Caching Placement Optimization in UAV-Assisted Cellular Networks: A Deep Reinforcement Learning-Based Framework

Yun Wang, Shu Fu, Changhua Yao, Haijun Zhang, *Fellow, IEEE*, and Fei Richard Yu, *Fellow, IEEE*

***Abstract*—Capable of delivering contents offloaded from the base station (BS) to users, unmanned aerial vehicle (UAV) has emerged as a crucial leverage to compensate for terrestrial BSs-based communication. However, the limited storage capacity of the UAV brings challenges to providing low-latency services for users. In this letter, we investigate the caching placement of the UAV for enhancing the timeliness of services. To overcome the unknown content popularity, proximal policy optimization (PPO) is adopted in the proposed algorithm. To be specific, we first propose a modified timeliness model, named effective age of information (EAoI), to comprehensively evaluate the timeliness of services. Then, we employ PPO to build a deep reinforcement learning framework for finding the optimal caching strategy adaptively. Extensive simulation results are provided to verify the superiority of the proposed scheme, in comparison with the conventional schemes.**

***Index Terms*—Caching placement, timeliness, proximal policy optimization, unmanned aerial vehicle.**

I. Introduction

In THE upcoming era of beyond fifth-generation (B5G), the demand for seamless coverage of networks and low latency of service has become a critical challenge to the traditional cellular networks with fixed topology among base stations (BSs) [1]. Unmanned aerial vehicle (UAV) is a promising approach to compensate for terrestrial BSs-based communication. Owing to its flexibility and cost-efficiency, UAV can deliver contents offloaded from BSs to users in rural areas or urban areas without sufficient ground infrastructures, especially, in emergency scenarios [2]. However, due to the limited storage capacity of UAV, how to satisfy users' large amount of requests for contents is still challenging for the UAV-assisted cellular networks. Caching placement provides an efficient method for allocating the cache resources of the UAV, serving as a promising manner to meet the challenge of B5G.

A large number of works have contributed to UAV-assisted cellular networks using caching placement [3], [4], [5]. The work in [3] proposed an optimization strategy for content placement at UAV according to content popularity attributes. In [4], the energy efficiency of the UAV was maximized by jointly optimizing UAV's trajectory, task offloading, and caching strategy. The work in [5] jointly optimized the association of virtual reality users with UAV base stations (UBSs), caching policy, computing capacity allocation, and locations of UBSs to minimize the maximum latency, which required UAV to cache the most popular contents. Most of the existing works investigated caching placement for UAV-assisted cellular networks with pre-determined content popularity. However, the popularity of contents is not always available in practice. To overcome this shortage, some works applied content popularity predicting algorithms before content placement policy. These works simplified the problem to a certain extent, but degraded the optimality of caching placement due to the deviation in content popularity predicting. Reinforcement Learning (RL) has been deemed a key leverage catering for complex environments without any prior information, which is adaptive for unknown content popularity [6], [7], [8]. Proximal policy optimization (PPO), a typical policy gradient algorithm based on actor-critic framework, is one of the most popular RL algorithms due to its stability and reliability.

As the requirement for low latency becomes increasingly urgent, how to increase the timeliness by leveraging caching placement at UAV has attracted ever-increasing attention. Recently, age of information (AoI) is proposed as a timeliness index to characterize the freshness of contents. In [9], to realize a better tradeoff the AoI and energy consumption, cache placement and content updating interval were jointly optimized. The work in [10] investigated to reduce the load of backhaul links and the AoI of contents with a joint cost function. In [11], the expected peak age of information (PAoI) of a cache-enabled UAV network was optimized by jointly studying the caching, trajectory, and transmission power of the UAV. We found that AoI cannot reflect the effect of packet error on the timeliness. However, the timeliness performance considering packet errors caused by poor channel quality remains vacant in the existing works.

