Deep Reinforcement Learning for Secure MEC Service in Vehicular Networks with Reconfigurable Intelligent Surfaces

Haoyu Wang*, Haowen Bai†, Ying Ju†, A. Lee Swindlehurst*
*Center for Pervasive Communications and Computing, University of California Irvine, CA, USA.
†School of Telecommunications Engineering, Xidian University, China.
Email: haoyuw30@uci.edu, juying@xidian.edu.cn, swindle@uci.edu.

2023 57th Asilomar Conference on Signals, Systems, and Computers | 979-8-3503-2574-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/IEEECONF59524.2023.10476897

***Abstract*—The broadcasting nature of wireless signals may result in the task offloading process of mobile edge computing (MEC) suffering serious information leakage. As a novel technology, physical layer security (PLS) combined with reconfigurable intelligent surfaces (RIS) can enhance transmission quality and security. This paper investigates the MEC service delay problem in RIS-aided vehicular networks under malicious eavesdropping. Due to the lack of an explicit formulation for the optimization problem, we propose a deep deterministic policy gradient (DDPG)-based communication scheme to optimize the secure MEC service. It aims to minimize the maximum MEC service time while reducing eavesdropping threats by jointly designing the RIS phase shift matrix and computing resource allocation in real-time. Simulation results demonstrate that 1) the DDPG-based scheme can help the base station make reasonable actions to realize secure MEC service in dynamic MEC vehicular networks; 2) deploying RIS can dramatically reduce eavesdropping threats and improve the overall MEC service quality.**

***Index Terms*—Deep Reinforcement Learning, Mobile Edge Computing, Vehicular Networks, Reconfigurable Intelligent Surfaces, Security.**

I. INTRODUCTION

EMERGING vehicle-to-everything (V2X) communication technology is expected to support numerous intelligent transportation services, requiring strong computing capability for data analysis [1]. To emancipate resource-limited vehicle users from heavy computing services, mobile edge computing (MEC) can take advantage of abundant computing resources at the network edge. However, the task offloading rate can be low due to the severe channel fading in congested urban environments, which prolongs the offloading delay. Moreover, due to the broadcast nature of wireless signals, wireless links are prone to security threats such as eavesdropping. It is crucial to promote service quality and data security in MEC vehicular networks from a secure communication perspective [1].

Reconfigurable intelligent surfaces (RIS) are regarded as a promising technique to enhance wireless transmission quality and coverage [2]. Previous studies have demonstrated that by exploiting the inherent randomness in wireless channels, physical layer security (PLS) can be an effective alternative or complementary solution for safeguarding the security of complicated wireless networks [3]. However, when the eavesdroppers are closer to the base station (BS) than the legitimate users or when the legitimate users and eavesdroppers have correlated channels, many PLS technologies will seriously deteriorate. For these severe challenges, RIS combined with PLS brings hope for designing a robust transmission mechanism given its ability to flexibly reconstruct the channel environment [4]. However, the techniques proposed therein require modifications in order to work in complicated dynamic communication scenarios such as MEC vehicular networks.

Inspired by advances in artificial intelligence (AI), various Deep Learning (DL) or Deep Reinforcement Learning (DRL) algorithms have been exploited to solve the optimization challenge due to the randomness, dynamism, and mathematical complexity in 6G networks [5]. Motivated by this, several advanced works use AI algorithms to jointly optimize the parameters in RIS-aided systems and realize secure communication. For example, the authors in [6] propose a DRL-based secure transmission method to resist eavesdroppers by jointly optimizing the beamforming and the RIS phase shifts.

Currently, researchers are focusing on the security threats inherent in MEC service applications and have developed methods for different communication scenarios [7]–[10]. In [7], the authors propose a RIS-assisted secure MEC service framework that aims to solve the max-min computation efficiency problem. In [8], the authors minimize the MEC energy consumption in a RIS-assisted MEC system, where the full-duplex BS emits artificial noise to resist eavesdroppers. The work of [9] concentrates on optimizing the secure MEC delay for a target vehicle, where the BS emits artificial noise (AN) to confuse the eavesdroppers. These MEC networks are proposed for static communication scenarios with fixed locations or ignore the MEC computing time, and thus they cannot be applied to dynamic MEC vehicular networks. While the authors of [10] propose a DRL-based secure scheme for MEC service in a dynamic Internet of Vehicles (IoV) setting, it does not explore the potential benefits of RIS.

