NOMA-Assisted Secure Offloading for Vehicular Edge Computing Networks with Asynchronous Deep Reinforcement Learning

Ying Ju, *Member, IEEE*, Zhiwei Cao, Yuchao Chen, *Graduate Student Member, IEEE*, Lei Liu, *Member, IEEE*, Qingqi Pei, *Senior Member, IEEE*, Shahid Mumtaz, *Senior Member, IEEE*, Mianxiong Dong, *Senior Member, IEEE*, and Mohsen Guizani, *Fellow, IEEE*

***Abstract*—Mobile edge computing (MEC) offers promising solutions for various delay-sensitive vehicular applications by providing high-speed computing services for a large number of user vehicles simultaneously. In this paper, we investigate non-orthogonal multiple access (NOMA) assisted secure offloading for vehicular edge computing (VEC) networks in the presence of multiple malicious eavesdropper vehicles. To secure the wireless offloading from the user vehicles to the MEC server at the base station, the physical layer security (PLS) technology is leveraged, where a group of jammer vehicles is scheduled to form a NOMA cluster with each user vehicle for providing jamming signals to the eavesdropper vehicles while not interfering with the legitimate offloading of the user vehicle. We formulate a joint optimization of the transmit power, the computation resource allocation and the selection of jammer vehicles in each NOMA cluster, with the objective of minimizing the system energy consumption while subjecting to the computation delay constraint. Due to the dynamic characteristics of the wireless fading channel and the high mobility of the vehicles, the joint optimization is formulated as a Markov decision process (MDP). Therefore, we propose an asynchronous advantage actor-critic (A3C) learning algorithm-based energy-efficiency secure offloading (EESO) scheme to solve the MDP problem. Simulation results demonstrate that the agent adopting the A3C-based EESO scheme can rapidly adapt to the highly dynamic VEC networks and improve the system energy efficiency on the premise of ensuring offloading information security and low computation delay.**

***Index Terms*—Vehicular edge computing, physical layer security, non-orthogonal multiple access (NOMA), secure offloading, deep reinforcement learning.**

I. Introduction

A. Background

The great progress of vehicular networks with their wide applications in recent years has motivated a large amount of data-hungry and delay-sensitive services, such as crash warning, traffic flow prediction and unmanned driving. Various computation-intensive tasks are generated by these services and require considerable communication and computation resources to process. Recently, mobile edge computing (MEC) has become a promising technology in the Internet of Vehicles (IoV) by deploying edge computing servers with sufficient computation resources at the edge of the networks, such as base stations or roadside units, that can address the task computation requirements. With the deployment of MEC in vehicular networks, multiple vehicles can offload their tasks to the MEC servers simultaneously and acquire high-speed computing services. Accordingly, the great potential of MEC has drawn lots of attention [1], [2]. In this field of research, energy consumption is a key concern since the MEC approach involves the transmission energy consumption in the offloading process, which is absent in the traditional local computing methods. The authors in [3] minimize the energy consumption of a multi-cell MEC system by jointly optimizing the computation and radio resources. Under the constraint of task time delay, the authors in [4] propose an offloading strategy to reduce the energy consumption of the users. By optimizing the power allocation and offloading decision, the authors in [5] reduce energy consumption and meet the requirements of low system delay.

So far, researchers mainly focus on various energy consumption performances and do not consider the security of the communication process from the vehicle users to the edge server. Due to the open feature of the wireless networks, during the offloading process, malicious eavesdroppers intercept the data transmitted through the wireless channel, which leads to confidential information leakage [6], [7]. In the past, encryption methods were often used to prevent eavesdroppers from decoding confidential messages correctly [8]–[10]. However,

The work of Ying Ju, Zhiwei Cao, Yuchao Chen, Lei Liu, and Qingqi Pei was supported in part by the National Key Research and Development Program of China under Grant 2020YFB1807500, in part by the National Natural Science Foundation of China under Grants 62102301, 62001357 and 62132013, in part by the Guangdong Basic and Applied Basic Research Foundation under Grants 2020A1515110772 and 2020A1515110079, and in part by the Key Research and Development Programs of Shaanxi under Grant 2021ZDLGY06-03. The work of Mianxiong Dong was supported in part by the JSPS KAKENHI under Grants JP20H04174 and JP22K11989, in part by the Leading Initiative for Excellent Young Researchers (LEADER), MEXT, Japan, and in part by JST, PRESTO, Japan, under Grant JPMJPR21P3. (*Corresponding authors: Lei Liu and Ying Ju.*)

Ying Ju, Zhiwei Cao, Yuchao Chen, Lei Liu and Qingqi Pei are with the State Key Laboratory of Integrated Services Networks, School of Telecommunications Engineering, Xidian University, Xi'an 710071, China, and also with the Hangzhou Institute of Technology, Xidian University, Hangzhou 311231, China (e-mail: juyingtju@163.com; 21011210393@stu.xidian.edu.cn; 28521-6759@qq.com; leiliu@xidian.edu.cn; qqpei@mail.xidian.edu.cn).

Shahid Mumtaz is with the Department of Applied Informatics, Silesian University of Technology Akademicka, Gliwice 16 44-100, Poland, and also with the Departement of Engineering, Nottingham Trent University, Nottingham NG1 4FQ, UK (E-mail: shahid.mumtaz@ntu.ac.uk).

Mianxiong Dong is with the Department of Sciences and Informatics, Muroran Institute of Technology, Muroran 0508585, Japan (e-mail: mx.dong@csse.muroran-it.ac.jp).

Mohsen Guizani is with the Machine learning Department, Mohamed Bin Zayed University of Artificial Intelligence (MBZUAI), Abu Dhabi, UAE (e-mail: mguizani@ieee.org).

with the rapid development of the computing performance of the devices, it is difficult to protect the security of information only by relying on the key.

To address the security issues, many works have investigated the potential of physical layer security (PLS) techniques to deteriorate the reception of eavesdroppers in wireless networks during signal transmission phases, exploiting the inherent physical characteristics of the wireless channel and thus enhancing the secrecy performance of the networks [11]–[17]. There are also a few kinds of research on PLS in MEC networks [18]–[20]. The authors in [18] design an artificial noise (AN) assisted beamforming method to improve the secrecy performance of the computation offloading. The authors in [19] use full-duplex communication technique to propose the corresponding physical layer assists privacy-preserving offloading scheme to prevent information interception. In [20], the authors propose a PLS method to safeguard the offloading process to the MEC server and minimize the weighted sum-energy consumption for all users.

To further improve the confidentiality performance and reduce the system latency, it is also necessary to utilize suitable access strategies. Non-orthogonal multiple access (NOMA) has been identified as one of the key technologies to achieve high-frequency spectral efficiency and a large number of connections in future networks. In the NOMA scheme, multiple users can access the same frequency band and exploit the successive interference cancellation (SIC) to mitigate the co-channel interference. The inherent feature of multi-user superposition transmission will inevitably cause interference depending on the decoding order of multiple users. Therefore, with proper design, we can leverage the interference of the NOMA scheme as jamming signals to deteriorate the reception of the eavesdroppers. In this context, PLS in NOMA-assisted networks has attracted a lot of research interest [21]–[26]. In [21], the authors consider a single-input single-output NOMA system and maximize the secrecy sum rate. In [22], a minimum transmit power scheme, considering a multiple-input single-output NOMA cognitive radio (CR) network, is proposed. The authors in [23] use the inter-user interference scheme to confuse eavesdroppers while improving the signal quality of legitimate users. In [24], the remaining idle user is used as a friendly jammer to confuse the eavesdroppers in the uplink NOMA system and enhance the secrecy performance of the system. In [25], the authors secure a large-scale downlink system by using the AN scheme. The authors in [26] introduce a half-duplex decode-and-forward relay to safeguard the NOMA networks.

In addition, the NOMA scheme is also used to secure the MEC networks [27], [28]. The authors in [27] optimize the transmit power, the offloaded workload and the NOMA-transmission duration to enhance the security and reduce the power consumption of the MEC network. In [28], the authors investigate the cooperative mechanism between NOMA user pairs to improve the security of the MEC system. Nevertheless, the schemes in the above works are all designed in static scenarios where the locations of the users remain unchanged and cannot be directly applied in the dynamic vehicular networks.

Due to the high mobility of the vehicles, the algorithm should make fast decisions according to the current environmental state. However, the traditional intelligent optimization algorithm needs a certain amount of time to iterate and obtain the relatively optimal solution, which will cause an intolerable delay in the actual scene. Therefore, the joint optimization of security and energy consumption requires an algorithm that can obtain the optimal solution in a short time. Deep learning (DL) is one of the efficient methods to solve the problem [29], [30]. In [29], the authors propose a deep learning-based workload orchestrator for mult-access multi-tier vehicular edge computing (VEC) architectures to Orchestrate the dynamic and heterogeneous resources in the VEC systems. In [30], the authors propose a deep learning-based vehicle-to-everything (V2X) wireless channel prediction model using a long short-term memory (LSTM) network. Reinforcement learning (RL) is the other way to solve the problem [31], [32]. In [31], the authors use the actor-critic RL algorithm to provide non-interfering resources to vehicles before they enter the out-of-coverage area. In [32], the authors use the RL algorithm to minimize the system consumption cost and maximize the transaction throughput of the blockchain system. DL requires large amounts of labeled data and RL has a small action space and state space.

Compared to DL and RL algorithms, deep reinforcement learning (DRL) algorithm can have a huge action and state space. Besides, the DRL algorithm does not need labeled data for training, which is convenient for training a model. Therefore, the DRL scheme attracts more interest in task offloading [33]–[36]. Once the model is completely trained, it can meet the latency and reliability requirements by making a quick decision, which is robust in the highly mobile vehicular network [33], [34]. In [35], the authors propose an asynchronous advantage actor-critic (A3C) learning algorithm-based scheme to minimize the cost of the cloud service center (CSC). In [36], the authors use A3C based learning algorithm to maximize the system utility by making the optimal task and resource scheduling policy. All the studies above do not consider the security of information during the offloading process.

