Federated Multi-Agent Reinforcement Learning for AoI Minimization in UAV-Enabled IoV-MEC Systems

1st Shoukun Xu
Academy of Computer Science and Artificial Intelligence
Changzhou University
Changzhou, China
xsk@cczu.edu.cn

2nd Yu Zhang
Academy of Computer Science and Artificial Intelligence
Changzhou University
Changzhou, China
s23150812064@smail.cczu.edu.cn

3rd Xueyuan Wang
Academy of Computer Science and Artificial Intelligence
Changzhou University
Changzhou, China
xywang@cczu.edu.cn

4th M. Cenk Gursoy
Department of Electrical Engineering and Computer Science
Syracuse University
Syracuse, NY, 13244 USA
mcgursoy@syr.edu

***Abstract*—With the advancement of Mobile Edge Computing (MEC), effective solutions for communication scenarios, including the industrial Internet of Things (IoT) and the Internet of Vehicles (IoV), are becoming increasingly feasible. Unmanned aerial vehicles (UAVs) can further enhance flexibility in delivering computational services within MEC contexts. Addressing the urgent need for information freshness, this paper proposes a three-tier IoV-MEC system, supported by multiple UAVs and a cloud center, aiming to minimize the system's average age of information (AoI). We propose a heterogeneous multi-agent reinforcement learning algorithm based on the actor-critic framework, where vehicles act as data sources, UAVs serve as edge devices, and the cloud acts as a control center. All three tiers learn interaction strategies cooperatively based on the observations. To further enhance system performance, we implement an efficient federated learning method, allowing same-tier agents to share learning parameters, thus improving system performance and convergence speed. Extensive simulation results demonstrate that the proposed algorithm outperforms baseline algorithms in terms of average AoI and convergence speed.**

***Index Terms*—multi-agent deep reinforcement learning, age of information, Federated learning, mobile-edge computing**

I. Introduction

A. Background

The Internet of Vehicles (IoV) [1] is transforming modern intelligent transportation systems, enhancing communication between vehicles and infrastructure to support applications like autonomous driving and smart city functionalities. Enhanced vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications facilitate real-time data exchange, vital for adaptive traffic systems. Additionally, Mobile Edge Computing (MEC) [2] positions computational resources close to data sources, reducing latency and improving responsiveness, thus alleviating strain on core networks and expediting urban traffic and autonomous vehicle decision-making processes. [3], [4]

The concept of the average Age of Information (AoI) [5] ensures the freshness of information in IoV, measuring the time since the last update to prevent inefficiencies or safety risks. Moreover, deep reinforcement learning [6] optimizes efficiency in dynamic environments like IoV. Federated learning [7] enhances data privacy and security by allowing collaborative model training across nodes without data exchange. Our study integrates these methodologies to address timeliness and data privacy, improving IoV system performance and security.

B. Related Work

Recent advances in UAV-assisted MEC have significantly enhanced mobile edge computing capabilities. For instance, Ndikumana et al in [8], developed a framework that optimizes communication, computation, caching, and control within MEC to efficiently handle big data challenges. Zhou et al. [9] built on this by focusing on UAV-enabled MEC systems, particularly on offloading optimization and trajectory design to improve service delivery. Liu et al. [10] extended these concepts into a cooperative setting, where they optimized various operational parameters to enhance energy efficiency. Additionally, Xu et al. [11] addressed security concerns in UAV-assisted MEC by optimizing both resources and trajectories to safeguard against potential threats. Finally, Yang et al. [12] tackled stochastic optimization problems to adjust UAV trajectories and resource allocations dynamically, ensuring reliable service under varying conditions.

This work was supported by Changzhou Sci & Tech Program (No. CJ20220241).

The concept of AoI has gained prominence as a crucial metric for ensuring data freshness in networked systems. Chi et al. [13] introduced the Age of Model as a new metric within the Edge Intelligence-enabled IoV, focusing on keeping AI models up-to-date to handle dynamic task requirements. Bai et al. [14] applied a deep reinforcement learning approach to intelligently manage AoI in vehicular networks, optimizing both scheduling and power allocation. Furthermore, Han et al. [15] explored AoI-aware UAV deployment strategies to optimize information freshness in intelligent transportation systems. They continued their exploration in another study [16], analyzing the performance impact of AoI in UAV-aided IoT systems, leading to enhanced data collection and processing efficiency. Lastly, Qin et al. [17] focused on AoI-aware scheduling in air-ground collaborative mobile edge computing networks, effectively integrating air and ground resources to minimize AoI across user equipment.

Inspired by these works, our study integrates real-time information processing with a three-layer IoV network model, accommodating multiple UAVs and system heterogeneity. Our objective is to minimize the AoI of the system under constraints related to stochastic computation offloading, resource allocation, and UAV trajectory. This provides a robust solution for dynamic and complex environments.

C. Contributions

The main contributions of this paper are summarized as follows:

1) We design a three-tier IoV-MEC system with multi-UAV support and a cloud center, optimizing UAV trajectory, bandwidth allocation, and computational offloading to minimize AoI within a cooperative game framework.
2) We propose a Collaborative Heterogeneous Federated Actor-Critic (CHFAC) framework that merges federated learning with multi-agent reinforcement learning, optimizing performance and privacy in dynamic environments.
3) We introduced an efficient federated learning solution that prioritizes impactful agent updates using a regulation term and a non-uniform sampling strategy based on probability P, aimed at enhancing system performance and data utilization efficiency.