The above research shows the great potential of RIS and the optimization capability of AI algorithms. In this paper, we demonstrate the potential of RIS for realizing secure MEC service in dynamic vehicular networks, a problem that has

The work of H. Wang and A. L. Swindlehurst were supported by the U.S. National Science Foundation under grants ECCS-2030029 and CNS-2107182. The work of H. Bai and Y. Ju were supported by the National Natural Science Foundation of China under Grant 62102301.

not yet been investigated. In particular, we propose a DRL-based communication scheme to realize secure MEC service in RIS-aided vehicular networks, where the BS designs RIS properties and allocates the computing resources to optimize the MEC service time while reducing the eavesdropping threat. It is worth noting that in most MEC studies, the BS does not begin allocating MEC resources until all tasks have been offloaded [7]–[9]. Fortunately, our proposed DRL-based scheme can help the BS to flexibly allocate computing resources to users once they complete the offloading process, which benefits the utilization of idle MEC resources.

II. System Model

A. RIS-aided MEC Vehicular Network

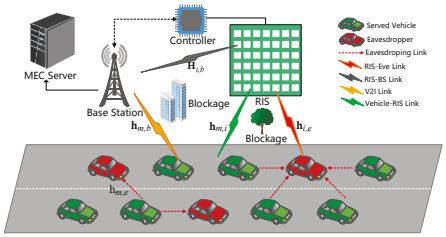

Fig. 1. Illustration of the RIS-aided secure MEC vehicular network.

We consider a RIS-aided MEC vehicular network, illustrated in Fig. 1, where a base station simultaneously establishes multiple communication links with vehicular users in different sub-bands to realize a high data-rate service [10]. In the MEC scenario, through the vehicle-to-infrastructure (V2I) links, the resource-constrained vehicles can offload their computational tasks to the BS equipped with a MEC server. The BS flexibly allocates the MEC resources for different task requests and then feeds back the results to the users. In this paper, we assume that the length of the feedback delay is negligible relative to the time scale required to meet the computational tasks [7]. Due to the limited resources at the BS, only a subset of the vehicles that send a computing service request can be serviced. The V2I links acquiring computing services are denoted by $\mathcal{M} = \{\mathrm{User}_1, \mathrm{User}_2, \cdots, \mathrm{User}_\mathrm{M}\}$. The vehicles that are not served are considered to be potential eavesdroppers and are denoted by the set $\mathcal{E} = \{\mathrm{Eve}_1, \mathrm{Eve}_2, \cdots, \mathrm{Eve}_\mathrm{E}\}$.

B. RIS-aided Secure Communication Model

In our assumed model, all vehicles are equipped with one omnidirectional antenna, while the BS has a K-antenna array. We assume the reflection coefficient for the n-th element of the RIS is given by $\theta_n = e^{j\phi_n}$, where $\phi_n \in [0, 2\pi)$, and we define the RIS reflection-coefficient matrix as $\mathbf{\Theta} = \mathrm{diag}([\theta_1, \theta_2, \cdots, \theta_\mathrm{N}])$. Since there is no in-band interference, the BS employs Maximum-Ratio Combining (MRC) for each V2I link. The matrix of MRC beamformers is denoted by $\mathbf{F} = [\mathbf{f}_1, \cdots, \mathbf{f}_M] \in \mathbb{C}^{K\times M}$, where $\mathbf{f}_m$ represents the beamforming vector with unit norm for the m-th V2I link.

