{k=1}^{K} E_F, \tag{23}$$

where E_T, E_C, $\sum_{k=1}^{K} E_F$ represent data transmission, VNF deployment and SFC calculation, and UAV flight energy consumption, respectively. b_1, b_2, and b_3 are the weight coefficients, which can be adjusted for different tasks.

2) Service provision Constraints: To guarantee services are completed on time, the delay violation degree is constrained, and the delay violation degree is less than the end-to-end service delay violations, it can be expressed as

$$\alpha_s \leqslant \alpha_{\max}. \tag{24}$$

3) Capacity bound: To guarantee the correctness of communication, the amount of data transmitted between nodes is not greater than the channel capacity in a wireless channel, that is

$$T_t \leqslant T_{e2e}. \tag{25}$$

4) Node Resource Constraints: The transmit power of each node is limited, so the optimized power cannot exceed its maximum value

$$P_{ij}^t \leqslant P_i, \forall i \in N, \forall j \in N. \tag{26}$$

To guarantee that the signal can be received correctly, the received power must be greater than the threshold power

$$P_{ij}^r \geqslant P_{thr}, \forall i \in N, \forall j \in N. \tag{27}$$

The energy of UAV is limited, so the overall energy consumption is less than the maximum energy consumption

$$E_T + E_C + \sum_{k=1}^{K} E_F \leqslant E_{\max}. \tag{28}$$

5) Constraint problem: Combining the above optimization goal and constraints, it can be written as the following optimization problem

$$\begin{aligned} &\max_{\{x(i),y(i)\}_{i=1}^{N}} \quad P = R - C \\ &s.t. (24) - (28) \end{aligned} \tag{29}$$

where $(x(i), y(i))$ is the horizontal location of the i^{th} UAV.

B. DQN-based algorithm for UAV trajectory planning

In this paper, it is necessary to determine the trajectory of UAVs, emphasizing real-time action selection rather than the final optimal solution. The RL method is suitable for solving this sequence decision problem, and the movement direction of the UAV in this problem is a discrete action, so we propose a DQN-based algorithm to solve this problem [14].

The agent, state, action, and reward are as follows:

1) agent: In general reinforcement learning, the agent is the main body responsible for the exploration of the environment. In this paper, the set of swarm UAVs is considered as an agent.

2) state: In this paper, the state is the location of users and the location of UAVs and can be represented as

$$s_k(\tau) = \left(l_1^{user}(\tau), l_2^{user}(\tau), l_1^{uav}(\tau), \cdots, l_N^{uav}(\tau)\right), \forall \tau \in K \tag{30}$$

where l_i^{user} represents the location of the i^{th} user, l_j^{uav} represents the location of the j^{th} UAV, K represents the total number of time slots for task completion.

3) action: Each agent can choose one of the following actions in the current s_k: "move forward", "move backward", "move right", "move left," or "do not move" and can be expressed as

$$a_i(\tau) \in \{\pm \Delta x, \pm \Delta y\}. \tag{31}$$

4) reward: Since the constraints (28) are difficult to reflect in reinforcement learning, considering that in this paper, the transmission energy consumption of the UAV network and the calculation energy consumption of the SFC have a small variation range, and the main change is the flight energy consumption of the UAV network, so we add constraints (28)

TABLE I
SIMULATION PARAMETERS

Parameters	P^t	σ	B	h	coverage	f_c
Value	$0.15W$	$-120dBm$	$5MHz$	$100m$	$500m$	$2.4GHz$

to the optimization objective function and modify E_F in cost C of the optimization objective function as

$$E_F = \sum_{i=1}^{N} \left(d_i^t \cdot P_i^f \cdot \frac{S_{\max}}{S_{\max} - S_t} \right), \tag{32}$$

where $S_{\max}$ represents max flight distance of UAV, S_t represents UAV's flight distance in t^{th} time slot. Then calculate the optimization goal according to formulas (20), (21), and (23), and use this optimization goal as the reward of the DQN-based algorithm.

