毫米波车联网多基站多用户下的安全传输方案

俱莹[1]，陈宇超[2]，田素恒[1]，刘雷[1]，李赞[1]，裴庆祺[1]，王明阳[3]

（1. 西安电子科技大学通信工程学院，陕西 西安 710077；2. 深圳市国电科技通信有限公司，广东 深圳 518000；
3. 北京跟踪与通信技术研究所，北京 100085）

摘　要：针对移动场景下毫米波安全波束形成的时效性和毫米波动态窃听场景下安全连接的鲁棒性等问题，提出了一种基于决斗双重深度Q网络（D3QN）–深度确定性策略梯度（DDPG）算法的多智能体安全协作通信方案。该方案利用路侧单元（RSU）辅助的协作干扰技术降低窃听者对合法信号的接收质量，并通过联合优化车辆用户（VU）的基站与波束连接控制、阻塞RSU的选择，以及RSU的协作干扰方向和发射功率控制，使得所有合法车辆的总保密传输速率最大化。在此基础上，针对车联网的高动态性，通过构建基于D3QN的VU智能体和基于D3QN-DDPG的RSU智能体，实现了实时的离散–连续混合决策。最后，通过多维度的性能分析和方案对比实验，验证了所提方案的有效性。

关键词：毫米波车联网；多智能体；物理层安全；深度强化学习

中图分类号：TN92

文献标志码：A

DOI：10.11959/j.issn.1000–436x.2024149

Secure transmission scheme for millimeter-wave Internet of vehicles with multiple base stations and users

JU Ying[1], CHEN Yuchao[2], TIAN Suheng[1], LIU Lei[1], LI Zan[1], PEI Qingqi[1], WANG Mingyang[3]

1. School of Telecommunications Engineering, Xidian University, Xi'an 710077, China
2. China Gridcom CO., LTD Shenzhen City, Shenzhen, 518000 China
3. Beijing Institute of Tracking and Communication Technology, Beijing 100085, China

Abstract: For the timeliness of millimeter-wave secure beamforming in mobile scenarios and the robustness of secure connections in millimeter-wave dynamic eavesdropping scenarios, a multi-agent secure cooperative communication scheme based on a dueling double deep Q network (D3QN)-deep deterministic policy gradient (DDPG) algorithm was proposed to address communication security issues. The scheme utilized road side unit (RSU)-assisted cooperative jamming technology to reduce the eavesdropper's reception quality of confidential signals. The optimization problem was formulated to maximize the total secrecy rate of all legitimate vehicles by optimizing the joint base station and beam connection control of the VUs, the selection of the jamming RSUs, and the cooperative jamming direction and transmit power of RSUs. Furthermore, for the challenges posed by the high dynamics of vehicular networks, the scheme achieved a fusion of real-time discrete and continuous decision-making by creating a VU agent grounded in D3QN and an RSU agent harnessing D3QN-DDPG capabilities. Finally, through multi-dimensional performance analysis and scheme comparison experiments, simulation results demonstrate the effectiveness of the proposed scheme.

Keywords: millimeter-wave Internet of vehicles, multi-agent, physical layer security, deep reinforcement learning

---

收稿日期：2024–03–12；**修回日期：**2024–07–26

通信作者：王明阳，wangmy2000@163.com

基金项目：国家自然科学基金资助项目（No.62102301, No.62132013）；国家杰出青年科学基金项目（延续资助）（No.62425103）；科学探索奖基金资助项目

Foundation Items： The National Natural Science Foundation of China (No.62102301, No.62132013), The National Science Foundation for Distinguished Young Scholars of China (No.6242510), XPLORER PRIZE

0　引言

车联网（IoV, Internet of vehicles）作为一种新兴的技术，实现了车辆与其他交通设备或互联网设施之间的互联互通[1-2]。通过信息网络平台对车辆以及道路信息的收集和有效利用，车联网能够为不同的用户提供个性化的服务，满足不同场景（如自动驾驶、交通协调、娱乐应用等）的业务需求。然而，在车联网不断发展过程中，无线通信数据量和终端设备数量呈指数级增长，并且微波频谱资源拥挤不堪，导致无法满足用户多样化的服务需求。拥有丰富频谱资源的毫米波（mmWave, millimeter wave）通信技术为此提供了解决方案[3-4]。

随着毫米波车联网通信技术的日益成熟，其安全问题（如保密信息泄露、非法截获信息、窃听攻击等）层出不穷[5]。这是由于车联网包含多种通信模式，且信息交互频率更高，而无线信道的开放性导致通信中断和机密信息泄露的可能性增大，对用户的服务质量和隐私安全构成了巨大的威胁[6-7]。因此，在设计毫米波车联网通信系统时，必须考虑信号传输的安全性。作为最底层的安全机制，物理层安全技术为保障车联网通信系统的安全传输提供了有效的解决方案[8-10]。协作干扰技术作为物理层安全的重要手段之一，通过向特定区域发射人工噪声或阻塞信号，能够有效地降低潜在恶意的窃听者接收机密信号的质量，使其难以解码，增强车联网通信过程的安全性[11]。物理层安全方案的设计依赖于物理层信道，而毫米波的传播特性和信道特点不同于传统的微波频段，这给车联网中安全方案的设计带来了新的挑战。首先，毫米波的窄波束定向通信需要发射端和接收端之间进行精确的波束对准，而车联网中节点移动速度快，波束需要快速变化以跟踪节点移动，安全策略也需要快速与高时变的波束对准相匹配，这使得设计难度进一步提升。其次，毫米波穿透性差，在高动态的车联网下直连链路的可靠性难以保证，可能无法提供用以对抗窃听者的容量冗余，使通信安全面临威胁，因此，兼顾链路稳定性和通信安全性要求，实现高动态场景下安全链路的稳定通信也极具挑战。

针对上述问题，研究者们对毫米波车联网中的物理层安全问题进行了研究。文献[12]在车辆对车辆（V2V, vehicle to vehicle）的通信场景下，提出了毫米波单射频链路下的天线子集调制安全方案和多射频链路下的人工噪声安全方案。然而，该文献仅考虑了点对点通信问题，未在整个网络场景下设计安全方案。基于此，文献[7]研究了蜂窝车联网中上行毫米波通信链路的物理层安全问题，提出了基于最小距离和最大功率的 2 种上行链路连接方案，并评估了上行链路传输的安全性。文献[13]提出了一种基于阻塞和功率的协作干扰选择策略，以解决毫米波蜂窝车联网中潜在的安全隐患。上述相关研究验证了人工噪声、协作干扰等物理层安全技术在毫米波车联网中的有效性，能够有效地对抗窃听攻击，提升系统的安全性。然而上述研究采用了泊松过程对场景进行静态建模，并没有考虑车辆的移动性和节点间的时空关联性。因此，毫米波车联网物理层安全方案设计中的挑战并未得到充分解决。

车辆快速移动下的毫米波窄波束切换和安全链路的选择挑战，使得在进行安全传输方案的设计时需要考虑如何满足实时决策的需求，而数学方法难以推导出低复杂度的最优解，传统的优化方法难以适应车联网的高动态性[14]。车辆固有的移动特性使得信道状态信息与当前的交通模式具有时间相关性和空间相关性。如果能够学习其中隐式的关联性，就可以利用历史的局部信息来预测未来的网络状态信息，从而实时地做出最优的传输决策[15]。近年来，许多研究人员利用机器学习、深度学习、强化学习等人工智能算法学习车联网场景中交通模式的相关性，解决了毫米波车联网中不同类型的问题[14,16-18]。文献[14]针对毫米波车联网中的多基站通信场景，提出了一种基于强化学习的毫米波车联网协同波束选择算法，将波束选择问题建模为多臂老虎机问题，并采用基于 Q 学习的算法学习如何协调波束选择决策。文献[16]提出了一种基于上下文感知的快速机器学习算法，旨在通过实时的最佳波束分配来最大化车联网系统总容量。针对毫米波单基站中的通信问题，文献[17]设计了一种基于深度 Q 网络（DQN, deep Q network）的联合波束分配和中继选择方法，以对抗车联网中的随机阻塞效应和提升毫米波车联网通信的有效性。针对车辆网络的多维资源管理问题，文献[18]设计了一种基于多智能体深度确定性策略梯度（DDPG, deep deter-

ministic policy gradient）的方法来求解分布式优化问题。上述相关研究验证了人工智能算法在毫米波车联网中是有效的，能够提供车联网中快速决策的有效手段，但仅考虑了通信的有效性和可靠性，并未针对该场景下的通信安全问题提出解决方案。

基于此，本文针对毫米波车联网中的通信安全问题，设计了一个基于协作干扰的多基站安全传输方案。该方案考虑了一个多窃听者的动态窃听场景，并通过设计基于路侧单元（RSU, road side unit）辅助的协作干扰策略降低多个窃听者对合法信号的接收质量，从而提高车联网的安全传输性能。相应地，本文采用系统保密传输速率和平均保密连接概率作为主要指标，对毫米波车联网进行全面的性能分析。本文的主要贡献总结如下。

1) 本文提出了一种毫米波车联网中的安全协作传输方案，旨在应对多基站多用户环境下的多窃听者动态窃听问题。该方案通过RSU辅助的协作干扰技术降低潜在窃听者的信号接收质量，使其解码困难，从而增强车联网通信的安全性。该方案联合优化了车辆用户（VU, vehicular user）的基站与波束连接控制、阻塞RSU的选择，以及RSU设备的协作干扰方向和发射功率，以实现合法车辆总保密传输速率的最大化。

2) 本文基于排队理论设计了车辆的动态模型，以更真实地模拟车辆的到达和行驶过程。本文通过搭建多用户安全通信模型，推导出系统保密传输速率的表达式和联合优化问题。将复杂的联合优化问题建模为多智能体联合决策问题，并搭建了一个基于多智能体的安全协作通信系统。系统中多个不同类型的智能体节点通过通信模块进行信息交互，相互协作，共同提升毫米波车联网通信的有效性和安全性。

3) 本文首先针对VU的离散决策空间特点设计了基于决斗双重深度Q网络（D3QN, dueling double deep Q network）的VU智能体，针对RSU的混合决策空间特点设计了基于D3QN-DDPG的RSU智能体。然后，针对不同的智能体构建独立的马尔可夫决策过程，实现离散-连续的混合决策控制。在此基础上，本文设计了基于个体车辆保密传输速率的奖励惩罚机制，以平衡系统整体优化目标与个体合法车辆用户的安全通信质量间的决策冲突，避免了不合理的系统优化。

