

VIL2C: Value-of-Information Aware Low-Latency Communication for Multi-Agent Reinforcement Learning

Qian Zhang¹, Zhuo Sun^{1*}, Yao Zhang¹, Zhiwen Yu^{1,2}, Bin Guo¹, Jun Zhang³

¹Northwestern Polytechnical University

²Harbin Engineering University

³The Hong Kong University of Science and Technology

qianz5152@mail.nwpu.edu.cn, zsun@nwpu.edu.cn, yaozh.g@nwpu.edu.cn, zhiwenyu@nwpu.edu.cn, guob@nwpu.edu.cn, eejzhang@ust.hk

Abstract

Inter-agent communication serves as an effective mechanism for enhancing performance in collaborative multi-agent reinforcement learning (MARL) systems. However, the inherent communication latency in practical systems induces both action decision delays and outdated information sharing, impeding MARL performance gains, particularly in time-critical applications like autonomous driving. In this work, we propose a **Value-of-Information aware Low-latency Communication (VIL2C)** scheme that proactively adjusts the latency distribution to mitigate its effects in MARL systems. Specifically, we define a Value of Information (VOI) metric to quantify the importance of delayed message transmission based on each delayed message’s importance. Moreover, we propose a progressive message reception mechanism to adaptively adjust the reception duration based on received messages. We derive the optimized VoI aware resource allocation and theoretically prove the performance advantage of the proposed VIL2C scheme. Extensive experiments demonstrate that VIL2C outperforms existing approaches under various communication conditions. These gains are attributed to the low-latency transmission of high-VoI messages via resource allocation and the elimination of unnecessary waiting periods via adaptive reception duration.

Introduction

Multi-Agent Reinforcement Learning (MARL) has shown impressive potential in solving complex multi-agent decision-making problems across real-world applications, such as autonomous driving (Palanisamy 2020), traffic control (Wei et al. 2019), and robot control (Gu et al. 2023). While pioneering studies devote considerable efforts to develop algorithms under the centralized training and decentralized execution (CTDE) paradigm (Yu et al. 2022), MARL still suffers from the challenges caused by the non-stationary (Oroojlooy and Hajinezhad 2023) and partially observable environments (Foerster et al. 2016).

To address these challenges, inter-agent communication is incorporated into MARL (Zhu, Dastani, and Wang 2022). Early works (Zhang, Zhang, and Lin 2019) adopt the broadcast mechanism, allowing agents to share messages with

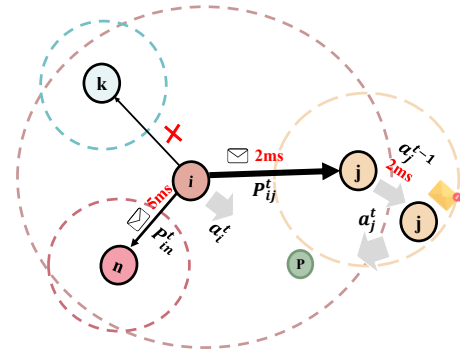


Figure 1: The toy example illustration of VIL2C. Agent i transmits messages to agents j and n to collaboratively capture target P . Based on the message importance relative to recipients j and n , agent i allocates transmission resources (represented by arrow thickness) to proactively adjust latency. Agent j terminates reception (represented by red dot), once receiving sufficient messages. Black arrows represent communication links and gray arrows represent movement.

each other, which incurs substantial communication overheads. Resource constraints in communication systems hinder the deployment of broadcast mechanisms in real-world multi-agent systems. Thus, several studies (Hu et al. 2023; Sun et al. 2024; Guo, Shi, and Fan 2023; Li and Zhang 2024; Wang et al. 2020a) explore effective communication protocols, where agents dynamically determine the communication time, partners, and messages based on demands. Although these protocols effectively reduce communication overheads, they ignore the inevitable and dynamic communication latency in real-world systems.

Communication latency may impair multi-agent cooperation through causing action decision delays (Chen et al. 2021, 2020) and outdated information sharing. Particularly, the latency of decision-critical messages dominates this impact. Previous work (Yuan et al. 2023) avoids unbounded waiting by dynamically adjusting the waiting time based on real-time network conditions. CoDe (Song et al. 2025) alleviates misguidance of outdated messages by inferring the long-term intent and prioritizing the most relevant recent

*Corresponding author: zsun@nwpu.edu.cn

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

messages. While these approaches mitigate the latency impact on cooperation performance via well-designed reception schemes, they rely on the channel-dependent inherent latency distribution but overlook the potentials of proactive latency distribution adjustment. Proactive and precise adjustment of latency distribution can be achieved by effective resource allocation in wireless communication systems (Sun et al. 2019). Unlike conventional wireless communication systems that focus on maximizing sum rates or minimizing packet-level latency (She, Yang, and Quek 2018; Liu et al. 2023; Meng et al. 2022), the communication design in MARL requires considering the heterogeneous importance of delayed messages on recipients’ decisions to improve cooperation performance. Thus, a novel communication mechanism is needed to adapt latency according to message importance under latency awareness.