Simulation results indicate that our proposed algorithm significantly outperforms three benchmark algorithms in reducing the AoI, demonstrating improved system metrics and faster convergence. This validates the effectiveness of our approach, confirming its superiority in managing information freshness more efficiently.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A three-tier IoV-MEC system, enhanced by the integration of multiple UAVs and a cloud center(CC), is depicted in Fig. 1. The first layer, referred to as the vehicle layer, includes multiple vehicle devices(VDs). These devices are mobile, capable of generating data packets periodically, and equipped with computational abilities. The second layer, the UAV layer, consists of mobile edge devices, with UAVs acting as the operational units. These UAVs navigate within designated areas, facilitate communication between the VDs and the CC, and process data offloaded from the vehicle devices. The third and topmost layer is the CC, which boasts extensive computational resources. It supports the UAVs with their computational tasks, stores data, and centrally manages the entire system. In order to meticulously analyze the operational dynamics within each specific period, the continuum of time is segmented into discrete intervals, denoted as t, representing individual time slots.

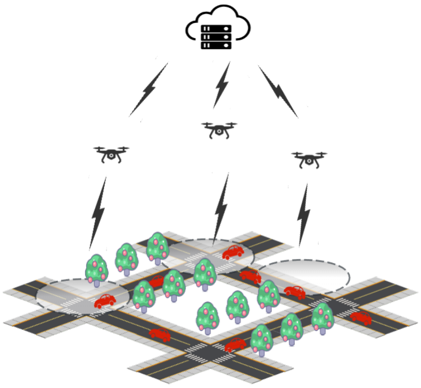


Fig. 1. An illustration of a multi-UAV and a cloud center enhanced three-tier IoV-MEC system.

A. System Model

1) Data Generation Model: In the vehicular network environment, data sources such as vehicle sensors and smart vehicular devices autonomously generate uniformly structured data packets. These packets are modeled as tuples consisting of packet size d, time elapsed w since generation, and a unique source index id. Each device is assigned a distinct index to aid in tracking and managing data packets from various sources. The packets are stored temporarily in a source buffer for subsequent processing or transmission.

2) Mobile Model: In our IoV-MEC system, we define a total of K UAVs, each indexed by k. UAVs operate at a prescribed altitude H above the data sources. Each UAV's position at time t is represented as $(X_k(t), Y_k(t), H)$. The position and movement of each UAV at the onset of the t-th time slot are modeled as follows:

$$\mathbf{p}_k(t+1) = \mathbf{p}_k(t) + \text{move}_k(t) \tag{1}$$

subject to the constraint:

$$\|\text{move}_k(t)\|_2 \leq r_{\text{move}}^k \tag{2}$$

Here, $\|\cdot\|_2$ denotes the Euclidean norm of the vector. Additionally, our model incorporates VDs, each indexed by m. VDs are equipped with mobility capabilities, enabling them to navigate through urban environments effectively. They can move forward, reverse, and make turns at intersections based on the traffic conditions and navigation requirements. Each vehicle maintains a constant speed throughout its journey, ensuring predictable and uniform motion across the network. The position of each VD at time t is represented as $(X_m(t), Y_m(t), 0)$. The dynamics of these VDs are formally described by the following movement model:

$$\mathbf{p}_m(t+1) = \mathbf{p}_m(t) + \mathrm{move}_m(t) \tag{3}$$

where $\mathbf{p}_m(t)$ represents the position of the vehicle at time t and $\mathrm{move}_m(t)$ is the vector describing the vehicle's movement during the slot.

3) Edge Processing and Offloading Model: The edge processing and offloading model can be constructed based on the size of the packets and the processing speeds of the edge devices. At each time slot t, the k-th UAV gathers data from the m-th VD, and the time required for preprocessing this data is calculated using the following formula:

$$\tau_n^k(t) = \frac{\sum_i \left[1_{\{\mathrm{id}^{\mathrm{col}}_{k,i}(t)=m\}} \times d^{\mathrm{col}}_{k,i}(t)\right]}{f_k^c(t)} \tag{4}$$

where f_k^c is the k-th UAV's data processing rate at time slot t, the indicator function $1_{\{\cdot\}}$ yields 1 if its condition is true, and 0 otherwise, $d^{\mathrm{col}}_{k,i}$ is the size of the i-th data packet in the k-th UAV's collected buffer, and $\mathrm{id}^{\mathrm{col}}_{k,i}$ is the index of the data source for that packet.

During each time slot t, the k-th UAV allocates computing resources to unprocessed data in its collected buffer, denoted by D_k^{col}. This buffer consists of data sourced externally but not yet processed at the edge. the processing decison can be expressed as:

$$\mathrm{exec}_k(t) = [\mathrm{exec}_k^1(t), \ldots, \mathrm{exec}_k^{B_k^{\mathrm{col}}}(t)] \tag{5}$$

where $\mathrm{exec}_k^i(t) \in \{0,1\}$ indicates whether the k-th UAV allocates resources to process the i-th data block in the collected buffer, and B_k^{col} represents the total number of blocks in the k-th UAV's collected buffer at time t.

Data processed by the UAVs is temporarily stored in executed buffer $D_k^{\mathrm{exe}}(t)$, before it is offloaded to the CC. At time slot t, the k-th UAV evaluates its executed buffer and decides the set of data to be transferred, which is referred to the offloading decision and can be expressed as follows:

$$\mathrm{off}_k(t) = [\mathrm{off}_k^1(t), \ldots, \mathrm{off}_k^{B_k^{\mathrm{exe}}}(t)] \tag{6}$$

where $\mathrm{off}_k^i(t) \in \{0,1\}$ indicates whether the i-th block in the executed buffer is selected for offloading, and B_k^{exe} represents the total number of blocks in the k-th UAV's executed buffer at time t.