IV. SIMULATION RESULTS

In this section, we evaluate the performance to solve the optimization problem proposed in this paper. The deep neural network includes two fully-connected layers with $(128, 128)$ neurons, and ReLU activation functions are used for all layers, the learning rate is 0.01, the discount factor is 0.9, the replay memory can store 500 transition samples, and the minibatch size of training transitions is set to be 32. Probability ϵ decreases linearly from 1 to 0.005 with the rate of $\epsilon_{decay} = 10^{-5}$ [15]. The settings of other parameters are summarized in Table I.

In this paper, to compare the performance of different algorithms, we set the users positions to $[0, 0]$, $[500, 500]$ and the UAVs start positions to $[125, 125]$, $[375, 125]$, and $[250, 375]$. User services are jointly provided by these three UAVs. To evaluate the performance of the proposed DQN-based algorithm for the UAV trajectory planning problem, we use four schemes for comparison:

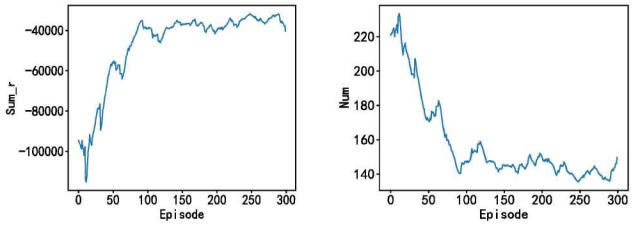
1. Fixed position: UAVs stay at the starting position and do not move, similar to the fixed base station.

2. Random action: UAVs randomly select an action from the action space.

3. Greedy algorithm: This greedy algorithm is greedy for the service completion time. By adding the service completion time after all actions are performed to an array, and selecting the action corresponding to the smallest value to execute, to guarantee that service completion time is as short as possible.

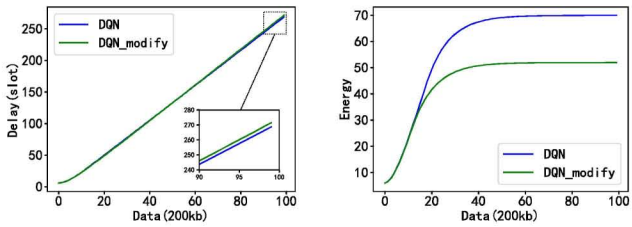
4. Heuristic algorithm: This algorithm finds the position of UAVs corresponding to the better solution of the optimization objective function, and then guides the flight of UAVs.

Fig. 2 shows the convergence of our proposed DQN-based algorithm, where Fig. 2(a) is the result of the change of the sum of r in each episode with the increase of the episode. Sum_r is the sum of the rewards calculated after each change of the state of the UAV network in an episode, that is, the sum of the rewards calculated by all the time slots in an episode. Therefore, when the convergence is poor, the number of the time slot of each episode is large, and the reward is a negative value, so Sum_r is smaller; when the convergence is better, Sum_r is larger. It can be seen from the figure that with the increase of the episode, Sum_r is an obvious increasing trend, which tends to be stable at the high point after about 80 episodes; Fig. 2(b) is the result of the change of the number of neural network learning in each episode with the increase of episodes. It can be seen from the figure that with the increase of episodes, the number of learning in each episode tends to decrease, and it also tends to be stable at the low point after about 80 episodes, which is consistent with the convergence trend in Fig. 2(a), which proves the convergence of our proposed DQN-based algorithm.



(a) sum_r of the episode (b) learning number of episode

Fig. 2. Convergence performance of the proposed DQN-based algorithm.



(a)delay vs the amount of data. (b)energy vs. the amount of data.

Fig. 3. Comparison before and after objective function modifications.