1 系统模型

1.1 多基站多用户的毫米波车联网模型

本文首先考虑一个多基站多用户的毫米波车联网架构，多个毫米波基站能够同时为多个车辆用户提供下行链路数据传输服务。场景中基站集合用 $B=\{B_i,i=1,\cdots,N_{\mathrm{B}}\}$ 表示，合法车辆用户集合用 $V_{\mathrm{T}}=\{v_k^{(\mathrm{T})},k=1,\cdots,N_{\mathrm{T}}\}$ 表示，其中 B_i 为第 i 个基站，$v_k^{(\mathrm{T})}$ 为第 k 个合法车辆，N_{B} 和 N_{T} 分别为基站和合法车辆用户的数量。此外，每个毫米波基站拥有多个定向窄波束，定向波束的宽度与基站所拥有的天线数目有关，同时能够为不同的用户提供服务，且毫米波基站最多能同时服务的车辆数量受到射频链路数量 N_{RF} 的限制。所有基站的波束集合用 $L^{(\mathrm{B})}=\{L_i^{(\mathrm{B})},i=1,\cdots,N_{\mathrm{B}}\}$ 表示，其中 $L_i^{(\mathrm{B})}=\{L_{i,l}^{(\mathrm{B})},l=1,\cdots,N_{i,\mathrm{L}}^{(\mathrm{B})}\}$ 表示第 i 个基站的波束集合，$L_{i,l}^{(\mathrm{B})}$ 表示第 i 个基站的第 l 个波束，$N_{i,\mathrm{L}}^{(\mathrm{B})}$ 为第 i 个基站的可分配波束总数。本文主要考虑一个动态窃听场景，道路上存在多个移动的恶意窃听车辆，每个窃听者配备单根天线进行全向信号接收，意图窃听合法通信链路的机密信息。多基站多用户的毫米波车联网模型如图1所示。为了保障车联网通信系统的安全性，路边配置有多个RSU设备，通过设计RSU辅助的协作干扰方案（如发射波束选择和发射功率控制），向场景中的特定区域（如图1中的虚线区域）发射阻塞信号，降低窃听者接收合法信号的质量。相应地，窃听车辆集合用 $V_{\mathrm{E}}=\{v_n^{(\mathrm{E})},n=1,\cdots,N_{\mathrm{E}}\}$ 表示，RSU集合用 $R=\{R_j,j=1,\cdots,N_{\mathrm{R}}\}$ 表示，其中 v_n^{E} 为第 n 个窃听者，R_j 为第 j 个RSU，N_{E} 和 N_{R} 分别为窃听者和RSU的数量。每个RSU拥有多个定向窄波束，所有RSU的波束集合用 $L^{(\mathrm{R})}=\{L_j^{(\mathrm{R})},j=1,\cdots,N_{\mathrm{R}}\}$ 表示，其中 $L_j^{(\mathrm{R})}=\{L_{j,l}^{(\mathrm{R})},l=1,\cdots,N_{j,\mathrm{L}}^{(\mathrm{R})}\}$ 表示第 j 个RSU的波束集合，$L_{j,l}^{(\mathrm{R})}$ 表示第 j 个RSU的第 l 个波束，$N_{j,\mathrm{L}}^{(\mathrm{R})}$ 表示第 j 个RSU的可选波束总数。

在每个时隙（服务周期的起始点），毫米波车联网的服务过程主要分为3步：① 根据先申请先服务的原则，从道路中确定合法车辆集合 V_{T}；② 合法车辆用户根据当前状态信息，如基站位置、当前

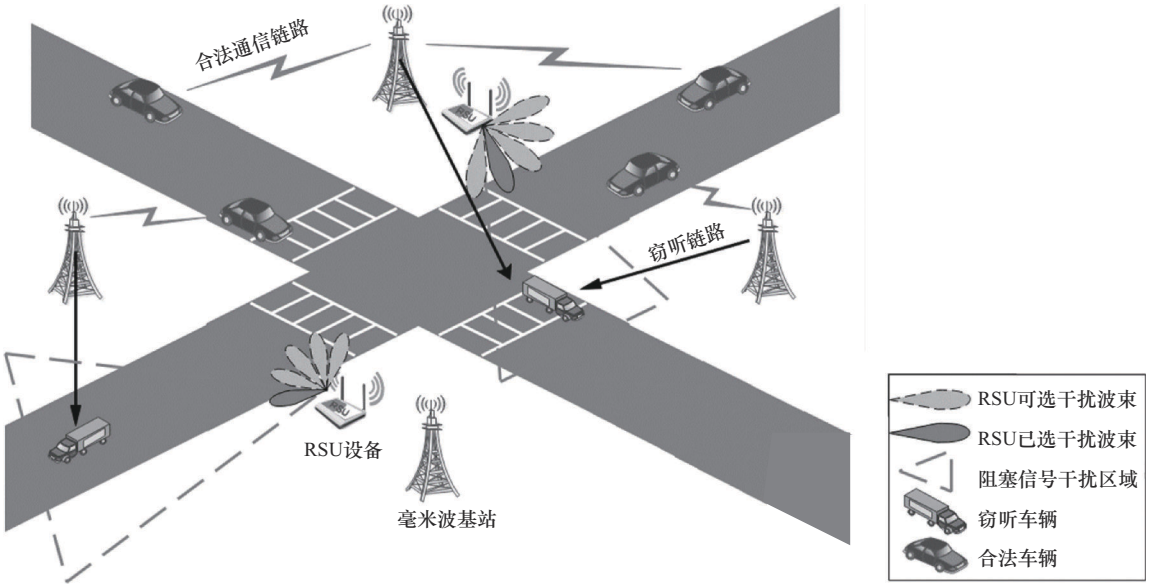


图1　多基站多用户的毫米波车联网模型

位置等，从多个基站和多个波束中选择一个基站和波束的组合进行连接；③ 位于道路旁的RSU设备也根据当前状态信息，如合法车辆和窃听者等，决定是否发射阻塞信号、选择哪个波束发射阻塞信号，以及发射阻塞信号功率的大小。

为了更好地描述实际的车联网场景，本文利用排队理论来模拟车辆的到达过程[19]，即车辆到达的时间间隔Δt服从负指数分布，其概率密度函数表示为

$$f(\Delta t)=\begin{cases}\lambda \mathrm{e}^{-\lambda \Delta \mathrm{t}}, \Delta t \geqslant 0 \\ 0,\text{其他}\end{cases} \tag{1}$$

其中，λ是车辆的平均到达率，通过灵活调整平均到达率能够模拟不同的交通流量密度。当λ较大时，车联网通信场景为高峰时间段；当λ较小时，车联网通信场景为空闲时间段。

1.2　通信模型

1) 天线增益。网络中发射端和接收端均部署了多天线阵列，以实现数据传输的波束形成，补偿毫米波信号的高路径损耗。为了便于分析，本文采用毫米波扇区模型来近似天线方向图[20-21]，所有波束具有相同的波束宽度。因此天线阵列增益的表达式为

$$G_u(\theta)=\begin{cases}M_u, |\theta|<\dfrac{\theta_u}{2} \\ m_u,\text{其他}\end{cases} \tag{2}$$

其中，$u \in \{\mathrm{B,V,R}\}$分别表示车联网中的毫米波基站、车辆节点和RSU设备，θ_u为天线阵列的主瓣波束宽度，M_u和m_u分别表示主瓣和旁瓣的波束增益。此外，窃听者全向接收，其天线增益用G_{E}表示。

2) 路径损失模型。从发射端（毫米波基站或RSU）到接收端（合法车辆或窃听车辆）的路径损耗计算式为

$$L_{\mathrm{t,r}}=\mu_1 \lg\left(f_{\mathrm{c}}\right)+\mu_2 \lg\left(d_{\mathrm{e}}\right)+\mu_3 \tag{3}$$

其中，f_{c}是载波频率，d_{e}表示发射端和接收端之间的距离，μ_1、μ_2和μ_3表示路径损失的强度[22]。

3) 信道增益。当使用窄波束发射时，毫米波信道的小尺度衰落小于微波信道的小尺度衰落，且在毫米波信号的传输过程中，大尺度衰落的影响远高于小尺度衰落的信号衰减[23]。因此本文在模型中忽略了小尺度衰落因素的影响。此外，毫米波信号具有低穿透性的特点，对传播环境的阻塞（如车联网场景中的建筑物、树木、广告牌等）十分敏感，即毫米波信号在穿透阻塞时的能量急剧衰减[24]。因此，本文考虑了阻塞效应对传输性能的影响。在每个服务期间，发射端和接收端间的信道增益$g_{\mathrm{t,r}}$表示为

$$g_{\mathrm{t,r}}=\alpha_{\mathrm{t,r}} L_{\mathrm{t,r}} \tag{4}$$

其中，$\alpha_{\mathrm{t,r}}$为发射端到接收端所历经的阻塞穿透系数。相应地，从第i个毫米波基站到第k个合法车

辆的信道增益表示为$g_{i,k}^{(\mathrm{B})}$，从第i个毫米波基站到第n个窃听车辆的信道增益表示为$g_{i,n}^{(\mathrm{B})}$，从第j个RSU到第k个合法车辆的信道增益表示为$g_{j,k}^{(\mathrm{R})}$，从第j个RSU到第n个窃听车辆的信道增益表示为$g_{j,n}^{(\mathrm{R})}$。

2　安全协作传输方案

在毫米波车联网场景中，由于存在多个动态的恶意窃听者，因此本文通过设计RSU设备的联合波束选择和发射功率控制方案，向特定的区域发送阻塞信号，与合法车辆相互协作共同提升系统的安全性能。在一个服务周期内，合法车辆可能受到来自其他合法链路的阻塞信号$I_{k,k'}$和来自RSU的阻塞信号$I_{k,\mathrm{R}}$的干扰，分别表示为

$$I_{k,k'}=\sum_{i=1}^{N_{\mathrm{B}}}\sum_{l=1}^{N_{i,\mathrm{L}}^{\mathrm{B}}}\sum_{k'=1,k'\neq k}^{N_{\mathrm{T}}}x_{i,l}[k']P_{\mathrm{B},i}g_{i,k}^{(\mathrm{B})}G_{i,l}^{(\mathrm{B})}G_{\mathrm{V},k}\tag{5}$$

$$I_{k,\mathrm{R}}=\sum_{j=1}^{N_{\mathrm{R}}}\sum_{l=1}^{N_{j,\mathrm{L}}^{\mathrm{R}}}y_{j,l}P_{\mathrm{R},j}g_{j,k}^{(\mathrm{R})}G_{j,l}^{(\mathrm{R})}G_{\mathrm{V},k}\tag{6}$$