In this work, we propose a Value-of-Information aware Low-Latency multi-agent Communication (VIL2C) scheme. Firstly, we define Value of Information (VoI) in MARL to quantify how delayed messages influence recipient agents’ action decisions, accounting for both their semantic importance and associated communication latency. Based on the defined VoI, we jointly design a message transmission scheme and a progressive reception strategy. The message transmission scheme dynamically optimizes both bandwidth and power allocation by jointly considering the heterogeneous VoI of different messages and time-varying wireless channels. Such resource allocation can reduce communication latency of messages that critically influence the recipient’s action decision, thereby mitigating the impact on multi-agent cooperation performance. Meanwhile, the progressive message reception strategy adaptively determines the recipient’s waiting time of communication, ceasing reception once sufficient information for reliable action decision is obtained. This strategy aims to minimize the waiting duration while making a reliable action decision. A toy example is shown in Fig. 1, where agent i prioritizes resource allocation for transmitting messages to agent j and agent j terminates its reception process upon the receipt of sufficient messages. Moreover, we provide theoretical analysis that establishes the VIL2C’s performance guarantee in enhancing multi-agent cooperation. We evaluate the proposed VIL2C scheme on three typical multi-agent environments: Predator-Prey (PP), Cooperative Navigation (CN) (Lowe et al. 2017), and Starcraft Multi-Agent Challenge (SMACv2) (Ellis et al. 2023). Extensive experiment results demonstrate that VIL2C outperforms all considered baselines. We also apply VIL2C and baselines under different resource budgets and channel conditions. The experiments show that VIL2C exhibits strong robustness against various communication conditions. The contributions of this work are summarized as follows:

- To the best of our knowledge, this work is the first attempt to proactively adjust communication latency distribution in MARL by VoI aware resource allocation.
- We propose a VoI aware low-latency communication system in MARL, named VIL2C, which innovatively introduces VoI into the designs of message transmission and

reception to improve cooperation performance with considering communication latency.

- We derive the optimized VoI aware resource allocation in VIL2C. We also provide theoretical analysis that establishes the VIL2C’s performance guarantee in enhancing multi-agent cooperation performance.
- We test the proposed VIL2C scheme in three typical multi-agent environments. VIL2C consistently demonstrates superior performance compared with baselines across various resource budgets and channel conditions.

Related Works

Learning communication protocols in collaborative MARL has drawn significant attention, as strategic information exchange enhances agents’ environmental awareness and cooperative performance. RIAL and DIAL (Foerster et al. 2016) introduced end-to-end communication strategies, followed by broadcast mechanisms (Sukhbaatar, Fergus et al. 2016; Peng et al. 2017; Das et al. 2019). Although simple to implement, broadcast mechanisms incurs high overheads, motivating the development of more effective protocols.

Effective multi-agent communication protocols primarily address three key aspects: when to communicate, whom to communicate with, and what information to transmit. For timing, Gated-ACML (Mao et al. 2020) and IC3Net (Singh, Jain, and Sukhbaatar 2019) employed gating mechanisms, while ETCNet (Hu et al. 2023) used event-triggered networks to decide whether and whom to engage. For recipient selection, I2C (Ding, Huang, and Lu 2020) and ToM2C (Wang et al. 2022) leveraged causal inference. TEM (Guo, Shi, and Fan 2023) defined Kullback-Leibler (KL) divergence based importance metric to select transmissions. For content, T2MAC (Sun et al. 2024) supported selective engagement and evidence-driven integration. IMAC (Wang et al. 2020a) applied information bottleneck for compact messages. The studies (Wang et al. 2020b; Yuan et al. 2022; Li and Zhang 2024) generated personalized messages tailored to recipients’ needs. While the protocols reduce communication overhead, they neglect latency’s impact.

Communication latency is unavoidable in practical multi-agent systems. DA-MG (Chen et al. 2021) modeled delay aware Markov games to assess action delays’ impact on decision quality. DACOM (Yuan et al. 2023) integrated latency into MARL with adaptive waiting time scheduling for dynamic networks. CoDe (Song et al. 2025) addressed latency across decision intervals via dual alignment for asynchronous message integration. These methods rely on channel-dependent latency distributions, whereas our work explores proactive adjustment through joint message transmission and reception design.

