Optimization of Resource Allocation for V2X Security Communication based on Multi-Agent Reinforcement Learning

Baofeng Ji, Bingyi Dong, Da Li, Lvxi Yang, Yi Wang, Charalampos Tsimenidis, Varun G Menon

***Abstract*—In order to address the data security and communication efficiency of vehicles during high-speed mobile communication, this paper investigates the problem of secure in-vehicle communication resource allocation based on slow-variable large-scale fading channel information, to meet the quality of service requirements of vehicular communication, i.e., to ensure the reliability of V2V communication and the time delay while maximizing the transmission rate of the cellular link. And an eavesdropping model is introduced to ensure the secure delivery of link information. Considering that the high mobility of vehicles causes rapid channel changes, we model the problem as a Markov decision process and propose a resource allocation optimization framework based on the Multi-Agent Reinforcement Learning Algorithm (MARL-DDQN), in which a large-scale neural network model is built to train vehicular intelligences to learn the optimal resource allocation strategy for optimal communication performance and security performance. Simulation results show that the load successful delivery rate and confidentiality performance of the vehicular communication network are effectively improved compared to the baseline and MADDPG strategies while ensuring link security. This study provides useful references and practical value for the optimization of secure communication resource allocation in vehicular networking.**

***Index Terms*—V2X, resource allocation, multi-agent reinforcement learning, MARL-DDQN**

I. Introduction

THE vehicle-to-everything (V2X) paradigm, as an extension of the internet of things (IoT) concept, can assist in the development of smart cities [1]. Due to the growing popularity of Internet of Things (IoT) user devices, researchers have been working on network optimization challenges to improve the energy or spectrum efficiency (EE/SE) of wireless networks to satisfy the users' demanding data rates and varied quality of service (QoS) requirements, e.g. [2] and [3]. These studies combine vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication in vehicular networking to improve the performance of intelligent transport systems (ITS). The full use of resource allocation and frequency sharing through the convergence of networks between V2V and V2I will determine the effectiveness and efficiency of future ITS.

In the cellular V2X [4] paradigm, have completed the standardization of LTE-V2X in 3GPP Release 14 and introduced two new communication modes especially designed for V2V communication, Centralized Resource Scheduling (Mode 3) and Distributed Resource Scheduling (Mode 4). And the architecture of V2X service is further enhaned in Release 15. By using two radio interfaces, Uu and PC5, the architecture supports V2I and V2V connectivity, realizing the need for reliable communication over long distances and greater ranges, as well as direct communication over short distances. This provides low-latency, high-capacity, and high-reliability communication capabilities for V2X communications, providing technical support for application areas such as intelligent transportation systems and vehicle safety. Specifically, the Uu interface is used to link communications between vehicles and base stations to achieve reliable communications over long distances and greater ranges, while the PC5 interface is used to link communications between vehicles, people and road infrastructure to achieve direct communications with low-latency, high-capacity and high-reliability communications through direct connection, broadcasting, and network scheduling. In 3GPP Release 16 [5], a number of new use cases and requirements are proposed and analyzed with the aim of enhancing 5G V2X technology. These include in-vehicle entertainment services, which require high-rate connectivity to base stations (BS) for high-rate data transfer, such as dynamic digital map updates. V2I and V2V communications coexist to achieve smarter, efficient transportation systems and vehicle safety. In this paradigm, V2V and V2I communications need to fulfill a number of QoS requirements, including low transmission latency, high reliability, higher data transfer capability, high bandwidth efficiency, and security and privacy protection. In recent years, the development of vehicular communication has attracted attention to physical layer security in the V2X [6]. Traditional communication security primarily relies on higher-layer cryptographic encryption [7]. However, cryptographic algorithms heavily depend on the secrecy of encryption keys. Once the encryption keys are compromised, the information security becomes vulnerable. In the vehicular communication environment, low latency, highly reliable connections, and a massive number of connected devices are required. The large number of encryption keys, along with the cost of key distribution and management, does not align well with the practicality of daily vehicular communication environments. In contrast to cryptographic encryption techniques, physical layer security offers new opportunities for wireless network security. It leverages the inherent randomness of wireless channels and utilizes techniques such as channel coding [8] and signal processing [9] to increase the difficulty for eavesdroppers to obtain information, thereby achieving secure transmission over

This paper was produced by the IEEE Publication Technology Group. They are in Piscataway, NJ.

Manuscript received April 19, 2021; revised August 16, 2021.

wireless channels. Physical layer security provides additional protection for vehicular communication security with lower complexity and overhead.

Moreover, vehicles are interconnected with hundreds of millions of devices, which generate huge data sets and model parameters that grow exponentially. Nowadays, machine learning models, neural networks, and other computational models have a large number of parameters and computational power. These large models(LMs) are used in collaborative and cooperative systems between vehicles and road infrastructure, and can provide a deeper understanding of complex traffic data in dealing with a variety of complex data and tasks, allowing their models to more accurately capture patterns and associations in the data, thus improving the quality of data understanding and the performance of subsequent tasks.

A. Related work

In V2V communication scenarios, various challenges such as mobility [10], ultra-dense networks [11], and non-line-of-sight (NLOS) situations [12] significantly contribute to the complexity of the problem. The dynamic nature of the system, coupled with the difficulty in establishing a fixed mathematical model, And many optimizing algorithms are difficult to achieve the optimal solution, which adds to the complexity of the problem. Reinforcement learning demonstrates excellent performance in terms of its ability to solve uncertain decision problems. For instance, in [13], the authors consider the joint optimization of networking, caching, and computation to enhance Telematics performance. Due to its computational complexity, an advanced reinforcement learning algorithm that approximates the Q-value-action function using a deep Q-network is employed for automatic resource allocation. Reinforcement learning effectively reduces computational complexity by iteratively interacting with the uncertain environment, enabling robust handling of environmental dynamics and sequential decision-making. Moreover, hard-to-optimize objective problems can be effectively solved within the reinforcement learning framework by designing training rewards that align with the ultimate objective. In [14], the authors address channel uncertainty caused by the channel state information (CSI) fed back to the base station (BS) and analyze the correlation of rapidly changing channels. They propose a joint channel assignment and power control algorithm to satisfy individual V2V link delay and reliability requirements while maximizing system throughput. A dual time-scale resource allocation scheme is proposed in [15], aiming to minimize the worst-case delay of V2V transmissions based on extensive road traffic information. Subsequently, the transmit power of each V2V link is optimized based on the small time-scale CSI. However, this scheme primarily focuses on worst-case delay minimization, and its performance in other scenarios needs further investigation. To address modeling accuracy concerns, [16] [17] model resource sharing as a multi-agent reinforcement learning problem and employ deep Q-learning algorithms for joint channel assignment and power allocation design. It is worth noting that the aforementioned works primarily concentrate on minimizing delay and maximizing throughput, No concern for link security on in-vehicle networks. Aiming at the decentralized joint optimization problem of channel selection and power control in V2V communication, a new federated multi-agent deep reinforcement learning (FedMARL) approach is proposed in [18] to simultaneously exploit the advantages of deep reinforcement learning (DRL) and federated learning (FL) to maximize the transmission rate of the cellular link. while meeting the reliability and delay requirements of V2V communication. In [19], a dual time-scale federated DRL algorithm is proposed to address the local limitations of DRL model training and improve model robustness. It leverages a graph-theoretic vehicle clustering algorithm to capture global information on longer time scales and combines it with a federated learning algorithm to enhance training efficiency on shorter time scales. Although federated learning itself can enhance data confidentiality, it is crucial to exercise caution in practice and implement additional privacy protection measures. Moreover, establishing robust security mechanisms is essential to ensure the confidentiality of data and models.