其中，$P_{\mathrm{B},i}$为第i个基站的发射功率，$P_{\mathrm{R},j}$为第j个RSU的发射功率；$G_{i,l}^{(\mathrm{B})}$为第i个基站的第l个波束增益，$G_{\mathrm{V},k}$为第k个合法车辆的波束增益；$x_{i,l}[k]$为二进制联合基站和波束选择的指示变量，当$x_{i,l}[k]=1$时表示第i个基站的第l个波束为第k个用户服务，否则$x_{i,l}[k]=0$；$y_{j,l}$为二进制RSU干扰区域的指示变量，当$y_{j,l}=1$时表示第j个RSU选择第l个波束发射阻塞信号，否则$y_{j,l}=0$。相应地，第k个合法车辆的信干噪比为

$$\xi_k=\frac{\sum_{i=1}^{N_{\mathrm{B}}}\sum_{l=1}^{N_{i,\mathrm{L}}^{\mathrm{B}}}x_{i,l}[k]P_{\mathrm{B},i}g_{i,k}^{(\mathrm{B})}M_{\mathrm{B}}M_{\mathrm{V}}}{I_{k,k'}+I_{k,\mathrm{R}}+\sigma^2}\tag{7}$$

其中，σ^2为信道噪声功率，M_{B}为毫米波基站主瓣的波束增益，M_{V}为车辆节点主瓣的波束增益。因此，第k个合法车辆的容量表达式为$C_{\mathrm{b}}^k=W\mathrm{lb}(1+\xi_k)$，其中$W$为带宽。

本文考虑窃听者拥有较强的窃听能力，能同时窃听多个不同位置的合法车辆。因此，对于第k个合法车辆，第n个窃听者受到的可能干扰主要来自RSU所发射的阻塞信号，其表达式为

$$I_{n,\mathrm{R}}=\sum_{j=1}^{N_{\mathrm{R}}}\sum_{l=1}^{N_{j,\mathrm{L}}^{\mathrm{R}}}y_{j,l}P_{\mathrm{R},j}g_{j,n}^{(\mathrm{R})}G_{j,l}^{(\mathrm{R})}G_{\mathrm{E}}\tag{8}$$

因此，对于第k个合法车辆，第n个窃听者的信干噪比为

$$\xi_{\mathrm{e},n}^k=\frac{\sum_{i=1}^{N_{\mathrm{B}}}\sum_{l=1}^{N_{i,\mathrm{L}}^{(\mathrm{B})}}x_{i,l}[k]P_{\mathrm{B},i}g_{i,n}^{(\mathrm{B})}G_{i,l}^{(\mathrm{B})}G_{\mathrm{E}}}{I_{n,\mathrm{R}}+\sigma^2}\tag{9}$$

相应地，窃听者的窃听信道容量为$C_{\mathrm{e},n}^k=W\mathrm{lb}\left(1+\xi_{\mathrm{e},n}^k\right)$。

本文主要考虑了一个非共谋的、多窃听者的窃听场景，因此第k个合法车辆的保密传输速率为

$$C_{\mathrm{s}}^k=\max\left\{0,C_{\mathrm{b}}^k-\max_{v_n^{\mathrm{E}}\in V_E}C_{\mathrm{e},n}^k\right\}\tag{10}$$

为了描述本文所优化的决策变量，本文定义了3个指示变量集合：$\boldsymbol{X}=\{x_{i,l}[k],k=1,2,\cdots,N_{\mathrm{T}},i=1,2,\cdots,N_{\mathrm{B}},l=1,2,\cdots,N_{i,\mathrm{L}}^{(\mathrm{B})}\}$用以描述合法车辆的联合基站和波束的连接控制，$\boldsymbol{Y}=\{y_{j,l},j=1,2,\cdots,N_{\mathrm{R}},l=1,2,\cdots,N_{j,\mathrm{L}}^{(\mathrm{R})}\}$用以描述发射阻塞RSU的选择和RSU发射波束的选择，$\boldsymbol{Z}=\{P_{\mathrm{R},j},j=1,2,\cdots,N_{\mathrm{R}}\}$用以描述RSU的阻塞信号发射功率控制，其中，$P_{\mathrm{R},j}$的取值范围为$\left[0,P_{\mathrm{R,max}}\right]$。当$P_{\mathrm{R},j}=P_{\mathrm{R,max}}$时，表明第$j$个RSU往特定区域以最大功率发射阻塞信号。

2.1　性能指标

为了衡量毫米波车联网的安全通信性能，本文采用了2个重要的系统安全性能指标。

1) 平均保密连接概率。为了保护机密信息不被窃听，发射端采用怀纳（Wyner）保密编码机制对机密数据进行编码[7]。当保密传输速率$C_{\mathrm{s}}\geqslant\eta_{\mathrm{s}}$时，窃听者无法从传输的码字中获得机密信息。因此本文采用平均保密连接概率作为性能指标之一，即瞬时保密传输速率大于保密速率阈值η_{s}的概率，其表达式为$\mathcal{P}_{\mathrm{s}}=P\left(C_{\mathrm{s}}\geqslant\eta_{\mathrm{s}}\right)$。

2) 系统保密传输速率。系统保密传输速率是在保证安全的前提下所有车辆用户的实际传输速率和，它反映了系统的保密吞吐量性能。因此，所有车辆用户的总保密传输速率也被认为是毫米波车联网的一个重要的安全通信性能指标，其表达式为$C_{\mathrm{s}}^{\mathrm{tot}}=\sum_{k=1}^{N_{\mathrm{T}}}C_{\mathrm{s}}^k$。

2.2　联合优化问题和约束条件

基于上述的性能指标，本文的目标是在保证每个合法车辆的通信质量和通信安全的前提下，

通过优化合法车辆的联合基站和波束的连接控制，发射阻塞RSU的选择，以及RSU设备的联合协作干扰方向和发射功率的决策，使得所有合法车辆的总保密传输速率最大化。优化目标和约束的表达式为

$$\max_{\{\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z}\}} \quad C_{\mathrm{s}}^{\mathrm{tot}}$$

$$\text{C1:} \quad \sum_{l=1}^{N_{i,\mathrm{L}}^{(\mathrm{B})}}\sum_{k=1}^{N_{\mathrm{T}}} x_{i,l}[k] \leqslant N_{\mathrm{RF}}, \forall i = 1,\cdots,N_{\mathrm{B}}$$

$$\text{C2:} \quad \sum_{i=1}^{N_{\mathrm{B}}}\sum_{l=1}^{N_{i,\mathrm{L}}^{(\mathrm{B})}} x_{i,l}[k] \leqslant 1, \forall k = 1,\cdots,N_{\mathrm{T}}$$

$$\text{C3:} \quad \sum_{l=1}^{N_{j,\mathrm{L}}^{(\mathrm{R})}} y_{j,l} \leqslant 1, \forall j = 1,\cdots,N_{\mathrm{L}}$$

$$\text{C4:} \quad C_{\mathrm{s}}^{k} \geqslant \eta_{\mathrm{th}}, \forall k = 1,\cdots,N_{\mathrm{T}}$$

$$\text{C5:} \quad x_{i,l}[k] \in \{0,1\}, y_{j,l} \in \{0,1\}, P_{\mathrm{R},j} \in \left[0, P_{\mathrm{R,max}}\right] \tag{11}$$

其中，约束条件C1表示每个毫米波基站能够同时服务的车辆数量受限于射频链路的数量，约束条件C2意味着每个合法车辆最多只能选择一个基站的一个波束来为自己提供服务，约束条件C3表明每个RSU设备只能选择一个特定方向的波束进行阻塞信号的发射，约束条件C4表明每个合法车辆用户的保密传输速率应大于最低门限阈值η_{th}，约束条件C5限制联合优化问题的决策变量$x_{i,l}[k]$和$y_{j,l}$为二进制离散决策变量，以及$P_{\mathrm{R},j}$为连续决策变量，且其值在$\left[0,P_{\mathrm{R,max}}\right]$。

在求解多基站多用户的联合优化问题时，存在以下几个方面的挑战。

1) 车辆的高速移动特性使得车联网信道状态快速地变化，合法车辆和RSU设备难以实时地做出最优的安全传输决策或所做出的决策难以有效地保障车联网通信的安全性。

2) 每个车辆用户的决策都会对其他车辆用户产生影响，如小区间干扰，每个RSU设备的协作干扰决策也会不同程度地影响车联网系统的安全传输性能。如何有效地协调多个车辆用户和多个RSU设备，使得车联网节点间相互协作，共同最大化系统保密传输速率。

3) 现有的针对毫米波车联网通信的相关研究主要对离散动作决策或连续动作决策进行策略设计。当决策动作空间为混合动作时，即同时包含离散动作和连续动作时，结合毫米波车联网的网络特点设计出合理的框架，实现混合决策控制并有效地求解联合优化问题。

3　基于D3QN-DDPG的多智能体安全协作通信策略

为了克服上述的困难和挑战，本文设计了一种基于D3QN-DDPG的多智能体安全协作通信策略来求解式(11)中的优化问题。本文首先搭建了一个多智能体安全协作通信系统，将车联网中的VU节点和RSU设备节点建模为智能体。其次，本文针对VU的决策空间特点设计了基于D3QN的联合基站选择和波束选择的安全传输方案，并部署在VU智能体上。最后，本文针对RSU设备的离散–连续的混合动作空间特点设计了基于D3QN-DDPG的联合波束选择和功率控制的安全传输方案，并部署在RSU智能体上。多个不同类型的智能体相互协作，共同提升毫米波车联网通信的有效性和安全性，详细的系统和智能体的设计介绍如下。

3.1　多智能体安全协作通信系统

现有针对毫米波车联网的智能算法研究主要集中于单智能体的设计。在单智能体场景下，智能体不需要建模或预测环境中其他智能体的行为。但是，在多智能体环境中，智能体需要同时学习其他智能体的策略，即多个智能体之间的决策和所获得的奖励相互影响。因此，单智能体算法在多智能体环境下不能保证收敛[25]。在多用户多基站的毫米波车联网场景中，如果采用单智能体的算法，每个智能体会忽略其他智能体对环境的影响，而将自己的信息保存在经验池中。然而，当其他智能体的策略发生改变时，可能会对智能体产生错误的引导，使得算法变得振荡而难以收敛[26]。因此，本文搭建了一个基于多智能体的安全协作通信系统来提高模型的鲁棒性和稳定性。

