

EMT-MAPPO: Task-Aware Multi-Agent Scheduling of Cooperative Awareness Messages and Video Streams in NR-V2X

Bin Wang*

College of Communication Engineering, Chongqing Polytechnic University of Electronic Technology, Chongqing, China

**Corresponding Author*

Abstract: The joint resource allocation problem for cooperative awareness messages (CAMs) and video streaming in NR-V2X Mode 2 is investigated, where safety-critical and bandwidth-intensive services compete for limited radio resources under dynamic fast-fading channel conditions. To address this challenge, we propose EMT-MAPPO, an enhanced multi-task multi-agent proximal policy optimization framework that augments each agent's local observation with task-aware CAM queue occupancy and video-frame transmission progress. This task-aware state augmentation enables the policy to better coordinate periodic safety message delivery with frame-based video transmission. We compare EMT-MAPPO with standard MAPPO, MAA2C, BP-SPS, and random scheduling and further evaluate convergence behaviour across 4, 8, and 16 agents. Experimental results show that EMT-MAPPO achieves the highest cumulative reward and the better overall trade-off between frame success rate (FSR) and CAM success rate (CSR).

Keywords: Cooperative Awareness Messages (CAMs); New Radio Vehicle-to-Everything (NR-V2X); Multi-Agent Reinforcement Learning (MARL); Enhanced Multi-Task MAPPO (EMT-MAPPO); Resource Allocation; Video Streaming

1. Introduction

With the rapid development of 5G mobile communication systems, intelligent transportation systems (ITS) are evolving toward highly connected, cooperative, and autonomous transportation [1]. Vehicle-to-everything (V2X) communication provides the fundamental information exchange capability among vehicles, infrastructure, pedestrians, and network entities [2]. To meet increasingly stringent latency,

reliability, and capacity requirements, the 3rd Generation Partnership Project (3GPP) has evolved from Long-Term Evolution vehicle-to-everything (LTE-V2X) to New Radio vehicle-to-everything (NR-V2X), introducing enhanced sidelink communication and flexible resource configurations [2, 3]. Recent surveys have systematically examined 3GPP features and optimization techniques for NR-V2X resource allocation, highlighting the challenges posed by high mobility, diverse QoS requirements, and dynamic interference [4]. In particular, NR-V2X Mode 2 supports distributed resource selection without continuous centralized scheduling, which is attractive in highly dynamic vehicular environments [3]. However, distributed resource selection relies heavily on local sensing, so individual vehicles cannot obtain complete knowledge of global interference and resource utilization. As vehicle density and traffic dynamics increase, resource collisions, interference, and spectrum inefficiency become more severe, especially when heterogeneous V2X services share the same resource pool [5]. Adaptive semi-persistent scheduling (SPS) and probabilistic rescheduling have been proposed to improve distributed allocation [6-7], but they remain constrained by local observations and periodic-traffic assumptions.

Among V2X services, Cooperative Awareness Messages (CAMs) are periodically broadcast safety messages carrying basic vehicle status information and require high reliability and timely delivery [8, 9]. High-precision cooperative perception, in contrast, increasingly requires high-volume sensor data, such as compressed camera streams, whose video frames are bandwidth-intensive, bursty, and accumulative because one frame may require multiple transmission opportunities. Recent work has begun to model video flows following a GOP cycle with one large I-frame and several small P-frames in NR sidelink [10]. When CAMs and

accumulative video frames share NR-V2X Mode 2 resources, vehicles face a dual-objective conflict: maintaining CAM reliability/timeliness while providing sufficient resources for video streaming throughput and efficiency. Existing CAM/CPM scheduling and value-of-information (VoI)-based studies [11] mainly focus on structured messages, while adaptive SPS mechanisms [6-7] are designed around periodic traffic and do not explicitly model heterogeneous quality-of-service (QoS) requirements. Even when heterogeneous QoS is considered, existing approaches either rely on network slicing with stationary queueing assumptions [12] or target Mode 1 configured grants [10], rather than distributed Mode 2 resource selection under mixed CAM/video traffic.

Learning-based methods offer greater adaptability. Deep reinforcement learning (DRL) has been applied to V2X resource allocation [13-14], and multi-agent reinforcement learning (MARL) with centralized training and decentralized execution (CTDE) is more suitable for distributed competition among vehicles [15-18]. Recent studies have extended MARL to NR-V2X Mode 2 through federated distributed training [19] and curriculum-assisted scalable learning [20]. MAPPO with CTDE has also been directly applied to cross-layer scheduling in NR-V2X Mode 2 [21], and heterogeneous-agent PPO has been used for fair perception message dissemination in C-V2X networks [22]. However, recent MARL studies either consider homogeneous or structured CAM/CPM traffic [22-24] or target C-V2X Mode 4 [25]. Even when video-aware scheduling is considered, existing work primarily targets Mode 1 configured grants [10] or relies on RAN slicing for heterogeneous QoS [12], rather than the coexistence of periodic safety-critical CAMs and accumulative video frames in NR-V2X Mode 2. This leaves an important gap: a dual-objective, partially observable multi-agent decision problem in which shared resources must be dynamically allocated to balance CAM reliability/timeliness and video streaming efficiency under local observations, interference, and varying channel conditions.