Background

Delay aware Decentralized Partially Observable Markov Decision Process

The Delay aware Decentralized Partially Observable Markov Decision Process (Da-Dec-POMDP) extends the standard Dec-POMDP framework to explicitly

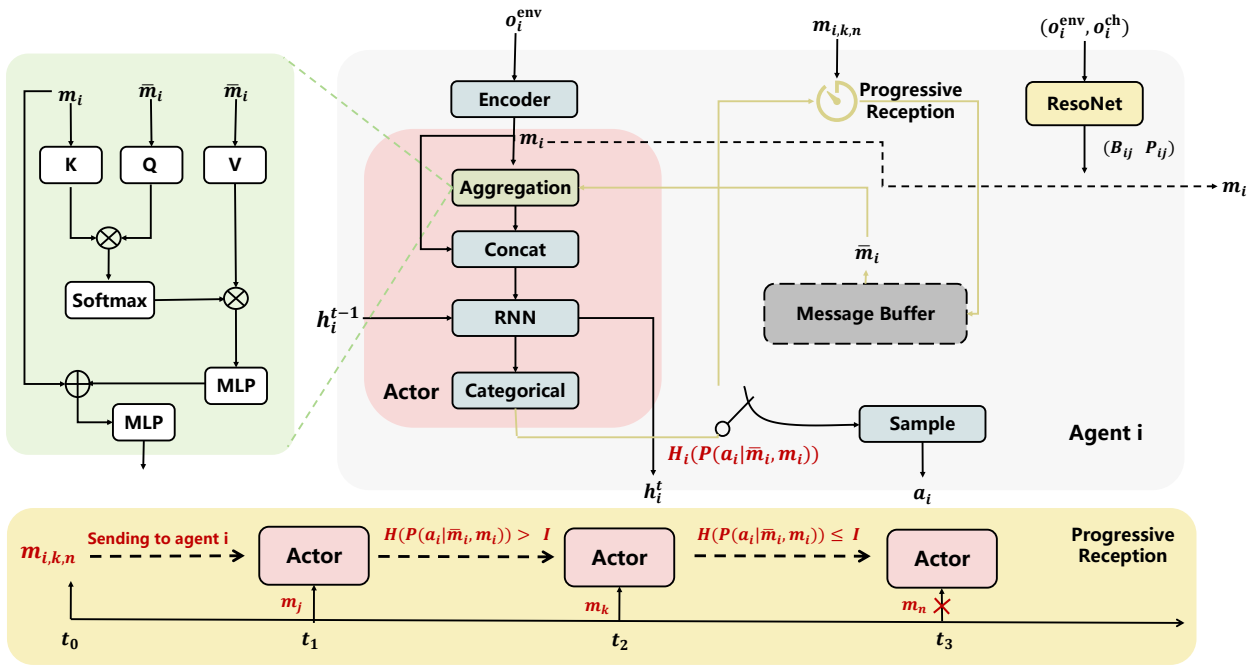


Figure 2: The framework of VIL2C. Each agent consists of six components: 1) an Encoder that generates the message from the local observation, 2) a ResoNet that performs the online resource allocation, 3) a Message Buffer that stores received messages, 4) an Actor that utilizes the own message and received messages to obtain an action probability distribution, 5) a Progressive Reception module that determines whether to stop receiving messages, illustrated by the yellow box, and the corresponding process is represented by yellow lines, 6) a Sample module to select actions from the action probability distribution.

model communication latency in multi-agent systems. Formally, the Da-Dec-POMDP is defined as $\mathcal{G} = \langle \mathcal{S}, \mathcal{A}, \mathcal{O}, P, R, \Omega, N, \gamma, \mathcal{T} \rangle$. Here, N is the number of agents in the system, \mathcal{S} is the global state space, \mathcal{A} is the joint action space of all agents, \mathcal{O} is the partial observation space for agents. P represents the transition function characterizing the effect of joint action on states, R represents the reward function used to assess the quality of action decisions in current states, and Ω represents the observation function mapping states to observations. Meanwhile, γ is the discount factor and \mathcal{T} is the communication latency distribution. At timestep t , agent i , $i \in [1, \dots, N]$, obtains the local observation $o_i^t = (o_i^{env}, o_i^{ch})$ that captures both environment and channel information. It then transmits its observation to the selected agent j with a random latency $\tau_{i,j} \sim \mathcal{T}_{i,j}$. Based on the local observation and received messages, agent i decides its action a_i^t according to the policy π_i . The joint action of all agents \mathbf{a}^t leads to state transition and generates the global reward $r^t = R(s^t, \mathbf{a}^t)$. The objective is to learn the joint policy π that maximizes the expected discounted return $J(\pi) = \mathbb{E}_\pi [\sum_{t=0}^{\infty} \gamma^t r^t]$ under the stochastic dynamics induced by policies, transitions, and communication latency.