多智能体安全协作通信系统如图2所示。本文首先将图1中的毫米波车联网建模为环境，其次将网络中的VU节点和RSU节点建模为智能体。通过与环境交互获取经验，多个VU智能体和RSU智能体相互协作，共同优化VU智能体的联合基站和波束的连接控制，发射阻塞RSU的选择，以及RSU的联合协作干扰方向和发射功率的决策，使得所有合法车辆的总保密传输速率最大化。在每个时隙t，

每个智能体首先从环境中获取当前状态信息$s_t \in S$，并根据状态信息选择需要执行的动作$a_t \in A$，其中，S和A分别代表状态空间和动作空间。多个智能体的动作共同作用于环境中，并引起当前环境的变化，即进入当下一个环境状态s_{t+1}。与此同时，环境评估每个智能体的动作，并通过奖励模块将奖励r_t反馈回智能体。至此，每个智能体获得训练样本$\left(s_t,a_t,r_t,s_{t+1}\right)$，并存入经验回放池，用作智能体模型的优化训练。对于每个智能体，其马尔可夫链的表达式为$\{s_t,a_t,s_{t+1},a_{t+1},\cdots\}$。

在实际的多智能体场景中，若每个智能体不知道其他智能体的信息，训练过程中可能会产生很大的波动，影响模型的收敛性和稳定性。例如，多个车辆用户可能重复选择同一个波束和基站，或者RSU智能体可能在干扰窃听者时对合法用户造成了较大的干扰，进而导致车联网安全传输性能的下降[27]。因此，为了使多个智能体相互协作共同提高系统整体性能，智能体不仅要考虑局部的观察信息，还要考虑全局其他智能体的信息。如图2所示，本文设计了一个通信模块用以收集和传递所有智能体的局部信息，如信道信息、位置信息等。多个VU智能体和多个RSU智能体通过通信模块定期交换信息，以帮助彼此做出更加合理的决策，共同提升系统的整体性能。

3.2　基于D3QN的VU智能体

合法车辆的决策变量$\boldsymbol{X}$为联合基站选择和波束选择的决策，是一个离散决策变量。基于DQN的深度强化学习算法利用深度神经网络来学习高维状态和动作空间的映射关系，能够有效地处理离散动作空间的决策问题[28]。然而，DQN中的最大运算器使用相同的值来进行动作选择和动作价值估计，这使得DQN模型更容易选择被高估的值，即DQN模型存在“过估计”问题。受到D3QN算法成功应用的启发[29-30]，本文采用基于D3QN的方法来构建VU智能体模型，基于D3QN的VU智能体如图3所示。D3QN算法采用训练网络和目标网络2种不同的网络进行动作选择和动作评估，即利用训练网络获取s_{t+1}状态下的最佳动作，并利用目标网络计算该动作的动作价值。通过2个网络的交互，有效避免了DQN算法中的“过估计”问题。详细的VU智能体设计介绍如下。

1) 状态信息。VU智能体的当前状态中包含当前时刻合法车辆、窃听者车辆、多个基站和RSU的位置信息，以更好地学习交通模式与VU智能体的联合基站和波束决策之间的关系。此外，VU智能体的状态还包含基站到VU的信道状态信息（CSI, channel state information）、基站到窃听车辆的CSI，上一时刻合法车辆的码字传输速率。因

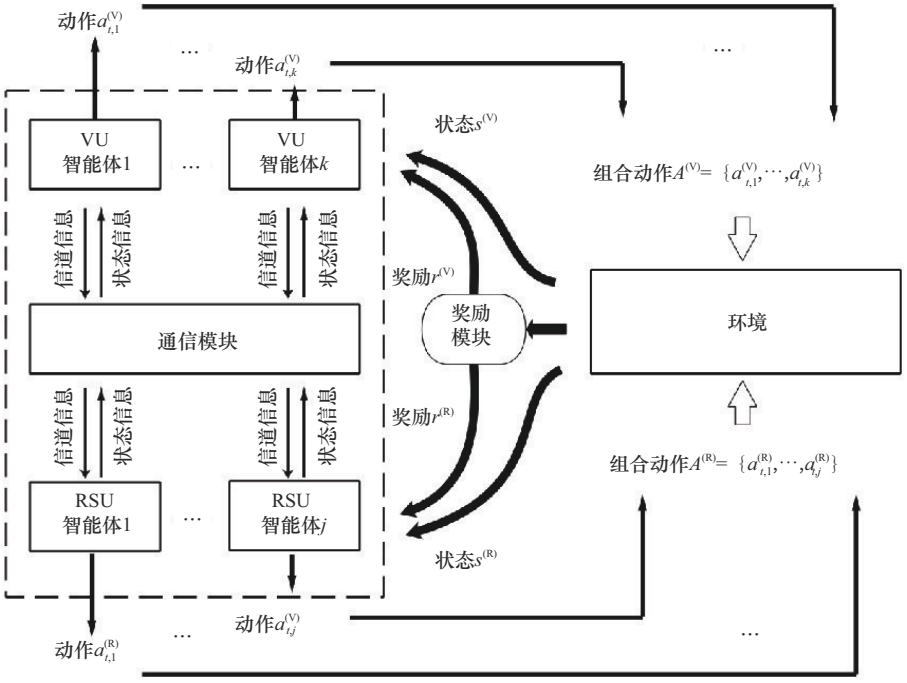


图2　多智能体安全协作通信系统

此，VU 智能体的状态为

$$s_t^{(\mathrm{V})}=\left[\Gamma_{\mathrm{V}},\Gamma_{\mathrm{B}},\Gamma_{\mathrm{E}},\Gamma_{\mathrm{R}},\left\{g_{i,k}^{(\mathrm{B})}\right\},\left\{g_{i,n}^{(\mathrm{B})}\right\},C_{\mathrm{b},t-1}^{k},\left\{C_{\mathrm{e},t-1}^{k}\right\}\right] \tag{12}$$

其中，$\Gamma_{\mathrm{V}},\Gamma_{\mathrm{B}},\Gamma_{\mathrm{E}},\Gamma_{\mathrm{R}}$ 分别为合法车辆、毫米波基站、窃听车辆和 RSU 设备的位置坐标信息。

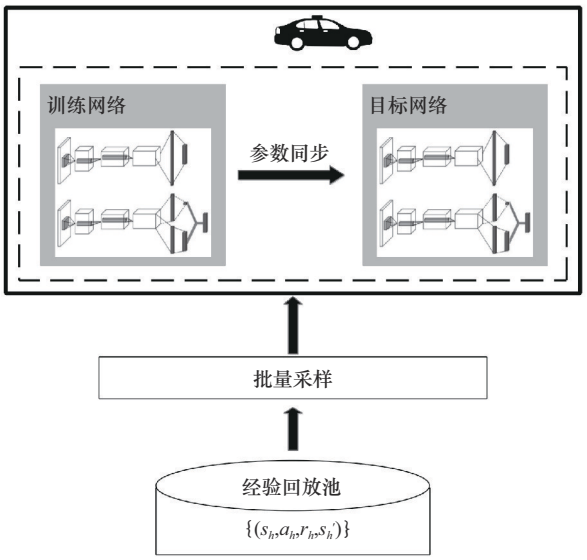


图3　基于 D3QN 的 VU 智能体

2) 动作空间。基于当前的状态信息，VU 智能体通过选择一个基站和波束的组合来获得基站提供数据传输服务。因此，VU 智能体的动作空间维度 $A^{(\mathrm{V})}$ 为 $N_{\mathrm{B}}\times N_{i,\mathrm{L}}^{(\mathrm{B})}$，每个动作 $a_t^{(\mathrm{V})}\in A^{(\mathrm{V})}$ 对应一个联合基站选择和波束选择的特定组合动作。D3QN 是以深度神经网络来拟合 Q 函数，以状态 s 作为网络输入，输出所有动作的估计 Q 值。之后，采用 ϵ-贪心策略来进行动作选择，以平衡动作的探索和动作的利用[31]。具体来说，它在所有动作中以 ϵ 的概率随机选择一个动作用以探索，或以 $1-\epsilon$ 的概率从 D3QN 模型中输出具有最大估计值的动作 a 用以利用。因此，VU 智能体的动作为

$$a_t^{(\mathrm{V})}=\begin{cases}\arg\max\limits_{a_t^{(\mathrm{V})}} Q\left(s_t^{(\mathrm{V})},a_t^{(\mathrm{V})};\Omega\right), & \text{以}1-\epsilon\text{的概率}\\ U\left(A^{(\mathrm{V})}\right), & \text{以}\epsilon\text{的概率}\end{cases} \tag{13}$$

其中，$Q\left(s_t^{(\mathrm{V})},a_t^{(\mathrm{V})};\Omega\right)$ 是动作-状态价值函数，$U\left(A^{(\mathrm{V})}\right)$ 是动作 $A^{(\mathrm{V})}$ 的离散均匀分布函数，$\epsilon\in[0,1]$ 是一个超参数，用于调整动作探索和利用之间的平衡，Ω 表示训练网络的权重，利用经验回放机制进行不断更新[32]。D3QN 网络能够选择具有高反馈奖励的动作，或者探索可能具有更高奖励但尚未被选择的动作，从而探索整个动作空间并更新 Q 值。

3) 奖励反馈机制。基于式(11)中的优化目标和约束条件，本文设计了基于车辆保密传输速率的奖励和基于系统保密传输速率的奖励。

①基于车辆保密传输速率的奖励。针对式(11)中的约束 C4，设计基于车辆保密传输速率的奖励，以保证每个合法车辆安全传输的通信质量，具体奖励设计为

$$r_{t,1}^{(\mathrm{V})}=\begin{cases}0, & C_{\mathrm{s}}^{k}>\eta_{\mathrm{s}}\\ \chi_1, & \text{其他}\end{cases} \tag{14}$$

其中，设置奖励 $\chi_1<0$ 以惩罚合法车辆的保密传输速率不满足最小传输速率要求，避免不合理的优化。

②基于系统保密传输速率的奖励。针对式(11)中的优化目标，本文还设计了基于系统保密传输速率的奖励，以最大化整个车联网系统的安全传输性能，具体奖励设计为

$$r_{t,2}^{(\mathrm{V})}=\left[\sum_{k=1}^{N_{\mathrm{T}}}C_{\mathrm{s}}^{k}-\zeta_{\min}\right]^{+} \tag{15}$$

其中，设置奖励 $\zeta_{\min}>0$ 为系统能容忍的最小安全传输速率，以保证系统的整体安全性能。基于此，对于每个 VU 智能体来说，其做出的动作所获得的总反馈奖励为

$$r_t^{(\mathrm{V})}=\omega_1 r_{t,1}^{(\mathrm{V})}+\omega_2 r_{t,2}^{(\mathrm{V})} \tag{16}$$