Motivated by the above observations, our work first establishes a novel modified timeliness model, referred to as effective age of information (EAoI), which is the timeliness index first taking into account the packet error. Then considering the distinguished performance of PPO, we investigate PPO based caching placement algorithm to optimize the timeliness in UAV-assisted cellular networks. The contributions of the work are two-fold:

Manuscript received 9 March 2023; accepted 4 May 2023. Date of publication 9 May 2023; date of current version 9 August 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 62271093, Grant 61971439, Grant U21A20448, and Grant U20A20157. The associate editor coordinating the review of this article and approving it for publication was D. B. da Costa. *(Corresponding author: Shu Fu.)*

Yun Wang and Shu Fu are with the College of Microelectronics and Communication Engineering, Chongqing University, Chongqing 400044, China (e-mail: 20173800@cqu.edu.cn; shufu@cqu.edu.cn).

Changhua Yao is with the School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 211544, China (e-mail: ych2347@163.com).

Haijun Zhang is with the Beijing Engineering and Technology Research Center for Convergence Networks and Ubiquitous Services, University of Science and Technology Beijing, Beijing 100083, China (e-mail: haijunzhang@ieee.org).

Fei Richard Yu is with the Department of Systems and Computer Engineering, Carleton University, Ottawa, ON K1S 5B6, Canada (e-mail: richardyu@cunet.carleton.ca).

Digital Object Identifier 10.1109/LWC.2023.3274535

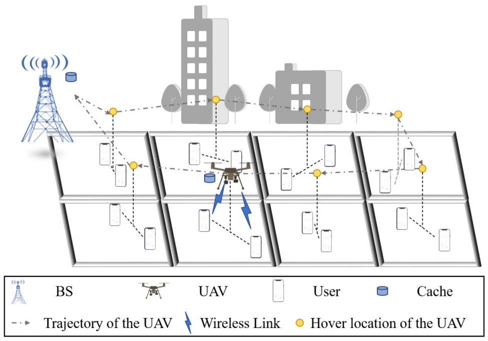

Fig. 1. System model.

- A modified timeliness model, named EAoI, is formulated considering the packet error caused by poor channel quality in this letter, which can illustrate the timeliness of the services provided by UAV more comprehensively.
- Considering the efficiency and robustness of PPO, we propose PPO based caching placement algorithm, which learns the content placement adaptively. Different from the traditional popularity-dependent caching placement algorithms, the proposed algorithm can be effectively applied to the environment with unknown content popularity.

II. System Model and Problem Formulation

A. System Overview

As illustrated in Fig. 1, we consider a system with C grids, denoted by $\mathcal{C} = \{1, \ldots, C\}$, where one periodically-moving UAV is deployed to deliver contents from the BS to terrestrial users via wireless links. The caching strategy is updated only when the UAV returns to the BS due to the large overhead of wireless backhaul link and latency between the UAV and the BS. We assume that the flying trajectory and speed of the UAV are pre-determined as in [12], where the UAV's flying time can be determined. N users are randomly scattered in each grid and served simultaneously when the grid is covered by the UAV. We adopt frequency division multiple access (FDMA) to avoid wireless interference.

Let $\mathcal{M} = \{1, \ldots, M\}$ denote the contents provided by the BS, which share the same size L. Limited by the storage capacity, the UAV can cache U kinds of contents from the BS each time, expressed as $\mathcal{U} = \{u_1, \ldots, u_U\}, (\mathcal{U} \subseteq \mathcal{M})$. Users upload their requests to the UAV when the UAV is hovering over to cover their cell. The t-th request of user n in grid c is recorded as $r_{c,n}^{m}(t) \in \{0,1\}, (c \in C, n \in \{1, \ldots, N\}, m \in \mathcal{M})$. Each user has only one request for each time, i.e., $\sum_{m=1}^{M} r_{c,n}^{m}(t) = 1, (c \in \mathcal{C}, n \in \{1, \ldots, N\})$. Set $s_{c,n}(t) = 1$ if the t-th request of user n in grid c is satisfied, otherwise $s_{c,n}(t) = 0$. For practical reasons, the users whose requests are not satisfied will generate new requests according to their preferences. Considering the influence of geographical location on user preferences, we assume the popularity of contents in different cells follows Zipf distribution with different parameters, while the users in the same cell share the same preferences. Thus, the probability that users in grid c require content m at their t-th requirement can be denoted by

$$P_{c,n}^{m}(t) = P\big(r_{c,n}^{m}(t) = 1\big) = \frac{\frac{1}{(\rho_c^m)^{\eta}}}{\sum_{i=1}^{M} \frac{1}{(\rho_c^i)^{\eta}}}, \forall n \in \{1, \ldots, N\}, \tag{1}$$

where $\rho_c^m \in \{1, \ldots, M\}$ denotes that content m is the ρ_c^m-th most popular content in grid c, and η is the parameter of the Zipf distribution to express the popularity skewness.