Motivated by this gap, this paper proposes an Enhanced Multi-Task Multi-Agent Proximal Policy Optimization (EMT-MAPPO) framework for multi-service resource allocation in NR-V2X Mode 2, building on PPO and MAPPO. The problem is formulated as a partially observable

Markov decision process (POMDP) with a CTDE architecture. Multiple vehicles act as cooperative agents and make decentralized decisions based on local observations, while a centralized critic exploits global information during training. To capture heterogeneous QoS requirements, a task-specific reward mechanism distinguishes CAM transmission from video streaming. A dynamic rate-threshold mechanism is further introduced to adaptively balance CAM delivery reliability and video transmission efficiency under changing resource competition and channel conditions.

The main contributions are summarized as follows:

We formulate the problem framework and system model: targeting the heterogeneous resource competition between periodic safety-critical CAMs and accumulative high-bandwidth video frames sharing the same resource pool in NR-V2X Mode 2, we model it as a partially observable multi-agent decision process. Unlike existing studies that consider only homogeneous safety messages or structured CAM/CPM traffic, this framework explicitly characterizes the coexistence and conflict between safety-oriented and streaming-oriented services within a shared resource pool.

We propose the EMT-MAPPO method: an enhanced multi-task multi-agent PPO framework that adopts a CTDE architecture for centralized training and decentralized execution. A task-specific reward mechanism is designed to distinguish the QoS requirements of CAMs and video streaming, and a dynamic rate-threshold mechanism is introduced to adaptively balance CAM delivery reliability and video transmission efficiency under dynamic interference and resource competition.

We experimentally validate the proposed method: a simulation framework integrating realistic vehicular mobility and wireless channel conditions is established, and EMT-MAPPO is compared with representative MARL, SPS, and random allocation strategies in terms of CAM delivery performance and video streaming efficiency.

The remainder of this paper is organized as follows. Section 2 presents the system model, including the network architecture, communication model, video streaming model, and formal optimization problem. Section 3 describes the proposed MAPPO-based resource allocation algorithm, covering algorithm principles, network structure, training procedure,

and key parameter settings. Section 4 provides the simulation setup, benchmark algorithms, performance metrics, and result analysis. Section 5 concludes the paper and discusses future research directions.

2. System Model

This section establishes the system model for NR-V2X sidelink vehicular networks. It is organized as follows. Section 2.1 describes the network model, including the V2I/V2V hybrid architecture and resource reuse. Section 2.2 presents the channel model, defining channel gains, SINR, and data rates. Section 2.3 formulates the problem, designing a reward function and posing a long-term reward maximization problem with QoS constraints. These models provide the basis for the MADRL-based resource allocation algorithm.

2.1 Network Model

We consider an NR-V2X sidelink vehicular communication network based on 3GPP Release 16 Mode 2, as illustrated in Fig. 1. The network consists of roadside units (RSUs) and multiple vehicles. Vehicles communicate with RSUs via vehicle-to-infrastructure (V2I) links and with neighbouring vehicles via vehicle-to-vehicle (V2V) links. Safety information, cooperative environmental perception data, and video streaming information are exchanged through V2V PC5 sidelinks, while V2I uplinks offload relevant applications to multi-access edge computing (MEC) servers or upload them to the cloud for processing.

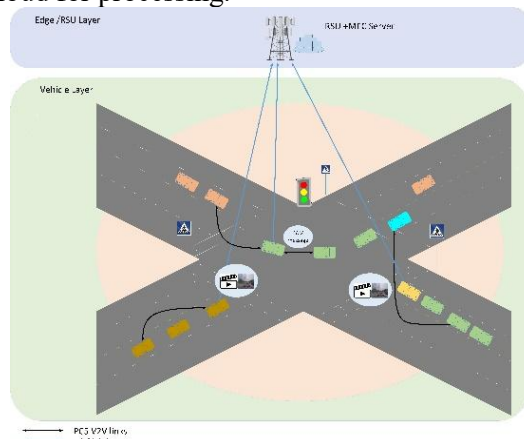


Figure 1. NR-V2X Vehicular Communication Network Architecture

We assume that the L V2V links share the uplink resources orthogonally allocated to the M V2I links. Each V2I link $m \in M = \{0, 1, \dots, M-1\}$ occupies a pre-allocated subchannel m with

constant transmit power P . One or more V2V links $i \in L = \{0, 1, \dots, L-1\}$ may reuse the subchannels of V2I links for CAM or video streaming transmission. In NR-V2X, Mode 1 relies on network-controlled resource allocation, while Mode 2 allows vehicles to autonomously select sidelink resources [20]. This paper focuses on joint resource scheduling based on Mode 2.

Scheduling time is discretized into equal-length control intervals indexed by $k \in K = \{0, 1, \dots\}$. At the beginning of each control interval k (i.e., time kT), each V2V link i samples its driving state to generate CAMs or obtains cooperative perception video streaming data for enhanced vehicular services, buffering them in transmission queues. According to 3GPP NR-V2X specifications, periodically transmitted CAM messages are small in size but require reliability $> 95\%$, whereas video streaming data volumes are significantly larger, aperiodic, frame-based, and can tolerate certain delays and packet losses. Each control interval k is further divided into T communication intervals indexed by $t \in T = \{0, 1, \dots, T-1\}$, operating at a faster time scale. The duration of each communication interval is Δt ms, corresponding to the subframe duration in NR-V2X communications. V2V transmitters make channel resource allocation and streaming frame rate decisions at time $kT + t\Delta t$, and the corresponding communication interval is denoted as (k, t) . We use $x(k, t) := x((kT + t)\Delta t)$ to denote any variable x at the beginning of communication interval (k, t) .