B. Communication Model

We conceptualize VD-UAV and UAV-CC links as air-to-ground (A2G) communications. The Euclidean distances between the m-th VD and the k-th UAV, and between the k-th UAV and the CC at coordinates $(X_c, Y_c, 0)$ at time t are calculated as:

$$d_{m,k}(t) = \sqrt{(X_k(t) - X_m)^2 + (Y_k(t) - Y_m)^2 + H^2}, \tag{7}$$

$$d_{k,c}(t) = \sqrt{(X_k(t) - X_c)^2 + (Y_k(t) - Y_c)^2 + H^2} \tag{8}$$

The path loss for the A2G link considers both line-of-sight (LoS) and non-line-of-sight (NLoS) components, with the probability of LoS influenced by environmental factors and UAV altitude [18]:

$$p^L = \frac{1}{1 + a \exp\left(-b\left(\arctan\left(\frac{H}{d(t)}\right) - a\right)\right)} \tag{9}$$

where a and b are constants specific to the environment.

Average path loss $PL(t)$ is then expressed as:

$$PL(t) = \left(\frac{4\pi f d(t)}{c}\right)^2 \left(\eta^L p^L + \eta^{NL}(1 - p^L)\right) \tag{10}$$

with f denoting the carrier frequency and c the speed of light.

Transmission rate $r_{m,k}(t)$ from the m-th VD to the k-th UAV, and $r_{k,c}(t)$ from the k-th UAV to CC are calculated using SINR, considering interference from other sources [19]:

$$r_{m,k}(t) = W_{VD} \log_2\left(1 + \frac{P_m^{tr}(t)}{PL_m(t)} \cdot \frac{1}{N_0 W_{VD} + I_m}\right) \tag{11}$$

$$r_{k,c}(t) = b_k(t) W \log_2\left(1 + \frac{P_k^{tr}(t)}{PL_k(t)} \cdot \frac{1}{N_0 b_k(t) W + I_k}\right) \tag{12}$$

where $P_m^{tr}(t)$ and $P_k^{tr}(t)$ are the transmission powers, N_0 represents the noise power spectral density, W_{VD} is the VD's bandwidth, and $b_k(t)W$ is the allocated bandwidth for UAV-to-CC links.

The total interference I_m in the link from the m-th VD to the k-th UAV and I_k in the link from the k-th UAV to CC are determined by summing the interference power from all other sources, excluding the main transmitter:

$$I_m = \sum_{m' \neq m} \frac{P_{m'}^{tr}(t)}{PL_{m'} N_0 W_{VD}} \tag{13}$$

$$I_k = \sum_{k' \neq k} \frac{P_{k'}^{tr}(t)}{PL_{k'} N_0 b_{k'}(t) W} \tag{14}$$

C. Problem Formulation

In addressing the critical needs of our IoV-MEC system, we prioritize the freshness of data, quantified through the AoI. For each VD at any given time t, AoI is defined as the time elapsed since the generation of the most recently processed data packet received by the CC, expressed as:

$$\Delta_m(t) = t - T_g^m(t) \tag{15}$$

where $T_g^m(t)$ is the timestamp of the last data packet generated by the m-th VD that was completed and received at the cloud.