1) Channel Model: In the MEC vehicular network, the channels include $\mathbf{h}_{m,b} \in \mathbb{C}^{K\times 1}$ for the m-th V2I link, $\mathbf{h}_{m,i} \in \mathbb{C}^{N\times 1}$ for the link between the m-th served vehicle and the RIS, $h_{m,e} \in \mathbb{C}$ for the m-th served vehicle to the e-th potential eavesdropper, $\mathbf{h}_{i,e} \in \mathbb{C}^{1\times N}$ for the RIS to the e-th potential eavesdropper, and $\mathbf{H}_{i,b} \in \mathbb{C}^{K\times N}$ for the RIS to BS. We assume the channel gain from the RIS to the BS follows a Rician distribution and can be denoted by

$$\mathbf{H}_{i,b} = \sqrt{\rho d_{i,b}^{-\alpha_{i,b}}}\left(\sqrt{\frac{\kappa_{i,b}}{1+\kappa_{i,b}}}\mathbf{H}_{i,b}^{LoS} + \sqrt{\frac{1}{1+\kappa_{i,b}}}\mathbf{H}_{i,b}^{NLoS}\right), \tag{1}$$

where $\kappa_{i,b}$ is the Rician factor, ρ is the path loss at the reference distance $d_0 = 1$m, $d_{i,b}$ is the distance between the RIS and BS, and $\alpha_{i,b}$ is the path loss exponent of the RIS-to-BS link. The line-of-sight (LoS) component $\mathbf{H}_{i,b}^{LoS} \in \mathbb{C}^{K\times N}$ is rank one, and each element of the non-LoS (NLoS) component $\mathbf{H}_{i,b}^{NLoS} \in \mathbb{C}^{K\times N}$ follows an i.i.d. complex Gaussian distribution with zero mean and unit variance. Likewise, the channels $h_{m,e}$, $\mathbf{h}_{m,b}$, $\mathbf{h}_{m,i}$, and $\mathbf{h}_{i,e}$ follow a Rician distribution similar to (1), except for $\mathbf{h}_{m,b}$ and $h_{m,e}$, where it is assumed that $\kappa_{m,b} = 0$ and $\kappa_{m,e} = 0$ due to the congested urban environment and the blocking effect between vehicles. We assume that all channels follow block-based fading and that the global CSI is known at the BS and remains invariant during each time slot but changes from one slot to another [7], [11].

2) Signal Receiving Process: The received signal at the BS from the m-th V2I link can be formulated as

$$y_m = \mathbf{f}_m^H\left[\sqrt{P_m}(\mathbf{H}_{i,b}\mathbf{\Theta}\mathbf{h}_{m,i} + \mathbf{h}_{m,b})s_m + \mathbf{n}_m\right], \tag{2}$$

where P_m is the transmit power of the m-th served vehicle and s_m represents a unit-energy signal sample associated with the computing task. The noise vector $\mathbf{n}_m$ is denoted by $\mathbf{n}_m = [n_1, \cdots, n_K]^T$, where $n_k \sim \mathcal{N}(0, \sigma^2)$. The uplink signal-to-interference-plus-noise ratio (SINR) of the m-th V2I link at the BS is thus given by

$$\eta_m = \frac{P_m|\mathbf{f}_m^H(\mathbf{H}_{i,b}\mathbf{\Theta}\mathbf{h}_{m,i} + \mathbf{h}_{m,b})|^2}{\sigma^2\|\mathbf{f}_m^H\|^2}. \tag{3}$$

Similarly, the eavesdropped signal at vehicle e from the m-th V2I link is expressed as

$$y_{e,m} = \sqrt{P_m}(\mathbf{h}_{i,e}\mathbf{\Theta}\mathbf{h}_{m,i} + h_{m,e})s_m + n_e, \tag{4}$$

where $n_e \sim \mathcal{N}(0, \sigma^2)$, and thus the SINR of the m-th V2I link at eavesdropper e can be expressed as

$$\eta_{e,m} = \frac{P_m|\mathbf{h}_{i,e}\mathbf{\Theta}\mathbf{h}_{m,i} + h_{m,e}|^2}{\sigma^2}. \tag{5}$$

Accordingly, the capacity of the m-th V2I link and the wiretapped capacity of eavesdropper e to the m-th V2I link are $C_m = \log(1+\eta_\mathrm{m})$ and $C_{e,m} = \log(1+\eta_{\mathrm{e,m}})$, respectively.