B. Motivations and Contributions

The importance of secure offloading and energy efficiency in VEC networks is demonstrated by all of the above studies. With the help of NOMA technology and resource allocation scheme, the user vehicles can improve their secrecy performance and reduce the overall energy consumption of the system to a lower level. However, due to the high dynamics of vehicles in VEC networks, it is difficult for traditional optimization algorithms to make real-time decisions according to the external environment. Therefore, we have the motivation to design a new DRL-based solution for the challenging dynamic NOMA-aided multi-user offloading scenario. To the best of our knowledge, the existing literature has not studied the scheme to reduce the overall energy consumption of the system while considering information security and computation delay with the NOMA technique in the VEC networks.

In this paper, we investigate a scenario where multiple users attempt to offload messages to the edge computing server when multiple eavesdroppers attempt to intercept them. Idle Vehicles are selected to form a NOMA cluster with the user vehicle to send jamming signals which can confuse the reception of the eavesdroppers. In addition to the security design, the energy efficiency is also maximized through a DRL-based approach in highly dynamic VEC networks. The main contributions of this work are summarized as follows.

1) Using NOMA and PLS techniques, we propose an energy-efficient secure offloading (EESO) scheme based on DRL in a multi-user offloading scenario. We minimize the overall energy consumption of the VEC network by jointly optimizing the transmit power of vehicles, the computation resource allocation and the NOMA cluster selection. The NOMA scheme is also designed to enhance the jamming signal strength received by the eavesdroppers, thus protecting the offloaded information from eavesdropping.
2) Considering the high dynamic characteristics of the VEC networks and high-dimensional systems with a large action space, we solve the joint optimization problem of the EESO scheme by adopting the A3C-based DRL algorithm, where two deep neural networks are deployed respectively in the critic and actor part. The on-policy A3C algorithm uses the Monte Carlo method to obtain an unbiased estimation of the current value function and hence promotes the stability of the decisions. The simulation results show that the agent adopting the EESO scheme adapts to the rapidly changing environment and improves energy efficiency on the premise of ensuring the information security.
3) To simulate a real VEC environment, we use queuing theory to model vehicle movements and consider a dynamic eavesdropping scenario, including multiple eavesdropper vehicles. In addition, we introduce the computation delay constraint in designing the EESO scheme, which can balance the energy consumption and the total task processing delay. By leveraging the powerful A3C-based decision-making capabilities, the agent can find an intelligent strategy that reduces the system delay while minimizing the energy consumption.

C. Organization of the Paper

The remainder of this paper is organized as follows. Section II illustrates the system model of the VEC networks. Section III describes the NOMA-assisted secure offloading scheme. In section IV, the details of the DRL-based EESO scheme are introduced. Simulation results are provided in Section V and a conclusion is drawn in section VI.

II. System Model

We consider a highly dynamic VEC network, as shown in Fig. 1. There are three types of vehicles in the network, namely user vehicles, jammer vehicles and eavesdropper vehicles. The user vehicles are denoted by $V_u = \{v_i^u, i = 1, 2, \cdots, N_u\}$, where N_u is the number of user vehicles. By using the cellular interface, the N_u user vehicles can offload the task to the base station. Jammer vehicles are denoted by $V_h = \{v_j^h, j = 1, 2, \cdots, N_h\}$, where N_h is the number of jammer vehicles. The user vehicles want to send the computation-intensive data to the MEC server for processing. The jammer vehicle performs NOMA pairing with the user vehicle, and the eavesdropper vehicles eavesdrop on the information transmitted from the user vehicle to the MEC server. In addition, we assume that there are multiple dynamic eavesdropper vehicles in the network. Eavesdropper vehicles are denoted by $V_e = \{v_n^e, n = 1, 2, \cdots, N_e\}$, where N_e is the number of eavesdropper vehicles. Besides, all transceivers use a single antenna to transmit information.

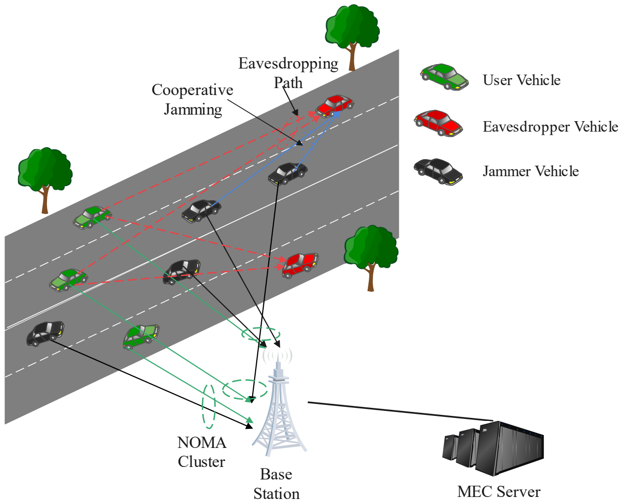


Fig. 1. Illustration of the VEC network model. Multiple user vehicles offload their computation tasks to the MEC server for execution.

TABLE I
NOTATIONS AND EXPLANATIONS

Notation	Explanation
N_u	Number of user vehicles
N_h	Number of jammer vehicles
N_e	Number of eavesdropper vehicles
λ	Average arrival rate
$g_{T,R}$	Channel gain
R_b^i	Channel capacity of user vehicle link
$R_{e,n}^i$	Channel capacity of eavesdropper vehicle
R_s^i	Secrecy rate of user vehicle
t_i^l	Local execution delay of the tasks
t_i^{exe}	Execution delay of the MEC server for computing tasks
t_i^{tr}	Secure transmission delay
E_i^{tr}	User vehicle transmission energy consumption
E_i^l	User vehicle local computing energy consumption
E_i^M	User vehicle MEC energy consumption
E_s	System energy consumption
T_s	System computation delay
P_s	Secrecy probability
r_e	Energy consumption-related reward
r_d	Computation delay-related reward
θ	Global actor network parameter
θ_c	Global critic network parameter
θ'	Actor network parameter in each worker thread
θ'_c	Critic network parameter in each worker thread

To practically simulate a dynamic traffic flow, we model the arrival process of the vehicles (including the user vehicles, jammer vehicles and eavesdropper vehicles) by utilizing Queuing Theory [37]. The time interval Δt between two arriving vehicles follows a negative exponential distribution. Therefore, the probability density function of the time interval Δt can be expressed as

$$f(\Delta t) = \begin{cases} \lambda e^{-\lambda \Delta t}, & \text{if } \Delta t \geq 0, \\ 0, & \text{Otherwise}, \end{cases} \tag{1}$$

where λ is the average arrival rate of vehicles. The different traffic flows can be simulated by adjustment of the value of λ.

In each time slot, multiple user vehicles simultaneously offload their computing tasks and select jammer vehicles for NOMA clusters. At this time, multiple eavesdroppers attempt to wiretap the transmission of user vehicles. Therefore, appropriate strategies should be designed to reduce information leakage. In this paper, in order to achieve secure offloading of all user vehicles and reduce the system energy consumption, we will jointly optimize the NOMA cluster selection, the transmit power and the computation resource allocation.

III. NOMA-Assisted Secure Offloading Scheme

In this section, we first propose a NOMA-assisted jamming scheme, followed by the computation task execution model for the VEC network. Then, the optimization problem to minimize the system energy consumption is formulated.

A. NOMA Jamming Scheme

Due to the open characteristics of the wireless channel, user vehicles of the offloading process to the base station risk eavesdropping. We consider a more practical eavesdropping scenario where the wireless offloading links are eavesdropped by multiple dynamic eavesdropper vehicles, compared to the traditional single eavesdropper scenario [18]–[20]. we adopt the NOMA technology in our EESO scheme to introduce jamming signals that can confuse the reception of eavesdroppers while not interfering with legitimate user vehicles. The idle vehicles in the network are used to play the role of jammer vehicles. We divide the user vehicles and jammer vehicles into N_u NOMA clusters. Each cluster contains one user vehicle and one or more jammer vehicles, resulting in no inter-user interference to the user vehicle. Different clusters use different frequency bands. This ensures that there is no inter interference in different NOMA clusters. In each cluster, the user vehicle and jammer vehicles share the same frequency band and leverage the NOMA scheme to transmit and decode messages. The mechanism of SIC decoding in the NOMA scheme determines that the signals of other users will not interfere with the last decoded user. Therefore, by properly designing the NOMA scheme, we can guarantee that the signals transmitted by the jammer vehicles will not interfere with the user vehicle in its own NOMA cluster. Thus, we set the user vehicle as the last decoded signal in each cluster at the base station. Different from the SIC reception of the base station, the eavesdroppers receive all the jamming signals generated by the jammer vehicles. Thus, the signal-to-interference-plus-noise ratio (SINR) at the eavesdroppers will significantly decrease.

During one coherence time period, the channel gain $g_{i,B}$ from the ith user vehicles to the base station can be expressed by

$$g_{i,B} = \alpha_{i,B} h_{i,B}, \tag{2}$$

where $\alpha_{i,B}$ is the large-scale fading effect, which represents the path loss. Besides, $h_{i,B}$ is the small-scale fading component and follows the exponential distribution with the unit mean. Similarly, $g_{i,n}$ represents the channel gain from the ith user vehicle to the nth eavesdropper vehicle, and $g_{j,n}$ represents the channel gain from the jth jammer vehicle to the nth eavesdropper vehicle.

The uplink NOMA allows an arbitrary decoding order. Besides, the base station can know which vehicle is the user vehicle and which one is the jammer vehicle. Therefore, we consider that the received signals of the ith NOMA cluster are decoded in the order of v_{i1}^h, v_{i2}^h, $\cdots$, $v_{iN_{hi}}^h$ and the user vehicle v_i^u finally. Therefore, the received SINR of the ith user vehicle can be described as

$$\xi_i^u = \frac{P_i^u \cdot g_{i,B}}{\sigma^2}, \tag{3}$$

where P_i^u is the transmit power of the ith user vehicle and σ^2 is the noise power. Then the channel capacity from the ith user vehicle to the base station can be formulated as

$$R_b^i = W \log(1 + \xi_i^u), \tag{4}$$

where W is the bandwidth.