Communication Channel Model

In real-world multi-agent cooperation scenarios, the agents exchange messages via a wireless communication network. In wireless networks, the point-to-point communication la-

tency between two agents is mainly determined by the transmission rate for a given message size. According to (Goldsmith 2005), the transmission rate from agent i to agent j is given by

$$R_{i,j} = B_{i,j} \log_2 \left(1 + \frac{P_{i,j}}{10^{\frac{PL_{i,j}}{10}} B_{i,j} N_0} \right), \quad (1)$$

where $B_{i,j}$ and $P_{i,j}$ refer to the bandwidth (Hz) and transmit power (W) allocated by agent i for transmission to agent j , respectively, and N_0 is the power spectral density of additive white Gaussian noise (AWGN). Here, $PL_{i,j}$ (in dB) refers to the distance-dependent path loss used to measure the signal propagation attenuation in wireless channels, modeled as

$$PL_{i,j} = \eta \log_{10} d_{i,j} + c, \quad (2)$$

where $d_{i,j}$ is the distance between agent i and agent j , and c is the offset term representing other fixed effects. η represents the path loss exponent determined by the communication environment, such as $\eta = 2$ for the free space and $\eta = 3.67$ for the urban area (Cho et al. 2010). It can be seen that the available transmission rate depends on the transmit agent's bandwidth and transmit power allocation under the given channel condition.

Methodology

In this section, we present the detailed design of VIL2C. As the communication module, VIL2C works with the action

decision module in multi-agent proximal policy optimization (MAPPO), which is a CTDE method. Since it is parallel to the action decision module, VIL2C is compatible to any CTDE method.

Value-of-Information Aware Low-latency Communication

In this part, we first define VoI to measure the importance of delayed messages on the recipient agent's action decision. Based on VoI, we propose the VIL2C module, which consists of a message transmission scheme and a progressive message reception strategy, as shown in Fig.2. We also derive a lower bound of cooperation performance in the presence of communication latency and theoretically prove the performance improvement of VoI aware delay distribution adjustment obtained by the proposed VIL2C.

Value of Information. In cooperative MARL scenarios, the message importance relative to a recipient agent is evaluated by the influence of its semantic information on this agent's action decision. A message is considered highly important when it significantly influences the recipient agent's action decision. Otherwise, it is deemed low importance. On the other hand, receiving the message incurs latency as a cost. We define VoI as the message importance per latency cost, given by

$$VoI_{i,j} = \frac{\xi_{i,j}}{\tau_{i,j}}, \quad (3)$$

where $\xi_{i,j}$ represents the influence of agent i 's message m_i on recipient agent j 's action decision and $\tau_{i,j}$ is communication latency of message m_i to recipient agent j . To quantify the influence of message m_i on recipient agent j 's action decision, we utilize KL divergence to measure the discrepancy in the action probability distribution of recipient agent j before and after receiving message m_i , expressed as

$$\xi_{i,j} = D_{KL}(P(a_j | o_j, m_i) || P(a_j | o_j)), \quad (4)$$

where $P(a_j | o_j)$ and $P(a_j | o_j, m_i)$ denote the action probability distribution of agent j before and after receiving message m_i , respectively. The communication latency of message m_i to agent j is

$$\tau_{i,j} = \frac{L_i}{R_{i,j}}, \quad (5)$$

where L_i is the size of message m_i in bits and $R_{i,j}$ is the available transmission rate from agent i to j . From Eqs. (1) and (5), it is observed that bandwidth and transmit power allocation can affect available transmission rates, thereby reshaping the inter-agent communication latency distribution.

VoI aware Transmission. To reshape the inter-agent communication latency distribution, i.e., reducing communication latency of important messages, we propose the VoI aware message transmission scheme. In this scheme, the transmit agent optimizes its bandwidth and power allocation to maximize the total VoI of its transmissions. This resource allocation problem is formulated as

$$\arg \max_{\{B_{i,j}, P_{i,j}\}} \sum_{j=1}^{N_i} VoI_{i,j}(B_{i,j}, P_{i,j}), \quad (6)$$

$$\text{s.t.} \quad \sum_{j=1}^{N_i} P_{i,j} \leq P_{\text{budget}}, \quad \sum_{j=1}^{N_i} B_{i,j} \leq B_{\text{budget}},$$

where N_i is the number of agent i 's recipients and $VoI_{i,j}(B_{i,j}, P_{i,j})$ is obtained by incorporating Eqs. (1), (4), and (5) into (3). Here, $B_{i,j}$ and $P_{i,j}$ are the bandwidth and transmit power allocated by agent i for the transmission to agent j , respectively. B_{budget} and P_{budget} are the transmit agent's available bandwidth and power budget, respectively. We first consider that each agent can obtain the KL divergence based importance of its message on recipients' action decisions before transmissions. In this case, the optimization problem is solved by using Karush-Kuhn-Tucker (KKT) (Kuhn and Tucker 2014) conditions in Proposition 1.