其中，和 $\omega_2>0$ 是正权重，用来描述每个合法车辆的服务质量和整个系统安全性能之间的权衡。

③D3QN 算法设计。基于 D3QN 的 VU 智能体期望找到一个最优策略来最大化长期累积奖励，其表达式为 $R_t^{(\mathrm{V})}=\sum_{\tau=0}^{\mathcal{T}}\gamma^{\tau}r_{t+\tau+1}^{(\mathrm{V})}$。其中 $\gamma\in[0,1)$ 是奖励的折扣率，当 $\mathcal{T}$ 有界时，设置 $\gamma=1$；当 $\mathcal{T}$ 无界时，设置 $\gamma<0$。最优策略能够通过贝尔曼方程进行求解，其要素包括离散的状态集合 $S^{(\mathrm{V})}$、离散的动作集合 $A^{(\mathrm{V})}$ 和状态转移概率 $\mathcal{P}\left(s'|s,a\right),\forall s,s_t\in S,a,a_t\in A$。因此对于时隙 t，VU 智能体的状态–动作函数（Q 函数）为

$$Q^{\pi}\left(s_t,a_t\right)=\mathrm{E}_{\pi}\left[R_t^{(\mathrm{V})\pi}|s_t=s,a_t=a\right]=\mathrm{E}_{\pi}\left[r_t+\gamma Q\left(s_{t+1},a_{t+1}\right)|s_t=s,a_t=a\right] \tag{17}$$

其中，π为t时隙下智能体的策略。相应地，Q函数的更新表达式为

$$Q_{t+1}(s_t,a_t) \leftarrow \sum_{s_{t+1}\in S} \mathcal{P}(s_{t+1}|s_t,a_t) \left[r_t^{(\mathrm{V})} + \gamma \arg\max_{a_{t+1}} Q_{t+1}(s,a) \right] \tag{18}$$

D3QN算法利用神经网络模型来拟合Q函数，由训练网络和目标网络构成。训练网络用于当前模型动作的选择，网络参数为Ω_t。目标网络用于动作的评估，网络参数为Ω_t^-。训练网络的目标为

$$\mathcal{Y}_t^{\mathrm{D3QN}} = r_t^{(\mathrm{V})} + \gamma Q(s_{t+1},\arg\max_{a_{t+1}\in A} Q(s_{t+1},a_{t+1}|\Omega_t)|\Omega_t^-) \tag{19}$$

因此，模型参数更新的依据是最小化每个时隙的损失函数$\mathcal{L}_h(\Omega_t)=\left[\mathcal{Y}_t^{\mathrm{D3QN}} - Q(s_t,a_t|\Omega_t)\right]^2$。目标值$\mathcal{Y}_t^{\mathrm{D3QN}}$与估计值$Q(s_t,a_t|\Omega_t)$间的误差$\delta$称为时间差分误差，表示为$\delta = \mathcal{Y}_t^{\mathrm{D3QN}} - Q(s_t,a_t|\Omega_t)$。因此D3QN依照$\Omega_{t+1} \leftarrow \Omega_t + \phi_l \delta \nabla_{\Omega\mathcal{L}_h}(\Omega_t)$更新其训练网络的参数。其中，$\phi_l$表示训练网络$\Omega$的学习率，$\nabla(\cdot)$表示一阶偏导数。在D3QN算法的实践中，采用随机批量样本数据$\mathcal{H}$进行模型的训练和参数的更新，其损失函数表达式为

$$L(\Omega_t) = \frac{1}{\mathcal{H}} \sum_{h=1}^{\mathcal{H}} \mathcal{L}_h \tag{20}$$

3.3 基于D3QN-DDPG的RSU智能体

RSU智能体的决策变量为$\boldsymbol{Y}$和$\boldsymbol{Z}$，其中$\boldsymbol{Y}$为联合RSU的选择和RSU协作干扰波束的选择，是一个离散的决策变量；$\boldsymbol{Z}$为RSU的协作干扰功率大小的选择，是一个连续的决策变量。D3QN算法在优化离散决策变量时拥有非常优异的性能，然而当问题中包含连续动作决策时，如信号的发射功率大小，D3QN算法的性能将不能很好地处理连续动作空间的决策。常见的操作为将连续动作进行离散化处理，然而在离散化过程中，会丢失部分的决策信息，使得算法优化的性能下降。与传统的根据ϵ概率生成动作的方法不同，DDPG根据参数策略生成确定的动作[33]。此外，该算法采用行动器-评判器（Actor-Critic）结构，使DDPG能够有效地处理连续动作问题。DDPG网络包含4个子网络：用于动作选择的行动器训练（Actor）网络、用于生成当前动作Q值的评判器训练（Critic）网络、用于生成目标值的行动器目标（Target-Actor）网络和评判器目标（Target-Critic）网络。此外，DDPG还利用经验回放机制和双重网络方法来提高原始框架的收敛性能。

基于上述分析，本文结合D3QN在处理离散动作空间的优异性能和DDPG在处理连续动作决策上的突出表现，设计了基于D3QN-DPPG混合模型的RSU智能体，如图4所示。每个RSU智能体由一个D3QN模型和一个DDPG模型组合构成，其中D3QN决定RSU发射的阻塞信号方向，DDPG决策RSU的阻塞信号功率大小，详细参数介绍如下。

1) 状态信息。RSU智能体的状态信息中包含合法车辆、窃听者车辆、毫米波基站和RSU的位置信息，以更好地学习交通模式与RSU智能体的协作干扰决策的关系。此外，VU智能体的状态还包含RSU到VU的CSI、RSU到窃听车辆的CSI，以及上一时刻合法车辆的保密传输速率和窃听者的窃听速率。因此，RSU智能体的状态表达式为

$$s_t^{(\mathrm{R})}=\left[\Gamma_{\mathrm{V}},\Gamma_{\mathrm{B}},\Gamma_{\mathrm{E}},\Gamma_{\mathrm{R}},\left\{g_{j,k}^{(\mathrm{R})}\right\},\left\{g_{j,n}^{(\mathrm{R})}\right\},C_{\mathrm{s},t-1}^{k},\left\{C_{\mathrm{e},t-1}^{k}\right\}\right] \tag{21}$$

2) 动作空间。基于当前的状态信息，RSU智能体选择一个干扰波束方向和干扰功率大小来发射阻塞信号。因此每个RSU的动作$a_t^{(\mathrm{R})}$由一个连续动作$a_{t,\mathrm{c}}^{(\mathrm{R})} \in A_{\mathrm{c}}^{(\mathrm{R})}$和一个离散动作$a_{t,\mathrm{d}}^{(\mathrm{R})} \in A_{\mathrm{d}}^{(\mathrm{R})}$组成。相应地，$a_{t,\mathrm{d}}^{(\mathrm{R})}$表示RSU波束方向的选择，其动作空间维度为$N_{j,\mathrm{L}}^{(\mathrm{R})}$；$a_{t,\mathrm{c}}^{(\mathrm{R})}$表示RSU功率大小，为连续决策变量，且决策空间为$\left[0,P_{\mathrm{R,max}}\right]$。

RSU智能体中的D3QN模块与3.2节中的VU智能体相同，因此其离散动作决策的$a_{t,\mathrm{d}}^{(\mathrm{R})}$获取与式(13)相似。而RSU智能体的连续决策动作$a_{t,\mathrm{c}}^{(\mathrm{R})}$主要由DDPG模型生成，其表达式为

$$a_{t,\mathrm{c}}^{(\mathrm{R})} = \kappa\left(s_t^{(\mathrm{R})}|\Theta_\kappa\right) + \mathcal{N}_{\mathrm{OU}}(0,v) \tag{22}$$

其中，$\mathcal{N}_{\mathrm{OU}}(0,v)$是均值为0的奥恩斯坦（OU, ornstein uhlenbeck）噪声，服从$\mathcal{N}_{\mathrm{OU}} \sim \mathrm{OU}(0,v)$，且$v$代表OU噪声的波动性；$\kappa\left(s_t^{(\mathrm{R})}|\Theta_\kappa\right)$为DDPG中的Actor网络，$\Theta_\kappa$为Actor网络的参数。

3) 奖励反馈机制。基于式(11)中的优化目标和约束条件，RSU智能体的奖励反馈机制与VU智能体中奖励反馈设计基本相同，包含基于车辆保密传

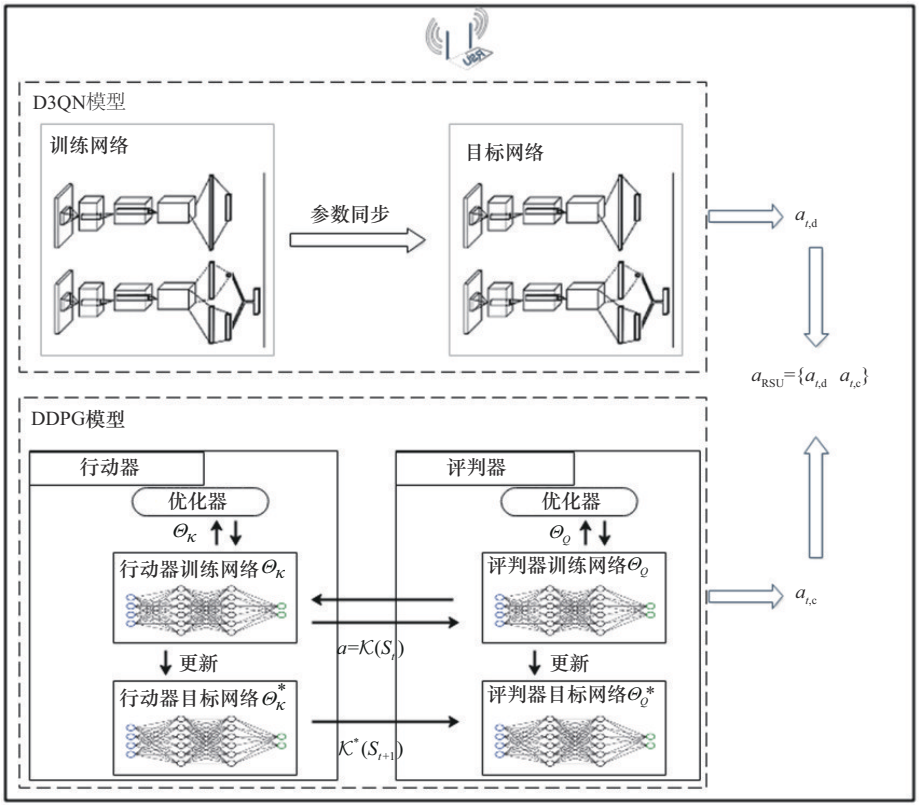


图4　基于D3QN-DDPG的RSU智能体

输速率的奖励$r_{t,1}^{(\mathrm{R})}$和基于系统保密传输速率的奖励$r_{t,2}^{(\mathrm{R})}$这2个部分，故在此不再赘述。因此，RSU智能体的反馈奖励为$r_t^{(\mathrm{R})}=\omega_1 r_{t,1}^{(\mathrm{R})}+\omega_2 r_{t,2}^{(\mathrm{R})}$。