We assume that the eavesdropper vehicles have strong eavesdropping capabilities and that all the information sent by user vehicles which they can eavesdrop. The reason for considering strong eavesdropping capabilities is that once the security of information is guaranteed, the offloading channel capacity is larger than the highest channel capacity of eavesdropper vehicles. Therefore, our proposed scheme can also guarantee the security of the information when one eavesdropper vehicle can only eavesdrop on one user vehicle. Then the channel capacity of the nth eavesdropper vehicle that is wiretapping the offloading information of the ith user vehicle can be expressed by

$$R_{e,n}^i = W \log(1 + \frac{P_i^u g_{i,n}}{\sigma^2 + I_{i,n}^e}), \tag{5}$$

where $I_{i,n}^e$ is the received interference, given by

$$I_{i,n}^e = \sum_{j=1}^{N_h} \rho_j[i] P_j^h g_{j,n}, \tag{6}$$

where the binary indicator $\rho_j[i]$ is exploited to indicate the selection of the NOMA cluster. $\rho_j[i] = 1$ represents that the jth jammer vehicle is selected by the ith user vehicle to transmit jamming signal in its NOMA cluster, while $\rho_j[i] = 0$ otherwise. Particularly, each jammer vehicle can be selected by only one user vehicle, i.e., $\sum_{i=1}^{N_u} \rho_j[i] \leq 1$.

To protect the wireless transmission of the offloading process, we leverage a physical layer wiretap coding scheme to encode the offloaded messages [38]. We define R_t and R_s as the codeword rate and the secrecy rate, and their difference $R_t - R_s$ describes the redundancy rate which is utilized to resist eavesdropping. If the redundancy rate exceeds the channel capacity of the eavesdropper vehicle, the decoding process at the eavesdropper vehicle can be completely confused, and the security is guaranteed [39]. Thus, the gap between the codeword rate and the channel capacity of the eavesdropper vehicles should be widened by an appropriate design. The base station can correctly decode the information sent by the user vehicles if the channel capacities of the user vehicles R_b^i are higher than the codeword rate. Therefore, to obtain the highest redundancy rate $R_t - R_s$ to confuse the eavesdropper vehicles, we set the codeword rate to its maximum value R_b^i, i.e., $R_t = R_b^i$.

There are two kinds of eavesdropping scenarios, namely the non-colluding eavesdropping and the colluding eavesdropping scenarios. In the non-colluding eavesdropper scenario, each eavesdropper vehicle individually decodes confidential messages, and the eavesdropper vehicle that has the best performance of the received signal is considered. In the colluding eavesdropper scenario, multiple eavesdropper vehicles adopt the maximal ratio combining to process the wiretapped confidential information. The expressions of these two eavesdropping scenarios are only different in the expression for the eavesdropping channel. This paper mainly focuses on the non-colluding eavesdropping scenario, where the Eve with the best performance of the received signal is considered [40], [41]. Therefore, the secrecy rate of the ith user vehicle can be given by

$$R_s^i = \max\{0, R_b^i - \max_{V_e} R_{e,n}^i\}. \tag{7}$$

When the offloading task at the ith user vehicle is B_i bits, the delay in the transmission to the base station securely can be given by

$$t_i^{tr} = \frac{B_i}{R_s^i}. \tag{8}$$

The challenge is how to design an efficient scheme that uses the NOMA technique to offload the computation task to the MEC server. To overcome this challenge, we design the following variables. The variable $\mathbf{X} = \{\rho_j[i] \mid i = 1, 2, \cdots, N_u, j = 1, 2, \cdots, N_h\}$ is used to represent the selection of the jammer vehicles by the user vehicle. We also select the transmit power level for the user vehicles and jammer vehicles. The variable $\mathbf{Y} = \{y_i^u[m], y_j^h[m] \mid i = 1, 2, \cdots, N_u, j = 1, 2, \cdots, N_h, m = 1, 2, \cdots, N_p\}$ is used to represent the transmit power selection. The transmit power is limited to N_p levels, which is given by $\{P_m \mid m = 1, 2, \cdots, N_p\}$. $y_i^u[m] = 1$ represents that the mth transmit power level P_m is selected by the ith user vehicle, while $y_i^u[m] = 0$ otherwise. $y_j^h[m] = 1$ represents that the mth transmit power level P_m is selected by the jth jammer vehicle, while $y_j^h[m] = 0$ otherwise. Then we can obtain the transmit power of the ith user vehicle and the jth jammer vehicle, which are given by $P_i^u = \sum_{m=1}^{N_p} y_i^u[m] P_m$ and $P_j^h = \sum_{m=1}^{N_p} y_j^h[m] P_m$, respectively.

Therefore, the transmission energy consumption of the ith user vehicle can be given by

$$E_i^{tr} = t_i^{tr} P_i^u. \tag{9}$$

Besides, the transmission energy consumption of the ith user vehicle with NOMA pairing to the jammer vehicles can be expressed as

$$E_i^J = t_i^{tr} \sum_{j=1}^{N_h} \rho_j[i] P_j^h. \tag{10}$$

B. Computation Task Execution Model

The user vehicles generate computation tasks and then execute them locally or offload them to the MEC server to get executed remotely. Thus, these two cases will be described separately.

1) Local Execution Model: In this case, the delay-sensitive tasks are executed locally. With the local computation capacity $f_{L,i}$ (in cycles/s) of the ith user vehicle, the local execution time can expressed as

$$t_i^l = \frac{\mu B_i}{f_{L,i}}, \tag{11}$$

where μ (in cycles/bit) represents the computation intensity of communication processing on processors. Therefore, the energy consumption for local computing can be described as

$$E_i^l = \eta_1 f_{L,i}^3 t_i^l, \tag{12}$$

where η_1 is the capacitance coefficient of the local computation central processing unit (CPU).

2) Remote Execution Model: In this case, the delay-sensitive tasks are sent to the MEC server and executed by it. The total delay in this process includes the time cost of offloading tasks to the MEC server and the time cost of execution by the MEC server. The former one is defined in (11).

The maximum computation capacity of the MEC server is assumed to be F_M^{max} (in cycles/s). To compute the delay-sensitive tasks quickly, the computation resource is divided into N_b parts with each resource block having a different size, i.e., $\{f_k \mid k = 1, 2, \cdots, N_b, i = 1, 2, \cdots, N_u\}$. Then, the execution time can be calculated by

$$t_i^{exe} = \frac{\mu B_i}{\sum_{k=1}^{N_b} z_i[k] \cdot f_k}, \tag{13}$$

where $z_i[k]$ is a binary indicator to represent the selection of computation resource. $z_i[k] = 1$ means that the ith user vehicle chooses the kth computation resource block and $z_i[k] = 0$ otherwise. We denote the variable $\mathbf{Z} = \{z_i[k] \mid i = 1, 2, \cdots, N_u, k = 1, 2, \cdots, N_b\}$.

Hence, the total delay time when the delay-sensitive tasks are offloaded to the MEC server can be expressed as

$$t_i^{total} = t_i^{tr} + t_i^{exe}. \tag{14}$$

The energy consumption of the ith user vehicle for MEC can be expressed as

$$E_i^M = \eta_2 f_{i,k}^3 t_i^{exe}, \tag{15}$$

where η_2 is the capacitance coefficient of the CPU at the MEC server. Accordingly, the total energy consumption for the ith user vehicle offloading the tasks to the MEC server can be obtained by

$$E_i^{total} = E_i^M + E_i^{tr} + E_i^J. \tag{16}$$

In this paper, we further consider specific scenarios where the eavesdropper vehicle has a higher SINR than the base station. If the secrecy rate equals zero, the task cannot be offloaded to the base station securely. Besides, if the secrecy rate is low, resulting in the remote execution consuming more energy than the local execution, the task will be executed locally. Then, the task energy consumption of the ith user vehicle can be expressed as

$$E_i = \min\{E_i^{total}, E_i^l\}. \tag{17}$$

C. Performance Metrics

We focus on minimizing the energy consumed by the system while guaranteeing the offloading tasks to the MEC server securely. Several significant performance metrics are used to evaluate the various performance of the NOMA-assisted VEC network. They are discussed in the following.

1) System Energy consumption: In our VEC network, multiple user vehicles offload their tasks simultaneously. Thus the sum energy consumption of all user vehicles is a significant performance metric to measure the overall energy cost of the vehicular system. We formulate the system energy consumption of the VEC network as $E_s = \sum_{i=i}^{N_u} E_i$.

2) System Computation Delay: In the multi-user offloading scenario, one of the key performance metrics is the system computation delay, which is the maximum computation delay among all user vehicles and can be represented by $T_s = \max_{(i)}\{t_i^{exe}\}$.

3) Secrecy Probability: As the secrecy rate depicts the capacity difference between the legitimate and eavesdropping channels, it represents the maximum rate at which confidential messages can be transmitted in perfect secrecy. The channel gain in secrecy rate R_s consists of a small-scale component part and a large-scale fading effect part, where the small-scale component is a random variable. Due to varying channel states, there may be a secrecy outage if the secrecy rate is smaller than the secrecy threshold. The probability that such events occur is described as secrecy outage probability denoting as P_{out}, and the secrecy probability can be expressed as $P_s = 1 - P_{out}$. Then, the secrecy probability can be described as

$$P_s = \mathbb{P}\{R_s > \beta_s\}, \tag{18}$$

where β_s is the secrecy rate threshold.