C. Problem Formulation

The MEC server will flexibly allocate the computing resources (e.g., CPU cycles) according to the size of the tasks once users finish the offloading process. Specifically, each

CPU cycle can process a certain number of data bits, and we assume that the total computing capability is ζ bits/s. To provide a stable service, the BS aims to minimize the service time for the whole MEC process while ensuring task offloading security for all users.

1) Secrecy Transmission Rate: We consider a worst-case security threat scenario in which any unserved vehicle can eavesdrop on any of the V2I links. To protect the task data from being wiretapped, the transmitters encode the confidential data (e.g., using the Wyner code [3]), and then two code rates must be determined before transmission, namely, the codeword rate R_b, and the target secrecy rate of the confidential information R_S. The redundancy for securing the confidential information is thus given by $\max\{0, R_b - R_S\}$. A secrecy outage occurs if the capacity of the eavesdropper C_e is larger than $R_b - R_S$. In practice, we approximate R_b with the capacity C_b. The secrecy transmission rate of the m-th V2I link can be defined as $R_{S,m} = [C_m - \max C_{e,m}]^+$ for $e \in \mathcal{E}$, where $[x]^+ = \max\{0, x\}$.

2) Optimization of MEC Service Time: Intuitively, the optimization target is to minimize the service time by designing the RIS reflection coefficient matrix $\mathbf{\Theta}$ and the MEC resource allocation ($\sum_{m=1}^{M} \zeta_m = \zeta$) for different computing tasks in each slot. Specifically, the secure MEC service time for m-th V2I link can be denoted by t_m^S, which contains the task offloading time and computing time. Considering that the entire MEC service period is determined by the maximum service time of all V2I links, we transform the above goal into the following min-max problem:

$$\min_{\mathbf{\Theta},\zeta} \max_{m\in\mathcal{M}} \left\{t_m^S\right\}, \tag{6}$$

$$C_1 : \sum_{m=1}^{M} \zeta_m = \zeta, \zeta_m \in [0, \zeta], \tag{6a}$$

$$C_2 : |\theta_n| = 1, \forall n = 1, 2, ..., N, \tag{6b}$$

where the constraint (6a) indicates the limit on the computing resource allocated for different tasks, and (6b) constrains the modulus of RIS reflection coefficients to be unity.

The joint design of the RIS reflection-coefficient matrix and MEC resource allocation for the entire MEC service can be modeled as a Markov Decision Process (MDP). It is composed of multiple time slots (and their specific actions), and each action impacts the future benefits. The optimization problem is not only non-convex but a long-term decision process with high dynamics and is intractable to formulate with an explicit mathematical expression. Consequently, we employ a DRL algorithm referred to as the Deep Deterministic Policy Gradient (DDPG) [12].

III. DDPG-Based Approach

In the DDPG-based algorithm, the environment consists of the RIS-aided MEC vehicular network, and the BS is regarded as an agent that interacts with the dynamic environment. In the MDP, at each time slot t, the BS obtains the current state S_t from the environment and selects the action a_t based on the current policy π_t. Each action includes the RIS phase shift matrix and the MEC resource allocation. Once the BS executes the action, it will lead to the new state S_{t+1} in the next time slot $t+1$, and the reward r_t for action a_t will be observed from the changing environment.

A. MDP Formulation

Based on the optimization problem proposed above, the state space, action space, and reward function in the MDP are explained in detail below.

*1) **Action Space**:* Based on the current state S_t, the BS agent takes action a_t corresponding to the RIS phase shift matrix and MEC resource allocation. At each time slot t, the action can be represented as $a_t = \{\mathbf{\Theta}_t, \zeta_t\}$, where $\zeta_t = \{\zeta_t^1, \zeta_t^2, \ldots, \zeta_t^M\}$ is the computing resource allocation.