B. Transmission Model

We denote the location of the BS is $\mathbf{q}_{BS} = \{0, 0, 0\} \in \mathbb{R}^3$, and the location of user n in grid c as $\mathbf{q}_{c,n} = \{q_{c,n}^x, q_{c,n}^y, 0\} \in \mathbb{R}^3$, where $\{q_{c,n}^x, q_{c,n}^y\}$ represents the fixed location of user n in horizontal dimension. Similarly, the hovering location of the UAV over grid c is denoted by $\mathbf{q}_c = \{q_c^x, q_c^y, h\} \in \mathbb{R}^3$, where h is a constant to denote the altitude of the UAV, and $\{q_c^x, q_c^y\}$ is the horizontal projection of the UAV, which are calculated by $q_c^x = \frac{\sum_n q_{c,n}^x}{n}$ and $q_c^y = \frac{\sum_n q_{c,n}^y}{n}$, respectively.

We adopt the channel model proposed in [13] for the communication link between the UAV and user n in grid c, which is composed of the probability of line-of-sight (LoS) and non-line-of-sight (NLoS) links as follows:

$$P_{c,n}^{\mathrm{LoS}} = \frac{1}{1 + \alpha \exp\Big(-\beta\Big[\frac{180}{\pi} \arcsin\Big(\frac{h}{||\mathbf{q}_c - \mathbf{q}_{c,n}||}\Big)\Big]\Big)}, \tag{2}$$

$$P_{c,n}^{\mathrm{NLoS}} = 1 - P_{c,n}^{\mathrm{LoS}}, \tag{3}$$

where α and β are constants depending on the environment. Therefore, the pathloss (in dB) between the UAV and user n in grid c can be calculated by

$$\zeta_{c,n} = \mathrm{FSPL}_{c,n} + P_{c,n}^{\mathrm{LoS}} \times \eta^{\mathrm{LoS}} + P_{c,n}^{\mathrm{NLoS}} \times \eta^{\mathrm{NLoS}}, \tag{4}$$

where η^{LoS} and η^{NLoS} represent the excessive pathlosses for LoS and NLoS links, respectively. $\mathrm{FSPL}_{c,n} = 20 \log||\mathbf{q}_c - \mathbf{q}_{c,n}|| + 10 \log(\frac{4\pi f_c}{c})$ denotes free space pathloss, where f_c is the carrier frequency, and c is the speed of light. Thus, the channel gain between the UAV and user n in grid c is

$$g_{c,n} = \frac{10^{-\frac{1}{10} \times [(\eta^{\mathrm{LoS}} - \eta^{\mathrm{NLoS}}) P_{c,n}^{\mathrm{LoS}} + \eta^{\mathrm{NLoS}} + 20 \log(\frac{4\pi f_c}{c})]}}{\|\mathbf{q}_c - \mathbf{q}_{c,n}\|^2}. \tag{5}$$

During the UAV's t-th cruise, only users whose t-th requested contents are cached by the UAV will establish communication links with the UAV. We use $N_c(t)$ to indicate the number of users in grid c building communication links with the UAV during its t-th cruise, which means $N_c(t) = \sum_n s_{c,n}(t)$. The system bandwidth B and the total transmission power of the UAV P are equally allocated to the users whose t-th requests are satisfied in grid c. Denoting the white Gaussian noise power of the system as σ^2, the white Gaussian noise power for users served by the UAV in grid c can be expressed as ${\sigma_c}^2(t) = \sigma^2 / N_c(t)$. Then the signal noise ratio (SNR) of user n in grid c can be calculated as

$$\gamma_{c,n}(t) = \begin{cases} \frac{P}{N_c(t)} \times \frac{g_{c,n}}{{\sigma_c}^2(t)}, & s_{c,n}(t) = 1 \\ 0, & s_{c,n}(t) = 0 \end{cases}. \tag{6}$$