D. Joint Optimization Problem

In this paper, our objective is to minimize the energy consumed by the system while ensuring the effectiveness of the NOMA-assisted jamming scheme. We jointly optimize the NOMA cluster selection $\mathbf{X}$, the transmit power $\mathbf{Y}$ and the computation resource allocation $\mathbf{Z}$. Then, the decision-making problem (DMP) is formulated as

$$\text{DMP}: \min_{\{\mathbf{X},\mathbf{Y},\mathbf{Z}\}} \sum_{i=1}^{N_u} E_i, \tag{19}$$

$$\text{Subject to}: \text{C1}: \sum_{i=1}^{N_u}\sum_{k=1}^{N_b} z_i[k] f_k \le F_M^{max}, \tag{19a}$$

$$\text{C2}: \sum_{i=1}^{N_u} \rho_j[i] \le 1, \forall j = 1, 2, \cdots, N_h, \tag{19b}$$

$$\text{C3}: \sum_{m=1}^{N_P} y_i^u[m] = 1, \forall i = 1, 2, \cdots, N_u, \tag{19c}$$

$$\text{C4}: \sum_{m=1}^{N_P} y_j^h[m] = 1, \forall j = 1, 2, \cdots, N_h, \tag{19d}$$

$$\text{C5}: \sum_{k=1}^{N_b} z_i[k] = 1, \forall i = 1, 2, \cdots, N_u, \tag{19e}$$

$$\text{C6}: \max_i t_i^{exe} \le T, \tag{19f}$$

$$\text{C7}: \rho_j[i], y_i^u[m], y_j^h[m], z_i[k] \in \{0, 1\}. \tag{19g}$$

The constraint (19a) indicates that all assigned computation resources at the MEC server should not exceed the maximum CPU-cycle frequency F_M^{max}. In the constraint (19b), $\sum_{i=1}^{N_u} \rho_j[i] = 1$ limits that one jammer vehicle can only choose one user vehicle to be a NOMA pairing. Besides, $\sum_{i=1}^{N_u} \rho_j[i] = 0$ shows that the jammer vehicle does not choose a user vehicle to be paired. The constraints (19c) and (19d) indicate that only one transmit power level can be selected by a user or jammer vehicle. The constraint (19e) limits that only one computation resource block can be selected by a user vehicle. The constraint (19f) limits the maximum computation delay of user vehicles. This constraint is to prevent the user vehicle from waiting too long due to excessive attention to the energy that the system consumes, and thus the experience of the user vehicle is degraded. The constraint (19g) shows that the decision-making variables are all binary indicators.

There are several challenges in solving the joint optimization problem in multi-user offloading scenarios:

1) The rapid movement of vehicles results in quick channel changes, making it difficult to make effective real-time decisions. An effective and fast decision should be made to securely offload the information to the base station and complete the task processing in the MEC server.
2) It is necessary to consider that the processing delay of the system should be smaller than the maximum tolerated delay of the user vehicles while reducing energy consumption. For example, larger tasks prefer to choose a lower computation speed resource block, which can reduce energy consumption. Therefore, how to coordinate the strategy to decrease the energy consumed by the system while meeting the delay constraint is a challenging issue.

Considering the above problems, we design a DRL-based EESO scheme to realize an energy-efficient secure offloading with proper resource allocation and NOMA scheme design in

VEC networks. This scheme will be elaborated in the next section.

IV. DRL-based EESO Scheme

DRL combines DL with RL and makes further development of both of them. However, there are certain differences between the RL and the DL. The RL is mainly used to make decisions. It can constantly update its decisions from the input complex environmental state, make different actions, get higher rewards and finally get an optimal strategy to solve the objective function. The DL mainly uses a multi-layer neural network to fit the objective function. It iteratively updates the parameters of the neural network through a large number of prepared data and finally gets a model that fits the objective function. A Single RL can only solve some relatively simple decision-making problems. The combination of both schemes gives rise to a new and more efficient technique called the "DRL". It has the advantages of both decision-making and deep neural network, so it can solve more complex decision-making problems.

Since the objective to minimize the system energy consumption, including the energy of vehicle transmission and the edge computation, is non-convex and non-linear, it is an exponentially difficult problem (NP-hard). This kind of problem can be solved by an intelligent optimization algorithm, but the solution generally takes a relatively long time. By the time the algorithm is finished, the vehicle position and channel information will have changed significantly. Therefore, the dynamic decision of our problem cannot be realized by using traditional intelligent optimization algorithms. On the contrary, the DRL can make dynamic decisions in a short time. It only needs to constantly interact with the environment, take actions in the training stage and constantly change its own decisions according to the designed rewards. Eventually, it can get a general decision model. When the training is finished, it can load the model to make a better decision in a short time.

The Deep Q-learning (DQN) algorithm is a classic deep reinforcement learning algorithm. However, in high-dimensional systems with a large action space, the DQN algorithm suffers from sample complexity, leading to a slow convergence and local maxima [42]. Therefore, we use the A3C algorithm to improve the convergence when facing a large action space. The A3C algorithm combines two parts, including the value and action part, which can solve the problem of minimizing energy optimization in the previous section better. The A3C algorithm belongs to the on-policy algorithms. The on-policy algorithms use the Monte Carlo method to obtain an unbiased estimation of the current value function $V(s)$ to improve stability. The A3C algorithm has a high stability and its stochastic strategy is more robust to external disturbances. Therefore, the A3C algorithm has a strong generalization ability [43]. Besides, the A3C algorithm discards the previous samples after each update of the network model. Therefore, its sample utilization rate is low. Our proposed scheme considers both the system energy consumption reward and the computational delay reward. We balance the two rewards to minimize the system energy consumption and limit the computation delay in a tolerable range. Besides, we use Adam Optimizer to normalize the parameter updates so that each parameter update has a similar magnitude, thus improving training.

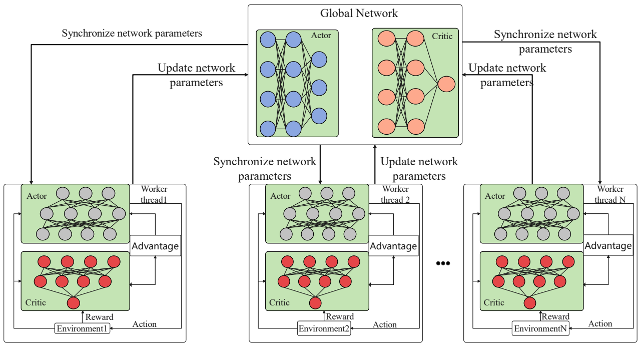


Fig. 2. A3C-based EESO scheme for VEC networks.

In the EESO scheme, the NOMA offloading scenario in Fig. 1 is modeled as an environment and the base station is modeled as the agent. We can model the optimization problem as a Markov Decision Process (MDP). The MDP can be expressed as a tuple $\{S, A, P_a, R\}$. S is the set of all possible state environments. A is the action space of all possible actions that the agent can take, which is the guidance of state transition. $P_a = \{p(s_{t+1}|s_t, a_t)\}$ stands for the set of transition probability. R is the reward for the agent which depends on the state and action.

A. State Space

The state s of the agent includes the wireless channel information, the NOMA jamming information and the task information, which are described in detail as follows.

1) Channel Information: the channel information includes the channel gain $g_{i,B}$ from the user vehicle to the base station, the channel gain $g_{j,B}$ from the jammer vehicle to the base station and the channel gain $g_{j,n}$ from the jammer vehicle to the eavesdropper vehicle. Therefore, we have the channel information

$$\mathcal{G} = \{g_{i,B}, g_{j,B}, g_{j,n} \mid i = 1, 2, \cdots, N_u, j = 1, 2, \cdots, N_h, \\ n = 1, 2, \cdots, N_e\}. \tag{20}$$

2) NOMA Jamming Information: The eavesdropper vehicles receive interference from the jammer vehicles. Therefore, the jamming information is introduced into the current state of the agent, which can be described as

$$\mathcal{I} = \{I^e_{i,n} \mid i = 1, 2, \cdots, N_u, n = 1, 2, \cdots, N_e\}. \tag{21}$$

3) Computation Task Information: The computation task B_i is introduced into the state space, which helps more reasonable decisions made by the agent to decrease the system energy consumption. The state of the computation task can be described as

$$\mathcal{B} = \{B_i \mid i = 1, 2, \cdots, N_u\}. \tag{22}$$

Therefore, the state of the agent can be described as

$$s = \{\mathcal{G}, \mathcal{I}, \mathcal{B}\}. \tag{23}$$

B. Action Space

The agent of each worker thread selects the transmit power, the computation resource block and the NOMA cluster according to the current state. Thus, the action space $\mathbf{A}$ has three main categories, including the NOMA cluster selection $\mathbf{X}$ (the jammer vehicle selects which NOMA cluster to send jamming signals to), the power selection of the user vehicles and jammer vehicles $\mathbf{Y}$, and the MEC server resource selection $\mathbf{Z}$. The output of the neural network is an action selection probability, and the current action can be obtained by random sampling according to this probability distribution.

The selection of the MEC resource block is obtained by preprocessing the resource block selection set $\mathbf{Z}'$. The resource block selection set $\mathbf{Z}'$ is obtained by all possible selections of MEC resources for user vehicle. The selection of a user vehicle to use a duplicate resource block causes other computing resource blocks to be idle and generates additional waiting latency. Therefore, we remove the same resource selection action from $\mathbf{Z}'$ and get the present MEC resource block selection set $\mathbf{Z}$.

C. Reward Design

The reward design is related to the model convergence. A proper reward design can greatly accelerate the convergence. In our scheme, the reward design includes the following two parts.

1) Energy Consumption-Related Reward: In order to achieve the objective in (19) and minimize the system energy consumption, we design an energy consumption-related reward, which is expressed as

$$r_e = \begin{cases} \zeta_1, & \text{if } E_s \leq e_1, \\ \frac{\zeta_1(e_2-E_s)}{e_2-e_1}, & \text{if } e_1 < E_s \leq e_2, \\ 0, & \text{otherwise}, \end{cases} \tag{24}$$

where ζ_1 is designed as a positive reward to encourage the computation task to be completed below a given energy consumption level. E_s is the system energy consumption, which can be given by $E_s = \sum_{i=1}^{N_u} E_i$. Besides, e_1 and e_2 are energy thresholds.

2) Delay-Related Reward: In the VEC network, the tasks are often delay-sensitive and the computation delay needs to be kept at a low level. Therefore, a computation delay limit is designed to prevent the agent from paying too much attention to the energy consumption and ignoring the computation delay, resulting in a large computation delay. Then, a delay-related reward is designed as

$$r_d = \begin{cases} \zeta_2, & \text{if } T_s \leq T, \\ 0, & \text{otherwise}, \end{cases} \tag{25}$$

where T is the maximum tolerance computation delay threshold. Once the computation delay is no larger than the threshold, the agent will get a positive reward ζ_2.