2.2 Channel model

We adopt the NR-V2X channel model specified in 3GPP TR 38.901, assuming that fading on different subchannels is independent and remains constant within one time slot. In communication interval (k, t) , the instantaneous channel gain of V2V link i on subchannel m , which is occupied by V2I link m , is denoted by $G_{\{i,m\}(k,t)}$. It remains constant within the interval and can be written as

$$G_{\{i,m\}(k,t)} = h_{\{i,m\}(k,t)} \alpha_{\{i,m\}(k,t)} \quad (1)$$

where $\alpha_{\{i,m\}(k,t)}$ captures frequency-independent large-scale fading effects, including path loss and shadowing, and $h_{\{i,m\}(k,t)}$ represents small-scale fading, modeled as identically distributed random variables with unit mean across all subchannels. We assume $G_{\{i,m\}(k,t)}$ follows a Markov process [6], i.e., its current value depends only on the previous communication interval. Similarly, we define:

$G_m(k,t)$: channel gain of V2I link m ;
 $G_{i,m}^l(k,t)$: interference channel gain from the transmitter of V2V link i to the receiver of V2I link m ;
 $G_{m,i}^l(k,t)$: interference channel gain from the transmitter of V2I link m to the receiver of V2V link i ;
 $G_{j,i}^l(k,t)$: interference channel gain from the transmitter of V2V link j to the receiver of V2V link i on subchannel m .

In communication interval (k,t) , the signal-to-interference-plus-noise ratios (SINRs) of V2I link m and V2V link i on subchannel m are respectively given by:

$$\gamma_m(k,t) = \frac{P_m(k,t)G_m(k,t)}{\sum_i P_{i,m}(k,t)G_{i,m}^l(k,t) + \sigma^2} \quad (2)$$

$$\gamma_i(k,t) = \frac{P_{i,m}(k,t)G_i(k,t)}{I_i(t) + \sigma^2} \quad (3)$$

where $P_{i,m}(k,t)$ denotes the transmit power of V2V link i on subchannel m , and σ^2 is the noise power. $I_i(t)$ represents the total interference power received by V2V link i on subchannel m :

$$I_i(t) = \sum_{j \neq i} P_{j,m}(k,t)G_{j,i}^l(k,t) + P_m(k,t)G_{m,i}^l(k,t) \quad (4)$$

The transmit power is discretized into $|P| = N$ levels:

$\mathcal{P} = \{23, 20, 17, 14, 11, 8, 5, -100\}$ dBm, where -100 dBm is treated as zero power. The instantaneous data rates of V2I link m and V2V link i are respectively:

$$r_m(k,t) = W \log_2(1 + \gamma_m(k,t)) \quad (5)$$

$$r_i(k,t) = W \log_2(1 + \gamma_i(k,t)) \quad (6)$$

where W is the subchannel bandwidth.

2.3 Problem formulation

To guide vehicles in adaptive resource allocation between CAM reliability and video throughput, we design a reward function. When a frame is successfully completed, the reward is proportional to its goodput, i.e., $\eta \cdot \text{Goodput}_i(t)$, where η is a positive scaling factor; no reward is given for incomplete frames. A successful CAM transmission yields a positive reward R , while a failure yields a negative reward R_M , with $R > |R_M|$ to reflect the high priority of safety messages. The objective is to maximize the long-term cumulative discounted reward subject to resource allocation and QoS constraints. The problem can be formulated as:

$$\max \mathbb{E} \left[\sum_{t=0}^T \gamma^t r_i(t) \right] \quad (7)$$

subject to:

$$C1: \sum_k \beta_{i,k} \leq 1, \forall i, t$$

$$C2: \beta_{i,k}, \theta_i \in \{0, 1\}, \forall i, k, t$$

$$C3: p_{i,k} \in \mathcal{P}, \forall i, k, t$$

$$C4: Q_i^{\text{CAM}}(t) \leq Q_{\max}, \forall i, t$$

where the immediate reward $r_i(t)$ is defined as

$$r_i(t) = \mathbb{1}_{\text{frame completion}} \cdot \eta \cdot \text{Goodput}_i(t) + (\mathbb{1}_{\text{CAM success}} R - \mathbb{1}_{\text{CAM failure}} R_M)$$

$\beta_{i,k}$ is the subchannel allocation indicator, θ_i is the communication mode selection (0 : V2I, 1 : V2V), $p_{i,k}$ is the transmit power, and $b_i(t)$ is the video bitrate level. The discount factor $\gamma \in (0, 1)$ balances current and future rewards. Due to the binary variables in the constraints and the non-convex interference coupling in the objective, this optimization problem is NP-hard and cannot be solved exactly by traditional methods. Therefore, we adopt multi-agent deep reinforcement learning to approximate the optimal policy in a distributed framework.

3. Enhanced Multi-Task MAPPO (EMT-MAPPO) Algorithm

In this section, we reformulate this problem as a multi-agent reinforcement learning environment, where each V2V link is treated as an agent. We then introduce the proposed Enhanced Multi-Task MAPPO (EMT-MAPPO) algorithm, with a particular focus on the Task-Aware State augmentation (TAS) that is shown to be the key to performance improvement.