Similarly, the transmission rate between the UAV and user n in grid c is calculated by

$$R_{c,n}(t) = \begin{cases} \frac{B}{N_c(t)} \log_2\big(1 + \gamma_{c,n}(t)\big), & s_{c,n}(t) = 1 \\ 0, & s_{c,n}(t) = 0 \end{cases}. \tag{7}$$

C. Timeliness Model

As indicated above, the caching placement strategy will be updated based on the timeliness of services only when the UAV returns to the BS. Therefore, we only consider the timeliness of services when the UAV ends its cruise. At the end of the UAV's *t*-th cruise, the EAoI of service for user *n* in grid *c* is defined as the latency from the last time the user demodulated a content to the time when the UAV returned to the BS for the *t*-th time, denoted by $\Delta_{c,n}(t)$. Different from the age of information, which decreases when providing services to the user, $\Delta_{c,n}(t)$ will decrease only when the following conditions are satisfied: 1) the user *n* in grid *c* is served by the UAV; 2) the packet of the requested content is successfully demodulated.

If user *n* in grid *c* is not served, i.e., the condition 1) is not satisfied, its EAoI will increase by the period of UAV's *t*-th cruise *T*(*t*), which is composed of the flying time and hovering time of the UAV. Let T_c^f denote the time of UAV flying from the hovering location over grid *c* − 1 (or the BS) to the hovering location over grid *c*. And the time of the UAV hovering over grid *c* is denoted by

$$T_c^s(t) = \begin{cases} \max_{n\in\{1,\ldots,N\}}\left\{\frac{L}{R_{c,n}(t)}\right\}, & N_c(t) > 0 \\ 0, & N_c(t) = 0. \end{cases} \tag{8}$$

Thus, in the case where the user *n* in grid *c* is not served during the UAV's *t*-th cruise, i.e., $s_{c,n}(t) = 0$, the EAoI can be formulated as

$$\Delta_{c,n}(t)\Big|_{s_{c,n}(t)=0} = \Delta_{c,n}(t-1) + \sum_{i=0}^{C} T_i^f + \sum_{i=1}^{C} T_i^s(t), \tag{9}$$

where T_0^f is the time of the UAV flying from the hovering location over grid *C* to the BS.

If user *n* in grid *c* is served by the UAV, its EAoI is extremely affected by the packet error $P_{c,n}^{\mathrm{PER}}(t)$. If the packet of the requested content is erroneous, i.e., the condition 2) is not satisfied, the EAoI of user *n* in grid *c* will increase as in the case where the user is not served. Otherwise, as the two conditions are satisfied, the EAoI will be updated as the latency from the time when the UAV left grid *c* to the time when the UAV returned to the BS. We introduce a random value φ according to a uniform distribution within [0, 1] to denote whether the packet is erroneous. $\varphi > P_{c,n}^{\mathrm{PER}}(t)$ denotes the packet is demodulated successfully, otherwise, the packet is erroneous. Therefore, in the case where the user *n* in grid *c* is served during the UAV's *t*-th cruise, i.e., $s_{c,n}(t) = 1$, we can formulate the EAoI as

$$\begin{aligned} &\Delta_{c,n}(t)|_{s_{c,n}(t)=1} \\ &\quad = \begin{cases} \sum_{i=c+1}^{C} T_i^f + T_0^f + \sum_{i=c+1}^{C} T_i^s(t), \varphi > P_{c,n}^{\mathrm{PER}}(t) \\ \Delta_{c,n}(t)|_{s_{c,n}(t)=0}, \varphi \leq P_{c,n}^{\mathrm{PER}}(t). \end{cases} \end{aligned} \tag{10}$$