相应地，对于RSU智能体中的D3QN模块，其算法设计与基于D3QN的VU智能体中的设计基本相同，故在此不再赘述。接下来介绍RSU智能体中的DDPG模块的原理和算法流程。

4) DDPG设计。DDPG网络由Actor网络$\kappa\left(s_t^{(\mathrm{R})}|\Theta_\kappa\right)$、Target-Actor网络$\kappa\left(s_t^{(\mathrm{R})}|\Theta_\kappa^*\right)$、Critic网络$\mathcal{Q}\left(s_t^{(\mathrm{R})},a_{t,\mathrm{c}}^{(\mathrm{R})}|\Theta_{\mathcal{Q}}\right)$和Target-Critic网络$\mathcal{Q}\left(s_t^{(\mathrm{R})},a_{t,\mathrm{c}}^{(\mathrm{R})}|\Theta_{\mathcal{Q}}^*\right)$组成。Actor网络主要用于动作的生成，Critic网络主要用于动作的评价。相应地，Θ_κ、Θ_κ^*、$\Theta_{\mathcal{Q}}$、$\Theta_{\mathcal{Q}}^*$分别对应4个网络的神经网络参数。类似地，每个DDPG模型都期望能为每个状态s_t选择动作a_t，使得累积期望奖励最大化，即$\max\sum_{\tau=t}^{\mathcal{T}}\gamma r_{\tau+1}^{(\mathrm{R})}$。DDPG模型期望找到最优的动作$a^*$，使得$Q$值最大，其最优动作对应的$Q$值表达式为$\mathcal{Q}^*\left(s_t,a_t\right)=\mathrm{E}\left[r_t^{(\mathrm{R})}+\gamma\mathcal{Q}^*\left(s_{t+1},a_{t+1}\right)\right]$。与D3QN中的经验回放机制相似，DDPG通过在经验池中的随机采样小批量$\mathcal{H}$的样本数据，通过最小化损失函数来训练DDPG模型[33]。

$$L\left(\Theta_{\mathcal{Q}}\right)=\frac{1}{\mathcal{H}}\sum_{h=1}^{\mathcal{H}}\left[y_t-\mathcal{Q}(s_t,a_t|\Theta_{\mathcal{Q},t})\right]^2 \tag{23}$$

为了避免训练过程中模型性能的振荡，y_t由目标网络生成，即Target-Critic网络。目标网络的结构与训练网络相同，但是每隔一段时间，目标网络的参数将从训练网络的参数中复制，以软更新的方式进行更新，并固定一段时间。目标网络参数更新时，以$1-\rho_2$的权重保留历史目标网络参数信息，并以ρ_2的权重复制训练网络参数来更新目标网络的参数，其中$\rho_2\in[0,1]$为Actor和Critic的目标网络参数更新率。相应地，D3QN中的目标 网络的参数更新方式和DDPG的参数更新方式相似，其更新率用$\rho_1\in[0,1]$表示。因此y_t的表达式为

$$y_t=r_t^{(\mathrm{R})}+\gamma\mathcal{Q}^*\left(s_{t+1},\kappa^*\left(s_t|\Theta_{\kappa,t}^*\right)|\Theta_{\mathcal{Q},t}^*\right) \tag{24}$$

因此，DDPG利用Critic网络计算得到的策略梯度来训练Actor网络中的参数，其表达式为

$$\nabla_{\Theta_\kappa}=\frac{1}{\mathcal{H}}\sum_{h=1}^{\mathcal{H}}\nabla_a\mathcal{Q}\left(s_h,a_h|\Theta_{\mathcal{Q}}\right)|_{s_h=s_t,a_h=\kappa\left(s_t\right)}\nabla_{\Theta_\kappa}\kappa(s_t|\Theta_\kappa|_{s_h=s_t}) \tag{25}$$

3.4　基于D3QN-DDPG的多智能体安全协作通信算法

本文所设计的安全协作通信算法主要包含2个阶段：学习阶段和应用阶段。

3.4.1　学习阶段

在学习阶段，多个智能体试图探索不同的基站选择、波束选择、干扰波束选择和干扰功率大小的不同组合，之后从环境中获取相应的反馈奖励和对应的训练样本，并将样本存入各自的经验回放池中，用以各自智能体模型的训练和参数更新。在此基础上，各个智能体通过不断更新D3QN模型或D3QN-DDPG混合模型，根据最优策略不断调整自己的动作。相应地，基于D3QN-DDPG的多智能体安全协作通信算法如算法1所示。

算法1　基于D3QN-DDPG的多智能体安全协作通信算法

初始化　环境参数和车辆数据信息，各个智能体的模型参数，包括训练网络参数、网络参数和经验回放池；

1) for $n_t=1,\cdots,N_{\max}^{\text{ite}}$
2)　更新所有车辆位置和环境信息；
3)　for $n_e=1,\cdots,N_{\max}^{\text{epi}}$
4)　　重置VU智能体的联合基站选择和波束选择决策，RSU智能体的联合波束选择和干扰功率控制决策。
5)　　for VU $v_k^{(\text{T})},k=1,\cdots,N_{\text{T}}$
6)　　　获取当前状态$s_{t,k}^{(\text{V})}$，并根据ϵ-贪心策略，每个合法车辆选择动作$a_{t,k}^{(\text{V})}$。
7)　　end for
8)　　for RSU $R_j,j=1,\cdots,N_{\text{R}}$
9)　　　获取当前状态$s_{t,j}^{(\text{R})}$。每个RSU利用D3QN模型选择离散动作$a_{t,\text{d}}^{(\text{R})}$，利用DDPG模型获得连续决策动作$a_{t,\text{c}}^{(\text{R})}$。
10)　　end for
11)　　执行所有智能体的动作，并作用于环境中。
12)　　各个智能体计算执行动作的反馈奖励$r_t^{(\text{V})}$和$r_t^{(\text{R})}$，并观察各自下一时刻的状态$s_{t+1}^{(\text{V})}$和$s_{t+1}^{(\text{R})}$。
13)　　智能体获得训练样本$\{s_h,a_h,r_h,s_h'\}$，并将训练样本存储到各自经验回放池中。
14)　　智能体根据式(19)或式(24)计算目标值，并利用随机梯度下降法式(20)、式(23)、式(25)更新各自模型的训练网络参数。
15)　　更新目标网络参数：$\Omega_t^-\leftarrow(1-\rho_1)\Omega_t^-+\rho_1\Omega_t,\Theta_{\mathcal{Q}}^*\leftarrow(1-\rho_2)\Theta_{\mathcal{Q}}^*+\rho_2\Theta_{\mathcal{Q}},\Theta_\kappa^*\leftarrow(1-\rho_2)\Theta_\kappa^*+\rho_2\Theta_\kappa$。
16)　　更新当前环境信息$s_{t+1}^{(\text{V})}\leftarrow s_{t+1}^{(\text{V})},s_{t+1}^{(\text{R})}\leftarrow s_{t+1}^{(\text{R})}$。
17)　end for
18) end for

3.4.2　应用阶段

模型训练完成之后，各个智能体进行应用阶段并停止模型训练以减小开销，之后各个智能体观察当前的环境状态并提取特征，作为状态输入所训练的智能体模型中，做出当前状态下合理的动作决策，在不需要进一步训练的情况下相互协作共同保证系统安全协作通信的性能。

4　仿真结果分析

4.1　仿真场景搭建和参数设置

根据排队理论模型，车辆以$\lambda=0.5$的到达率进入通信区域，并以36 km/h的速度匀速行驶。本文以0.1 s为时间切片，获得各个时刻下的原始数据信息，并进行特征提取。网络拓扑结构的快速变化使得各个合法车辆用户和RSU设备需要根据当前局部的环境状态信息，迅速做出最佳决策，以此来验证方案在城市车联网场景中的有效性和鲁棒性。本文在仿真实验中选取了更加复杂的十字路口的通信场景。场景在4个方位部署$N_{\text{B}}=4$个毫米波基站，且每个毫米波基站配备$N_{i,\text{L}}^{(\text{B})}=9$个不同方向的定向窄波束，覆盖系统中的所有通信区域。如图1所示，在每个服务周期，4个基站同时为$N_{\text{T}}=4$辆合法车辆提供数据传输服务。同时，本文在毫米波车联网场景的道路旁边部署$N_{\text{R}}=4$个RSU设备，且每个RSU预先配有$N_{j,\text{L}}^{(\text{R})}=12$个不同方向的波束码本，能够选择其中一个特定的方向发射阻塞信号。本文还考虑一个多窃听者的动态窃听场景，在场景中随机生成$N_{\text{E}}=3$个窃听车辆，并沿着当前道路方向行驶。仿真参数[22,34]如表1所示。

表1　　仿真参数

参数名称	仿真数值
波束宽度θ_B、θ_V、θ_R	15°
基站传输功率P_B/W	1
基站天线高度/m	25
RSU天线高度/m	10
车辆天线高度/m	1.6
RSU最大干扰功率$P_{R,max}$/W	1
主瓣天线增益M/dB	13
旁瓣天线增益m/dB	0.05
载波频率/GHz	28
带宽/GHz	2
噪声功率/dBm	−70

在构建D3QN和DDPG模型时，本文采用一个全连接的神经网络架构，它由一个输入层、3个隐藏层和一个输出层组成。D3QN隐藏层的神经元数量分别为128、256和128，DDPG隐藏层的神经元数量分别为64、128和64。在D3QN网络中，本文采用ReLU作为激活函数。在构建DDPG网络时，本文采用Sigmoid函数作为神经网络的激活函数，并利用Adam优化器更新各个网络参数[35]。此外，在D3QN网络训练过程中，探索率ϵ是一个从1到0.02的线性递减变量，并且达到最低值后保持不变，以保证模型在收敛后仍以较小的概率进行探索。此外，学习率和奖励折扣系数分别设置为0.01和0.9。