3.1 Multi-Agent Environment Modeling

We consider an NR-V2X sidelink network with N V2V links, as illustrated in Fig. 2.

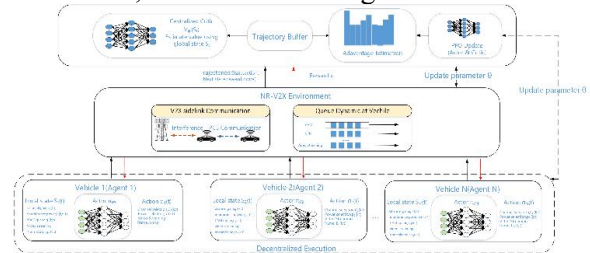


Figure 2. NR-V2X sidelink Multi-Agent Environment

Each V2V link i maintains two service queues: a CAM queue $Q_i^{\text{CAM}}(t)$ and a video queue $Q_i^{\text{video}}(t)$. These two services have distinct arrival processes and QoS requirements. CAM messages are generated periodically and require high reliability and single-shot success. The CAM queue is updated as

$Q_i^{\text{CAM}}(t+1) = \max(0, Q_i^{\text{CAM}}(t) + \mathbb{1}_{\text{gen}} - \mathbb{1}_{\text{suc}})$ (8) where $\mathbb{1}_{\text{gen}}$ indicates whether a new CAM is generated and $\mathbb{1}_{\text{suc}}$ indicates whether the CAM was successfully received. A maximum queue length Q_{max} is imposed to prevent infinite growth. A CAM transmission is successful if the instantaneous rate $r_i(t) \geq R_m$, where R_m is the minimum rate threshold.

Video streaming is frame-based. The frame size depends on the selected bitrate level $b \in \{1, \dots, B\}$, denoted as S_b (typically 30–480 KB). Frames arrive periodically with period $T_{\text{video}} = 100$ ms. The video queue state is represented by the remaining bytes of the current frame:

$Q_i^{\text{video}}(t+1) = \max(0, Q_i^{\text{video}}(t) - r_i(t)\delta t \cdot \mathbb{1}_{\text{vtrans}})$ (9) where $\mathbb{1}_{\text{vtrans}}$ indicates whether the current slot is allocated for video transmission. A frame is successfully completed when $Q_i^{\text{video}}(t)$ reaches zero within the maximum tolerable delay D_0 . To ensure CAM timeliness while preventing video starvation, we adopt a simple dynamic priority rule: at each slot, if the CAM queue is non-empty, CAM is transmitted with priority; otherwise, if video data exists, video is transmitted; otherwise, the link remains idle. This rule guarantees that safety messages are not delayed by video traffic. Standard MAPPO, while effective for cooperative multi-agent tasks, faces two key challenges in this heterogeneous service scenario. First, the state does not explicitly model CAM urgency and video frame progress, causing the policy to favor a single service type. Second, video frame rewards require multi-step accumulation, while CAM failure penalties discourage agents from risking CAM transmissions. To address these challenges, we propose EMT-MAPPO, which introduces Task-Aware State augmentation (TAS) to enable agents to jointly optimize CAM success rate and video frame transmission efficiency.

3.2 EMT-MAPPO Algorithm

This section presents the design of the proposed EMT-MAPPO algorithm. It begins with the core innovation, Task-Aware State augmentation (TAS), which explicitly provides each agent with task-specific urgency and progress information to overcome the limitations of standard MAPPO. The section then defines the action space and reward function, and finally describes the network architecture and training procedure under the centralized training with decentralized execution (CTDE) paradigm.

3.2.1 Task-Aware State Augmentation (TAS)

The core innovation of EMT-MAPPO is the Task-Aware State augmentation (TAS), which explicitly provides each agent with task-specific urgency and progress information. Standard MAPPO observations contain only general information such as channel quality (SINR, interference vectors) and queue lengths, lacking explicit modeling of the differentiated urgency of the two service types. This makes it difficult for the policy to determine when to prioritize CAM and when to focus on video accumulation. TAS addresses this by adding the following scalar features to each agent's local observation:

CAM urgency indicator $u_i^{\text{CAM}} = \min\left(\frac{q_i^{\text{CAM}}}{Q_{\text{max}}}, 1\right)$

where q_i^{CAM} is the current CAM queue length and Q_{max} is the maximum queue capacity (set to 4). When the queue is nearly full, u_i^{CAM} approaches 1, prompting the agent to send CAM immediately to avoid overflow.

Video frame progress $\rho_i = \frac{a_i}{F_i}$

where a_i is the accumulated bytes of the current frame and F_i is the target frame size determined by the selected bitrate. This ratio reflects the completion status: when close to 1, the frame is nearly complete and resources should be continued to obtain the frame completion reward; when close to 0, the frame has just started and can tolerate brief interruptions.

Normalized frame transmission time t_i^{frame} , which helps the agent perceive real-time frame delay.

Normalized global time step: $\tau = t/T_{\text{ep}}$,

which helps the policy distinguish early and late training stages. The final local observation vector for each agent is:

$$o_i = [\text{SINR}_i, I_i, q_i^{\text{CAM}}, u_i^{\text{CAM}}, \rho_i, t_i^{\text{frame}}, \tau] \in \mathbb{R}^{13},$$

where I_i denotes the measured interference power on each subchannel. By explicitly differentiating CAM urgency and video progress, TAS directly reduces the learning difficulty of the policy and enables adaptive resource allocation between the two services.