The formula (32) in this paper modifies the objective function of the original optimization problem and adds the constraint of flight energy consumption to the objective function. Fig. 3 presents a comparative analysis of this modification. Fig. 3(a) is the change curve of the service completion time with the increase in the amount of task data; Fig. 3(b) is the change curve of the flight energy consumption required to complete service with the increase in the amount of task data. The unit of energy consumption in this paper is defined as the energy consumption of a UAV flying a unit distance. Fig. 3(b) shows that the modified DQN, namely DQN_modify, its flight energy consumption of UAVs is significantly lower than the original DQN, and Fig. 3(a) shows that the service completion time of these two methods is almost the same, DQN_modify is slightly higher. This is because, in the original optimization problem, the constraint (28) is relatively loose and cannot play a good role in constraining, but adding this constraint to the optimization objective has a certain constraining effect on the flight of UAVs, and the results show that the modified DQN has a certain improvement in energy consumption. Therefore, we use the modified objective function (32). In the subsequent simulation results, DQN represents DQN_modify.

Fig. 4 shows the variation curve of the service completion time of different algorithms when the amount of task data increases. In the figure, the greedy algorithm performs the best on the indicator of service completion time. In contrast, the DQN algorithm is slightly inferior to the greedy algorithm in this indicator, obviously better than other algorithms. This is because the greedy algorithm is designed for the single indicator of service completion time, while the DQN algorithm comprehensively considers multi-dimensional indicators such as energy consumption and delay.

Fig. 5 shows the change curve of UAVs' flight energy consumption of different algorithms when the amount of task data increases. The UAVs flight energy consumption of the fixed algorithm is 0, but the service completion time is paid in the case of a large amount of task data. In addition to the fixed algorithm, the energy consumption of the DQN algorithm is the lowest, although, in the case of a small amount of task data, the energy consumption index of the DQN algorithm is slightly inferior to the heuristic algorithm. However, considering that in the future 6G scenario, the amount of task date required by users will generally not be less than $10Mb$, it can be considered that in the scenario of 6G on-demand services, the DQN algorithm is better than the heuristic algorithm. Combining Fig. 4 and Fig. 5, we can get that compared with other UAVs trajectory optimization algorithms proposed in this paper, the DQN algorithm has great advantages in terms of delay and energy consumption.

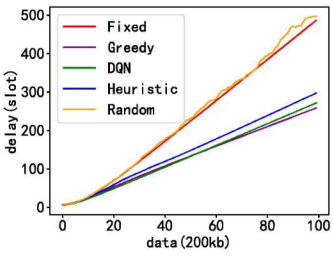


Fig. 4. Delay with the amount of task data.

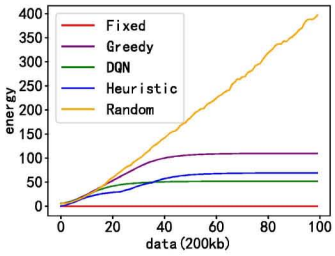


Fig. 5. Energy consumption with the amount of task data.

V. Conclusion

In this paper, we have studied how to reconfigure the UAV network to establish SFC dynamically. Firstly, we propose a framework for establishing the SFC of the UAV network to provide end-to-end on-demand services. Then, based on the proposed framework, we formulate an optimization problem to maximize the net benefit of constructing the SFC of an end-to-end communication service by programming the UAV trajectory for reconfiguring the topology of the UAV network. Finally, to solve the trajectory planning problem, we transform the primal optimization problem into an RL problem and we proposed a DQN-based algorithm for the real-time decision-making of multi-UAV actions. Results show that our proposed algorithm can obtain lower delay and energy consumption than other benchmark algorithms, and the gains increases as the task data. For future research, we will explore how to reconfigure the UAV network to serve multiple users.

Acknowledgement

This work was supported by the National Key Research and Development Program of China (2020YFB1807700), the National Natural Science Foundation of China (NSFC) under Grant No. 62071356, the Fundamental Research Funds for the Central Universities under Grant No. JB210113, the Fundamental Research Funds for the Central Universities of the Ministry of Education of China under Grant XJS221501, and the National Natural Science Foundation of Shaanxi Province under Grant 2022JQ-602.