Thus, with the objective described earlier, the return reward r_t of the agent can be obtained by

$$r_t = \omega_1 r_e + \omega_2 r_d, \tag{26}$$

where ω_1 and ω_2 are the positive weights that balance different constraints and objectives.

D. A3C-based Learning Algorithm

The DRL is learning through the continuous interaction of the agent with the environment and adjusting its behavior through its reward. The A3C learning algorithm is an improvement of the Actor-Critic (AC) algorithm. Compared to the AC algorithm, the A3C algorithm allows multiple worker threads to interact with the environment and asynchronously trains the worker threads. The interaction between each worker thread and the environment is independent, which means that the correlation is weak and there is no interference. When the maximum operation index or terminal state is reached, each worker thread computes its gradient and sends it to the global network. To ensure that each worker thread can share the same policy, the global network updates the global parameters and distributes them to each worker thread.

In this way, training duration consumes less time for convergence due to its updating method. Besides, because of its policy characteristics, A3C can solve problems that have huge states or action spaces. Then, we will discuss our A3C-based EESO scheme in detail.

When the environment state of the VEC network is s_t, the estimated state value is $V(s_t, \theta_c)$. The agent chooses an action a_t according to the policy $\pi(a_t|s_t;\theta)$ and gets a reward r_t. θ_c and θ are the parameters of the global critic and actor network. The state value function of A3C is expressed as

$$V(s_t, \theta_c) = E(G_t|s_t, \pi), \tag{27}$$

where $G_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k}$ is the discounted accumulated reward with $\gamma \in [0, 1]$ describing the discount factor. When considering n steps, the Q value function $Q(s_t, a_t)$ is given by

$$Q(s_t, a_t) = \sum_{k=0}^{n-1} \gamma^k r_{t+k} + \gamma^n V(s_{t+n}, \theta_c). \tag{28}$$

To decrease the variance of the estimation, A3C uses the advantage function $A(s_t, a_t)$ under the environment s_t and the action a_t, which is expressed as

$$A(s_t, a_t) = Q(s_t, a_t) - V(s_t, \theta_c). \tag{29}$$

The agent reduces the occurrence of overestimation or underestimation and improves its learning ability by using the advantage function.

The loss function of the actor network in each worker thread is given by

$$\begin{aligned} loss_{actor} = & \log \pi(a_t|s_t, \theta')(Q(s_t, a_t) - V(s_t, \theta'_c)) \\ & + \beta H\pi(s_t, \theta'), \end{aligned} \tag{30}$$

where θ'_c and θ' are the parameters of the critic and actor network in the worker thread. Besides, $H\pi(s_t, \theta')$ is an entropy item that is used to explore more possible actions instead of paying attention only to a few actions. β is a weight parameter used for the action entropy item. The loss function of critic network is given by

$$loss_{critic} = (Q(s_t, a_t) - V(s_t, \theta'_c))^2, \tag{31}$$

Algorithm 1 A3C-based EESO scheme

```
1: Initialize : The global actor network parameters $\theta$ and critic network parameters $\theta_c$ .
2: Initialize : The actor network parameters $\theta'$ and critic network parameters $\theta'_c$ in each worker thread.
3: Initialize : The global maximum number of iteration $N_{ep}^{\max}$ epochs and maximum length of single iteration $t_{worker}$ in each worker thread.
4: Set global counter epoch= $0$, worker thread step= $1$.
5: Set number of worker threads $N_w$.
6: for epoch= $0$ to $N_{ep}^{\max}$ do
7:   for $n = 1$ to $N_w$ do
8:     Reset the global network gradients: $d\theta \leftarrow 0$, $d\theta_c \leftarrow 0$.
9:     Synchronous parameters of each worker thread with global parameters: $\theta' = \theta$, $\theta'_c = \theta_c$.
10:    Reset $t_{start} = t$ and obtain the initial VEC network state $s_t$, including the channel information, the NOMA jamming information and the computation task information.
11:    Perform $a_t$ according to policy $\pi(a_t|s_t;\theta)$.
12:    Execute action $a_t$, obtain reward $r_t$ and next new state$s_{t+1}$.
13:    $t \leftarrow t + 1$.
14:    if $t - t_{start} = t_{worker}$ then
15:      Calculate the $Q(s_t, a_t)$ of the last time series state $s_t$.
16:    else
         Return Step 11.
17:    end if
18:    for $i = t - 1$ to $t_{start}$ do
19:      Calculate the Q value by (28).
20:      Update actor network gradient by (32).
21:      Update critic network gradient by (33).
22:    end for
23:  end for
24: end for
```

Therefore, the gradient of actor network and critic network is updated to

$$\begin{aligned} d\theta \leftarrow d\theta + \nabla_\theta \{ & \log \pi(a_t|s_t, \theta')(Q(s_t, a_t) - V(s_t, \theta'_c)) \\ & + \beta H \pi(s_t, \theta')\}, \end{aligned} \tag{32}$$

$$d\theta_c \leftarrow d\theta_c + \frac{\partial \{Q(s_t, a_t) - V(s_t, \theta'_c)\}^2}{\partial \theta'_c}. \tag{33}$$

The agent interacts with the unknown environment and learn from it. Then, after fully training the A3C model, it can solve the problem in the environment state that has never appeared before. Therefore, it can reduce the overhead after full training.

V. Numerical Results

In this section, the proposed A3C-based EESO scheme for the VEC networks is evaluated through simulations. The simulation environment and relevant parameters are set up in the following, which are selected according to the 3GPP technical specifications. Then, the benchmark algorithms and performance metrics are given. Finally, numerical experiments demonstrate that the proposed scheme is robust and effective.

A. Simulation Setup

TABLE II
SIMULATION PARAMETERS

Parameter	Value
Number of user vehicle N_u	2
Number of Jammer vehicle N_h	2
Number of eavesdropper vehicle N_e	2
Base station antenna height (m)	25
Vehicle antenna height (m)	1.5
Noise power σ^2 (dBm)	-114
Carrier frequency (GHz)	2
Bandwidth of sub-band W (MHz)	1
Task data size B (Mbit)	[0.5, 2]
Computation time limit T (s)	0.21
Speed of vehicles (km/h)	72
Arrival rate	0.5
Antenna gain of the base station (dBi)	10
Antenna gain of vehicles (dBi)	3
CPU cycles required to process a bit μ (cycles/bit)	1000
Maximum computation capacity of user vehicle F_L^{max} (GHz)	2
Maximum computation capacity of MEC server F_M^{max} (GHz)	30
Coefficient depending on local chip architecture η_1 (watts/s^3)	10^{-27}
Coefficient depending on MEC chip architecture η_2 (watts/s^3)	10^{-29}
Transmit power of user vehicle and jammer vehicle (dBm)	[10,5,1,0]
Computation resource block (GHz)	[8,10,12]

TABLE III
CHANNEL PARAMETERS [44]

Parameter	V2I Link	V2V Link
Path loss model	$128.1 + 37.6 \log_{10}(d)$	LOS in WINNER + B1 Manhattan
Shadowing distribution	Log-normal	Log-normal
Shadowing standard deviation	8 dB	3 dB
Transmit power	[10, 5, 1, 0] dBm	10 dBm

In the simulations, the vehicles drive on the four-lane road in two opposite directions and arrive at the road with the arrival rate λ [45]. The MEC server is deployed near the base station. We set the simulator parameters according to 3GPP [46], [47], including the channel model, traffic model, vehicle model, antenna model, etc. The More detailed simulation parameters are shown in Table II and Table III. In the A3C algorithm, we use three fully connected layers and the Adam adaptive optimizer. In addition, the discount coefficient is 0.99, the action entropy weight is 0.1, the number of threads is 2, and the CPU used is i7-10700.

We use the deep neural network for modeling the A3C. The A3C consists of actor network and critic network. The structures of actor network and critic network are not the same. Both two networks consist of one input layer, three fully connected hidden layers, and one output layer. The actor network's dimensions of the input layer are related to the dimensions of the state space in Sec. IV-A and the dimensions of the output layer are related to the dimensions of the action space in Sec. IV-B. The actor network has three hidden layers containing 256, 512, and 2048 neurons, respectively. As for the critic network, its dimension of the output layer is one and has three hidden layers containing 128, 256, and 32 neurons, respectively. Besides, we use Adam optimizer and the learning rate is 0.001.

B. Benchmark schemes and Metrics

In order to conduct a comprehensive performance analysis, the A3C-based EESO scheme will be compared with other benchmark schemes. The comparison schemes are as follows.

- **Optimal scheme**. The scheme traverses all possible combinations of the computation resource block selection, transmit power selection and NOMA cluster selection. Then, it selects the optimal decision to achieve the minimum energy consumed by the system. The performance of non-adaptive and greedy solutions is demonstrated using this scheme.
- **Without NOMA-Jamming (WNJ) scheme**. In this scheme, there is no jammer vehicles working in the NOMA mode to deteriorate the reception of the eavesdropper vehicles. The WNJ scheme is used to prove that the architecture using the NOMA cluster can decrease the channel capacities of the eavesdropper vehicles and enhance the security of VEC networks.
- **DQN scheme**. DQN is another kind of DRL method. This scheme is used to show the advantage of our A3C-based learning scheme.

To fully evaluate the performance of our proposed scheme, three important performance indicators are exploited, namely **the system energy consumption**, **the system computation delay** and **the secrecy probability**. More details of the numerical results are discussed below.

C. Numerical Evaluation

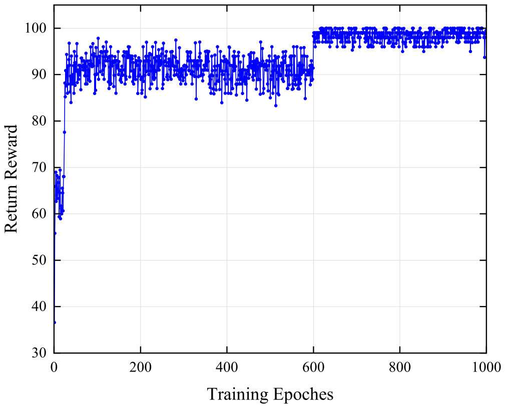


Fig. 3. Cumulative reward for each training epoch.