We adopt square 16-Quadrature amplitude modulation (QAM) in our system, where the bit error rate (BER) [14] of the packet delivered to user *n* in grid *c* can be calculated by

$$P_{\mathrm{c,n}}^{\mathrm{BER}}(t) = \frac{3}{4}Q\left(\sqrt{\frac{\gamma_{c,n}(t)}{5}}\right), \tag{11}$$

where $Q(x) = \int_x^\infty \frac{2}{\sqrt{2\pi}}\exp(-\frac{t^2}{2})dt$. Packaging each content as a packet with a header length of *b* bit, the packet error rate (PER) [15] of the packet delivered to user *n* in grid *c* is

$$P_c^{\mathrm{PER}}(t) = 1 - \left(1 - P_c^{\mathrm{BER}}(t)\right)^b. \tag{12}$$

Then the EAoI of user *n* in grid *c* is given as

$$\begin{aligned} \Delta_{c,n}(t) = {}& s_{c,n}(t) \times \Delta_{c,n}(t)\Big|_{s_{c,n}(t)=1} \\ &+ \left(1 - s_{c,n}(t)\right) \times \Delta_{c,n}(t)\Big|_{s_{c,n}(t)=0}. \end{aligned} \tag{13}$$

Our objective is to minimize the average EAoI of all users. Based on the formula (9) and (10), the average EAoI of all users can be expressed as

$$\Delta_{\mathrm{avg}}(t) = \frac{\sum_{i=1}^{C}\sum_{j=1}^{N}\Delta_{i,j}(t)}{N \times C}. \tag{14}$$

Therefore, the optimization problem can be formulated as

$$(\mathcal{P}1)\ \min_{\mathcal{U}}\ \Delta_{\mathrm{avg}}(t) \tag{15a}$$

$$\text{s.t.}\ \sum_{m\in\mathcal{M}} r_{c,n}^m(t) = 1, \forall c, \forall n, \tag{15b}$$

$$s_{c,n}(t) \in \{0,1\}. \tag{15c}$$

Constraint (15b) ensures that each user contains request for only one content. Unfortunately, $\mathcal{P}1$ is challenging due to the lack of information (the popularity of contents in each grid and the packet error rate). Besides, caching strategy is updated at the end of the UAV's each cruise, which is hard to solve within polynomial time by traditional convex optimization methods. In the next section, a PPO based caching placement algorithm will be introduced to cope with the optimization problem $\mathcal{P}1$.

III. PPO Based Caching Placement Algorithm

Let $\mathcal{D}(t)$ denote the EAoI set of the users who don't receive their required contents during the UAV's *t*-th cruise. According to the definition of the EAoI, we can see that the EAoIs in $\mathcal{D}(t)$ sharply increase, and induce heavy effect on the average EAoI. Assuming that $\Delta_{c,n}(t)$ is in $\mathcal{D}(t)$, based on formula (9) and (10), $\Delta_{c,n}(t)$ extremely relies on $\Delta_{c,n}(t-1)$. Thus, to solve the optimization problem $\mathcal{P}1$, a long-term caching strategy should be proposed. As one of the most popular RL algorithms, PPO has been applied successfully to problems with the objective of obtaining long-term rewards. Therefore, in this section, we introduce a PPO based caching placement algorithm to solve $\mathcal{P}1$.

For our design, the UAV is regarded as an agent, which will learn through interactions with the environment. Thus, the optimization problem $\mathcal{P}1$ can be modeled as a Markov Decision Process (MDP). The agent chooses an action a_t according to its current state s_t, then obtains its next state s_{t+1} and exclusive reward r_t from the environment. The state s_t of the agent is formulated as the

EAoI set of users at the end of the UAV's t-th cruise, denoted by $s_t = \{\{\Delta_{1,1}(t), \ldots, \Delta_{1,N}(t)\}, \ldots, \{\Delta_{C,1}(t), \ldots, \Delta_{C,N}(t)\}\}$. The action of the agent a_t is the cache decisions at the end of the UAV's t-th cruise, referred to as $a_t = \mathcal{U}(t)$. According to the optimization problem $\mathcal{P}1$, the reward of the agent r_t is denoted by the average EAoI for all users, i.e., $r_t = -\Delta_{avg}(t)$.

Let $\pi_\theta(a|s)$ present the probability that the agent with current state $s_t = s$ selects action $a_t = a$, where θ is the parameter of the neural network. The target of PPO is to get an optimized strategy $\pi_\theta{}^*$ with optimal θ^*. To be specific, PPO algorithm contains three neural networks: the actor network with parameter θ, the old actor network with parameter θ_{old}, and the critic network with parameter φ.