*2) **State Space**:* At time slot t, the state S_t^m of the m-th V2I link includes the global channel state information $\mathcal{H}_t^m = \left\{\mathbf{h}_t^{m,b}, \mathbf{h}_t^{m,i}, h_t^{m,e}, \mathbf{h}_t^{i,e}, \mathbf{H}_t^{i,b}\right\}$, the secrecy rate $R_{t-1}^{S,m}$, the amount of remaining offloading tasks $\mathcal{K}_t^{o,m}$, the amount of remaining computing tasks $\mathcal{K}_t^{c,m}$, and the amount of occupied MEC resources ζ_{t-1}^m. We denote the state of the m-th V2I as

$$S_t^m = \left\{\mathcal{H}_t^m, R_{t-1}^{S,m}, \mathcal{K}_t^{o,m}, \mathcal{K}_t^{c,m}, \zeta_{t-1}^m\right\}. \tag{7}$$

The overall state S_t of the environment can be expressed as:

$$S_t = \{S_t^m, m = 1, 2, \ldots, M\}. \tag{8}$$

*3) **Reward Design**:* At time slot t, the reward corresponding to the current action a_t can be denoted by

$$r_t = -\max_{m\in\mathcal{M}} \{t_m^{exp}\} + \mu_1, \tag{9}$$

where $t_m^{exp} = t_{m,1} + t_{m,2}$ indicates the estimated secure MEC service time for the m-th V2I link at time slot t, $t_{m,1}$ is the current elapsed time, and $t_{m,2}$ is the estimated residual time based on the current action. To enhance the secrecy rate, we set the penalty factor $\mu_1 = \sum_{m=1}^{M} \nu_m$ with $\nu_m \in \{0, \nu^*\}$. If the current action can satisfy the secrecy rate requirement r_m^s for the m-th link, ν_m equals 0, otherwise it will be a negative value ν^*. Based on guidance from the reward function, the DDPG-based algorithm will continuously learn action policies in the direction of reducing the maximal secure MEC service time within the given constraints. The total reward is defined cumulatively with the discount rate γ:

$$R_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1}, 0 \le \gamma \le 1. \tag{10}$$

B. DDPG Architecture

DDPG is model-free and off-policy with an actor-critic structure. The actor network is for action prediction, and the critic network is for evaluating the future benefit of the action for the current state. The actor and critic networks both consist of two DNN networks: the training network and the target network. The parameters of the actor training and target networks are denoted by $\boldsymbol{\theta}^a$ and $\boldsymbol{\theta}^{a'}$, respectively, while

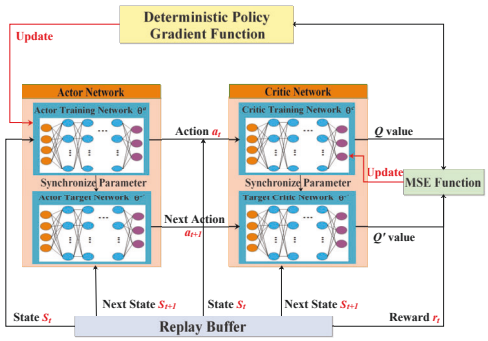

Fig. 2. DDPG Framework.

the parameters of the critic training and target networks are denoted by $\boldsymbol{\theta}^c$ and $\boldsymbol{\theta}^{c'}$, respectively.

At time slot t, the actor training network takes S_t as input and outputs action a_t, the critic training network takes S_t and a_t as input and outputs the state-action value $Q_\pi\left(S_t, a_t \mid \boldsymbol{\theta}^c\right)$, which can be expressed as

$$Q_\pi\left(S_t, a_t \mid \boldsymbol{\theta}^c\right) = E_\pi\left[R_t \mid S_t, a_t, \pi\right], \tag{11}$$

where $E[\cdot]$ denotes the expectation function, π represents the actor training network policy, and the reward R_t is defined in (10). When sufficient transition tuples (S_t, a_t, r_t, S_{t+1}) have been accumulated in the replay buffer D, the optimizer updates the actor and the critic training network by randomly sampling mini-batches of size N_d from the replay buffer. The target Q value Q' of the k-th transition tuple y_k is given by

$$y_k = r_k + \gamma Q'_{\pi'}\left(S_{k+1}, \pi'\left(S_{k+1} \mid \boldsymbol{\theta}^{a'}\right) \mid \boldsymbol{\theta}^{c'}\right), \tag{12}$$

where π' represents the actor target network policy.