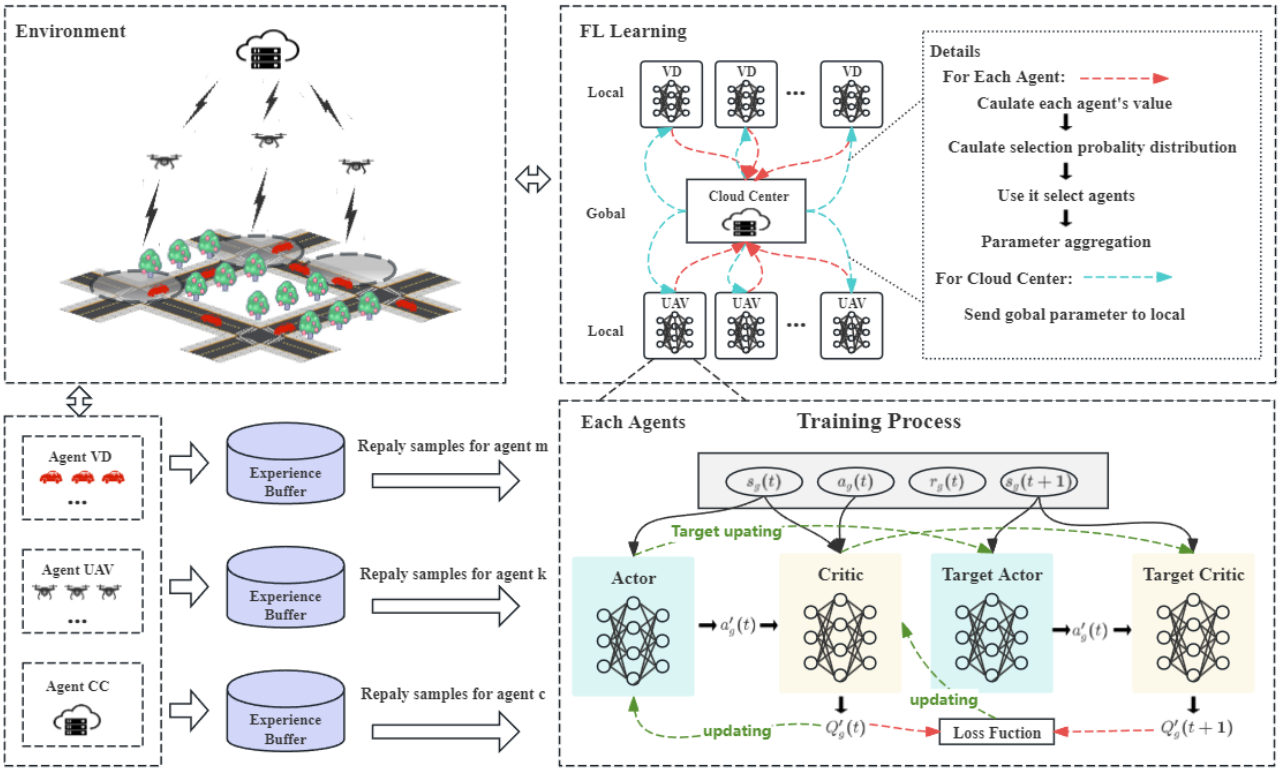


Fig. 2. Architecture of CHFAC.

The average AoI across all VDs provides a crucial measure of system performance, emphasizing the system's ability to handle and update data efficiently and timely. Thus, the average AoI of the system at any time t is given by:

$$\hat{\Delta}(t) = \frac{1}{M} \sum_{m=1}^{M} \Delta_m(t) \tag{16}$$

The goal of this work is to minimize the average AoI of the considered three-tier IoV-MEC system. Therefore, the optimization problem $\mathcal{P}$ can be formulated as:

$$\begin{aligned}
&\mathcal{P}: \min_{\{d_m(t),\text{move}_k(t),\text{exec}_k(t),\text{off}_k(t),b(t)\}} \hat{\Delta}(t) \\
&\text{s.t.} \\
&\text{C1: } \|\text{move}_k(t)\|_2 \leq r_{\text{move}}^k, \\
&\text{C2: } \sum_{i=1}^{B_{\text{col}}^k} \text{exec}_k^i(t) = 1, \quad \forall k = 1, \ldots, K, \\
&\text{C3: } \sum_{i=1}^{B_{\text{exe}}^k} \text{off}_k^i(t) = 1, \quad \forall k = 1, \ldots, K.
\end{aligned} \tag{17}$$

where $\{d_m(t), \text{move}_k(t), \text{exec}_k(t), \text{off}_k(t), b(t)\}$ denote the action set of each VD, UAV, and CC.

III. Edge-Based Federated Actor-Critic Approach for Enhancing Multi-Agent Cooperation

To address the optimization problem $\mathcal{P}$, we model it as a Distributed Partially Observable Markov Decision Process (DEC-POMDP) using a reinforcement learning (RL) approach. While value-based RL methods are useful for quantifying action utilities, they face challenges with complex action spaces due to computational demands. Policy-based approaches like DDPG and A2C are advantageous as they directly estimate actions and their values, simplifying decision-making in extensive action spaces. However, in multi-agent settings, traditional single or centralized methods face scalability issues due to high computational loads and complex parameter management [20].To overcome these challenges, we introduce the Collaborative Heterogeneous Federated Actor-Critic (CHFAC) framework, which uses federated learning principles to improve cooperation and system performance, depicted in Fig. 2. This framework supports the development of a novel collaborative learning algorithm to optimize the AoI across the IoV-MEC system.

A. Enhanced DEC-POMDP Framework for IoV-MEC Systems

As an advanced iteration of Distributed Partially Observable Markov Decision Processes, our cooperative DEC-POMDP framework encompasses a set of heterogeneous agents $G =$

$\{1, 2, \ldots, M+K+1\}$, consisting of M VDs, K UAVs, and 1 CC, each interacting within an environment. The model is defined by a state space S, where $S = \{S_O^g\}_{g\in G}$ represents the collection of observation spaces for each agent. Each agent has its own action sets A_g, and operates based on a reward mechanism R, where actions are determined by local, partial observations $o_g(t) \in O_g(t)$ from the environment, shaped by individual agent's policies. This framework supports dynamic decision-making in IoV-MEC systems under uncertainty.

1) State and Observation Dynamics: Each agent in the system has a state representation that includes its spatial location and resource utilization status. The state $S(t)$ at any given time t combines the position and buffer statuses of all VDs and the position, buffer, and bandwidth statuses of all UAVs:

$$\begin{aligned} S(t) = \{(\mathbf{p}_m(t), z_m(t)) \mid m \in M\} \\ \cup\{(\mathbf{p}_k(t), b_k(t), z_k(t)) \mid k \in K\} \end{aligned} \tag{18}$$

where M represents the set of all VDs, and K represents the set of all UAVs.

Each agent has a partial and localized view of the system, defined by its observable subset of states. Specifically:

a) VDs: For each VD, the observation is:

$$O_m(t) = (\mathbf{p}_m(t), z_m(t)) \tag{19}$$

where $\mathbf{p}_m(t)$ is the location and $z_m(t)$ is the buffer state of the m-th VD.

b) UAVs: For each UAV, the observation set comprises:

$$\begin{aligned} O_k(t) =& (\mathbf{p}_k(t), b_k(t), z_k(t), \\ & \{\mathbf{p}_m(t)\}_{m\in Cov_k}, \{b_m(t)\}_{m\in Cov_k}) \end{aligned} \tag{20}$$

where $\mathbf{p}_k(t)$ denotes the location, $b_k(t)$ represents the bandwidth of the k-th UAV, and $z_k(t)$ indicates the buffer capacity. The term Cov_k refers to the set of VDs covered by the k-th UAV, while $\{\mathbf{p}_m(t)\}$ and $\{b_m(t)\}_{m\in Cov_k}$ detail the states of these covered VDs.

c) CC: The central controller's observation encompasses:

$$O_{CC}(t) = \bigcup_{k\in K} (\mathbf{p}_k(t), b_k(t), z_k(t)) \tag{21}$$

aggregating the location, buffer, and bandwidth of all UAVs.