During each episode, the agent first adopts π_θ to interact with the environment and stores four-tuple $\langle s_t, a_t, r_{t+1}, s_{t+1}\rangle$ into the experience buffer for certain times. The parameter of the old actor network θ_{old} is updated as $\theta_{\text{old}} = \theta$ after the training of each episode, which is set to limit the variation of the new policy. The parameter of the actor network θ is updated by

$$\theta = \arg\max_{\theta} E[\min(r_\theta A_t, \mathit{clip}(r_\theta, 1-\varepsilon, 1+\varepsilon)A_t)], \quad (16)$$

where $r_\theta = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)}$ is the probability ratio, $A_t = R_t + \gamma V_\varphi(s_{t+1}) - V_\varphi(s_t)$ is the advantage function, $V_\varphi(s_t) = E\{\sum_{\tau=t}^{+\infty} \gamma^{\tau-t} r_\tau | s_t\}$ is state-value function calculated by the critic network with parameter φ, and γ is the discount factor. $\mathit{clip}(\cdot)$ is the clip function to clip the probability ratio of the advantage to the interval $[1-\varepsilon, 1+\varepsilon]$, where ε is a hyperparameter. The parameter of the critic network φ is updated by

$$\varphi = \arg\min_{\varphi} E\left[\left(V_\varphi(s_t) - \sum_{\tau=0}^{\infty} \gamma^\tau r_{t+1+\tau}\right)^2\right]. \quad (17)$$

The implementation steps of PPO are presented in Algorithm 1.

IV. Simulation Results

In the simulations, the area served by the UAV is divided into $C = 8$ grids, which are squares with side lengths of 0.5 km. The number of users in each grid is set as $N = 2$. The flight speed of the UAV is 10 m/s, and the hovering height is set as $h = 0.09$km. We assume the UAV can cache $U = 4$ kinds of contents each time. The size of each content is $L = 1$Mbit, and the length of header is $b = 50$bit. We set urban environment [13] with parameters $\alpha = 9.61$, $\beta = 0.16$, $\eta^{\text{LoS}} = 1$, and $\eta^{\text{NLoS}} = 20$. The carrier frequency is set as $f = 2$GHz. The system bandwidth is $B = $2MHz. Unless otherwise specified, the transmission power of the UAV is $P = 0.08$W. And the parameters of the PPO are set as $\gamma = 0.995$, $\varepsilon = 0.45$, and $T = 512$. Adam optimizer is adopted with learning rate of 0.00009.

The performance of PPO is illustrated in Fig. 2, where Advantage Actor Critic (A2C) and Deep Q-Network (DQN) are adopted to prove the superiority of PPO. Obviously, PPO obtains the highest cumulative rewards, and DQN performs worst. Although A2C introduces critic network to calculate the advantage function, the lack of old actor network results in the strategy π_θ being sensitive to θ, which causes the problem of low efficiency and local optimum. As for DQN, without critic network, its advantage function cannot reflect the environment, which reduces the calculation accuracy. In addition, the accumulative rewards curve of DQN first decreases sharply, then grows slowly. It is because inappropriate strategy makes the EAoI in $\mathcal{D}(t)$ increase linearly. Owing to its superiority, we design the deep reinforcement learning based caching placement framework according to PPO, which is denoted by PPO-EAoI. To verify that EAoI can illustrate the timeliness more comprehensively, we take PPO-based method with AoI and hit rate as its reward to compare performances in terms of the EAoI, denoted by PPO-AoI and PPO-HR, respectively. A traditional method, Least Frequently Used method (selects the contents most frequently used), is introduced as a baseline scheme and denoted by LFU.