3.2.2 Action Space and Reward Function

The action of each agent i is defined as a triple:

$$a_i(kt) = (c_i(kt), p_i(kt), b_i(kt)) \quad (10)$$

where $c_i(kt) \in \{0, \dots, M-1\}$ is the selected subchannel index, $p_i(kt) \in \mathcal{P}$ is the transmit power level, and $b_i(kt) \in \{0, \dots, 4\}$ is the video bitrate level (valid only when the communication mode is V2V).

The immediate reward of agent i is:

$$r_i(t) = \mathbb{1}_{\text{frame completion}} \cdot \eta \cdot \text{Goodput}_i(t) + (\mathbb{1}_{\text{CAM success}} R - \mathbb{1}_{\text{CAM failure}} R_M),$$

where $\text{Goodput}_i(t)$ is the throughput upon video frame completion, $R = 10$, and $R_M = 2$. The reward is designed to reflect the high priority of safety messages while encouraging video throughput.

3.2.3 Algorithm Architecture and Training

EMT-MAPPO follows the centralized training with decentralized execution (CTDE) paradigm. All agents share the same Actor and Critic networks. The Critic takes the global state s_t as input, which is the concatenation of all local observations augmented with global timing information, and outputs the state value $V_\phi(s_t)$. The Actor $\pi_\theta(a_i|o_i)$ relies only on local observations, ensuring deployability during decentralized execution.

The Actor loss is defined as:

$$L_{\text{actor}} = -L_{\text{CLIP}}(\theta) - \beta \mathbb{E}_t[H(\pi_\theta(\cdot|o_t))] \quad (11)$$

where $L_{\text{CLIP}}(\theta) = \mathbb{E}_t[\min(r_t(\theta)\hat{A}_t \text{clip}(r_t(\theta), 1 - \epsilon, 1 + \epsilon)\hat{A}_t)]$, ϵ is the clipping hyperparameter, β is the entropy coefficient, and $H(\cdot)$ is the policy entropy. The Critic is updated by minimizing the mean squared error:

$$L_{\text{critic}} = \mathbb{E}_t[(V_\phi(s_t) - \hat{R}_t)^2] \quad (12)$$

where $\hat{R}_t = \sum_{l=0}^{T-t} \gamma^l r_{t+l} + \gamma^{T-t} V_\phi(s_T)$ is the discounted cumulative return. Generalized Advantage Estimation (GAE) is employed to reduce the variance of advantage estimates:

where $\delta_t = r_t + \gamma V_\phi(s_{t+1}) - V_\phi(s_t)$ is the temporal-difference error, γ is the discount factor, and λ is the GAE parameter.

The complete training procedure of EMT-MAPPO is summarized in Algorithm 1. The algorithm follows the standard MAPPO pipeline with TAS integrated into the observation construction. At each training iteration, all agents interact with the environment in parallel to collect trajectories. The collected data are then used to update the Actor and Critic networks via stochastic gradient descent.

Algorithm 1: EMT-MAPPO Training Procedure

1. Initialization: Initialize shared Actor network π_θ and Critic network V_ϕ with random parameters. Set hyperparameters: clipping ϵ , entropy coefficient β , discount factor γ , GAE parameter λ , learning rates, and training episodes

E .

2. for episode $e = 1, 2, \dots, E$ do

3. Reset environment and obtain initial local observations o_i for each agent i .

4. for time step $t = 0, 1, \dots, T - 1$ do

5. For each agent i , augment local observation with TAS features:

$$o_i = [\text{SINR}_i, I_i, q_i^{\text{CAM}}, u_i^{\text{CAM}}, \rho_i, t_i^{\text{frame}}, \tau]$$

6. Each agent selects action $a_i = (c_i, p_i, b_i) \sim \pi_\theta(\cdot|o_i)$.

7. Execute joint actions, observe rewards $r_i(t)$ and next observations o'_i .

8. Store transition (o_i, a_i, r_i, o'_i) in the replay buffer.

9. end for

10. Compute global state s_t as the concatenation of all local observations.

11. Compute discounted cumulative returns $\hat{R}_t$ and GAE advantages $\hat{A}_t$.

12. for each training epoch do

13. Update Actor: $\theta \leftarrow \theta - \alpha_\theta \nabla_\theta L_{\text{actor}}$.

14. Update Critic: $\phi \leftarrow \phi - \alpha_\phi \nabla_\phi L_{\text{critic}}$.

15. end for

16. Clear replay buffer.

17. end for

18. Output: Trained Actor network π_θ for decentralized execution.

In the execution phase, each agent independently selects actions based on its local observation o_i using the trained Actor network, without requiring the global state or other agents' information. This ensures that the algorithm is fully deployable in a distributed manner, making it suitable for real-world NR-V2X vehicular networks where centralized coordination is impractical.

4. Experimental Setup and Analysis

This section presents the simulation setup and performance analysis of the proposed EMT-MAPPO algorithm. We first describe the simulation environment and parameter configuration, then introduce the benchmark algorithms and evaluation metrics, and finally discuss the experimental design for validating the effectiveness of the proposed method.