2) Actions and System Dynamics: Actions within the system are defined for each agent type—VDs, UAVs, and CC.

a) VDs: For each VD, actions include computation offloading decisions:

$$a_m(t) = \{d_m(t)\} \tag{22}$$

where $d_m(t)$ denotes the offloading decisions at time t.

b) UAVs: UAVs perform multiple tasks. For each UAV:

$$a_k(t) = \{\text{move}_k(t), \text{exec}_k(t), \text{off}_k(t)\} \tag{23}$$

Here, $move_k(t)$ represents the movement decisions, $exec_k(t)$ encapsulates computational, and $off_k(t)$ captures the offloading decisions.

c) CC: The CC coordinates system-wide resources with a focus on bandwidth management:

$$a_c(t) = \{b(t)\} \tag{24}$$

where $b(t)$ denotes the bandwidth allocations to UAVs at time t.

3) State Transition and Reward Mechanism: State transitions in the IoV-MEC system are governed by the probability $P(S(t+1) \mid S(t), A(t))$, where $A(t)$ denotes the set of actions taken by heterogeneous agents at time t. Considering the focus on age-sensitive IoV-MEC system optimization where all agents collaboratively minimize the average age of data sources, the reward for each agent can be described as the of the current AoI at any given slot t:

$$r_g(t) = \Delta(t) \quad \text{for } g = 1, \ldots, M+K+1. \tag{25}$$

To enhance the strategic decisions across longer time frames and encourage sustained improvements in AoI, a long-term reward function with decay is employed:

$$R_g(t) = \sum_{i=0}^{T} \gamma^i \cdot \Delta(t+i) \tag{26}$$

where γ is the decay coefficient and T is the length of the time window. This setup allows the agents to consider the future impact of their actions, integrating a more strategic approach to minimize AoI in the network.

B. CHFAC Framework Design

In the CHFAC framework, each agent operates with dedicated actor and critic networks to interact with the environment and learn their respective optimal strategies. The framework's strength lies in its heterogeneity—each agent's neural network design is specifically tailored to its role, enhancing its ability to effectively interact with different agent types within the network. Federated learning is also leveraged to ensure collaborative enhancement and synchronization across different agents.

Enhancements with Target Networks Target networks, which mirror the structure of primary actor and critic networks, are updated less frequently to stabilize training via temporal-difference learning:

$$\theta_g^{t+1} = \kappa\theta_g^t + (1-\kappa)\theta_g^t \tag{27}$$

$$\phi_g^{t+1} = \kappa\phi_g^t + (1-\kappa)\phi_g^t \tag{28}$$

Updates are applied every T^{up} training epochs.

Online Distributed Training-Execution Our system employs an online distributed training-execution mode, allowing agents to adaptively learn during operations:

$$a_g(t) = \begin{cases} \text{random action}, & \text{with probability } \epsilon, \\ \arg\max_a Q_g(O_g(t), a_g), & \text{with probability } 1-\epsilon \end{cases} \tag{29}$$

The exploration rate ϵ decreases exponentially, adjusting as $\epsilon = \epsilon \cdot 0.9^{N_e}$ where N_e denotes the episode number.

During the interaction phases, agents store transitions in their experience buffer to facilitate learning. The training involves the updating of both critic and actor networks based on their respective loss functions. The critic network updates are governed by the loss function $\ell_{\mathcal{C}}$, defined as:

$$\ell_{\mathcal{C}}(\phi_g) = \mathbb{E}\left[(Q_g(t) - y_g)^2\right] \tag{30}$$

where y_g is defined as:

$$y_g = r_g(t) + \gamma Q_g(t+1) \tag{31}$$

The actor network updates involve optimizing the loss function $\ell_{\mathcal{A}}$, which is influenced by the critic's evaluation:

$$\ell_{\mathcal{A}}(\theta_g) = \mathcal{C}(O_g(t), \mathcal{A}_g(O_g(t); \theta_g); \phi_g) \tag{32}$$

Network parameters are updated using SGD optimizer:

$$\phi_g^{t+1} = \phi_g^t - \eta_C \nabla_{\phi_g} \ell_{\mathcal{C}}(\phi_g) \tag{33}$$

$$\theta_g^{t+1} = \theta_g^t - \eta_A \nabla_{\theta_g} \ell_{\mathcal{A}}(\theta_g) \tag{34}$$

Federated Learning: Additionally, to improve system performance, accelerate convergence during the training process, and effectively address data heterogeneity and device heterogeneity, an enhanced FL mode is integrated into multi-agent DRL. This FL framework involves two distinct groups of agents: VDs and UAVs. Both groups participate independently in the federated learning process, each contributing to the global model based on their local data updates. This FL mode selects more important devices for parameter aggregation.

Specifically let us focus on UAVs as an example: During the local update steps of each agent, a loss function with an additional regularization term is minimized:

$$h_k(\mathbf{w}, \mathbf{w}^t) = \ell_k(\mathbf{w}^t) + \frac{\mu}{2}\|\mathbf{w} - \mathbf{w}^t\|^2 \tag{35}$$

where $\ell_k(\mathbf{w}^t)$ represents the original loss function, and $\mathbf{w}$ denotes the global weights. The second term is the regularization term, which helps in preventing overfitting and stabilizes the learning process by smoothing the parameter updates.