However, the aforementioned research work does not address the practical challenge of eavesdropping resistance. In real-world environments, where network resources such as bandwidth and energy are often limited, the primary objective of security resource allocation is to utilize these scarce resources efficiently to satisfy security performance metrics, including secrecy rate, secrecy outage probability, and power consumption. Most of the existing research focuses on fundamental aspects of security resource allocation, such as subcarrier selection and power control. Subcarrier allocation aims to identify the optimal choice that improves spectrum utilization efficiency and enhances security performance. Additionally, adaptive power allocation is another crucial approach to enhance secrecy performance. For instance, in [20], a Q-learning-based secure power allocation strategy is proposed, which mitigates the eavesdropper's ability to decode the transmitted signal while ensuring the desired signal-to-noise ratio. In the context of device-to-device (D2D) communications, [21] presents a physical layer security optimization model to address the security concerns associated with multiple eavesdroppers. The proposed model employs a novel access algorithm to manage the interference caused by D2D users, thus enhancing the confidentiality performance of cellular users and aiming to improve the security and confidentiality of D2D communications. Similarly, [22] discusses the security of D2D spectrum sharing networks, utilizing the interference generated by D2D users to enhance the security performance of cellular users. To ensure secure access to cellular spectrum, [23] proposes a secure and efficient Priority-based Power and Resource Management (PPRM) scheme. This scheme grants the Vehicle User Equipment (VUE) the same priority as the Cellular User Equipment (CUE), and the mixed integer nonconvex PPRM problem is efficiently solved by transforming it into power allocation and resource block allocation sub-problems. Moreover, in [24], Reconfigurable Intelligent Surfaces (RIS) are employed to ensure secure in-vehicle communications in the presence of eavesdroppers, demonstrating the potential improvement in confidentiality

under V2V and V2I communications with the utilization of RIS. It should be noted that all the aforementioned works primarily focus on addressing security requirements and do not explicitly consider the QoS requirements of vehicles in Vehicular Networking, including reliability and latency issues.

B. Motivations and Contribution

Although the spectrum sharing problem has been studied extensively, the comprehensive literature review reveals that only a limited number of studies have effectively integrated the physical layer security of V2X with the spectrum sharing problem in highly mobile vehicular networks. Recently multi-agent algorithm research has made significant progress in areas such as robot collaboration, UAV teams, self-driving vehicles, and resource allocation. However, the issues of spectrum sharing, physical layer security, distributed deep reinforcement learning algorithms require full knowledge of channel gain information, interference information from other vehicles, and remaining time to meet delay constraints, among others. However, very little research has been done on how to simultaneously address the QoS requirements of spectrum sharing, physical layer security, and high mobility in such complex environments.

This vehicle network scenario optimization problems are widely considered to be instantaneous optimization type of problems considering the high mobility of vehicles, the random variation of vehicular channel conditions, and the high computational complexity. However, this type of problem is very difficult to solve. To overcome this challenge, the instantaneous optimization problem can be transformed into a long-term reward accumulation optimization problem and modeled using the framework of Markov Decision Process (MDP). In which each vehicle as an agent that selects the best action at each time step based on the current environmental state, such as spectral subband selection and power control, to optimize the long-term reward accumulation. This transformation into an MDP problem allows the application of algorithms such as reinforcement learning to solve the vehicular channel optimization problem, thus adapting to the changing channel environment and vehicle mobility. Moreover, the use of pre-trained models improves the efficiency of the reinforcement learning system, reduces the complexity of the samples, and also improves the generalization ability of the intelligences. Especially this approach has potential when facing large-scale scenarios and complex tasks, and provides new possibilities for the optimization of V2X Communication Systems. The main points of our contribution can be summarized as follows:

1) We propose a novel communication architecture for vehicular networks that incorporates an eavesdropping model into the existing cooperative communication scenario between V2V links and V2I links. This architecture takes into account the spectrum sharing problem, eavesdropping resistance of links, and the QoS requirements of vehicles to enhance the overall performance of V2X.

2) To ensure the QoS of both V2V and V2I communication while meeting the security requirements, we introduce security outage probability constraints that characterize the reliability requirements of vehicular networking. Additionally, we present a secure resource allocation framework based on a multi-agent reinforcement learning algorithm. This framework models the problem as a Markov process, reducing computational complexity while maintaining performance.

3)The spectrum access problem for multiple V2V links is formulated as a multi-agent problem, where each V2V link transmitter acts as an agent aiming to maximize the secrecy rate of V2I links. This is achieved through appropriate reward design and training mechanisms. By adopting a distributed approach to resource allocation, we mitigate centralized latency and overhead. The optimization objectives and constraints are designed to enhance V2V confidential transmission delay and V2I system confidentiality performance.

The proposed scheme enables secure resource allocation in complex vehicular communication environments. By constructing Markov models and employing reinforcement learning algorithms, we analyze and optimize the reliability, low latency, and total rate of cellular users for V2V links. Simultaneously, the multi-agent framework and distributed resource allocation algorithms enhance the confidentiality rate of V2I links and address security concerns in dynamic V2X environments. These approaches hold significant academic and practical importance in improving the performance and security of V2X communications.

II. System Model

With the rapid development of the Internet of Things (IoT) and the Internet of Vehicles (IoV), the widespread use of numerous terminal devices has placed significant strain on existing spectrum resources. Consequently, spectrum resource sharing has become a crucial approach to alleviate this problem and improve spectrum utilization efficiency. In this context, this paper considers a scenario where V2I communication and V2V communication coexist within cellular vehicular networking. The objective is to explore methods that maximize the transmission rate of cellular links by leveraging the multiplexing of spectrum resources through the V2V link, while ensuring the reliability and confidentiality of V2V communication. In this scenario, the V2I link utilizes the Uu interface to establish communication between vehicles and the base station (BS), facilitating the provision of high data rate services. On the other hand, the V2V link periodically propagates security messages through the PC5 interface. In the considered vehicular network, we make the assumption that all transceivers utilize a single antenna. The set of V2I links and V2V links in this network are denoted as $m \in \{1, 2, \cdots, M\}$ and $k \in \{1, 2, \cdots, K\}$, respectively.

The system model is shown in Fig 1, V2V links are vehicle-to-vehicle communications and each V2I link connects to a BS for communication purposes and there exists an eavesdropper, Eve, that passively eavesdrops on both V2I links and V2V links. Frequency selective wireless channels are converted into multiple parallel flat channels on different subcarriers using orthogonal frequency division multiplexing (OFDM). Several consecutive subcarriers are grouped together to form a spectral sub-band, assuming that the V2I link uploading data has been pre-assigned orthogonal spectral sub-bands, i.e.,

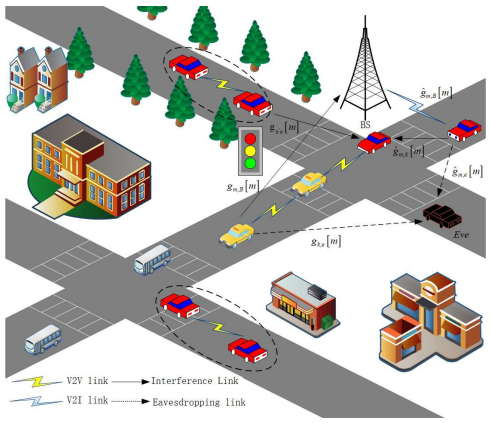

Fig. 1. System model

the mth V2I link occupies the mth sub-band. In order to achieve the goals of V2V links and V2I links with minimum signaling overhead in high-speed mobile environments and to improve spectrum utilization, we have designed an efficient spectrum sharing scheme. This scheme works by grouping consecutive subcarriers into spectral sub-bands and pre-allocating orthogonal spectral sub-bands for V2I links. In this scheme, the V2V link can share part of the spectrum resources occupied by the V2I link when transmitting packets. To ensure physical layer security. we introduce eavesdropping vehicles. consider calculating the confidentiality capacity to measure the maximum rate of confidential information that can be transmitted in the presence of an eavesdropping threat. Through this spectrum sharing scheme. We achieve effective utilization of spectrum resources and meet the communication requirements of V2V links and V2I links in high-speed mobile environments.