4.1 Simulation Parameters

Table 1. NR-V2X Environment Parameters

Parameter	Value
Communication interval	0.5 ms
Carrier frequency	5.9 GHz
Channel bandwidth	720 kHz

Bandwidth per RB	1 MHz
RBs per subchannel	2
Number of V2I links/subchannels	4
Number of V2V links/agents	4
V2I transmit power	23 dBm
V2V transmit power levels	[23, 20, 17, 14, 11, 8, 5] dBm
BS antenna height	25 m
BS antenna gain	8 dBi
BS noise figure	9 dB
Vehicle antenna gain	3 dBi
Vehicle noise figure	9 dB
Fast fading mode	3GPP TR 38.901 (UMi/UMa)
CAM size	300 bytes
Frame sizes	[30, 60, 120, 240, 480] KB

Table 2. EMT-MAPPO Hyperparameters

Hyperparameter	Value
Actor/Critic hidden dimension (FC)	128
Discount factor (γ)	0.99
Batch size	256
Mini-batches	4
PPO epochs	10
Entropy coefficient	0.001
Parameter sharing	True

All experiments are conducted on a PC equipped with Ubuntu 22.04, an Intel Xeon w5-2455 processor at 3.2 GHz, and 64 GB RAM. The simulation environment follows the evaluation methodology for the freeway scenario specified in Annex A of 3GPP TR 38.901, including road configuration, V2V, and V2I channel models. Vehicle positions are generated by SUMO, simulating a 2 km-long highway with three lanes in each direction (six lanes total). Three representative scenarios are considered, following existing work [15]: Scenario 1 (35 vehicles, 250 km/h), Scenario 3 (123 vehicles, 70 km/h), and Scenario 6 (500 vehicles, 50 km/h). The key parameters of the NR-V2X environment are summarized in Table 1. The hyperparameters of the proposed EMT-MAPPO algorithm are listed in Table 2.

4.2 Benchmark Algorithms and Evaluation Metrics

We compare EMT-MAPPO with MAPPO [23], MAA2C [24], BP-SPS [6], and random scheduling. MAPPO is the primary component-control baseline, as it differs from EMT-MAPPO only by removing the task-aware state features.

MAA2C represents a multi-agent actor-critic alternative, while BP-SPS serves as a traditional scheduling baseline. Random scheduling is included as a non-learning reference. All learning-based methods are evaluated with the same complete action space, training budget, evaluation traces, and random seeds, ensuring a fair comparison.

We report three metrics: (i) frame success rate (FSR), defined as the fraction of initiated video frames that are completed; (ii) CAM success rate (CSR), defined as the fraction of attempted CAM transmissions that succeed; (iii) episode reward, defined as the cumulative multi-agent reward over one episode. For each method, we report the mean, sample standard deviation, and 95% confidence interval across five independent random seeds.

4.3 Overall Performance

Table 3 summarizes the performance comparison between EMT-MAPPO and four baseline methods for heterogeneous service scheduling in NR-V2X Mode 2. The evaluation metrics include cumulative reward (Reward), frame success rate (FSR), and CAM success rate (CSR), reflecting the overall optimization performance, video transmission efficiency, and safety message reliability, respectively.

Table 3. Performance comparison of different scheduling algorithms in terms of reward, FSR, and CSR

Method	Reward	FSR	CSR
EMT-MAPPO	5166.05	90.33%	92.26%
MAPPO	3346.12	86.53%	87.81%
MAA2C	1218.90	74.26%	88.37%
BP-SPS	1639.98	77.83%	77.09%
Random	591.24	55.65%	88.14%

EMT-MAPPO outperforms all baselines on reward, FSR, and CSR. The fundamental reason is that Task-Aware State augmentation (TAS) explicitly provides the policy with CAM queue urgency and video frame progress, enabling the agent to distinguish the real-time requirements of the two services and thereby allocate resources in a more rational manner that balances the timeliness of safety messages against the cumulative efficiency of video transmission. Compared with MAPPO, the reward improves by approximately 54.4%, and FSR and CSR increase by 3.80 and 4.45 percentage points, respectively, validating the effectiveness of TAS. MAA2C and Random exhibit high CSR but low FSR because their policies tend to prioritize CAM transmission

and lack awareness of video frame accumulation progress, causing video transmission to be starved of sufficient resources over time. BP-SPS struggles to accommodate both services under dynamic fast-fading channels, failing to guarantee CAM timeliness while also failing to complete video frame transmissions efficiently. EMT-MAPPO is the only method that maintains CSR above 92% while raising FSR above 90%, indicating that TAS enables the policy to proactively reserve transmission opportunities for video frames while ensuring the reliability of safety messages, thereby achieving joint optimization of safety and throughput.

4.4 Balancing CAM and Video Services

Figure 3 illustrates the trade-off between video reliability and CAM delivery reliability in the four-agent setting. EMT-MAPPO occupies the upper-right region, achieving the highest frame success rate (90.33%) and CAM success rate (92.26%) among all evaluated methods. Compared with MAPPO, EMT-MAPPO improves FSR by 3.80 percentage points and CSR by 4.45 percentage points, demonstrating that the task-aware state representation more effectively coordinates video-frame completion with periodic CAM transmission. Although BP-SPS yields a higher FSR than random scheduling, its lower CSR indicates a weaker balance between the two services. Overall, EMT-MAPPO achieves the most favorable trade-off between CAM and video services.