To calculate the influence I_k^t for each client dynamically based on the previous training round, the formula is adjusted to integrate γ_k, which represents the relative gradient norm change between the current and the previous model weights. The modified formula is given by:

$$\begin{aligned} I_k^t =& \langle \nabla f(\mathbf{w}^t), \nabla \ell_k(\mathbf{w}^t) \rangle \\ &- \psi \left(\frac{\|\nabla h(\mathbf{w}_k^{t+1}, \mathbf{w}_k^t)\|}{\|\nabla h(\mathbf{w}_k^t, \mathbf{w}_k^t)\|} \right) \|\nabla \ell_k(\mathbf{w}^t)\|^2 \end{aligned} \tag{36}$$

where ψ is a hyperparameter, $\nabla f(\mathbf{w}^t)$ represents the global gradient. This approach allows the CC to weigh each client's update relevance more effectively, enhancing the utility of the aggregated model updates.Each parameter aggregation tends to select devices with high I_k^t values. The calculation of the optimal selection probability distribution is as follows:

$$P_t^k = \frac{I_k^t}{\sum_{k'=1}^{N} I_{k'}^t} \tag{37}$$

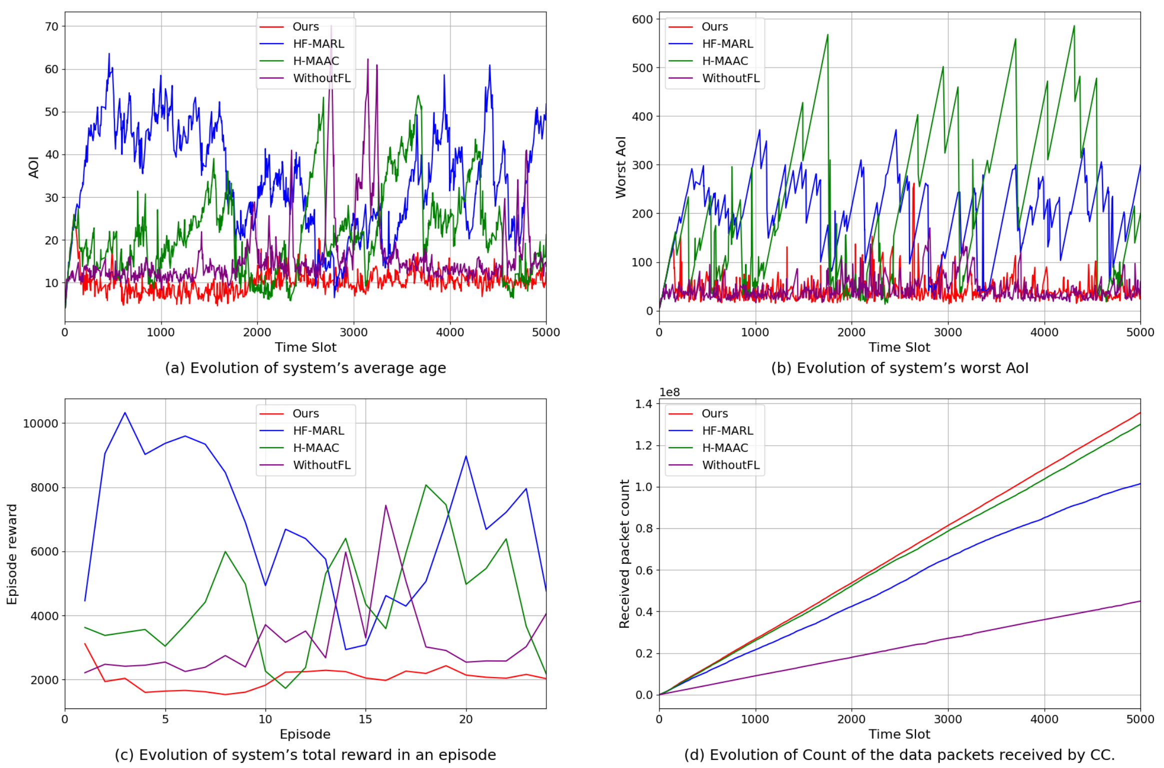
Algorithm 1 Online CHFAC

```
Initialization:
Initialize system parameters and neural network parameters $\theta_m, \theta_k, \theta_c, \phi_m, \phi_k, \phi_c$ for learning.
Initialize experience buffer $B$.
for epoch = 1 to MAXEPOCH do
  for agent $g$ in $G$ do
    Observe $O_g(t)$.
    Select action $a_g(t)$ following Eq.(29).
    Execute $a_g(t)$ in the environment.
    Store $\{s_g(t), a_g(t), r_g(t), s_g(t+1)\}$ in $B[g]$.
    Sample a batch from $B[g]$.
    Select actions for $t+1$ and evaluate Q-values
    Calculate loss $\ell_A(\theta_g)$ and $\ell_C(\phi_g)$ using Eq.(30) and Eq.(32).
    Update actor and critic networks using SGD by Eq.(33) and Eq.(34).
    if $t \mod T_{\text{up}} == 0$ then
      Update target networks by Eq.(27) and Eq.(28).
    end if
    if $t \mod E_f == 0$ then
      Each agent $g$ (UAV or VD) independently conducts federated learning:
      Calculate gradients $\nabla\theta_g$ and $\nabla\phi_g$ by Eq.(30) and Eq.(32).
      Send gradients and network weights to the CC.
      Cloud calculates $I_t^k$ using Eq.(36) and optimal selection $P_t^k$ by Eq.(37).
      Sample $N$ devices based on $P_t^k$ and aggregate using Eq.(38) to get global model $w^{t+1}$.
      Send global model $w^{t+1}$ back to the devices.
    end if
  end for
end for
```