4.2　基准方案

在实验中，本文通过将所提方案与以下不同传输方案进行对比，多维度分析系统性能。

1) 最大保密容量方案。该方案通过遍历所有合法车辆的联合基站和波束选择，以及RSU的所有波束和干扰功率的组合，选择能够实现总保密传输速率最大化的最优决策。此外，由于功率选择为连续动作变量，无法实现遍历，故将干扰功率进行离散化处理后遍历获得最优决策。

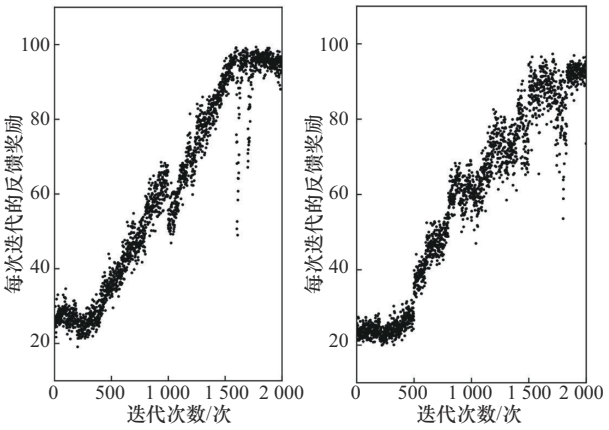
2) 邻近选择方案。在该方案中，所有合法车辆根据当前的位置选择最近的基站和波束，各个RSU根据窃听者的位置选择距离最近的窃听者所在的波束进行干扰的发射，且以最大发射功率发射阻塞信号，以最大程度降低距离RSU最近的窃听者的信号接收质量。

3) 基于DQN算法的方案。在该方案中，所有的智能体利用DQN模型进行智能体的搭建和训练，并且将连续决策动作进行离散化处理，与文献[17]中的描述相似。

4) 随机选择方案。在该方案中，合法车辆随机选择基站和波束作为当前决策，同时RSU也随机选择干扰方向和干扰功率大小进行阻塞信号的发射，并以此作为基准对比算法。

4.3　数值结果分析

图5为VU智能体和RSU智能体在经过大量的探索和训练后的模型收敛曲线，其中，每个点代表经过一次训练迭代后模型所获得的奖励。在最初训练时，模型通过随机探索各个动作的可能组合，获得更多的模型训练样本，此时模型做出的决策所获得的奖励较低。而随着训练迭代次数的不断增加，模型每次迭代所获得的奖励在不断增加，最终在第1 500次迭代时，模型开始收敛且性能保持平稳，这验证了所提方案的模型在多次迭代训练中逐渐收敛并趋于稳定。而在模型训练过程中，由于车辆位置的不断移动和周围环境的变化，使得在训练过程中，模型的奖励性能出现波动。同时在模型收敛之后，模型仍以小概率探索其他动作的组合，使得所提方案能够更好地适应场景的动态性，避免局部优化。基于此，本文将进一步分析所提方案的性能。



(a) 基于D3QN的VU智能体　(b) 基于D3QN-DDPG的RSU智能体

图5　VU智能体和RSU智能体的模型收敛曲线

图6为在不同传输策略下的系统保密传输速率，其中横坐标为交通图案的序号。在该实验中，本文通过随机生成多个不同的交通图案，并计算不同交通图案下4个合法车辆的总保密传输速率，随机选取其中10个交通图案来分析和说明系统保密传输性能。由图6可知，本文方案在各个交通图案下均能拥有很好的系统保密传输性能，并且其性能均优于其他方案策略下的性能，接近于最大保密容量方

案，这也验证了所提方案的有效性和鲁棒性，证明了其能够很好地适应车联网动态场景。同时，对于邻近选择方案，所有VU智能体在选择基站和波束，以及RSU在选择波束和干扰功率时，仅考虑了局部的信息和当前自身的性能，而忽略其他车辆的性能和其他窃听者的影响，如RSU仅选择离自己最近的窃听者进行干扰而忽略了其他窃听者和合法车辆的位置分布关系，车辆都选择离自身最近的基站和波束而使得发生分配冲突，这导致系统整体性能的下降。对于基于DQN算法的方案，由于DQN自身的过估计问题以及在对干扰功率进行离散化处理时，使得其最终的决策产生精度缺失，进而导致系统性能的下降。值得一提的是，当涉及决策制定时，最大保密容量方案由于需要详尽考察所有可能的动作组合，因此其决策过程需要长达360 s才能得出最优结果。然而，本文方案则显著提升了决策效率，它能够在仅10 ms内做出一个性能接近最优的决策，这不仅极大地缩短了决策时间，同时也能满足车联网对实时决策的高要求。

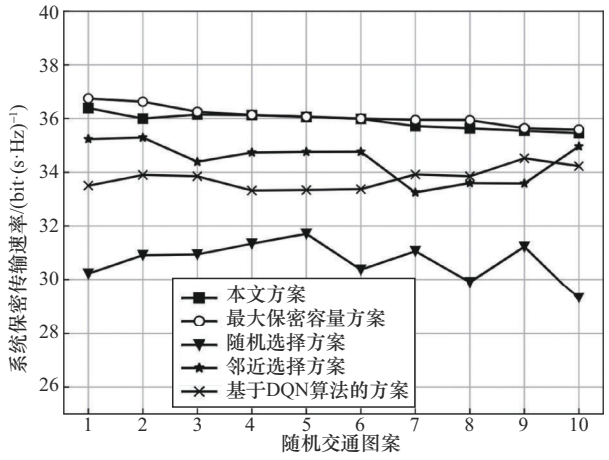


图6　不同传输策略下的系统保密传输速率

图7对比了不同基站策略和天线方案下的系统保密传输速率。由图7可知，在天线数$M=32$且单基站服务场景下，即使通过遍历所有的可能组合，其性能也远小于多基站服务场景的性能，这说明了多基站服务场景能够有效地提升系统的整体性能。在不同的基站服务天线数下，如$M=24$或$M=32$时，本文方案都拥有很好的系统安全传输性能，这从另一方面验证了所提方案的鲁棒性和多基站场景在提升系统性能的有效性。

图8分析了在不同的保密传输条件下，不同传输策略的平均保密连接概率。随着保密速率阈值的不断提高，随机选择方案的性能急剧下降，邻近选择方案的性能先保持缓慢下降后急剧下降。基于DQN算法的方案、本文方案和最大保密容量方案在保密速率阈值小于8.0 Gbit时，依旧能保持很高的保密连接概率，为车辆用户提供一个稳定的安全通信服务。并且本文方案的保密连接概率也优于基于DQN算法的方案，能够达到95%以上的保密连接概率，而基于DQN的算法能保证82%以上的保密连接概率。即使在阈值为8.5 Gbit的条件下，本文方案也能保证78%的保密连接概率，满足基本的安全通信的需求。这说明本文方案在提升车联网安全通信服务方面的鲁棒性，保障了毫米波车联网服务的安全稳定。

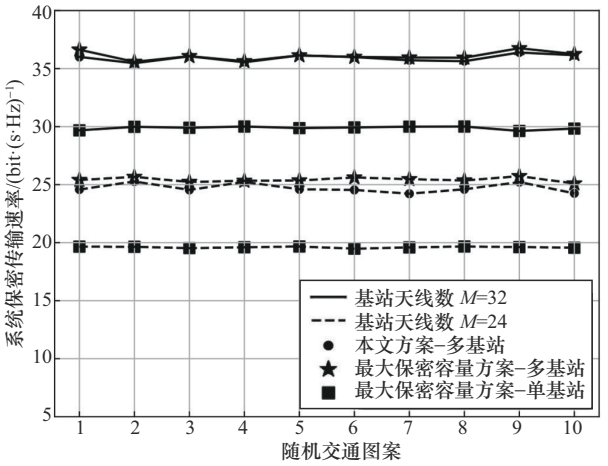


图7　不同基站策略和天线方案下的系统保密传输速率

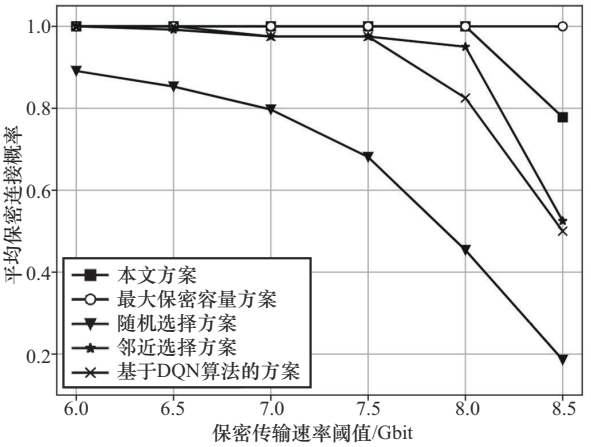


图8　不同传输策略的平均保密连接概率

图9为在不同传输策略下毫米波车联网在一段连续时间内的累积保密传输数据量，其中车辆的位置更新时间为100 ms。随着时间的推移，所有方案

的累积保密传输数据量稳步增加，但随机选择方案、邻近选择方案和基于DQN算法的方案这3种方案与本文方案和最大保密容量方案的差距也在不断增加。并且，在第10 s的时候，这3种方案的累积保密传输数据量与本文方案的差值分别达到52.2 Gbit、14.5 Gbit、21.3 Gbit。而本文方案与最大保密容量方案的差值仅为1.8 Gbit，且决策时间远小于最大保密容量方案，这反映出本文方案在实际的车联网动态通信场景中拥有优异的安全传输性能，并且对动态场景拥有很好的适应性。

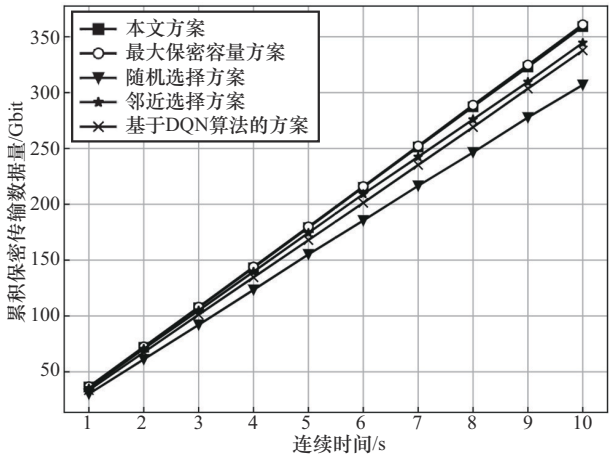


图9　不同传输策略下毫米波车联网在一段连续时间的累积保密传输数据量

5　结束语

本文研究了在多基站多用户的毫米波车联网通信场景下的安全协作通信问题，考虑了一个多窃听者的动态窃听场景，并通过对RSU进行协作干扰方案设计，向场景中特定区域发射阻塞信号降低多个窃听者对合法信号的接收质量，从而提高毫米波车联网的安全传输性能。本文通过联合设计车辆用户的联合基站和波束的连接控制，RSU的选择，以及RSU设备的联合协作干扰方向和发射功率的选择，使得所有合法车辆的总保密传输速率最大化。为了克服求解优化问题的困难和挑战，本文搭建了一个多智能体安全协作通信系统，并设计了基于D3QN的VU智能体和基于D3QN-DDPG的RSU智能体。在构建多智能体系统时，本文设计了基于个体车辆保密传输速率的奖励惩罚机制，以保证每个合法车辆用户的安全传输性能和通信过程的强连接性。最后，本文设计了多种对比方案，如邻近选择方案、基于DQN算法的方案、最大保密容量方案、随机选择方案等，进行了保密传输速率、平均保密连接概率、累积保密传输数据量等不同维度的性能分析，验证了所提方案在毫米波车联网动态场景下的有效性和鲁棒性。

参考文献：

[1] PENG H X, LE L, SHEN X M, et al. Vehicular communications: a network layer perspective[J]. IEEE Transactions on Vehicular Technology, 2019, 68(2): 1064-1078.