4.5 Training Stability and Convergence

Figure 4 presents the training trajectories of the learning-based methods, namely EMT-MAPPO, MAPPO, and MAA2C, under the 4-, 8-, and 16-agent settings. Non-learning baselines such as Random and BP-SPS are not included, as they do not involve a training process or convergence behaviour. Across all scales, EMT-MAPPO exhibited smooth reward growth in the late training stage and ultimately achieved the highest reward, whereas MAPPO improved more rapidly in the early phase but entered a lower-reward plateau. EMT-MAPPO overtook MAPPO at approximately episodes 65, 65, and 83 for the 4-, 8-, and 16-agent settings, respectively, indicating that the task-aware state augmentation required an initial adaptation period but produced a more effective policy as training progressed. Its FSR increased steadily and reached the highest final level in all three settings, while the curves

remained stable without pronounced oscillations. The CSR trajectories of all methods were relatively stable throughout training; notably, EMT-MAPPO maintained high CAM reliability while improving video-frame delivery, although its CSR was not uniformly higher than that of MAPPO in the 4-agent setting. In contrast, MAA2C showed substantially lower reward and FSR across all agent scales. Overall, the results support the improved late-stage learning stability and scalability of EMT-MAPPO, but the continuing upward reward trend at the end of training suggests that longer training horizons and multi-seed confidence intervals are needed before full convergence can be claimed.

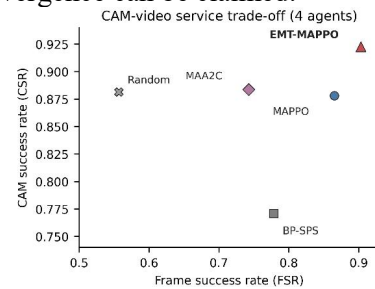


Figure 3. Performance Comparison of CAM Reliability and Video Reliability

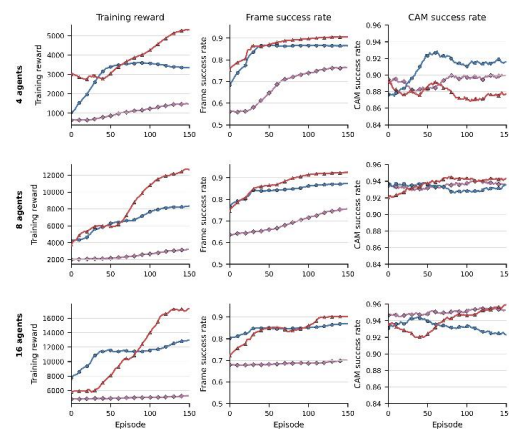


Figure 4. Training Reward, Frame Success Rate, and CAM Success Rate of EMT-MAPPO, MAPPO, and MAA2C under Different Agent Scales

5. Conclusion and Future Work

This paper addressed the joint resource allocation problem for CAM and video streaming in NR-V2X Mode 2, where safety-critical and bandwidth-intensive services compete for limited radio resources under dynamic fast-fading channel conditions. To tackle this challenge, we proposed EMT-MAPPO, a MAPPO-based framework that augments each local policy observation with task-aware CAM queue

occupancy and video-frame transmission progress. Experimental results demonstrated that EMT-MAPPO consistently outperformed standard MAPPO, MAA2C, BP-SPS, and random scheduling, achieving a 54.4% higher reward, a 3.80 percentage point improvement in frame success rate, and a 4.45 percentage point improvement in CAM success rate over MAPPO in the four-agent setting. Furthermore, we analysed the Pareto trade-off between video reliability and CAM reliability in the four-agent setting and compared the convergence behaviour across 4, 8, and 16 agents. EMT-MAPPO achieved the most favourable trade-off and consistently reached the highest reward with stable convergence at all scales, whereas MAPPO and MAA2C exhibited lower reward plateaus or slower convergence.

Despite these promising results, several limitations remain. First, the centralized training framework relies on global information at the critic, raising communication and privacy concerns in practical vehicular networks. Future work will investigate federated multi-agent reinforcement learning to reduce reliance on centralized information and enhance data privacy. Second, the current study focuses on a relatively simple straight-highway topology. More complex scenarios, including urban intersections, multi-lane roads, and heterogeneous road users such as pedestrians and cyclists, will be considered to evaluate the generalization capability of EMT-MAPPO. Finally, multi-objective optimization and Pareto-based decision mechanisms will be explored to provide more flexible trade-offs among safety, latency, reliability, and video throughput. These extensions may further improve the applicability of multi-agent reinforcement learning to heterogeneous and highly dynamic NR-V2X networks.

Acknowledgements

This research was supported by the Science and Technology Research program of Chongqing Municipal Education Commission (Grant No. KJQN202303107).

References

[1] Kandu SS, Prakash A, Tiwari M, Tripathi S, Gupta N, Pal R. C-V2X and 5G-V2X: Challenges and Improvement Strategies. In: Proceedings of the 2025 Third International Conference on Microwave, Antenna and Communication (MAC); 2025; Bhopal, India.

Piscataway, NJ: IEEE; 2025. p. 1-6.

[2] Shin C, Farag E, Ryu H, et al. Vehicle-to-everything (V2X) evolution from 4G to 5G in 3GPP: Focusing on resource allocation aspects. *IEEE Access*. 2023; 11: 18689-18703.

[3] 3GPP. Study on NR Vehicle-to-Everything (V2X). 3GPP TR 38.885, Release 16. Sophia Antipolis: 3GPP; 2019.