The critic training network uses the MSE function to update the network, which is given as

$$L\left(\boldsymbol{\theta}^c\right) = \frac{1}{N_d}\sum_{j=1}^{N_d}\left(y_k - Q_\pi\left(S_k, a_k \mid \boldsymbol{\theta}^c\right)\right)^2, \tag{13}$$

$$\boldsymbol{\theta}^c = \boldsymbol{\theta}^c - \alpha \nabla_{\boldsymbol{\theta}^c} L\left(\boldsymbol{\theta}^c\right). \tag{14}$$

The actor training network uses the deterministic policy gradient function to update the network, as follows:

$$\nabla_{\boldsymbol{\theta}^a} J = \frac{1}{N_d}\sum_{k=1}^{N_d}\nabla_a Q_\pi\left(S_k, a_k \mid \boldsymbol{\theta}^c\right)\nabla_{\boldsymbol{\theta}^a}\pi\left(S_k \mid \boldsymbol{\theta}^a\right), \tag{15}$$

$$\boldsymbol{\theta}^a = \boldsymbol{\theta}^a - \alpha' \nabla_{\boldsymbol{\theta}^\pi} J, \tag{16}$$

where N_d is the mini-batch size and α (α') is learn rate. The updates to the actor and the critic target networks are

$$\begin{aligned} \boldsymbol{\theta}^{c'} &= \tau_c \boldsymbol{\theta}^c + (1-\tau_c)\boldsymbol{\theta}^{c'} \\ \boldsymbol{\theta}^{a'} &= \tau_a \boldsymbol{\theta}^a + (1-\tau_a)\boldsymbol{\theta}^{a'}, \end{aligned} \tag{17}$$

where $\tau_c, \tau_a \in [0, 1]$ are soft update coefficients. The details of the DDPG-based scheme are summarized in Algorithm 1.

Algorithm 1 DDPG-based Approach

Initialization: Randomly initialize the training actor network and the training critic network with parameters $\boldsymbol{\theta}^a$ and $\boldsymbol{\theta}^c$, the target actor network with parameter $\boldsymbol{\theta}^{a'}$=$\boldsymbol{\theta}^a$, and the target critic network with parameter $\boldsymbol{\theta}^{c'}$=$\boldsymbol{\theta}^c$. Empty the experience replay buffer D with size D_B.
for each episode $i = 1, 2, \ldots, N$ **do**
  Initialize a random noise set for action exploration.
  Observe the initial state S_1
  for each time slot t **do**
    Obtain the action $a_t = \{\boldsymbol{\Theta}_t, \zeta_t\}$ from the actor network for M V2I links.
    Obtain the next state S_{t+1} given action a_t, and calculate the reward r_t. Store transition tuple (S_t, a_t, r_t, S_{t+1}) in the experience replay buffer D.
    Obtain the Q value function $Q_\pi\left(S_t, a_t \mid \boldsymbol{\theta}^c\right)$ from the critic network.
    Sample a mini-batch of transition tuples from D with size N_d.
    Update parameters of the critic network $\boldsymbol{\theta}^c$ and actor network $\boldsymbol{\theta}^a$ according to (14) and (16), respectively.
    Update the parameters of the target critic network $\boldsymbol{\theta}^{c'}$ and target actor network $\boldsymbol{\theta}^{a'}$ according to (17).
  end for
end for

IV. Simulations

In this section, we evaluate the system performance via numerical simulations. We utilize SUMO to simulate dynamic vehicle patterns for the congested urban environment [13]. For the vehicular network, we set $K = 32$, $N = 30$, $M = 6$, and $E = 2$. The origin of the three-dimensional coordinate is set at the beginning of the first lane, and the BS and RIS are located at (-10m, 150m, 25m) and (-15m, 170m, 15m), respectively. The path loss exponents of the Vehicle-BS, Vehicle-Eavesdropper, Vehicle-RIS, and BS-RIS channels are 3, 3, 2.2, and 2, respectively [11]. The path loss ρ is -20 dB, and the Rician factors for all channels are equal to 3 dB. Additional parameters are listed in Table I.