Fig. 3 depicts the relation between the training iterations and the cumulative reward, which directly demonstrates the convergence of the EESO scheme. As the number of iterations increases, the agent learns from different training sets, and the cumulative reward gradually increases over the first 600 iterations. After that, although there are certain fluctuations, the cumulative reward is basically greater than 95. Based on this, we further verify the performance of our proposed EESO scheme in the VEC network.

To prove the effectiveness of the EESO scheme, we compare the energy consumption performance of the EESO scheme with optimal, WNJ and DQN scheme in Fig. 4. We randomly generate ten different traffic patterns. Then the energy consumption of the system in each traffic pattern under different schemes is calculated. Obviously, the DQN scheme and WNJ scheme consume much more energy than the EESO scheme. Besides, Compared to the other two schemes, the gap between the EESO scheme and the Optimal scheme is small. In our simulation, we find that without the help of the jammer vehicle, the WNJ scheme cannot always guarantee a positive secrecy rate, so the information cannot be securely offloaded. Thus the local computation mode is often selected in the WNJ scheme, which causes longer computation time and higher energy consumption.

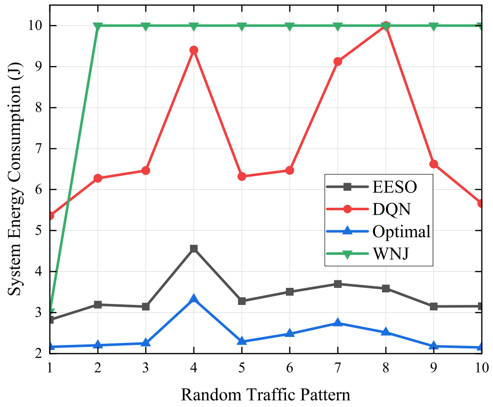


Fig. 4. System energy consumption performance under different schemes.

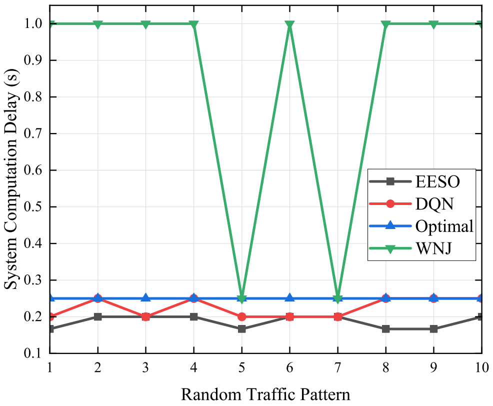


Fig. 5. System computation delay under different schemes.

Fig. 5 compares the system computation delay T_s, i.e., the maximum computation delay of all user vehicles, under different schemes. The computation delay of the Optimal scheme and the WNJ scheme is larger than that of the EESO scheme. Besides, computation delays of the EESO scheme in ten random traffic patterns are all less than the delay constraint of 0.21 seconds. Only the third, sixth and seventh points are the same as that of the DQN scheme. In other traffic patterns, the computation delay of the EESO scheme performs better than the other three schemes. It shows that the EESO scheme does not unduly sacrifice the delay performance when making decisions to minimize the energy consumption of the system, while the Optimal scheme always sacrifices the delay to achieve a smaller energy consumption. This is because we introduce a delay-related reward in the design of the EESO scheme. Under most traffic patterns, the computation delay of the WNJ scheme is kept at a particularly high level of 1 second. This reveals the superiority of our NOMA-assisted scheme in increasing the secrecy rate. In the absence of well-designed cooperative jamming, the WNJ scheme will fall into the local execution mode with a very large computation delay.

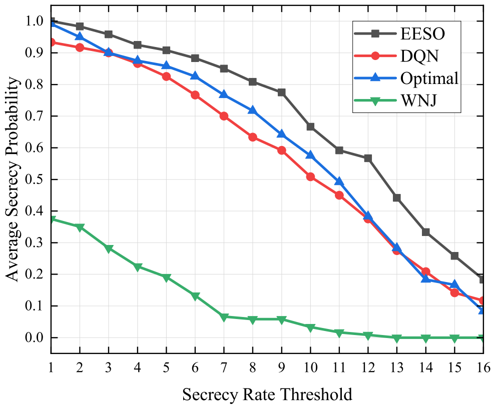


Fig. 6. Average secrecy probability under different secrecy rate thresholds.

Fig. 6 compares the average secrecy probability under four different schemes with different secrecy rate thresholds. As shown in the figure, as the secrecy rate threshold is continuously increased, there is a decline in the EESO scheme, the DQN scheme and the Optimal scheme are relatively slow before it is less than 6, and only the WNJ scheme drops rapidly. This is because more information can be intercepted by the eavesdropper vehicles in the WNJ scheme due to the lack of cooperative jamming. In the whole range of the secrecy rate threshold, the average secrecy probability of the EESO scheme is better than that of the DQN, optimal and WNJ schemes. This is because they mainly focus on how to make decisions to minimize the overall energy consumption of the system, which leads to the smaller transmit power they choose. As a result, the gap between the channel capacities of the user vehicle and the eavesdropper vehicles will decrease, which depresses the secrecy rate. In addition to the energy consumption, our proposed EESO scheme also takes the constraint of the computation delay into account, which leads to a rise in the transmit power and thus increases the secrecy rate.

Fig. 7 describes the system processing rate of ten different traffic patterns. The system processing rate is defined as the ratio of the total data volume of all user vehicles to the system processing delay, i.e., $R_p = \sum_i B_i / t_i^{total}$. We find that the system processing rate of the EESO scheme is better than the other three schemes. This is mainly because the EESO scheme adds a limit to the computation delay, which reduces the computation delay. Moreover, due to the reasons explained in Fig. 6 above, the EESO scheme exhibits the best secrecy performance, which results in the transmission delay being smaller than the other three schemes. Therefore, the system processing rate of the EESO scheme performs better. On the contrary, the secrecy rate of the WNJ scheme has the worst performance among all schemes because it lacks the help of the jammer, which results in the lowest system processing rate.

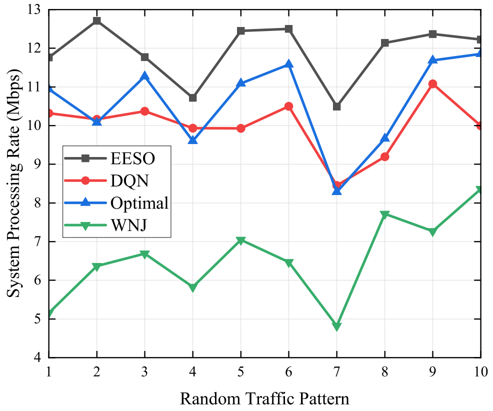
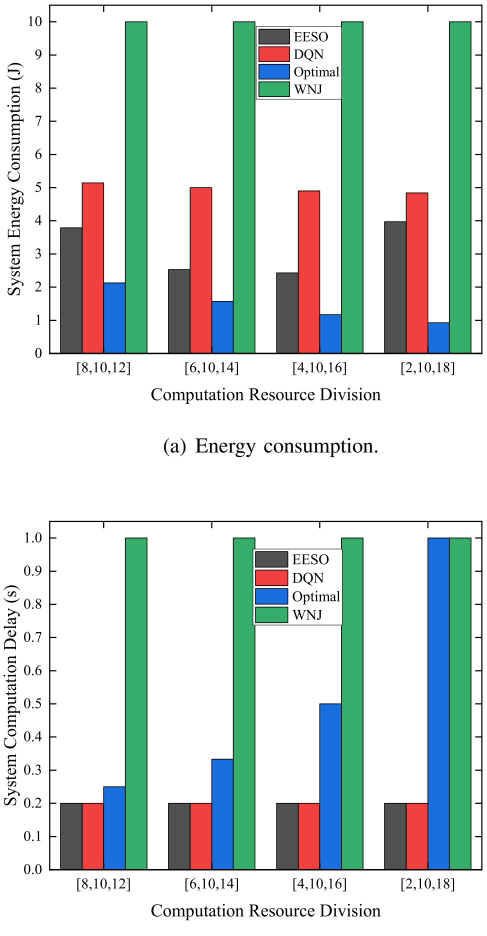


Fig. 7. System processing rate of the VEC network under different schemes.



(a) Energy consumption.

(b) Computation delay.

Fig. 8. Energy consumption and computation delay performance versus computation resource division.

Fig. 8 shows the energy consumption and computation delay performance versus the computation resource division under different schemes. As shown in Fig. 8(a), when the minimum value of the computation resource block is reduced from 8 to 4, the energy consumption of all schemes decreases except the WNJ scheme. The WNJ scheme costs the highest energy because it cannot provide secure offloading to the MEC server, which can only process tasks locally. However, when the minimum value of the resource block is reduced to 2, the EESO scheme has an increase in the energy consumption. This is because the computation delay is larger than the maximum limit. Therefore, the larger resource block with a faster computation speed is chosen to satisfy the constraint of the computation delay, resulting in a higher energy consumption. As shown in Fig. 8(b), as the gap between the maximum and minimum resource blocks continues to increase, the computation delay of the Optimal scheme becomes higher. This is because the Optimal scheme only has an energy consumption in its target and there is no constraint on the computation delay. On the contrary, our proposed EESO scheme considers the computation delay constraint. Therefore, the computation delay is limited. The Optimal scheme has a computation delay of 1 second in the worst situation, which is intolerable for user vehicles.