It is assumed that the channel fading within each sub-band is the same and is independent across channels. The channel power gain $g_k[m]$ of the kth V2V link on the mth sub-band in one unit of coherence time can be expressed as:

$$g_k[m] = \alpha_k h_k[m], \tag{1}$$

where $h_k[m]$ is the power small-scale fading component, band dependent, varying between successive time slots, and assumed to follow a unit mean exponential distribution [25]. And α_k captures the large-scale fading effect, including path loss and shadowing, assumed to be frequency independent. The fading model of the article is modeled as Rayleigh fading according to the urban scenario in 3GPP Technical Report TR 36.885[26], The V2I link fading model is based on the path loss model defined in 3GPP TR 36.885 as $128.1+37.6\log_{10}d$, where d denotes the distance between the vehicle and the BS. The shadowing due to obstacle effects obeys independent log-normal shadowing with standard deviation of 8 dB. And the V2V link road loss model is referenced from IST-4-027756 WINNER II D1.1.2 V1.2 WINNER II, and the shadow fading model is referenced from the "A-1.4 Channel model" in [26].

The interference channel from the transmitter vehicle of the $k'th$ V2V link to the receiving vehicle of the kth V2V link on the mth sub-band is denoted as $\mathrm{g}_{k'k}[m]$; the channel gain from the transmitter vehicle of the mth V2I link to the BS in the mth sub-band is denoted as $\hat{g}_{m,B}[m]$; the interference channel from the transmitter vehicle of the mth V2I link to the kth V2V receiver vehicle on the mth sub-band is denoted as $\hat{g}_{m,k}[m]$; the interference channel from the transmitter vehicle of the kth V2V link to the Eve in the mth sub-band is denoted by as $g_{k,e}[m]$; the interference channel from the sending vehicle of the mth V2I link to the Eve in the mth sub-band is denoted as $\hat{g}_{m,e}[m]$.

Then the signal-to-noise ratio of the mth V2I link at the base station $\gamma_m^I[m]$ is:

$$\gamma_m^I[m] = \frac{P_m^I \hat{g}_{m,B}[m]}{\sigma_B^2 + I_m^I}, \tag{2}$$

where P_m^I is the transmit power of the mth V2I transmitter; σ_B^2 is the noise power at the BS; and the power I_m^I of the mth V2I transmitter subject to interference is:

$$I_m^I = \sum_k \rho_k[m]\, g_{k,B}[m], \tag{3}$$

where $\rho_k[m]$ is the spectrum allocation factor. $\rho_k[m] = 1$ denotes that the mth sub-band is occupied by the mth V2V link, and conversely $\rho_k[m] = 0$, denotes that the mth sub-band is not occupied. Assuming that each V2V link occupies only one sub-band, i.e., $\sum_m \rho_k[m] \le 1$.

The signal-to-noise ratio of the kth V2V link receiving vehicle $\gamma_k^V[m]$ is:

$$\gamma_k^V[m] = \frac{P_k^V g_k[m]}{\sigma_v^2 + I_k^V}, \tag{4}$$

where P_k^V is the transmit power of the kth V2V transmitter, σ_v^2 is the noise power at the V2V link receiver, and the kth V2V transmitter is interfered with by the power I_k^V is:

$$tI_k^V = P_m^I \hat{g}_{m,k}[m] + \sum_{k' \neq k} \rho_{k'}[m]\, g_{k'k}[m], \tag{5}$$

The signal-to-noise ratio $\gamma_{m,e}^I[m]$ when Eve eavesdrops on the mth V2I link and the signal-to-noise ratio $\gamma_{k,e}^V[m]$ when Eve eavesdrops on the kth V2V link, can be expressed as:

$$\gamma_{m,e}^I[m] = \frac{P_m^I \hat{g}_{m,e}[m]}{\sigma_e^2 + \sum_k \rho_k[m] P_k^V[m]\, g_{k,e}[m]}, \tag{6}$$

$$\gamma_{k,e}^V[m] = \frac{P_k^V g_{k,e}[m]}{\sigma_e^2 + \sum_{k' \neq k} \rho_{k'}[m] P_{k'}^V g_{k',e}[m]}, \tag{7}$$

where σ_e^2 are the noise power at Eve. Then the secrecy capacity $C_m^I[m]$ of the mth V2I link and the secrecy capacity $C_k^V[m]$ of the kth V2V link can be expressed as:

$$C_m^I[m] = W\left[\log_2\left(1+\gamma_m^I[m]\right) - \log_2\left(1+\gamma_{m,e}^I[m]\right)\right]^+, \tag{8}$$

$$C_k^V[m] = W\left[\log_2\left(1+\gamma_k^V[m]\right) - \log_2\left(1+\gamma_{k,e}^V[m]\right)\right]^+. \tag{9}$$

where $[x]^+ = \max\{0, x\}$, W denotes the channel bandwidth.

In the cellular V2X, V2V and V2I communicate together and need to fulfill a set of QoS requirements. These requirements aim to ensure communication performance and reliability. The main QoS requirements considered in this paper:

1) Data rate: The latency issue has been one of the most critical requirements in the case of cooperative communication between V2V and V2I links. Every packet generated by a vehicle should be transmitted within a limited time. So the issue of packet rate is crucial. Vehicles need to ensure that the QoS meets certain standards when using network services, which can be parameterized as the minimum data transfer rate (in Mbps) to ensure a smooth service experience for the users. The achievable secrecy rate can be described as the difference between the data rate achievable by the legitimate channel and the eavesdropping channel. Let R_m^t be the achievable secrecy rate of the mth cellular link, i.e.

$$R_m^t = [R_{\mathrm{m}} - R_e]^+ = C_m^I, \tag{10}$$

where $[x]^+ = \max\{0, x\}$, R_m represents the data rate of the legitimate channel from the sender to the legitimate receiver, and R_e represents the data rate of the eavesdropping channel between the transmitter and the receiver. Satisfying $R_m^t \geq R_m^{\min}$. We assume that each cellular link has the same minimum data rate requirement.

2) Secure outage probability: The reliable transmission of V2V communication can be measured by the outage probability in SINR [25]. The outage probability is considered an important indicator for evaluating the reliability of V2V transmission. The security outage probability refers to the probability that eavesdroppers can successfully intercept data during the data transmission process. The lower the security outage probability, the safer and more reliable the data transmission, and the less likely eavesdroppers are to intercept the data. Then the security outage probability P_k^{out} of the kth V2V link is:

$$P_k^{out} = 1 - \Pr\{\gamma_{k,e}^V \leq 2^{-\mu}(1+\gamma_k^V) - 1\}, \tag{11}$$

where $\Pr\{\cdot\}$ denotes the probability of the event occurring, μ denotes the difference between the secrecy rate and the channel capacity. The security outage probability can be determined by whether the signal-to-noise ratio of the interfering signal injected by the eavesdropping is less than or equal to $2^{-\mu}(1+\gamma_k^V) - 1$. If this condition is satisfied, it means that the eavesdropper successfully steals the data. By calculating the complement of this probability, the probability that the eavesdropper fails to successfully steal the data, i.e., the security outage probability, can be obtained.