The CC selects a subset D_N consisting of N devices by performing K sampling operations according to the probability distribution P_t^k. This non-uniform sampling strategy not only prioritizes more significant devices for participation in refining the global model but also optimizes communication efficiency by reducing the number of necessary communication rounds and lowering the associated costs. The update to the global model is then computed through weighted aggregation of the parameters as per the equation below:

$$\mathbf{w}^{t+1} = \mathbf{w}^t + \sum_{k \in D_N} \frac{I_t^k}{\sum_{k' \in D_N} I_t^{k'}} \Delta \mathbf{w}_k^{t+1} \tag{38}$$

where $\Delta \mathbf{w}_k^{t+1} = \mathbf{w}_k^{t+1} - \mathbf{w}^t$. Following this, the CC dispatches the updated global model $\mathbf{w}^{t+1}$ to the devices within D_N for subsequent network synchronization. The proposed CHFAC framework is summarized in Algorithm 1.



(a) Evolution of system's average age

(b) Evolution of system's worst AoI

(c) Evolution of system's total reward in an episode

(d) Evolution of Count of the data packets received by CC.

Fig. 3. The performance evaluation of proposed CHFAC algorithm and other baselines.

IV. PERFORMANCE ANALYSIS

This research presents a multi-UAV-supported age-sensitive IoV-MEC system, utilizing a standard gym module for its framework. The system deploys six UAVs and thirty diverse VD types across a virtual 200x200 grid. Data from sensors is generated following a Poisson distribution, with an arrival rate of 1 kb/slot. The UAVs operate within a mobility radius of 6 meters and a surveillance radius of 60 meters, maintaining a constant altitude while gathering data from sources on the grid.

Data processing speed on the UAVs is set at 8 kb/slot, and each UAV has a dropout probability of 0.005 per time slot, implying an average of one UAV dropout every 200 slots. System rewards are controlled with a decay coefficient of 0.85, and the learning rates for actor and critic networks are configured at 0.03 and 0.05, respectively. The update interval and weight retention parameters are adjusted to 8 and 0.8.

In the CHFAC framework, VD networks employ three-layer MLPs with 128, 256, and 128 neurons respectively. UAV networks integrate spatial analysis via CNN layers, followed by MLPs with 256 and 128 neurons for trajectory and resource management. CC networks feature dense MLPs with two 256-neuron layers for bandwidth distribution, with a federation cycle set to 8 to ensure efficient operation.

The performance of the proposed CHFAC is compared against three key benchmarks:

1) The Edge Federated Multi-Agent Actor-Critic (H-MAAC) algorithm as outlined in a recent study [21].
2) The heterogeneous federated multi-agent reinforcement learning (HF-MARL) algorithm following [22].
3) All agents are trained in the same way as our algorithm but the federated updating is not performed.

These comparisons aim to demonstrate advancements in processing efficiency and data management within UAV-assisted IoV-MEC environments.

Fig. 3 illustrates the performance of the proposed algorithm against three benchmarks across different metrics. Fig. 3(a) shows that the proposed CHFAC algorithm significantly reduces the system's average AoI over time, indicating rapid convergence and better handling of network heterogeneity. Fig. 3(b) emphasizes our method's ability to consistently maintain lower worst-case AoI peaks compared to benchmarks. Fig. 3(c) illustrates the total accumulated AoI within a single episode, where the proposed CHFAC algorithm demonstrates a clear reduction in the cumulative AoI, underscoring enhanced efficiency. Finally, in Fig. 3(d), our approach shows higher data throughput, processing an increasing count of packets efficiently.

Table I offers a numerical comparison of AoI metrics across several approaches over 1,000 test rounds. Our proposed approach markedly outperforms existing methods in terms of both average and worst-case AoI. This underscores its superior efficacy in minimizing information delay and variability

TABLE I
NUMERICAL COMPARISON ON AOI METRICS

Approach	AoI		Worst Age	
	mean	std	mean	std
HF-MARL	24.2	7.0	188.65	78.52
H-MAAC	16.0	2.9	36.14	15.30
Without FL	17.2	5.5	181.528	118.49
CHFAC (proposed)	**10.3**	**1.6**	**34.22**	**9.35**

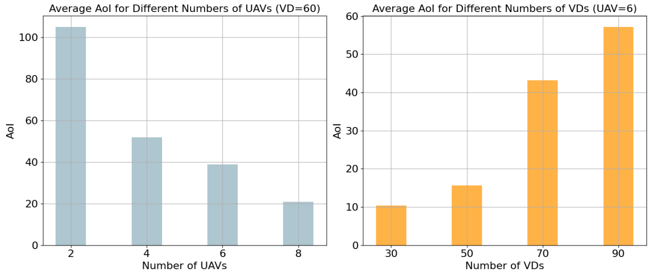


Fig. 4. Impact of UAV and VD Numbers on Average AoI.

across the network.

Fig. 4 depicts the test outcomes assessing the impact of different numbers of UAVs and VDs on the average AoI. The tests, conducted with a fixed set of 60 VDs for the UAV trials and 6 UAVs for the VD trials, demonstrate how variations in the number of devices influence system performance. Notably, an increase in the number of UAVs correlates with a reduction in AoI, suggesting enhanced performance with a larger fleet of UAVs. On the contrary, a higher number of VDs tends to increase AoI, indicating scalability challenges and potential inefficiencies with larger fleets.