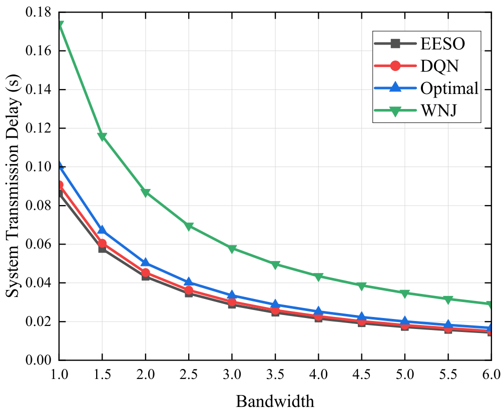
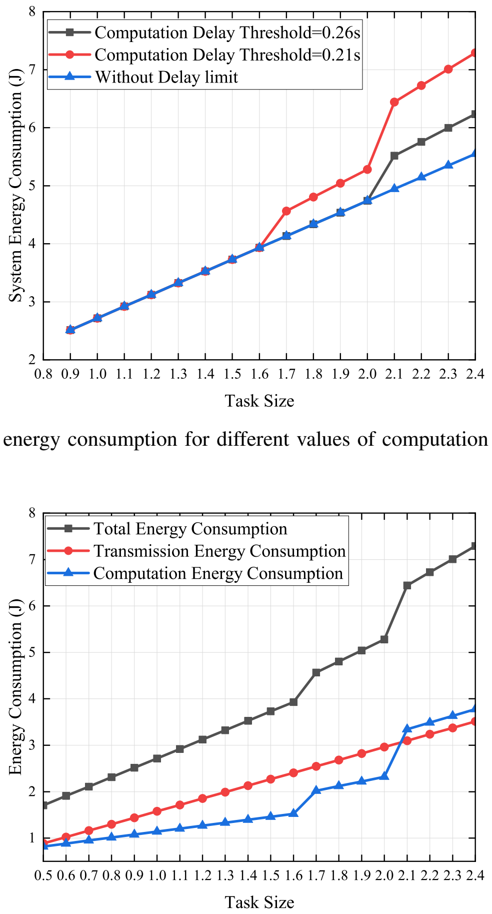


Fig. 9. Transmission delay of the secure offloading under different bandwidths.

Fig. 9 reflects the impact of the bandwidth on the security offloading transmission delay. As the bandwidth increases, the transmission delay of all schemes decreases. The transmission delay of the DQN, EESO and Optimal schemes are basically the same when the bandwidth is greater than 5.5. The transmission delay of the WNJ scheme decreases more sharply than the other three schemes. This shows that as the bandwidth increases, the difference in the transmission delay among different schemes will continue to shrink. In addition, our proposed EESO scheme has the smallest transmission delay in all bandwidths. Thus, the transmission delay of the EESO scheme performs better than the other three schemes.

Fig. 10(a) shows the system energy consumption versus the task size for different values of the computation delay limits. As the amount of task data increases, the energy consumed by the system also increases. In addition, we can see that the scheme with lower computation delay constraints consumes more energy as the task size increases. This is because the stricter delay constraint requires a shorter computation time. Thus, the computation resource with the faster computation speed is selected to use, which results in more energy consumption. The detailed energy consumption of the scheme with the computation delay constraint of 0.21s is shown in Fig. 10(b). As the size of the computation task increases, the offloading to the base station and the computation will consume more energy. When the task size is increased to 1.7M and 2.1M, the growth rate of the computation energy consumption exceeds the previous growth rate. This is because the original resource block selection cannot meet the computation delay requirements at this time. Therefore, it is necessary to update the resource block selection.



(a) Total energy consumption for different values of computation delay limits.

(b) Transmission, computation and total energy consumption with the computation delay threshold $T = 0.21$s.

Fig. 10. Energy consumption versus the task size under different computation delay limitations.

VI. Conclusion

In this paper, an A3C-based EESO scheme was proposed in the NOMA offloading scenario using PLS technology. Our goal is to minimize the energy consumed by the system and limit the computation delay to a suitable range. We have jointly optimized the transmit power, the computation resources and the NOMA pairing between user vehicles and jammer vehicles. We have solved the optimization problem by using the A3C algorithm. With the proper training, the agent was successfully adapted to highly dynamic VEC networks, which reduces the energy consumed by the system and protects confidential information from eavesdropping. Comprehensive analysis has demonstrated the robustness and effectiveness of the EESO scheme which have been proven. Besides, the system energy consumption and average secrecy probability of the A3C-based EESO scheme are improved compared to the other three schemes. On the basis of the current work done in this paper, there are several points that need to be further considered. For example, only a single-cell base station scenario has been considered and a multi-cell base station scenario should be considered in the future.

References

[1] 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.

[2] N. Abbas, Y. Zhang, A. Taherkordi, and T. Skeie, “Mobile edge computing: A survey,” *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 450–465, 2018.

[3] S. Sardellitti, G. Scutari, and S. Barbarossa, “Joint optimization of radio and computational resources for multicell mobile-edge computing,” *IEEE Transactions on Signal and Information Processing over Networks*, vol. 1, no. 2, pp. 89–103, 2015.

[4] C. You, K. Huang, and H. Chae, “Energy efficient mobile cloud computing powered by wireless energy transfer,” *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 5, pp. 1757–1771, 2016.

[5] C. Jiang, X. Cheng, H. Gao, X. Zhou, and J. Wan, “Toward computation offloading in edge computing: A survey,” *IEEE Access*, vol. 7, pp. 131 543–131 558, 2019.

[6] W. Wu, F. Zhou, R. Q. Hu, and B. Wang, “Energy-efficient resource allocation for secure NOMA-enabled mobile edge computing networks,” *IEEE Transactions on Communications*, vol. 68, no. 1, pp. 493–505, 2020.

[7] X. He, R. Jin, and H. Dai, “Physical-layer assisted secure offloading in mobile-edge computing,” *IEEE Transactions on Wireless Communications*, vol. 19, no. 6, pp. 4054–4066, 2020.

[8] Y. Ju, G. Zou, H. Bai, L. Liu, Q. Pei, C. Wu, and S. A. Otaibi, “Random beam switching: A physical layer key generation approach to safeguard mmwave electronic devices,” *IEEE Transactions on Consumer Electronics*, pp. 1–1, 2023.

[9] B. Brecht, D. Therriault, A. Weimerskirch, W. Whyte, V. Kumar, T. Hehn, and R. Goudy, “A security credential management system for V2X communications,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 12, pp. 3850–3871, 2018.

[10] J. Zhou, X. Dong, Z. Cao, and A. V. Vasilakos, “Secure and privacy preserving protocol for cloud-based vehicular DTNs,” *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 6, pp. 1299–1314, 2015.

[11] A. D. Wyner, “The wire-tap channel,” *The Bell System Technical Journal*, vol. 54, no. 8, pp. 1355–1387, 1975.

[12] M. Bloch, J. Barros, M. R. D. Rodrigues, and S. W. McLaughlin, “Wireless information-theoretic security,” *IEEE Transactions on Information Theory*, vol. 54, no. 6, pp. 2515–2534, 2008.

[13] Y. Liu, H.-H. Chen, and L. Wang, “Physical layer security for next generation wireless networks: Theories, technologies, and challenges,” *IEEE Communications Surveys & Tutorials*, vol. 19, no. 1, pp. 347–376, 2017.

[14] J. M. Hamamreh, H. M. Furqan, and H. Arslan, “Classifications and applications of physical layer security techniques for confidentiality: A comprehensive survey,” *IEEE Communications Surveys & Tutorials*, vol. 21, no. 2, pp. 1773–1828, 2019.

[15] Y. Ju, M. Yang, C. Chakraborty, L. liu, Q. Pei, M. Xiao, and K. Yu, “Reliability-security tradeoff analysis in mmwave ad hoc based CPS,” *ACM Trans. Sen. Netw.*, 2023.

[16] L. Sun and Q. Du, “Physical layer security with its applications in 5G networks: A review,” *China Communications*, vol. 14, no. 12, pp. 1–14, 2017.

[17] K. Cao, B. Wang, H. Ding, and J. Tian, “Adaptive cooperative jamming for secure communication in energy harvesting relay networks,” *IEEE Wireless Communications Letters*, vol. 8, no. 5, pp. 1316–1319, 2019.

[18] 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.

[19] X. He, R. Jin, and H. Dai, “Physical-layer assisted privacy-preserving offloading in mobile-edge computing,” in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, 2019, pp. 1–6.

[20] J. Xu and J. Yao, “Exploiting physical-layer security for multiuser multicarrier computation offloading,” *IEEE Wireless Communications Letters*, vol. 8, no. 1, pp. 9–12, 2019.

[21] Y. Zhang, H.-M. Wang, Q. Yang, and Z. Ding, “Secrecy sum rate maximization in non-orthogonal multiple access,” *IEEE Communications Letters*, vol. 20, no. 5, pp. 930–933, 2016.

[22] F. Zhou, Z. Chu, H. Sun, R. Q. Hu, and L. Hanzo, “Artificial noise aided secure cognitive beamforming for cooperative MISO-NOMA using SWIPT,” *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 4, pp. 918–931, 2018.

[23] X. Chen, Z. Zhang, C. Zhong, D. W. K. Ng, and R. Jia, “Exploiting interuser interference for secure massive non-orthogonal multiple access,” *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 4, pp. 788–801, 2018.

[24] K. Cao, B. Wang, H. Ding, L. Lv, J. Tian, and F. Gong, “On the security enhancement of uplink NOMA systems with jammer selection,” *IEEE Transactions on Communications*, vol. 68, no. 9, pp. 5747–5763, 2020.

[25] C. Gong, X. Yue, Z. Zhang, X. Wang, and X. Dai, “Enhancing physical layer security with artificial noise in large-scale NOMA networks,” *IEEE Transactions on Vehicular Technology*, vol. 70, no. 3, pp. 2349–2361, 2021.

[26] B. Chen, R. Li, Q. Ning, K. Lin, C. Han, and V. C. Leung, “Security at physical layer in NOMA relaying networks with cooperative jamming,” *IEEE Transactions on Vehicular Technology*, vol. 71, no. 4, pp. 3883–3888, 2022.