The security outage probability is regarded as an important metric for assessing the reliability of V2V transmission. If the outage probability of a V2V link is below a certain threshold ξ, the reliability constraint can be expressed as:

$$\Pr\{\gamma_{k,e}^V \leq 2^{-\mu}(1+\gamma_k^V) - 1\} \leq \xi, \tag{12}$$

This constraint means that both the probability of outage of the legitimate channel and the probability of outage of the eavesdropping link should be less than or equal to a given threshold to ensure the reliability of the link. If both are greater than the given threshold, the link is unreliable and appropriate measures need to be taken to improve the reliability, such as increasing the signal transmission power, improving the modulation scheme, enhancing encryption, etc.

Let $\gamma_0 = 2^{-\mu}(1+\gamma_k^V) - 1$, the constraint (12) can be written as:

$$\begin{aligned}&\Pr\{\gamma_{k,e}^V \leq \gamma_0\} = \Pr\{P_k^d g_{k,e}[m] \leq \gamma_0 \sigma_e^2 + \\ &\gamma_0 \sum_{k' \neq k} \rho_{k'}[m] P_{k'}^d g_{k',e}[m]\} \leq \xi,\end{aligned} \tag{13}$$

Lemma 1, according to the literature [26], if $z_1, \cdots, z_n$ are all independently exponentially distributed random variables, then $E[\mathrm{z}_i] = \frac{1}{\lambda_i}, i = 1, \cdots, n$,then

$$\Pr\left\{z_1 \leq \sum_{i=2}^{n} z_i + c\right\} = 1 - e^{-\lambda_{1c}} \prod_{i=2}^{n} \frac{1}{1+\frac{\lambda_1}{\lambda_i}}, \tag{14}$$

where c is a constant, λ denotes the parameter of the random variable z_i, which controls the scale and distributional properties of the exponential distribution.

In the time slot, the path loss and transmit power remain constant and only the fast fading has a unit-mean exponential distribution in the time slot. This gives the exponential distribution at time slot t:

$$E[P_k^d g_{k,e}[m]] = P_k^d \alpha_k^t = \frac{1}{\lambda_1}, \tag{15}$$

$$E[\gamma_0 P_{k'}^d g_{k',e}[m]] = \gamma_0 P_{k'}^d \alpha_{k',e}^t = \frac{1}{\lambda_i}, \quad i \neq 1 \ , \tag{16}$$

According to Lemma 1, we can express the outage constraint (13) as follows:

$$\begin{aligned}&\Pr\{P_k^d g_{k,e}[m] \leq \gamma_0 \sigma_e^2 + \gamma_0 \sum_{k' \neq k} \rho_{k'}[m] P_{k'}^d g_{k',e}[m]\} \\ &= 1 - \exp(-\tfrac{\gamma_0 \sigma_e^2}{P_k^d \alpha_k^t}) \prod_{k' \neq k} \tfrac{\rho_{k'}}{1+\frac{\gamma_0 P_{k'}^d \alpha_{k',e}^t}{P_k^d \alpha_k^t}} \leq \xi,\end{aligned} \tag{17}$$

Further applying the inequality of [27]: $e^k \prod_i^n (1+z_i) \leq e^k + \sum_i^n z_i$, the constraint(16) is then expressed as[17]:

$$\begin{aligned}&1 - \exp(-\tfrac{\gamma_0 \sigma_e^2}{P_k^d \alpha_k^t}) \prod_{k' \neq k} \tfrac{\rho_{k'}}{1+\frac{\gamma_0 P_{k'}^d \alpha_{k',e}^t}{P_k^d \alpha_k^t}} \leq \\ &1 - \exp(-\tfrac{\gamma_0 \sigma_e^2}{P_k^d \alpha_k^t}) \prod_{k' \neq k} \rho_{k'}(1 + \tfrac{\gamma_0 P_{k'}^d \alpha_{k',e}^t}{P_k^d \alpha_k^t}) \\ &\leq 1 - \exp(-\tfrac{\gamma_0 \sigma_e^2}{P_k^d \alpha_k^t} - \sum_{k' \neq k} \rho_{k'} \tfrac{\gamma_0 P_{k'}^d \alpha_{k',e}^t}{P_k^d \alpha_k^t}) \\ &= 1 - \exp(-\frac{\gamma_0 \sigma_e^2 + \sum_{k' \neq k} \rho_{k'} \gamma_0 P_{k'}^d \alpha_{k',e}^t}{P_k^d \alpha_k^t}) \leq \xi,\end{aligned} \tag{18}$$

The reliable confidentiality constraint for the kth V2V pair can be written as:

$$\frac{P_k^d \alpha_k^t}{\sigma_e^2 + \sum_{k' \neq k} \rho_{k'} P_{k'}^d \alpha_{k',e}^t} \geq \frac{\gamma_0}{\ln(\frac{1}{1-\xi})}. \tag{19}$$

III. Problem formulation

Our goal is to maximize the time-averaged confidential transmission rate of V2I links while ensuring the reliability and latency requirements of V2V links, with a maximum V2V transmission delay of T=100ms as defined in 3GPP TR 36.885. Our main scenario is decentralized channel access and power control for V2V pairs in the case of V2I, a centrally allocated orthogonal OFDM sub-channel at the base station. Since priority is usually given to V2I links, it can be assumed that V2I channel assignment is a known parameter in the considered model and is not part of the optimization variables. Therefore the problem under consideration can be formulated as:

$$\max_{\rho_k[m],P_k^d[m]} \frac{1}{T}\frac{1}{M}\sum_{t=1}^{T}\sum_{\mathrm{m}=1}^{M} C_m^I, \tag{20}$$

$$\text{s.t.} \sum_m C_m^I > C_0, \tag{21a}$$

$$R_m^t \geq R_m^{\min}, \tag{21b}$$

$$\frac{P_k^d \alpha_k^t}{\sigma_e^2 + \sum\limits_{k' \neq k} \rho_{k'} P_{k'}^d \alpha_{k',e}^t} \geq \frac{\gamma_0}{\ln(\frac{1}{1-\xi})}, \tag{21c}$$

$$0 \leq P_k^d[m] \leq P^{\max}, \tag{21d}$$

$$\sum_{m \in M} \rho_k^t \leq 1 \ , \quad \rho_k^t \in \{0,1\}. \tag{21e}$$

The problem is a time-rate combinatorial optimization problem, which is more challenging to compute and solve due to the mathematical model being idealized in relation to the subsequent states, plus the typical centralized solution is inadequate due to the high mobility of the vehicles. To address these issues, we solve the problem through reinforcement learning. In the paper we transform the joint channel selection and power allocation problem into a multi-agent problem, where the transmitter vehicle of each V2V link is an agent that executes independently and updates its resource policy. Constraint (21a) represents the minimum channel capacity requirement for the reliability of a V2I link, constraint (21b) represents the minimum rate requirement for a V2V link, constraint (21c) represents the secrecy outage constraint, constraint (21d) represents the maximum range of transmit power that should not be exceeded, and constraint (21e) represents that each V2V link can only occupy one RB.

IV. Resource Allocation for Based Multi-agent Reinforcement Learning

In resource allocation problems, the optimization problem is often NP-hard and has high computational complexity. In addition, obtaining the global optimal solution becomes more difficult due to the stochastic variation of vehicle channel conditions. To solve this problem, it can be transformed into a common Markov decision process with appropriate deep reinforcement learning algorithms.