TABLE I
SIMULATION PARAMETERS [14], [15]

Parameter	Value
Vehicle transmit power P_m	20 dBm
Vehicle speed	$[36, 72]$ km/h
Carrier frequency	28 GHz
Bandwidth of each target vehicle W	10 MHz
Noise power σ^2	-104 dBm
Max service time T_{max} or number of steps S_{max}	5s or 50
Total computing resource ζ	200 Mb/s

A. Benchmark Schemes and Metrics

We provide a thorough performance analysis by comparing our proposed DRL-based approach to the following benchmark schemes:

* **DDPG-NRIS**: No RIS is present. The DDPG approach is adapted to minimize the maximum MEC time by dynamically allocating the MEC resources to users once they complete the task offloading.
* **Random-CVX**. The RIS reflection coefficients are randomly generated. The vehicular users first complete all task offloading, and then the MEC server pursues the task computing process. The MEC resource allocation is then found by minimizing the maximum computation time, which is a convex optimization.

To adequately evaluate the MEC service performance, we use the following performance metrics: **Average Maximum**

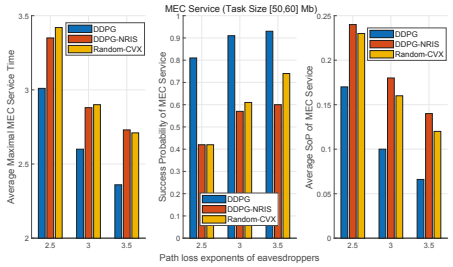

Fig. 3. MEC service under different levels of eavesdropping.

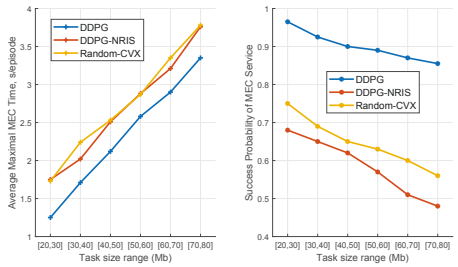

Fig. 4. MEC service performance under different ranges of task size.

MEC Service Time and **Average MEC Service Secrecy Outage Probability (SoP)**. In many situations, the BS will face the challenge that the tasks cannot be completely offloaded in an adequate amount of time. As a result, we set a service time threshold T_{max} or a maximum number of steps S_{max} within which the MEC service must occur for all users. If the maximum MEC service time is smaller than these values, the MEC service episode is considered a success; otherwise, it is labeled a failure. Consequently, we also adopt the **MEC Success Probability** as a metric to evaluate the service quality.

B. Numerical Results

Fig. 3 illustrates the network performance under different levels of eavesdropping capability, which is controlled by the path-loss exponent α_{me}. Compared with the two benchmarks, the DDPG-based approach can reasonably design RIS and achieve a significant reduction in the average maximum MEC service time while enhancing the MEC success probability, particularly for severe eavesdropping threats (e.g., $\alpha_{me} = 2.5$). A higher MEC success probability means that the BS can allow more vehicular users to obtain secure MEC service within the specified service time threshold, which is vital for the robustness of the network.

From Fig. 4, When the task size increases, the average maximum MEC time rapidly increases, while the MEC success probability decreases for all three methods. This is because eavesdropping significantly limits the achievable secrecy rate of the vehicular users, resulting in the inability of the tasks to be quickly offloaded. Fortunately, the DDPG-based approach allows more users to successfully obtain secure MEC services. From the left sub-figure, we see that the average MEC service time for our DDPG-based approach is significantly lower than the benchmark methods, although the percentage decrease in service time is lower for large task sizes. However, the benefit of our DDPG-based algorithm is more clear from the right-hand subfigure, where we see a dramatic increase in the MEC success probability. This demonstrates that our proposed approach is able to successfully learn effective strategies in complicated and very dynamic communication scenarios.