V. CONCLUSION

In this paper, we utilize a novel FL method in an IoV-MEC system to optimize the AoI. This method, termed CHFAC, integrates a MARL framework that dynamically evaluates and selects devices for federated aggregation based on each agent's significance. Our innovative approach not only enhances the efficiency and relevance of the learning process but also adapts effectively to the inherent heterogeneity and dynamic conditions of vehicular networks. The CHFAC method, with its robust decision-making capabilities, ensures optimal agent selection, significantly reducing the system's average AoI and improving data packet utilization. Our simulations demonstrate these enhancements, confirming the robustness and effectiveness of our method in managing the complex interactions and real-time demands of IoV-MEC environments.

REFERENCES

[1] Yang F, et al., ”An overview of internet of vehicles,” *China Communications*, vol. 11, no. 10, pp. 1–15, 2014.

[2] P. Mach and Z. Becvar, “Mobile edge computing: A survey on architecture and computation offloading,” *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1628–1656, 2017.

[3] T. Ren, J. Niu, B. Dai, X. Liu, Z. Hu, M. Xu, and M. Guizani, “Enabling efficient scheduling in large-scale multi-assisted mobile edge-computing systems via hierarchical reinforcement learning,” *IEEE Internet of Things Journal*, vol. 9, no. 10, pp. 7957–7970, 2021.

[4] Liu B, Wan Y, Zhou F, Wu Q, Hu RQ, ”Resource allocation and trajectory design for miso UAV-assisted MEC networks,” *IEEE Transactions on Vehicular Technology*, vol. 71, no. 5, pp. 4933–4948, 2022.

[5] Abd-Elmagid M. A., Pappas N., Dhillon H. S., ”On the role of age of information in the Internet of Things,” *IEEE Communications Magazine*, vol. 57, no. 12, pp. 72–77, Dec. 2019.

[6] Arulkumaran K, et al., ”Deep reinforcement learning: A brief survey,” *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 26–38, 2017.

[7] Li Q, et al., ”A survey on federated learning systems: Vision, hype and reality for data privacy and protection,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 4, pp. 3347–3366, 2021.

[8] Ndikumana, Anselme, et al. ”Joint Communication, Computation, Caching, and Control in Big Data Multi-Access Edge Computing.” IEEE Transactions on Mobile Computing 19.6 (2019): 1359-1374.

[9] Zhou, Fuhui, et al. ”UAV-enabled Mobile Edge Computing: Offloading Optimization and Trajectory Design.” 2018 IEEE International Conference on Communications (ICC). IEEE, 2018.

[10] Liu, Yuan, et al. ”UAV-assisted Wireless Powered Cooperative Mobile Edge Computing: Joint Offloading, CPU Control, and Trajectory Optimization.” IEEE Internet of Things Journal 7.4 (2019): 2777-2790.

[11] Xu, Yu, et al. ”Joint Resource and Trajectory Optimization for Security in UAV-Assisted MEC Systems.” IEEE Transactions on Communications 69.1 (2020): 573-588.

[12] Yang, Zheyuan, Suzhi Bi, and Ying-Jun Angela Zhang. ”Online Trajectory and Resource Optimization for Stochastic UAV-Enabled MEC Systems.” IEEE Transactions on Wireless Communications 21.7 (2022): 5629-5643.

[13] Chi, Xiaoyu, et al. ”Age of Model: A Metric for Keeping AI Models Up-to-date in Edge Intelligence-enabled IoV.” IEEE Transactions on Vehicular Technology (2024).

[14] Bai, Guangming, et al. ”AoI-Aware Joint Scheduling and Power Allocation in Intelligent Transportation System: A Deep Reinforcement Learning Approach.” IEEE Transactions on Vehicular Technology (2023).

[15] Han, Rui, et al. ”Age of Information Aware UAV Deployment for Intelligent Transportation Systems.” IEEE Transactions on Intelligent Transportation Systems 23.3 (2021): 2705-2715.

[16] Han, Rui, et al. ”Age of Information and Performance Analysis for UAV-Aided IoT Systems.” IEEE Internet of Things Journal 8.19 (2021): 14447-14457.

[17] Qin, Zhen, et al. ”AoI-Aware Scheduling for Air-Ground Collaborative Mobile Edge Computing.” IEEE Transactions on Wireless Communications 22.5 (2022): 2989-3005.

[18] Al-Hourani A, Kandeepan S, Lardner S. ”Optimal LAP altitude for maximum coverage.” in *IEEE Wireless Communications Letters*, 2014;3(6):569-572.

[19] Luong NC, Hoang DT, Wang P, Niyato D, Kim DI, Han Z. ”Applications of deep reinforcement learning in communications and networking: A survey.” in *IEEE Communications Surveys & Tutorials*, 2019;21(4):3133-3174.

[20] Zhang K, Zhu Y, Letaief KB. ”Centralized deep reinforcement learning for large-scale cellular network optimization.” in *IEEE Transactions on Wireless Communications*, 2020;19(8):5578-5592.

[21] Z. Zhu, S. Wan, P. Fan and K. B. Letaief, ”Federated Multiagent Actor–Critic Learning for Age Sensitive Mobile-Edge Computing.” in *IEEE Internet of Things Journal*, 2022.

[22] Li H, Zhang J, Zhao H, Ni Y, Xiong J, Wei J. ”Joint optimization on trajectory, computation and communication resources in information freshness sensitive MEC system.” in *IEEE Transactions on Vehicular Technology.*, 2023.