At time step t, the agent observes its surroundings and obtains observations, i.e. states Then, the action $\mathrm{s} \in S$ takes the action $a \in A$ is determined by the policy $\pi(a|s)$:

$$\pi(a|s) = \Pr(A_t = a | S_t = s), \tag{22}$$

After the environment is evolve to the next state $s' \in S$, while receives the immediate reward that evaluates the effect of its action, which can then be used to adjust the individual strategy. This interaction with the environment at time step t forms an experience described by the tuple (s, a, r, s'). Through the utilization of prior experiences, an agent can acquire a strategy that enables it to select the optimal action within a given state, leading to the maximization of the long-term cumulative reward. Furthermore, deep reinforcement learning algorithms have the capacity to explore and exploit trade-offs, resulting in a reduction of risk and uncertainty during decision making processes, and enhancing adaptability to varying environmental conditions.

Through the utilization of prior experiences, an agent can acquire a strategy that enables it to select the optimal action within a given state, leading to the maximization of the long-term cumulative reward. Furthermore, deep reinforcement learning algorithms have the capacity to explore and exploit trade-offs, resulting in a reduction of risk and uncertainty during decision-making processes, and enhancing adaptability to varying environmental conditions.

A. State space

Under distributed resource allocation conditions, the transmitting vehicle of each V2V link can be considered as an agent that needs to perform resource allocation, including spectrum selection and power control. This problem can be solved by modeling it as a Markov process. Each intelligence can only observe local information and take actions based on the observation space. The actions of all agents constitute joint actions, and the agent adjusts their strategies based on rewards. The true environmental state contains all channel conditions and all actions of all agents, but is unknown to each agent. Local observations of each agent include information such as channel gain and interference. At each time slot, the state space includes various channel conditions and temporal information. Assuming that the location of the eavesdropper is known and the distribution of the Eve channel is known, the state space S in time slot t is:

$$S_{\mathrm{t}}^k = \{B_k, T, I_k[m], G_k[m]\}, \tag{23}$$

where $I_k[m]$ denotes full-band interference, B_k denotes the remaining V2V load that should be transmitted by the V2V link, T denotes the remaining time to satisfy the delay, and $G_k[m]$ denotes the channel conditions for all links, and:

$$\begin{aligned} G_k[m] = \{&g_k\left[\mathrm{m}\right], g_{k',k}\left[\mathrm{m}\right], g_{k,B}\left[\mathrm{m}\right], \\ &\hat{g}_{m,k}\left[m\right], g_{k,e}[m], \hat{g}_{m,e}[m]\}. \end{aligned} \tag{24}$$

B. Action space

In the distributed resource allocation in Mode 4, each agent selects an action based on the local observation to form a set

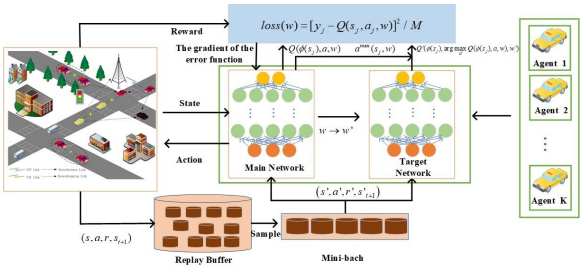

Fig. 2. Agorithm model

a, and the actions adopted by all agents form a joint action A_t. The set of actions is the set of actions that each agent selects based on the local observation to form a joint action. Assuming that there are M resource blocks in total, the V2V transmit power can be discretized and set to four levels such as [5,15,23,-100]dBm, so that each agent has $4 \times M$ actions, i.e., the dimension of the action space is $4 \times M$. At the same time the output layer of each neural network can be viewed as $4 \times M$. In the later simulation tests, the dimension of the actions of each DDQN itself in the multi-agent is drastically reduced compared to that of the single-agent strategy.

C. Reward function

In V2X communication, a reasonable reward function design can effectively optimize the performance and efficiency of the V2X communication system and guarantee the security and timeliness of data transmission. To ensure the quality and confidentiality of the communication, it is necessary to consider the interference between the frequency band and messages selected by each vehicle on other V2I links and other vehicles, and it is necessary to satisfy the time delay constraint. In order to maximize the total rate of all cellular links while guaranteeing the QoS requirements of V2V pairs. Therefore the reward function R_t can be defined as:

$$\begin{aligned} R_t = &\lambda_1 \sum_{m \in M} U(R_m^t - R_m^{\min}) \\ &+\lambda_2 \sum_{k \in K} U\left(\frac{P_k^d \alpha_k^t}{\sigma_e^2 + \sum_{k \neq k} \rho_{k'} P_{k'}^d \alpha_k^t} - \frac{\gamma_0}{\ln(\frac{1}{1-\xi})}\right), \end{aligned} \tag{25}$$

where $U(x)$ provides a penalty if the reward function is not satisfied, as follows:

$$U(x) = \begin{cases} x, & if \quad x \geq 0; \\ \beta, & if \quad x < 0. \end{cases} \tag{26}$$

where$\beta < 0$, λ_1, λ_2 are the weight coefficients of the rewards, respectively. At the time slot t, all the agents act independently, the global reward is evaluated at the BS, and each intelligence receives an evaluation of the global reward at the end of the task, which is broadcast by the BS with a smaller signaling.

V. Multi-agent reinforcement learning algorithms

In the scenario in this paper, as shown in Figure 2, each V2V link sending vehicle is used as an agent to construct a deep Q network to simulate the action value function. In the starting phase, the environment is initialized, vehicles and channels are randomly generated, and two neural networks are built for each agent, one for the main network and one for the target network, with the starting parameters set randomly and the same parameters for both neural networks. At the beginning of each episode, the vehicle position and large-scale fading are updated, and the vehicle position is updated every Tms. Each agent has an independent DDQN network, the input to the neural network is immediate local observations, and the output of the neural network is the Q-value of all actions. After all the agent have performed their actions, the system environment changes, generating a global reward while updating the state of all the agent, allowing the observation of the next state to be obtained. We follow a DDQN with a replay buffer, where the generated data is stored in memory. Each sample includes the(s_j, a_j, r_j, s_{j+1}). Small batches of data used to update the main network are sampled from memory in each iteration.

During the learning process, the target values are obtained from the target network, and it is worth noting here that instead of finding the Q values of individual actions directly in the target network as in DQN, the actions corresponding to the maximum Q values are first found inside the main network, i.e.

$$a^{\max}(s_j, w) = \arg\max_{a'} Q(\phi(s_j), a, w), \tag{27}$$

Using this selected action $a^{\max}(s_j, w)$ in the target network to calculate the target Q value, i.e.

$$y_j = R_j + \gamma Q'(\phi(s_j), a^{\max}(s_j, w), w'), \tag{28}$$

where w' denotes the parameters of the target network, w denotes the parameters of the main network. During the training process, the parameters of the main network are updated by minimizing the loss function, which is calculated as the mean square error between the target Q value and the Q value obtained from the main network. To update the network parameters, small batches of data samples are uniformly sampled from the buffer in each training session. The stochastic gradient descent (SGD) method is then employed to iteratively minimize the loss function:

$$loss(w) = [y_j - Q(s_j, a_j, w)]^2 / H. \tag{29}$$

where H is the size of the mini-batch.The network parameters are continuously updated, aiming to decrease the value of the loss function. As the loss function approaches a global minimum, the corresponding optimal policy for selecting modes and power levels in each V2V link can be derived. Initially, the strategy for mode and power level selection is random, but it gradually improves as the state-action value model is updated through training iterations. After completing a certain number of training iterations, the parameters of the behavioral network model are synchronized with the target network for the next learning phase. The training procedure is summarized in Algorithm 1.