Proposition 1. Consider the importance $\xi_{i,j}$ of agent i 's message on recipient j . The optimized bandwidth $B_{i,j}^*$ and transmit power $P_{i,j}^*$ allocated by agent i to agent j satisfy

$$\frac{\xi_{i,j}}{L_i} \left(\log_2(1 + \gamma_{i,j}) - \frac{\gamma_{i,j}}{1 + \gamma_{i,j}} \right) = \lambda, \quad (7a)$$

$$\frac{\xi_{i,j} \gamma_{i,j} B_{i,j}^*}{L_i P_{i,j}^* (1 + \gamma_{i,j})} = \mu. \quad (7b)$$

respectively, and $B_{i,j}^*$ and $P_{i,j}^*$ are proportional to $\xi_{i,j}$. Here, $\gamma_{i,j} = \frac{P_{i,j}^*}{10^{\frac{PL_{i,j}}{10}} B_{i,j}^* N_0}$, $\tau_{i,j}$ is communication latency from agent i to j , defined as Eq. (5). μ and λ are Lagrange multipliers for bandwidth and power constraints, respectively.

It is concluded from Proposition 1 that it preferentially allocates bandwidth and power to critical messages, thereby reducing their communication latency.

In practice, it is difficult for agents to obtain the importance of their messages before transmission. To address this, we design a neural network within VIL2C to optimize resource allocation, called **ResoNet**. By exploiting the CTDE method, ResoNet is first trained to maximize the total VoI based on global information and message importance in a centralized way. Then, ResoNet uses local state information, including local observation and wireless channels to other agents, as input and decides the optimized bandwidth and power allocation in a decentralized way, as shown in Fig. 2. The detailed network design will be presented next.

Progressive Reception. The agent waits for receiving messages from multiple agents. Due to the existence of communication latency, messages arrive asynchronously. The reception of more messages means longer waiting duration, which may cause unaffordable action delays and task failure. Conversely, if the waiting duration is too short, it may miss many messages and degrade multi-agent cooperation performance. Thus, we propose a progressive reception strategy to minimize the waiting duration while ensuring a reliable action decision. Specifically, the recipient agent immediately processes this message and obtains the corresponding action probability distribution, once one message is received. When the uncertainty of its action probability distribution is smaller than a predefined threshold, the agent terminates its reception and takes an action. Otherwise, the agent continues waiting for more messages. We adopt the entropy to

quantify the uncertainty of an action probability distribution. The termination condition of message reception is

$$H(P(a | \bar{m}_i, m_i)) \leq I, \quad (8)$$

where $H(\cdot)$ is the entropy function and $P(a | \bar{m}_i, m_i)$ is agent i 's action probability distribution conditioned on messages of the buffer \bar{m}_i and its own message m_i . Here, I is a predefined entropy threshold, indicating the maximum acceptable uncertainty of action probability distribution to obtain a satisfactory action decision. It is obvious that a larger entropy threshold leads to the reception of more messages and longer waiting, and vice versa. Then, there exists an optimal entropy threshold to balance two effects and to achieve the maximum reward. This optimal entropy threshold is obtained by the grid search from 0 and $\log_2 n$, where n is the dimension of action space in the considered task. In the proposed VIL2C, the message transmission scheme significantly reduces communication latency of important messages via preferential resource allocation. Thus, important messages are received earlier with a high probability and the termination condition can be satisfied in advance, thereby reducing the waiting duration. Note that, a maximum waiting time is set to prevent unbounded waiting. The effect of maximum waiting time will be evaluated in the experiments. We also note that when the recipient agent decides to terminate reception, it feeds back an ACK to notify other agents to halt their message transmission, thereby releasing communication resources (Ahmed et al. 2021).

As illustrated in Fig. 2, agents transmit messages to agent i at time t_0 . Due to the presence of communication latency, agent i receives the agent j 's message at time t_1 . After processing the message, it checks whether the termination condition is satisfied. When the termination condition is not satisfied, agent i continues waiting for messages from other agents. At time t_2 , agent i receives the agent k 's message and processes it to check the termination condition. When it is satisfied, agent i terminates reception and takes an action.

Theoretical analysis. We use the cumulative reward as a performance metric to theoretically illustrate the advantage of VoI aware latency distribution adjustment in VIL2C. And, we first derive a lower bound for the cumulative reward under an average communication latency in Proposition 2.

Proposition 2. Consider a cumulative reward J_{opt} of MARL without considering communication latency. We have the following lower bound for the cumulative reward of communication latency aware MARL:

$$J \geq J_{opt} - C \cdot \bar{\tau}, \quad (9)$$

where $\bar{\tau}$ is the mean of communication latency distribution and $C > 0$ is a task-dependent constant representing the latency sensitivity.