In Fig. 3, the EAoI performance with η is illustrated. It can be observed that compared with PPO-AoI, PPO-HR and LFU, the performance of PPO-EAoI method is the best. The reason is that PPO-HR and LFU only focus on the most popular

Algorithm 1 PPO Based Caching Placement Algorithm

```
Input: Contents proposed by the BS $\mathcal{M}$, locations of the BS
    and users $\{\mathbf{q}_{BS}, \mathbf{q}_{1,1}, \cdots, \mathbf{q}_{1,N}, \cdots, \mathbf{q}_{C,1}, \cdots, \mathbf{q}_{C,N}\}$, the
    UAV's hovering location $\{\mathbf{q}_1, \cdots, \mathbf{q}_C\}$, system bandwidth $B$,
    and transmission power of the UAV $P$.
Output: Caching strategy, $\mathcal{U}$.
Initialize the environment and the state.
Initialize the parameters $\theta$, $\varphi$, let $\theta_{old} \leftarrow \theta$.
Clear the experience buffer.
for each episode do
    Interact with the environment for $T$ time steps, store four-tuple
        $\langle s_t, a_t, r_{t+1}, s_{t+1}\rangle$ into the experience buffer.
    Calculate advantage function according to $A_t = R_t +$
        $\gamma V_\varphi(s_{t+1}) - V_\varphi(s_t)$.
    for $1 : F$ do
        Update the parameters of actor network and critic network
            according to (16) and (17), respectively.
    end for
    $\theta_{old} \leftarrow \theta$
    Clear the experience buffer.
end for
return Caching strategy, $\mathcal{U}$
```

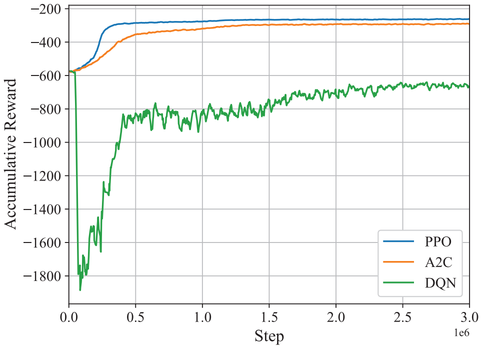

Fig. 2. The training process of RL algorithms.

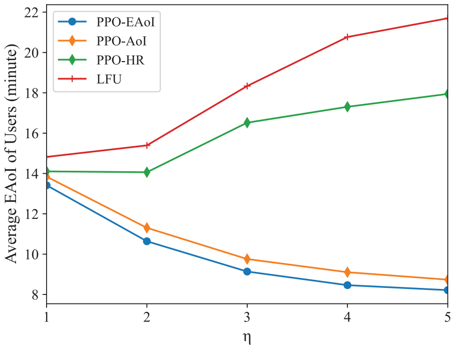

Fig. 3. Impact of η on average EAoI.

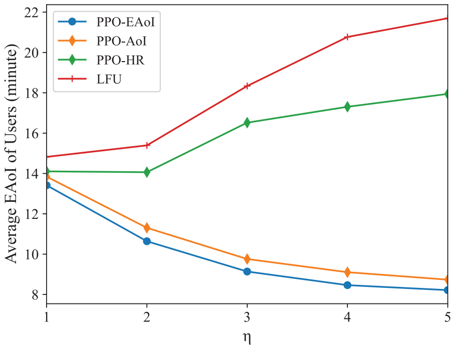

Fig. 4. Impact of the transmission power on average EAoI.

contents. These lead to the increase of the EAoIs in $\mathcal{D}(t)$ and heavily effect on the average EAoI. With the increase of η, i.e., the popularity skewness is becoming obvious, the gaps between PPO-HR, LFU and PPO-EAoI increase. Although PPO-AoI overcomes the shortcomings of PPO-HR and LFU, its performance is still not satisfactory. Owing to its neglect of packet error, PPO-AoI misestimates the timeliness of services, which further degrades the optimality of caching placement. Fig. 3 proves EAoI can indicate the timeliness more accurately and comprehensively.

In Fig. 4, we plot the EAoI curves of methods with the transmission power of the UAV P. The increase of P means the decrease of PER in (12). Obviously, the performance of PPO-AoI, PPO-HR and LFU in unknown environment with large PER are degraded, because the packet error makes the deviation in content popularity. Fig. 4 illustrates that the proposed PPO-EAoI method is more robust in complex environments.