And the algorithm can be trained offline, using stored historical data to simulate the behavior and environmental situations of multiple V2V links. And a training set is generated for training the transmitter-side intelligences of each V2V link, including combinations of states, actions, rewards, etc., for reinforcement learning training. This approach avoids the

Algorithm 1 MARL-DDQN-based training algorithm for security resource allocation

Initialize the environment
Initialize Replay Buffer Memory D
Initialize parameters of the main network and target network $w = w'$
for each episode **do**
  Reset the environment to obtain the initial state S
  for each step t **do**
    for each agent k **do**
      if In the current state s **then**
        Selecting actions with probability ε Randomness
      else
        use Eq.(27) Choose action
      end if
      Take an action in the current state s, receive a reward r and obtain
      Add (s, a, r, s_{t+1}) to the replay buffer memory D
    end for
  end for
  for each agent k **do**
    Randomly sample a batch of size batch size from $D(s_j, a_j, r_j, s_{j+1})$
    Calculate the target Q value
    Calculate the loss function according to the Eq.(29)
    Update the weights using gradient descent to minimize its loss function.
  end for
  Update the target network every C step.
end for

need for real-time training in real applications, thus reducing computational cost and latency.

VI. Simulation Results

A. Simulation environment

In this section, the proposed scenario optimization is simulated and validated for analysis. The design is based on the urban scenario specified in 3GPP TR 36.885. The MARL-DDQN for each agent consists of three fully connected hidden layers containing 500, 250 and 120 neurons respectively. The rectified linear unit Relu was used as the activation function and the RMSprop algorithm optimizer was used to update the training parameters. The remaining simulation parameters are shown Table I and Table II, and the channel mode are shown Table III and Table IV.

TABLE I
Simulation parameters

Parameters	Value
V2I transmit power P_{m}^{I}	0.2W
V2V transmit power P_{k}^{V}	[5,15,23,-100]dBm
Carrier frequency	2GHz
Bandwidth	4MHz
Noise power σ^2	-114dBm
BS antenna gain	8dBi
Vehicle antenna gain	3dBi
Time constraint	100ms

TABLE II
Neural Network Parameters

Parameters	Value
Replay Buffer size	200,000
Minni-batch size	2000
Start exploration rate	1
Final exploration rate	0.02
Initial learning rate	0.001
momentum	0.95
Active function	Relu
Optimizer	RMSprop

TABLE III
Channel Model For V2V Link

Parameters	V2V link
Path loss model	LOS in WINNER +B1 Manhattan[29]
Shadowing distribution	Log-narmal
Shadowing standard deviationψ	3dB
Decorrelation distance	10m
Path loss and shadowing update	A.1.4 in [26] every 100ms
Fast fading	Rayleigh fading
Fast fading update	Every 1 ms

TABLE IV
Channel Mode For V2I Link

Parameters	V2I Link
Path loss model	128.1+37.6log$_{10}d$
Shadowing distribution	Log-narmal
Shadowing standard deviationψ	8dB
Decorrelation distance	50m
Path loss and shadowing update	A.1.4 in [26] every 100ms
Fast fading	Rayleigh fading
Fast fading update	Every 1 ms

B. Baseline Algorithm

We iterated each agent's Q-network 1000 times, with an exploration rate linearly annealed from 1 to 0.02 over the first 800 iterations. Throughout this process, we fixed large-scale decay during iteration to make the algorithm more stable. To verify the effectiveness of our proposed algorithm, we compared it with different strategies:

1) Upper: maximizes communication efficiency between vehicles within a time limit without considering communication between vehicles and infrastructure. This means that the V2V link does not consider communication needs between vehicles and infrastructure when transmitting, but rather treats the problem as transferring B bytes of data through multiple steps within a given time frame, which can be regarded as the upper limit of performance achieved.

2) SARL-DDQN: It is indicated that the base station acts as a computational center and allocates resources based on the location of vehicles in the environment, channel, eavesdropper information, and traffic conditions. In this case, only the base station acts as an agent with the ability to make intelligent decisions. The base station gets the current moment state s_t, selects the action a_t, selects the resource block and transmit

power for all the transmitting vehicles of the V2V link based on the environment.There are four power levels for each V2V link, then there are 4^K actions for k V2V links, and the agent has $K! \times 4^K$ actions.

3) MADDPG: This algorithm uses the same centralized training and distributed execution scheme as MADDQN in the same scenario and has the same hyperparameters and discount factors. However, the difference between them lies in the neural network structure.The neural network of MARL-DDQN consists of a policy-based Actor network and a value-based Critic network.The Actor network evaluates the goodness of the Actor network's choice of action according to the state-action value function by means of the collected environment states.Since MADDPG applies to continuous action space and MARL-DDQN applies to discrete action space, we determine the range of values for continuous actions after mapping based on the discrete action space in MARL-DDQN. For discrete actions, the mapping is used to map them into a set of continuous actions to ensure that the action space variables are consistent, making the optimization objectives of the two algorithms consistent.

C. Simulation Results

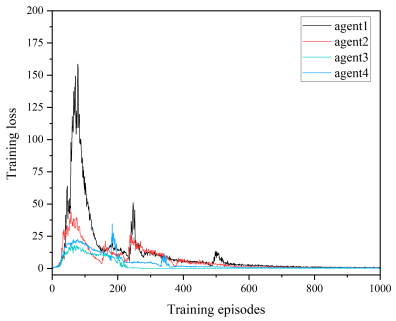

Fig. 3. The training loss of the MARL-DDQN algorithm

Fig 3 represents the change process of the loss function of each agent as the number of iterations increases under the simulation conditions of both the V2V link and the V2I link being 4, transmission payload B=3180bytes, and time delay T=100ms. From the Fig 3, we can see that the loss function gradually approaches 0 as the number of rounds increases, indicating that our proposed algorithm is convergent. And the loss function of each agent is different, indicating that the decision-making strategies of the agent are different, and each agent can pick different strategies according to its observations, trying to avoid vicious competition and make its overall performance develop towards optimization. When the network starts training, the loss gradually rises because the learning samples are relatively small, the neural network is updating, and less effective experience can be obtained. As the number of training sessions increases, the loss value gradually rises and then rapidly decreases, after which the training loss tends to stabilize, indicating that our proposed algorithm is capable of automatically updating the decision strategy and converging to the optimal solution according to the dynamics of the network.

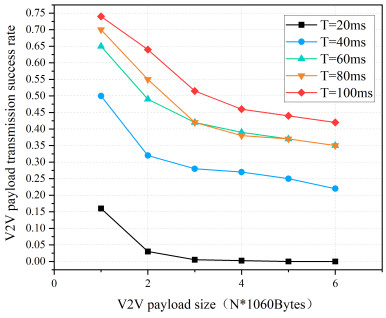

Fig. 4. Variation of transmission Security Rate with transmission payload within a Limited Time.

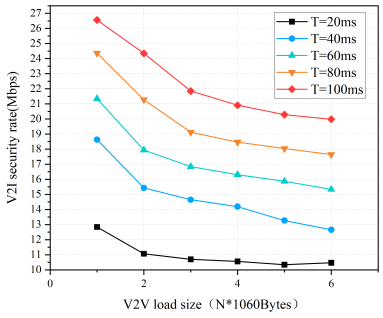

Fig. 5. Variation of V2I link occupancy with transmission payload for a limited period of time

The variation of the number of transmission payloads and the amount of completed transmission payloads under different time delay conditions of our proposed multi-agent training optimization algorithm can be seen in Fig 4. The success rate of the transmission payloads of the V2V link is the highest when T=100ms. so the time delay constraint set by this algorithm is 100ms. This is because as the number of payloads increases, the transmission success rate decreases due to the large number of tasks transmitted and the limited resources occupied by the spectrum under the constraint of limited time. In particular, at T=20ms, the payload transmission completions are below 15%.