Then, we analyze the advantage of VoI aware latency distribution adjustment in VIL2C. In previous multi-agent communication works, where resources are equally allocated to transmission links and the channel-dependent inherent latency distribution is used, the cumulative reward is modeled as J_r . Compared to J_r , the cumulative reward improvement achieved by our latency distribution adjustment is derived in Proposition 3.

Proposition 3. Given the average communication latency $\bar{\tau}$ and the reward R at a time step, the cumulative reward difference between proposed distribution adjustment J_v and the channel-dependent inherent latency distribution J_r is

$$J_v - J_r = \text{Cov}(R, e^{-\lambda\bar{\tau}}), \quad (10)$$

and $\text{Cov}(R, e^{-\lambda\bar{\tau}}) > 0$. Here, $\text{Cov}(\cdot, \cdot)$ is covariance function and λ is a latency-dependent reward discount factor.

It is observed from Proposition 3 that the cumulative reward difference increases with the latency-dependent reward discount factor λ . It implies that the cumulative reward improvement achieved by our distribution adjustment is more obvious in environments with high latency sensitivity.

Network Design

The schematics of VIL2C is illustrated in Fig. 2. At each time step, agent i obtains its local environment observation o_i^{env} and generates message m_i via the Encoder. At the same time, agent i uses the ResoNet, which inputs local environment observation o_i^{env} and wireless channel condition o_i^{ch} , to obtain the optimal bandwidth and power allocation $(B_{i,j}^*, P_{i,j}^*)$.

As a receiver, agent i stores incoming messages in buffer \bar{m}_i and uses an attention-based aggregator to fuse them with its own message m_i , yielding aggregated message m_i^a . Specifically, fully connected layers generate keys k and values v from \bar{m}_i , and query q from m_i . Attention weights α are computed from k and q to capture correlations, then multiplied with v to produce m_i^a via a final fully connected layer. This is concatenated with m_i to form m_i^c . An RNN processes m_i^c and previous hidden state h_i^{t-1} for actions, outputting to a Categorical module for action probability distribution. If the action distribution's entropy falls below the threshold, reception terminates and an action is sampled. Otherwise, the reception continues for more messages.

Overall Training Objective

Building upon the MAPPO framework (Yu et al. 2022), our model is trained by optimizing two separate loss functions in an alternating fashion. Specifically, the loss of the action network in MAPPO is defined as

$$\mathcal{L}_a(\theta) = \sum_{i=0}^N [\min(r^i A^i, \text{clip}(r^i, 1-\epsilon, 1+\epsilon)A^i)], \quad (11)$$

where $r^i = \frac{\pi(a^i|o^i)}{\pi_{\text{old}}(a^i|o^i)}$ and A^i is the advantage function. In each training iteration, after updating the joint policy network, each agent's ResoNet is also updated. The training objective for ResoNet is to maximize the total VoI in Eq. (6). To achieve this, we define the loss function for agent i as the negative of the total VoI: $\mathcal{L}_v^i(\theta) = -\sum_{j=1}^{N_i} \text{VoI}_{i,j}$. The loss of critic network is same as MAPPO.

Experiments

We compare the performance of VIL2C with baselines on three well-known partially-observed multi-agent cooperative tasks: PP, CN(Lowe et al. 2017), and SMACv2 (Ellis

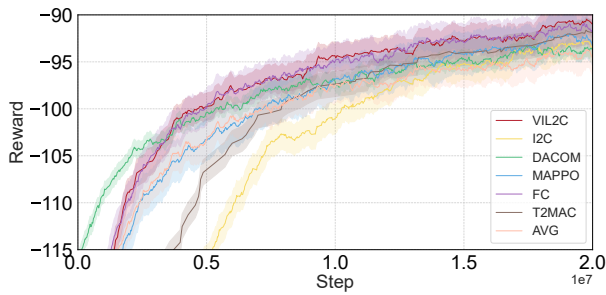


Figure 3: Learning curves for Predator Prey (PP)

et al. 2023). We also demonstrate the performance of VIL2C under various communication conditions. In addition, as a crucial hyperparameter in VIL2C, the effect of maximum waiting time on performance is explored.

The baselines include MAPPO, I2C (Ding, Huang, and Lu 2020), DACOM (Yuan et al. 2023), T2MAC (Sun et al. 2024), latency-free full communication (FC), and equal resource allocation with fixed communication waiting (AVG). MAPPO is the CTDE backbone without inter-agent communication. I2C is the individual communication and each agent has a fixed waiting duration. DACOM dynamically adjusts the waiting duration based on the network status. In T2MAC, agents craft personalized messages, select communication windows and partners. In FC, each agent receives all agents’ messages without latency. In AVG, agents equally allocate resources and the waiting duration is fixed. We set the fixed waiting time as 30% of total time in baselines.