V. Conclusion

In this paper, we have designed a DRL-based communication algorithm to realize secure MEC service in vehicular networks. Our proposed approach aims to minimize the maximum MEC service time while ensuring secure task offloading. Simulation results demonstrate the feasibility and robustness of the proposed DDPG-based approach and validate the great potential of RIS for reducing eavesdropping threats and improving the overall MEC service quality.

References

[1] R. Lu, L. Zhang, J. Ni, and Y. Fang, "5G vehicle-to-everything services: Gearing up for security and privacy," *Proceedings of the IEEE*, vol. 108, no. 2, pp. 373–389, 2020.

[2] C. Pan, H. Ren, K. Wang, J. F. Kolb, M. Elkashlan, M. Chen, M. Di Renzo, Y. Hao, J. Wang, A. L. Swindlehurst, X. You, and L. Hanzo, "Reconfigurable intelligent surfaces for 6G systems: Principles, applications, and research directions," *IEEE Communications Magazine*, vol. 59, no. 6, pp. 14–20, 2021.

[3] Y. Ju, H. Wang, Q. Pei, and H.-M. Wang, "Physical layer security in millimeter wave DF relay systems," *IEEE Transactions on Wireless Communications*, vol. 18, no. 12, pp. 5719–5733, 2019.

[4] M. Hua, Q. Wu, W. Chen, O. A. Dobre, and A. L. Swindlehurst, "Secure intelligent reflecting surface aided integrated sensing and communication," *arXiv e-prints*, p. arXiv:2207.09095, Jul. 2022.

[5] K. B. Letaief, Y. Shi, J. Lu, and J. Lu, "Edge artificial intelligence for 6G: Vision, enabling technologies, and applications," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 1, pp. 5–36, 2022.

[6] H. Yang, Z. Xiong, J. Zhao, D. Niyato, L. Xiao, and Q. Wu, "Deep reinforcement learning-based intelligent reflecting surface for secure wireless communications," *IEEE Transactions on Wireless Communications*, vol. 20, no. 1, pp. 375–388, 2021.

[7] S. Mao, L. Liu, N. Zhang, M. Dong, J. Zhao, J. Wu, and V. C. M. Leung, "Reconfigurable intelligent surface-assisted secure mobile edge computing networks," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 6, pp. 6647–6660, 2022.

[8] B. Li, W. Wu, Y. Li, and W. Zhao, "Intelligent reflecting surface and artificial-noise-assisted secure transmission of MEC system," *IEEE Internet of Things Journal*, vol. 9, no. 13, pp. 11 477–11 488, 2022.

[9] Y. Liu, W. Wang, H.-H. Chen, F. Lyu, L. Wang, W. Meng, and X. Shen, "Physical layer security assisted computation offloading in intelligently connected vehicle networks," *IEEE Transactions on Wireless Communications*, vol. 20, no. 6, pp. 3555–3570, 2021.

[10] Y. Ju, Y. Chen, Z. Cao, H. Wang, L. Liu, Q. Pei, and N. Kumar, "Learning based and physical-layer assisted secure computation offloading in vehicular spectrum sharing networks," in *IEEE INFOCOM 2022 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2022.

[11] Y. Chen, Y. Wang, J. Zhang, and M. D. Renzo, "QoS-driven spectrum sharing for reconfigurable intelligent surfaces (RISs) aided vehicular networks," *IEEE Transactions on Wireless Communications*, vol. 20, no. 9, pp. 5969–5985, 2021.

[12] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," *arXiv preprint arXiv:1509.02971*, 2015.

[13] D. Krajzewicz, J. Erdmann, M. Behrisch, and L. Bieker, "Recent development and applications of SUMO-simulation of Urban MObility," *International journal on advances in systems and measurements*, vol. 5, no. 3&4, 2012.

[14] Z. Li, L. Xiang, X. Ge, G. Mao, and H.-C. Chao, "Latency and reliability of mmwave multi-hop V2V communications under relay selections," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 9, pp. 9807–9821, 2020.

[15] Y. Zhu, B. Mao, and N. Kato, "A dynamic task scheduling strategy for multi-access edge computing in IRS-aided vehicular networks," *IEEE Transactions on Emerging Topics in Computing (Early Access)*, 2022.