References

[1] N. Cheng, W. Quan, W. Shi, H. Wu, Q. Ye, H. Zhou, W. Zhuang, X. Shen, and B. Bai, "A comprehensive simulation platform for space-air-ground integrated network," *IEEE Wireless Communications*, vol. 27, no. 1, pp. 178–185, 2020.

[2] N. Cheng, J. He, Z. Yin, C. Zhou, H. Wu, F. Lyu, H. Zhou, and X. Shen, "6g service-oriented space-air-ground integrated network: A survey," *Chinese Journal of Aeronautics*, vol. 35, no. 9, pp. 1–18, 2022.

[3] G. Wang, S. Zhou, S. Zhang, Z. Niu, and X. Shen, "SFC-based service provisioning for reconfigurable space-air-ground integrated networks," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 7, pp. 1478–1489, 2020.

[4] Y. Cheng, L. Yang, and H. Zhu, "Deployment of service function chain for NFV-enabled network with delay constraint," *2018 International Conference on Electronics Technology (ICET)*, pp. 383–386, IEEE, Chengdu, China, 2018.

[5] J. Li, W. Shi, H. Wu, S. Zhang, and X. Shen, "Cost-aware dynamic SFC mapping and scheduling in SDN/NFV-enabled space–air–ground-integrated networks for internet of vehicles," *IEEE Internet of Things Journal*, vol. 9, no. 8, pp. 5824–5838, 2022.

[6] Z. Yin, M. Jia, N. Cheng, W. Wang, F. Lyu, Q. Guo, and X. Shen, "UAV-Assisted physical layer security in multi-beam satellite-enabled vehicle communications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 3, pp. 2739–2751, 2022.

[7] N. Cheng, F. Lyu, W. Quan, C. Zhou, H. He, W. Shi, and X. Shen, "Space/aerial-assisted computing offloading for iot applications: A learning-based approach," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 5, pp. 1117–1129, 2019.

[8] L. Ruan, J. Wang, J. Chen, Y. Xu, Y. Yang, H. Jiang, Y. Zhang, and Y. Xu, "Energy-efficient multi-UAV coverage deployment in UAV networks: A game-theoretic framework," *China Communications*, vol. 15, no. 10, pp. 194–209, 2018.

[9] C. Zhou, H. He, P. Yang, F. Lyu, W. Wu, N. Cheng, and X. Shen, "Deep RL-based trajectory planning for AoI minimization in UAV-assisted IoT," *2019 11th International Conference on Wireless Communications and Signal Processing (WCSP)*, pp. 1–6, Xi'an, China, 2019.

[10] Y. Zeng and R. Zhang, "Energy-efficient UAV communication with trajectory optimization," *IEEE Transactions on Wireless Communications*, vol. 16, no. 6, pp. 3747–3760, 2017.

[11] H. Bayerlein, M. Theile, M. Caccamo, and D. Gesbert, "Multi-UAV path planning for wireless data harvesting with deep reinforcement learning," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 1171–1187, 2021.

[12] A. Al-Hourani, S. Kandeepan, and S. Lardner, "Optimal lap altitude for maximum coverage," *IEEE Wireless Communications Letters*, vol. 3, no. 6, pp. 569–572, 2014.

[13] X. Tan, S. Su, X. Guo, and X. Sun, "Application of MIMO-OFDM technology in UAV communication network," *2020 2nd World Symposium on Artificial Intelligence (WSAI)*, pp. 1–4, Guangzhou, China, 2020.

[14] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, et al., "Human-level control through deep reinforcement learning," *nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[15] Z. Huang and X. Xu, "DQN-Based relay deployment and trajectory planning in consensus-based Multi-UAVs tracking network," *2021 IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 1–7, Montreal, QC, Canada, 2021.