To incorporate the influence of communication latency on action execution, we modify the original environment by dividing the execution time into two parts: the first part is the time for waiting messages and the second part is the time for action execution. In the waiting time, the agent performs the action selected at last time step. In the execution time, the agent selects and executes a new action.

Performance in Different Environments

Predator Prey. In PP experiments, we consider that four slower predators chase two faster preys and two landmarks obstruct their paths. Each predator has an unique observation range, in which it observes landmarks’ positions, the positions and velocities relative to other predators and preys. The reward is the sum of all predators’ negative distances to their closest preys or landmarks. Predators are penalized for colliding with others. In this environment, we compare VIL2C with baselines, as shown in Fig. 3. It is seen that the average reward of VIL2C is close to FC that is latency-free and outperforms other baselines. It implies that VIL2C can effectively alleviate the impact of communication latency on cooperation performance. Moreover, it is observed that the rewards of I2C and AVG are lower than that of MAPPO. It indicates that without a judiciously designed low-latency communication system in MARL, the impact of communication latency may overwhelm benefits of communication.

Cooperative Navigation. In CN experiments, we consider

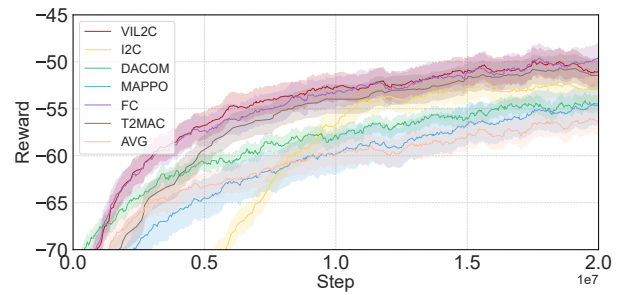


Figure 4: Learning curves for Cooperative Navigation (CN)

that four agents aim to occupy four stationary landmarks individually, where the positions of landmarks and agents are randomly initialized. Each agent has an uniquely limited observation range. Due to the limitation of observations, agents need to cooperatively reschedule their targeted landmarks, when more than one agent choose the same landmark as the target. We compare the training performance of VIL2C with baselines in this environment, as shown in Fig. 4. It is seen that VIL2C approaches FC that is latency-free and outperforms other baselines.

SMACv2. In SMACv2 experiments, we consider that five agents controlled by the learned algorithm aim to cooperatively attack five enemies via some strategies. Three different combat scenarios are employed, i.e., 10gen_terrari, 10gen_protoss, and 10gen_zerg. We compare VIL2C with baselines in the three scenarios, in terms of the win rate, as shown in Fig. 5. Here, the win rate is averaged over 32 test games. It can be seen that FC achieves the highest win rate, due to the reception of messages from all other agents with no communication latency. Moreover, VIL2C outperforms other baselines. This is owing to the communication latency distribution adjustment in VIL2C. Note that, we transform continuous latency into discrete game steps in this environment, where eight game steps equal one MARL step.

Performance Under Different Communication Conditions

We evaluate the performance of VIL2C under different communication conditions, as shown in Fig. 6. Here, we employ the PP task as an example and similar results can be obtained in other tasks. The considered communication conditions in-