[4] Wael CBA, Dogheche EH, Armi N, Subekti A, Dayoub I. Leveraging 3GPP Features and Optimization Techniques for 5G NR-V2X Resource Allocation: A Survey. *IEEE Open J Intell Transp Syst*. 2025; 6: 967-994.

[5] Zhou X, Hui F, Liu J, Wang W, Zhang J. Resource Reservation in C-V2X Networks for Dynamic Traffic Environments: From Vehicle Density-Driven to Deep Reinforcement Learning. *IEEE Trans Veh Technol*. 2025; 74(11):17429-17444.

[6] Dayal A, Shah VK, Choudhury B, Marojevic V, Dietrich C, Reed JH. Adaptive Semi-Persistent Scheduling for Enhanced On-road Safety in Decentralized V2X Networks. In: Proceedings of the 2021 IFIP Networking Conference; 2021; Espoo and Helsinki, Finland. Piscataway, NJ: IEEE; 2021. p. 1-9.

[7] Hyeon D, Lee C, Kim H, Cho S, Paek J, Govindan R. Probabilistic resource rescheduling for C-V2X based on delivery rate estimation. *J Commun Netw*. 2024; 26(2):239-251.

[8] Jung S, Kim JH, Kim J. Intelligent extra resource allocation for cooperative awareness message broadcasting in cellular-V2X networks. *IEEE Trans Netw Serv Manag*. 2025; 22 (2):1677-1689.

[9] Fu Q, Liu J, Wang J, et al. Toward throughput optimization: A congestion control scheme for CAM delivery via NR sidelink. *IEEE Internet Things J*. 2025; 12 (9):12206-12218.

[10] S. Heo and H. Kim. Two-Dimensional Compound Configured Grants for Video-Aware NR Sidelink Mode 1 Scheduling. *IEEE Trans Veh Technol*. 2026.

[11] Wolff VA, Xhoxhi E, Schiegg FA, et al. Uncertainty and prioritization: Empirical evaluation of a VoI-based CPM generation pipeline using real-world data. In: Proceedings of the 2025 IEEE Vehicular Networking Conference (VNC); 2025; Piscataway, NJ: IEEE; 2025. p. 1-8.

[12] Yordanov N, Çavuşoğlu B. Transient, deadline-aware RAN slicing for NR-V2X

- using finite-capacity time-varying slice queues. *Veh Commun.* 2026;101064.
- [13] Huan Z, Sun J, Chen Z, et al. Resource Allocation in V2X Networks: A Double Deep Q-Network Approach with Graph Neural Networks. *Comput Mater Contin.* 2025; 84(3):5427-5443.
- [14] Song S, Zhang Z, Wu Q, Fan P, Fan Q. Joint optimization of age of information and energy consumption in NR-V2X system based on deep reinforcement learning. *Sensors.* 2024; 24(13):4338.
- [15] Wang S, Lei L, Maheshwari P, Bellefeuille S, Zheng K, Niyato D. Multi-Agent DRL for V2X Resource Allocation: Disentangling Challenges and Benchmarking Solutions. *arXiv preprint arXiv:2603.06607.* 2026.
- [16] Wang C, Guan Y, Peng S, et al. Data prioritization aware resource allocation in internet of vehicles using multi-agent deep reinforcement learning. *Neural Netw.* 2025; 190: 107671.
- [17] Gao A, Zhu Z, Zhang J, et al. Matching combined heterogeneous multi-agent reinforcement learning for resource allocation in NOMA-V2X networks. *IEEE Trans Veh Technol.* 2024; 73(10):15109-15124.
- [18] Hegde A, Song R, Festag A. Radio resource allocation in 5G-NR V2X: A multi-agent actor-critic based approach. *IEEE Access.* 2023; 11: 87225-87244.
- [19] M. M. Saad, M. A. Tariq, M. Ajmal, D. Kim and G. Srivastava. Federated Multiagent Reinforcement Learning for Resource Allocation in NR-V2X Mode 2. *IEEE Internet Things J.* 2025; 12(13):23402-23417.
- [20] Morshed M, Zaman M. Curriculum-assisted multi-agent reinforcement learning for scalable V2X resource allocation. *Phys Commun.* 2026; 76.
- [21] I KP, Venkatesh TG. Integrated Sensing, Communication, and Computing for NR-V2X: A Cross-Layer Resource Allocation Framework Using Multi-Agent Reinforcement Learning. *arXiv preprint arXiv:2605.24972.* 2026.
- [22] Fair Resource Allocation in C-V2X Networks: A Reinforcement Learning Approach for Perception Message Dissemination. In: *Proceedings of the IEEE International Conference on Cloud and Big Data Computing (CBDCom); 2025 Oct; Hakodate, Japan.* Piscataway, NJ: IEEE; 2025.
- [23] Basset L, Bouriga A, Dhaou R. Learning to Talk: Emergent CAM/CPM Scheduling via MAPPO in Constrained V2X Networks. In: *Proceedings of the 2025 International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM); 2025; Piscataway, NJ: IEEE; 2025.* p. 767-770.
- [24] Khan I, Haladappa MS. Risk-Aware Multi-Agent Advantage Actor-Critic Based Resource Allocation for C-V2X Communication in Cellular Networks. *Proc Eng Technol Innov.* 2025; 29:47-60.
- [25] Gu B, Chen W, et al. Multiagent reinforcement learning-based semi-persistent scheduling scheme in C-V2X mode 4. *IEEE Trans Veh Technol.* 2022; 71(11):12044-12056.