[2] CHEN S Z, HU J L, SHI Y, et al. Vehicle-to-everything (V2X) services supported by LTE-based systems and 5G[J]. IEEE Communications Standards Magazine, 2017, 1(2): 70-76.

[3] VA V, SHIMIZU T, BANSAL G, et al. Millimeter wave vehicular communications: a survey[J]. Foundations and Trends® in Networking, 2016, 10(1): 1-113.

[4] 马小博, 彭嘉豪, 薛磊, 等. 5G时代车联网信息物理融合系统综合安全研究[J]. 中国科学(信息科学), 2019, 49(12): 1640-1658.
MA X B, PENG J H, XUE L, et al. Integrated security of cyber-physical vehicle networked systems in the age of 5G[J]. Scientia Sinica (Informationis), 2019, 49(12): 1640-1658.

[5] 范茜莹, 黄传河, 朱钧宇, 等. 无人机辅助车联网环境下干扰感知的节点接入机制[J]. 通信学报, 2019, 40(6): 90-101.
FAN X Y, HUANG C H, ZHU J Y, et al. Interference-aware node access scheme in UAV-aided VANET[J]. Journal on Communications, 2019, 40 (6): 90-101.

[6] HE X, YENER A. MIMO wiretap channels with unknown and varying eavesdropper channel states[J]. IEEE Transactions on Information Theory, 2014, 60(11): 6844-6869.

[7] ZHENG T X, WEN Y T, LIU H W, et al. Physical-layer security of uplink mmWave transmissions in cellular V2X networks[J]. IEEE Transactions on Wireless Communications, 2022, 21(11): 9818-9833.

[8] SHIU Y S, CHANG S Y, WU H C, et al. Physical layer security in wireless networks: a tutorial[J]. IEEE Wireless Communications, 2011, 18 (2): 66-74.

[9] POOR H V, SCHAEFER R F. Wireless physical layer security[J]. Proceedings of the National Academy of Sciences of the United States of America, 2017, 114(1): 19-26.

[10] JU Y, YANG M J, CHAKRABORTY C, et al. Reliability-security tradeoff analysis in mmWave Ad Hoc-based CPS[J]. ACM Transactions on Sensor Networks, 2024, 20(2): 1-23.

[11] WANG H Y, JU Y, ZHANG N, et al. Resisting malicious eavesdropping: physical layer security of mmWave MIMO communications in presence of random blockage[J]. IEEE Internet of Things Journal, 2022, 9(17): 16372-16385.

[12] ELTAYEB M E, CHOI J, AL-NAFFOURI T Y, et al. Enhancing secrecy with multiantenna transmission in millimeter wave vehicular communication systems[J]. IEEE Transactions on Vehicular Technol-

ogy, 2017, 66(9): 8139-8151.

[13] YANG M J, JU Y, LIU L, et al. Secure mmWave C-V2X communications using cooperative jamming[C]//Proceedings of the GLOBECOM 2022 - 2022 IEEE Global Communications Conference. Piscataway: IEEE Press, 2022: 2686-2691.

[14] WANG L, GE S X, ZHOU X B, et al. Multi-agent reinforcement learning-based cooperative beam selection in mmWave vehicular networks[C]//Proceedings of the 2021 IEEE 18th International Conference on Mobile Ad Hoc and Smart Systems (MASS). Piscataway: IEEE Press, 2021: 510-517.

[15] TAN J J, LIANG Y C, ZHANG L, et al. Deep reinforcement learning for joint channel selection and power control in D2D networks[J]. IEEE Transactions on Wireless Communications, 2021, 20(2): 1363-1378.

[16] ASADI A, MÜLLER S, SIM G H, et al. FML: fast machine learning for 5G mmWave vehicular communications[C]//Proceedings of the IEEE Conference on Computer Communications. Piscataway: IEEE Press, 2018: 1961-1969.

[17] JU Y, WANG H Y, CHEN Y C, et al. Deep reinforcement learning based joint beam allocation and relay selection in mmWave vehicular networks[J]. IEEE Transactions on Communications, 2023, 71(4): 1997-2012.

[18] PENG H X, SHEN X M. Multi-agent reinforcement learning based resource management in MEC- and UAV-assisted vehicular networks[J]. IEEE Journal on Selected Areas in Communications, 2021, 39(1): 131-141.

[19] LI Y H, CHEN H K, FENG M Y. A novel model for the traffic of urban roads based on queuing theory[C]//Proceedings of the 2020 International Conference on Intelligent Computing, Automation and Systems (ICICAS). Piscataway: IEEE Press, 2020: 190-194.

[20] CHETLUR V V, DHILLON H S. Poisson line cox process: asymptotic characterization and performance analysis of vehicular networks[C]// Proceedings of the 2019 IEEE Global Communications Conference (GLOBECOM). Piscataway: IEEE Press, 2019: 1-6.

[21] VENUGOPAL K, VALENTI M C, HEATH R W. Interference in finite-sized highly dense millimeter wave networks[C]//Proceedings of the 2015 Information Theory and Applications Workshop (ITA). Piscataway: IEEE Press, 2015: 175-180.

[22] 3GPP. Study on evaluation methodology of new vehicle-to-everything (V2X) use cases for LTE and NR: TR 37.885V15.3.0[R]. 2019.

[23] RAPPAPORT T S, SUN S, MAYZUS R, et al. Millimeter wave mobile communications for 5G cellular: it will work![J]. IEEE Access, 2013, 1: 335-349.

[24] WANG Y Y, VENUGOPAL K, MOLISCH A F, et al. Blockage and coverage analysis with MmWave cross street BSs near urban intersections[C]//Proceedings of the 2017 IEEE International Conference on Communications (ICC). Piscataway: IEEE Press, 2017: 1-6.

[25] LI T X, ZHU K, LUONG N C, et al. Applications of multi-agent reinforcement learning in future Internet: a comprehensive survey[J]. IEEE Communications Surveys & Tutorials, 2022, 24(2): 1240-1279.

[26] LIANG L, YE H, LI G Y. Multi - agent reinforcement learning for spectrum sharing in vehicular networks[C]//Proceedings of the 2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). Piscataway: IEEE Press, 2019: 1-5.

[27] JU Y, CHEN Y C, CAO Z W, et al. Joint secure offloading and resource allocation for vehicular edge computing network: a multi-agent deep reinforcement learning approach[J]. IEEE Transactions on Intelligent Transportation Systems, 2023, 24(5): 5555-5569.

[28] ARULKUMARAN K, DEISENROTH M P, BRUNDAGE M, et al. Deep reinforcement learning: a brief survey[J]. IEEE Signal Processing Magazine, 2017, 34(6): 26-38.

[29] CHEN G, SUN J L, SHEN F, et al. Joint edge computing and caching based on D3QN for lunar Internet of things[C]//Proceedings of the 2022 14th International Conference on Wireless Communications and Signal Processing (WCSP). Piscataway: IEEE Press, 2022: 1-6.

[30] HU H, WU D G, ZHOU F H, et al. Dynamic task offloading in MEC-enabled IoT networks: a hybrid DDPG-D3QN approach[C]//Proceedings of the 2021 IEEE Global Communications Conference (GLOBECOM). Piscataway: IEEE Press, 2021: 1-6.

[31] SUTTON R S, BARTO A G. Introduction to reinforcement learning [M]. Cambridge: MIT Press, 1998.

[32] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529-533.

[33] PENG H X, SHEN X S. DDPG-based resource management for MEC/UAV-assisted vehicular networks[C]//Proceedings of the 2020 IEEE 92nd Vehicular Technology Conference (VTC2020-Fall). Piscataway: IEEE Press, 2020: 1-6.

[34] LI Z P, XIANG L, GE X H, et al. Latency and reliability of mmWave multi-hop V2V communications under relay selections[J]. IEEE Transactions on Vehicular Technology, 2020, 69(9): 9807-9821.

[35] DOGO E M, AFOLABI O J, NWULU N I, et al. A comparative analysis of gradient descent-based optimization algorithms on convolutional neural networks[C]//Proceedings of the 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS). Piscataway: IEEE Press, 2018: 92-99.

[作者简介]

俱莹（1986– ），女，陕西宝鸡人，博士，西安电子科技大学副教授、博士生导师，主要研究方向为无线通信网络安全、毫米波通信、物理层安全传输、车联网安全、区块链等。

陈宇超（1997– ），男，广东揭阳人，深圳市国电科技通信有限公司工程师，主要研究方向为通信类产品研发、管理及设计等。

田素恒（1999– ），男，山东济宁人，西安电子科技大学硕士生，主要研究方向为无线通信物理层安全、太赫兹通信、通信感知一体化等。

刘雷（1987– ），男，河南南阳人，博士，西安电子科技大学副教授，主要研究方向为车联网、边缘智能、算力网络和区块链等。

李赞（1975– ），女，陕西西安人，博士，西安电子科技大学教授、博士生导师，主要研究方向为智能隐蔽通信、通信信号处理等。

裴庆祺（1975– ），男，广西玉林人，博士，西安电子科技大学教授、博士生导师，主要研究方向为数据安全与隐私保护、区块链、边缘计算及安全等。

王明阳（1977– ），男，四川南充人，博士，北京跟踪与通信技术研究所教授，主要研究方向为网络和系统安全、信息保护和数据安全等。