Fig 5 again shows the best performance of the secrecy rate with a delay constraint of T=100ms. The Fig 5 shows that as the payload increases, the system needs to allocate more resources to ensure the reliable transmission of the V2V link, reducing the performance of the V2I system. As the delay constraint is reduced from 100ms to 40ms, the system secrecy rate of V2I gradually decreases, indicating that the system needs more resources for the V2V link to ensure that it can complete the transmission of the payload within the constraint, sacrificing some of the performance of the V2I link for this purpose. From the above analysis, it can be seen that the delay constraint has an important impact on the performance of

V2V and V2I links, and the appropriate selection of delay constraints can balance the performance of V2V and V2I systems while ensuring the success rate of load transfer.

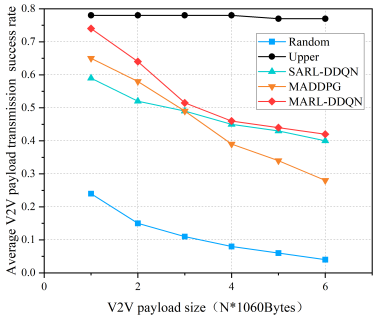

Fig. 6. Relationship between transmission success rate and transmission payload

Fig 6 shows the relationship between the transmission payload and the transmission success rate of the V2V link. In order to accurately demonstrate the better of the proposed algorithm, the baseline Random, representing the lower bound of algorithm performance. The experimental results show that the transmission success rate of the three optimization strategies will decrease as the payload increases, which is consistent with the trend expressed by the baseline, verifying the correctness of our algorithm. The overall downward trend indicates that the more data the V2V link transmits, the more likely the transmission failure will occur, resulting in a decline in the transmission success rate. The experimental results demonstrate that our proposed MARL-DDQN algorithm outperforms the single-agent SARL-DDQN strategy and MADDPG strategy under different payload sizes.

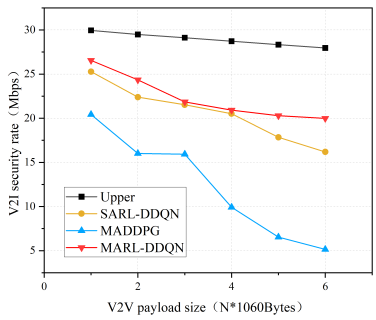

Fig. 7. Security rate in relation to transmitted payload

Fig 7 represents the differences between different algorithms in optimizing the rate and performance of V2I links under different V2V link payload sizes. The experimental results show that the trend of our proposed method and the baseline direction are the same, which in turn proves the effectiveness of our proposed method. As the V2V link payload increases, all optimized V2I secrecy rates are trending downwards. The reason for the decrease is that as more payloads are required to be transmitted, a longer transmission time is needed, which may cause an increase in V2V to transmit power to optimize the performance of the V2V link, so the agent needs to make an optimal choice of resource blocks and power and to improve the transmission success of the V2V payload, the increase in V2V payload leads to stronger interference with the V2I link for a longer period of time, affecting communication capacity to become smaller, making the communication rate decrease.

Among the three compared algorithms, it can be seen that our proposed algorithm has certain superiority. And it can be seen from the Fig 7 that the proposed algorithm slows down the drop rate when the payload reaches B=3180bytes. Combined with the analysis in Fig 6, it can be seen that the V2V transmission payload increases, the average V2V link's payload transmission success rate decreases to less than 50% when the payload reaches B=3180bytes, and the V2V link do not produce interference to the V2I link after its payload transmission is completed interference, so the secrecy performance of the V2I link is buffered, resulting in a slower decline.

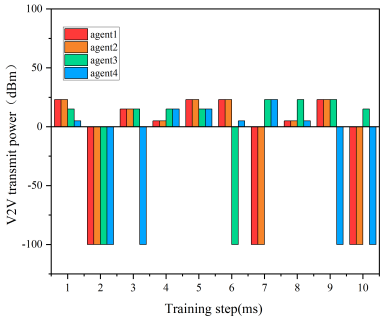

Fig. 8. Power selected for each step for each agent

Fig 8 shows the action (V2V transmit power) selected at each step for each link instantaneous rate. From this Fig 8, it can be seen that each link as an agent selects the transmit power in real-time at each step in the episode according to the Markov process training, and in this experiment four power levels of [5,15,23,-100] dBm were set. It can be seen from that each agent can choose the appropriate transmit power level for itself at each step according to the current state, further demonstrating the effectiveness of the distributed collaboration of multiple intelligences in this paper's algorithm.

In Fig 9, we further demonstrate the average V2I sub-band rate occupied by each vehicle and its neighbors at each step for the instantaneous rates of the four links. The resource allocation strategies of MARL-DDQN and the random policy are shown in Fig 9 and Fig 10, respectively. From Fig 9, it can be observed that by adopting the approach proposed in this study, Agent 1 and Agent 4 achieve high transmission rates, fully utilizing the good channel conditions of the channel, while Agent 2 and Agent 4 maintain lower rates. When the transmission payload of Agent 2 increases, its rate also increases accordingly. Each link has an agent strategy that adapts to the current state and flexibly occupies the V2I link bandwidth. The comparison between the two Figs shows that the overall rate of our proposed method is superior to

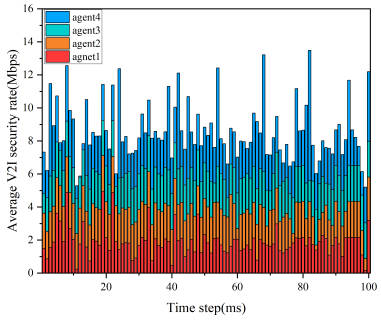

Fig. 9. MARL-DDQN

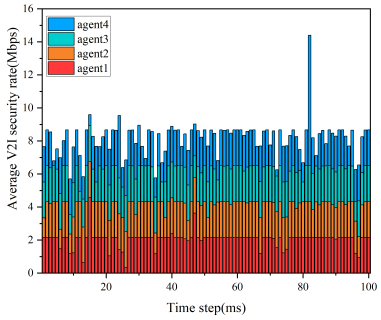

Fig. 10. Random

that of the random policy. In contrast, Fig 10 illustrates that when each vehicle needs data transmission upon arrival at the intersection, the random policy equally occupies the frequency sub-bands of the V2I link. This situation may result in a sudden change as shown in Fig 9 when the amount of transmitted data is particularly large, which would lead to an increase in both the rate and bandwidth utilization. From the above analysis, it can be seen that the method proposed in this paper is able to flexibly and adaptively occupy the frequency band of the V2I link according to the current state in terms of resource allocation, which enables each vehicle to obtain a higher rate when transmitting. In contrast, the random strategy leads to poor resource utilization.

VII. Conclusion

In this paper, we have investigated the resource allocation problem in vehicular communications. Considering the different QoS requirements of V2X communication, our objective is to maximize the sum rate of V2I link occupancy while ensuring the maximum transmission rate of cellular links and meeting the reliability and delay requirements of V2V communication. We also introduced an eavesdropping model to ensure secure communication between links, modeled it using the Markov decision process, and proposed a resource allocation algorithm based on multi-agent reinforcement learning (MARL-DDQN) to address continuous action and power control issues. This work has important implications for improving the QoS of vehicular communications, particularly in enhancing the anti-eavesdropping ability in complex environments.