V. Conclusion

In this letter, we investigated a UAV-assisted cellular network, where caching placement was applied to increase the timeliness of contents. We proposed EAoI as the timeliness metric. Aiming to minimize the average EAoI of the users, we used PPO to make the caching strategy. In simulations, the proposed scheme was compared with traditional schemes, and the performance of our scheme was adequately proved.

References

[1] N. Huang, T. Wang, Y. Wu, Q. Wu, and T. Q. S. Quek, “Integrated sensing and communication assisted mobile edge computing: An energy-efficient design via intelligent reflecting surface,” *IEEE Wireless Commun. Lett.*, vol. 11, no. 10, pp. 2085–2089, Oct. 2022.

[2] X. Feng, S. Fu, F. Fang, and F. R. Yu, “Optimizing age of information in RIS-assisted NOMA networks: A deep reinforcement learning approach,” *IEEE Wireless Commun. Lett.*, vol. 11, no. 10, pp. 2100–2104, Oct. 2022.

[3] E. Wang, Q. Dong, Y. Li, and Y. Zhang, “Content placement considering the temporal and spatial attributes of content popularity in cache-enabled UAV networks,” *IEEE Wireless Commun. Lett.*, vol. 11, no. 2, pp. 250–253, Feb. 2022.

[4] H. Mei, K. Yang, J. Shen, and Q. Liu, “Joint trajectory-task-cache optimization with phase-shift design of RIS-assisted UAV for MEC,” *IEEE Wireless Commun. Lett.*, vol. 10, no. 7, pp. 1586–1590, Jul. 2021.

[5] A. A. Nasir, “Latency optimization of UAV-enabled MEC system for virtual reality applications under Rician fading channels,” *IEEE Wireless Commun. Lett.*, vol. 10, no. 8, pp. 1633–1637, Aug. 2021.

[6] A. S. Kumar, L. Zhao, and X. Fernando, “Mobility aware channel allocation for 5G vehicular networks using multi-agent reinforcement learning,” in *Proc. IEEE Int. Conf. Commun.*, 2021, pp. 1–6.

[7] J. Huang, Y. Yang, L. Yin, D. He, and Q. Yan, “Deep reinforcement learning-based power allocation for rate-splitting multiple access in 6G LEO satellite communication system,” *IEEE Wireless Commun. Lett.*, vol. 11, no. 10, pp. 2185–2189, Oct. 2022.

[8] A. S. Kumar, L. Zhao, and X. Fernando, “Multi-agent deep reinforcement learning-empowered channel allocation in vehicular networks,” *IEEE Trans. Veh. Technol.*, vol. 71, no. 2, pp. 1726–1736, Feb. 2022.

[9] R. Sun, Y. Zhang, N. Cheng, R. Chai, T. Yang, and M. Qin, “AoI-oriented content caching and updating in maritime Internet of Things,” in *Proc. IEEE Global Commun. Conf.*, 2022, pp. 4794–4799.

[10] G. Ahani and D. Yuan, “Accounting for information freshness in scheduling of content caching,” in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2020, pp. 1–6.

[11] P. Yang, K. Guo, X. Xi, T. Q. S. Quek, X. Cao, and C. Liu, “Fresh, fair and energy-efficient content provision in a private and cache-enabled UAV network,” *IEEE J. Sel. Topics Signal Process.*, vol. 16, no. 1, pp. 97–112, Jan. 2022.

[12] T. Zhang, Z. Wang, Y. Liu, W. Xu, and A. Nallanathan, “Caching placement and resource allocation for cache-enabling UAV NOMA networks,” *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 12897–12911, Nov. 2020.

[13] A. Al-Hourani, S. Kandeepan, and S. Lardner, “Optimal LAP altitude for maximum coverage,” *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 569–572, Dec. 2014.

[14] C. Chen, W.-D. Zhong, and D. Wu, “Non-hermitian symmetry orthogonal frequency division multiplexing for multiple-input multiple-output visible light communications,” *J. Opt. Commun. Netw.*, vol. 9, no. 1, pp. 36–44, 2017.

[15] M. Pundir and J. K. Sandhu, “A systematic review of quality of service in wireless sensor networks using machine learning: Recent trend and future vision,” *J. Netw. Comput. Appl.*, vol. 188, Aug. 2021, Art. no. 103084.