[27] L. Qian, W. Wu, W. Lu, Y. Wu, B. Lin, and T. Q. S. Quek, “Secrecy-based energy-efficient mobile edge computing via cooperative non-orthogonal multiple access transmission,” *IEEE Transactions on Communications*, vol. 69, no. 7, pp. 4659–4677, 2021.

[28] B. Li, W. Wu, W. Zhao, and H. Zhang, “Security enhancement with a hybrid cooperative NOMA scheme for MEC system,” *IEEE Transactions on Vehicular Technology*, vol. 70, no. 3, pp. 2635–2648, 2021.

[29] C. Sonmez, C. Tunca, A. Ozgovde, and C. Ersoy, “Machine learning-based workload orchestrator for vehicular edge computing,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 4, pp. 2239–2251, 2021.

[30] G. Liu, Y. Xu, Z. He, Y. Rao, J. Xia, and L. Fan, “Deep learning-based channel prediction for edge computing networks toward intelligent connected vehicles,” *IEEE Access*, vol. 7, pp. 114 487–114 495, 2019.

[31] T. Şahin, R. Khalili, M. Boban, and A. Wolisz, “Reinforcement learning scheduler for vehicle-to-vehicle communications outside coverage,” in *2018 IEEE Vehicular Networking Conference (VNC)*, 2018, pp. 1–8.

[32] X. Ye, M. Li, P. Si, R. Yang, E. Sun, and Y. Zhang, “Blockchain and MEC-assisted reliable billing data transmission over electric vehicular network: An actor—critic RL approach,” *China Communications*, vol. 18, no. 8, pp. 279–296, 2021.

[33] Y. Ju, H. Wang, Y. Chen, T.-X. Zheng, Q. Pei, J. Yuan, and N. Al-Dhahir, “Deep reinforcement learning based joint beam allocation and relay selection in mmwave vehicular networks,” *IEEE Transactions on Communications*, vol. 71, no. 4, pp. 1997–2012, 2023.

[34] Y. Ju, Y. Chen, Z. Cao, L. Liu, Q. Pei, M. Xiao, K. Ota, M. Dong, and V. C. M. Leung, “Joint secure offloading and resource allocation for vehicular edge computing network: A multi-agent deep reinforcement learning approach,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 5, pp. 5555–5569, 2023.

[35] H. Zhou, Z. Wang, H. Zheng, S. He, and M. Dong, “Cost minimization-oriented computation offloading and service caching in mobile cloud-edge computing: An A3C-based approach,” *IEEE Transactions on Network Science and Engineering*, vol. 10, no. 3, pp. 1326–1338, 2023.

[36] L. Liu, J. Feng, X. Mu, Q. Pei, D. Lan, and M. Xiao, “Asynchronous deep reinforcement learning for collaborative task computing and on-demand resource allocation in vehicular edge computing,” *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–14, 2023.

[37] Y. Li, H. Chen, and M. Feng, “A novel model for the traffic of urban roads based on queuing theory,” in *2020 International Conference on Intelligent Computing, Automation and Systems (ICICAS)*, 2020, pp. 190–194.

[38] S. Yan, N. Yang, G. Geraci, R. Malaney, and J. Yuan, “Optimization of code rates in SISOME wiretap channels,” *IEEE Transactions on Wireless Communications*, vol. 14, no. 11, pp. 6377–6388, 2015.

[39] Y. Ju, H.-M. Wang, T.-X. Zheng, and Q. Yin, “Secure transmissions in millimeter wave systems,” *IEEE Transactions on Communications*, vol. 65, no. 5, pp. 2114–2127, 2017.

[40] T.-X. Zheng, Y. Wen, H.-W. Liu, Y. Ju, H.-M. Wang, K.-K. Wong, and J. Yuan, “Physical-layer security of uplink mmwave transmissions in cellular V2X networks,” *IEEE Transactions on Wireless Communications*, vol. 21, no. 11, pp. 9818–9833, 2022.

[41] C. Wang, Z. Li, X.-G. Xia, J. Shi, J. Si, and Y. Zou, "Physical layer security enhancement using artificial noise in cellular vehicle-to-everything (C-V2X) networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 12, pp. 15 253–15 268, 2020.

[42] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," in *International conference on machine learning*, 2018, pp. 1861–18 970.

[43] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. P. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, "Asynchronous methods for deep reinforcement learning," 2016. [Online]. Available: https://arxiv.org/abs/1602.01783

[44] M. Döttling, W. Mohr, and A. Osseiran, *WINNER II Channel Models*, 2010, pp. 39–92.

[45] X. Hou, J. Wang, Z. Fang, Y. Ren, K.-C. Chen, and L. Hanzo, "Edge intelligence for mission-critical 6G services in space-air-ground integrated networks," *IEEE Network*, vol. 36, no. 2, pp. 181–189, 2022.

[46] *TR 22.886: Technical Specification Group Radio Access Network; Study enhancement 3GPP Support for 5G V2X Services; (Release 15).* Technical Specification Group Radio Access Network 3rd Generation Partnership Project (3GPP), Mar. 2017.

[47] *TR 36.885: Technical Specification Group Radio Access Network; Study on LTE-based V2X services; (Release 14).* Technical Specification Group Radio Access Network 3rd Generation Partnership Project (3GPP), Jun. 2016.

Ying Ju (Member, IEEE) received the B.S. and M.S. degrees from the School of Electronic Information Engineering, Tianjin University, Tianjin, China, in 2008 and 2010, respectively, and the Ph.D. degree from the School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, China, in 2018. From 2016 to 2017, she was a Visiting Scholar at the Department of Computer Science, University of California, Santa Barbara, USA. From 2010 to 2018, she was a Senior Engineer at the State Radio Monitoring Center, Xi'an. She is currently an Associate Professor with the Department of Telecommunications Engineering, Xidian University, Xi'an. Her research interests include physical layer security of wireless communications, millimeter wave communication systems, and the Internet of Things.

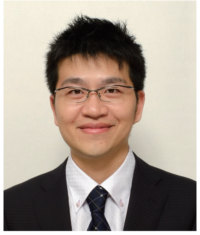
Zhiwei Cao received the B.S. degree in communication engineering from Shanghai University, Shanghai, China, in 2021. He is currently pursuing the M.S. degree in electronic information engineering, Xidian University, Xi'an, China. His research interests include edge computing, NOMA, and deep reinforcement learning.

Yuchao Chen (Graduate Student Member, IEEE) received the B.S. degree in communication engineering from Beijing Jiaotong University, Beijing, China, in 2020. He is currently pursuing the M.S. degree in information and communication engineering, Xidian University, Xi'an, China. His research interests include physical layer security of wireless communication and AI optimization in millimeter wave communication systems.

Lei Liu (Member, IEEE) received the B.Eng. degree in electronic information engineering from Zhengzhou University, Zhengzhou, China, in 2010, and the M.Sc. and Ph.D. degrees in communication and information systems from Xidian University, Xi'an, China, in 2013 and 2019, respectively. From 2013 to 2015, he was employed by a subsidiary of China Electronics Corporation, Beijing, China. From 2018 to 2019, he was supported by the China Scholarship Council to be a Visiting Ph.D. Student at the University of Oslo, Oslo, Norway. He is currently a Lecturer with the State Key Laboratory of Integrated Service Networks, Xidian University, and also with the Xidian Guangzhou Institute of Technology, Xi'an. His research interests include vehicular ad hoc networks, intelligent transportation, mobile-edge computing, and the Internet of Things.

Qingqi Pei (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees in computer science and cryptography from Xidian University, in 1998, 2005, and 2008, respectively. He is currently a Professor and a member of the State Key Laboratory of Integrated Services Networks, a Professional Member of ACM, and a Senior Member of Chinese Institute of Electronics and China Computer Federation. His research interests include digital contents protection and wireless networks and security.

Shahid Mumtaz (Senior Member, IEEE) is a Nottingham Trent University (NTU), UK professor. He is an IET Fellow, founder, and EiC of IET "Journal of Quantum Communication," Vice Chair: Europe/Africa Region- IEEE ComSoc: Green Communications & Computing Society. He authorizes four technical books, 12 book chapters, and 300+ technical papers (200+ IEEE Journals/transactions, 100+ conferences, 2 IEEE best paper awards) in mobile communications. Most of his publication is in the field of Wireless Communication. He is a Scientific Expert and Evaluator for various research funding agencies. In 2012, he was awarded an "Alain Bensoussan fellowship." China awarded him the young scientist fellowship in 2017.



Mianxiong Dong (Senior Member, IEEE) received B.S., M.S. and Ph.D. in Computer Science and Engineering from The University of Aizu, Japan. He is the Vice President and Professor of Muroran Institute of Technology, Japan. He was a JSPS Research Fellow with School of Computer Science and Engineering, The University of Aizu, Japan and was a visiting scholar with BBCR group at the University of Waterloo, Canada supported by JSPS Excellent Young Researcher Overseas Visit Program from April 2010 to August 2011. Dr. Dong is the Vice President and Professor at Muroran Institute of Technology. His research interests include large-scale network systems such as mobile networks, wireless sensor networks, vehicle networks, cyber-physical systems, and IoT. He is also engaged in research on a wide range of cutting-edge information technologies, such as edge computing, AI technology, SDN, and big data analysis. His laboratory combines knowledge in the field of large-scale network systems with physical layer technologies to create new value.

Mohsen Guizani (Fellow, IEEE) received his BS (with distinction), MS, and Ph.D. degrees in Electrical and Computer engineering from Syracuse University, Syracuse, NY, USA in 1985, 1987, and 1990, respectively. He is currently a Professor of Machine Learning at Mohamed Bin Zayed University of Artificial Intelligence (MBZUAI), Abu Dhabi, UAE. Previously, he worked in different institutions in the USA. His research interests include applied machine learning and artificial intelligence, the Internet of Things (IoT), intelligent systems, smart city, and cybersecurity. He was elevated to IEEE Fellow in 2009 and was listed as a Clarivate Analytics Highly Cited Researcher in Computer Science in 2019, 2020, and 2021. His research interests include applied machine learning and artificial intelligence, the Internet of Things (IoT), intelligent systems, smart city, and cybersecurity.