We believe that the application of IRS technology in Telematics has great potential in future research. Intelligent Reflective Surface (IRS), as an intelligent reflective technology, will become one of the important trends in the field of intelligent transportation and wireless communication in the future joint research with Telematics. By intelligently reflecting and controlling the propagation path of electromagnetic waves, IRS can improve the utilization efficiency of the spectrum, thus achieving higher data rate and capacity and meeting the demand for high rate communication for future 6G communication. Specifically, we can explore how to combine IRS with V2V and V2I communications to improve communication quality, data rate, and network capacity. In addition, we can also investigate how to optimize the deployment strategy of IRS to achieve the best signal coverage and link performance. However, the current research faces some challenging issues. First, how to effectively design and deploy IRS networks is an important issue that needs to take into account the dynamic changes in vehicle motion, channel characteristics, and network topology. Second, how to optimize the resource allocation and power control strategies of IRS to maximize the communication performance is also a challenging issue. In addition, issues such as signal interference and privacy security associated with IRSs need to be addressed. By addressing the challenging issues currently faced, the field of intelligent transportation and wireless communication can be further advanced.

References

[1] Contreras-Castillo J, Zeadally S, Guerrero-Ibañez J A. Internet of Vehicles: Architecture, Protocols, and Security[J]. IEEE Internet of Things Journal, 2018, 5(5): 3701-3709.

[2] Chen Y, Wang Y, Zhang J, et al. QoS-Driven Spectrum Sharing for Reconfigurable Intelligent Surfaces (RISs) Aided Vehicular Networks[J]. IEEE Transactions on Wireless Communications, 2021, 20(9): 5969-5985.

[3] Alnasser A, Sun H, Jiang J. QoS-Balancing Algorithm for Optimal Relay Selection in Heterogeneous Vehicular Networks[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(7): 8223-8233.

[4] Zhou H, Xu W, Chen J, et.al. Evolutionary V2X Technologies Toward the Internet of Vehicles: Challenges and Opportunities[J]. Proceedings of the IEEE, 2020, 108(2): 308-323.

[5] Cao X, Liu L, Cheng Y, et.al. Towards Energy-Efficient Wireless Networking in the Big Data Era: A Survey[J]. IEEE Communications Surveys Tutorials, 2018, 20(1): 303-332.

[6] Makarfi A U, Rabie K M, Kaiwartya O, et.al. Toward Physical-Layer Security for Internet of Vehicles: Interference-Aware Modeling[J]. IEEE Internet of Things Journal, 2021, 8(1): 443-457.

[7] Hooshmand R, Aref M R. Efficient Polar Code-Based Physical Layer Encryption Scheme[J]. IEEE Wireless Communications Letters, 2017, 6(6): 710-713.

[8] Chen D, Zhang N, Cheng N, et.al. Physical Layer based Message Authentication with Secure Channel Codes[J]. IEEE Transactions on Dependable and Secure Computing, 2020, 17(5): 1079-1093.

[9] Çatak E, Ata L D, Mantar H A. Enhanced physical layer security by OFDM signal transmission in fractional Fourier domains[C]//2015 23nd Signal Processing and Communications Applications Conference (SIU). 2015: 1881-1884.

[10] Oh H, Yoo J, Kim C kwon, et.al. A novel mobility management for seamless handover in vehicle-to-vehicle/vehicle-to-infrastructure (V2V/V2I) networks[C]//2009 9th International Symposium on Communications and Information Technology. 2009: 259-260.

[11] Wang L, Yang C, Hu R Q. Autonomous Traffic Offloading in Heterogeneous Ultra-Dense Networks Using Machine Learning[J]. IEEE Wireless Communications, 2019, 26(4): 102-109.

[12] Pan S, Zhang X, Sung D K. Intelligent Reflecting Surface-Aided Centralized Scheduling for mmWave V2V Networks[C]//2022 International Conference on Computer Communications and Networks (ICCCN). 2022: 1-10.
[13] Liang L, Kim J, Jha S C, et.al. Spectrum and Power Allocation for Vehicular Communications With Delayed CSI Feedback[J]. IEEE Wireless Communications Letters, 2017, 6(4): 458-461.
[14] Yang H, Zheng K, Zhao L, et.al. Twin-Timescale Radio Resource Management for Ultra-Reliable and Low-Latency Vehicular Networks[J]. IEEE Transactions on Vehicular Technology, 2020, 69(1): 1023-1036.
[15] Liang L, Ye H, Li G Y. Spectrum Sharing in Vehicular Networks Based on Multi-Agent Reinforcement Learning[J]. IEEE Journal on Selected Areas in Communications, 2019, 37(10): 2282-2292.
[16] Miao J, Chai X, Song X, et.al. A DDQN-based Energy-Efficient Resource Allocation Scheme for Low-Latency V2V communication[C]//2022 IEEE 5th International Electrical and Energy Conference (CIEEC). 2022: 53-58.
[17] Li X, Lu L, Ni W, et.al. Federated Multi-Agent Deep Reinforcement Learning for Resource Allocation of Vehicle-to-Vehicle Communications[J]. IEEE Transactions on Vehicular Technology, 2022, 71(8): 8810-8824.
[18] Zhang X, Peng M, Yan S, et al. Deep-Reinforcement-Learning-Based Mode Selection and Resource Allocation for Cellular V2X Communications[J]. IEEE Internet of Things Journal, 2020, 7(7): 6380-6391.
[19] Jameel F, Javed M A, Zeadally S, et.al. Secure Transmission in Cellular V2X Communications Using Deep Q-Learning[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(10): 17167-17176.
[20] Wang L, Liu J, Chen M, et.al. Optimization-Based Access Assignment Scheme for Physical-Layer Security in D2D Communications Underlaying a Cellular Network[J]. IEEE Transactions on Vehicular Technology, 2018, 67(7): 5766-5777.
[21] Wang W, Teh K C, Li K H. Enhanced Physical Layer Security in D2D Spectrum Sharing Networks[J]. IEEE Wireless Communications Letters, 2017, 6(1): 106-109.
[22] Liu Y, Wang W, Chen H H, et.al. Secrecy Rate Maximization via Radio Resource Allocation in Cellular Underlaying V2V Communications[J]. IEEE Transactions on Vehicular Technology, 2020, 69(7): 7281-7294.
[23] Ai Y, deFigueiredo F A P, Kong L, et.al. Secure Vehicular Communications Through Reconfigurable Intelligent Surfaces[J]. IEEE Transactions on Vehicular Technology, 2021, 70(7): 7272-7276.
[24] Liang L, Li G Y, Xu W. Resource Allocation for D2D-Enabled Vehicular Communications[J]. IEEE Transactions on Communications, 2017, 65(7): 3186-3197.
[25] Ji B, Huang J, Wang Y, et.al. Multi-Relay Cognitive Network With Anti-Fragile Relay Communication for Intelligent Transportation System Under Aggregated Interference[J]. IEEE Transactions on Intelligent Transportation Systems, 2023: 1-10.
[26] Technical Specification Group Radio Access Network; Study LTE-Based V2X Services; (Release 14), document 3GPP TR 36.885 V14.0.0, 3rd Generation Partnership Project, Jun. 2016
[27] Kandukuri S, Boyd S. Optimal power control in interference-limited fading wireless channels with outage-probability specifications[J]. IEEE Transactions on Wireless Communications, 2002, 1(1): 46-55.
[28] Papandriopoulos J, Evans J, Dey S. Optimal power control for Rayleigh-faded multiuser systems with outage constraints[J]. IEEE Transactions on Wireless Communications, 2005, 4(6): 2705-2715.
[29] Kyösti P, Meinilä J, Hentila L, et.al. IST-4-027756 WINNER II D1.1.2 v1.2 WINNER II channel models[J]. Inf. Soc. Technol, 2008, 11.