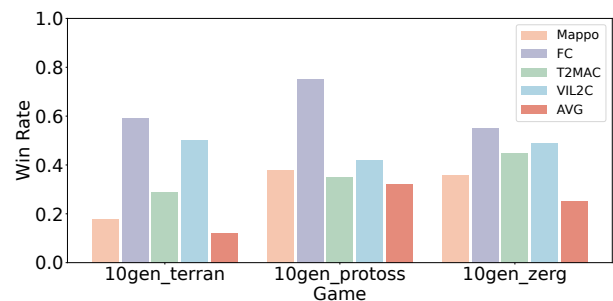


Figure 5: Win Rates for SMACv2

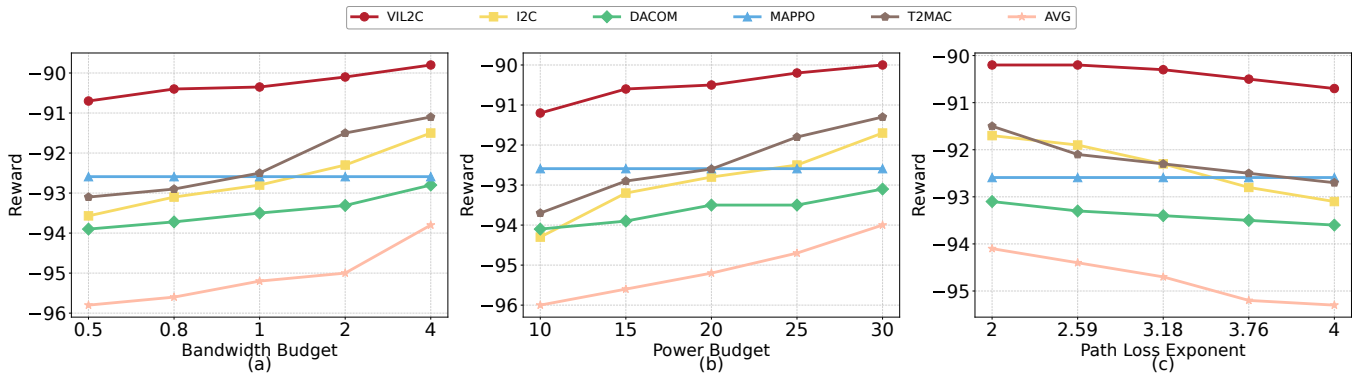


Figure 6: Performance comparisons under different communication conditions.

clude each agent’s bandwidth budget, power budget, and the path loss exponent. It can be seen from Figs. 6(a) and 6(b) that VIL2C achieves a significant improvement over baselines for the whole range of bandwidth/power budget. Moreover, as the bandwidth/power budget decreases, the performance of VIL2C suffers the slighter degradation than baselines, especially I2C and T2MAC. It implies that VIL2C exhibits great robustness against limited resource budgets. This is because that VIL2C can adaptively adjust the bandwidth/power allocation and the waiting duration for different communication latency. It is also observed from Fig. 6(c) that VIL2C outperforms baselines during the range of path loss exponent. As the path loss exponent increases, i.e., the channel condition becomes worse, VIL2C illustrates the small performance degradation.

Effect of Maximum Waiting Time

We further explore the effect of maximum waiting time, as shown in Fig. 7. As a crucial hyperparameter in MARL with communication latency, the maximum waiting time strikes the balance between the benefit of communication and the cost of message waiting. It can be seen that the performance of VIL2C first rises and then drops, as the maximum waiting time increases. This is because that the increase of waiting time allows agents to receive more messages to improve the performance, while the too large waiting time makes the impact of communication latency overwhelm benefits of communication.

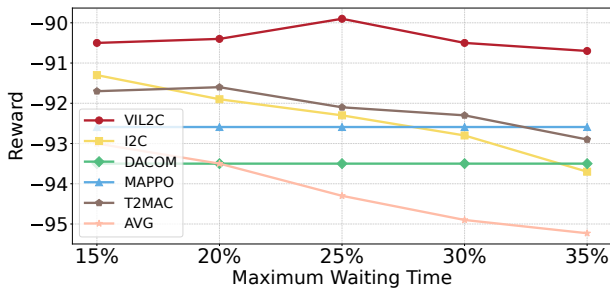


Figure 7: The effect of maximum waiting time.

Maps	VIL2C	VIL2C w/o P	VIL2C w/o R
10gen_terrán	0.55	0.35	0.28
10gen_protoss	0.42	0.32	0.26
10gen_zerg	0.48	0.28	0.41

Table 1: Ablation for ResoNet and Progressive Reception.

Ablation

To quantify the individual contributions of ResoNet (R) and Progressive Reception (P), we conduct the ablation study in three combat scenarios of SMACv2 environment, as reported in Tab.1. It is observed that removing either module of VIL2C reduces the win rate, confirming that both are essential. In the scenarios of 10gen_terrán and 10gen_zerg, the removal of ResoNet demonstrates more severe performance degradation than the removal of progressive reception. Because in these two environments, the agents have smaller observation ranges and the communication gain, especially from important messages, is more obvious. In this case, using resource allocation can reduce the latency of important messages, thereby significantly improving multi-agent cooperation performance.

Conclusion

This work investigates proactive communication latency distribution adjustment to mitigate the latency impact on cooperation performance in MARL. To achieve this, we propose VIL2C. Transmitter agents optimize bandwidth and power to maximize the total VoI, while recipient agents adaptively determine the waiting time based on whether sufficient information is obtained. We demonstrate the advantage of VIL2C through theoretical analysis and experiments. However, our focus is on point-to-point communications, where latency is primarily determined by wireless signal propagation, and the latency caused by network conditions, such as network congestion, is not considered in this work. Moreover, a more generalized VoI definition for tasks with various latency sensitivities will be considered for future work.

Acknowledgments

This work was supported by the National Key R&D Program of China (No. 2024YFB4505502), the National Natural Science Foundation of China (No. 62532009, No. 62302396), the Hong Kong Research Grants Council under the Areas of Excellence Scheme Grant AoE/E-601/22-R and the NSFC/RGC Collaborative Research Scheme Grant CRS_HKUST603/22.

References

- Ahmed, A.; Al-Dweik, A.; Iraqi, Y.; Mukhtar, H.; Naeem, M.; and Hossain, E. 2021. Hybrid automatic repeat request (HARQ) in wireless communications systems and standards: A contemporary survey. *IEEE Communications Surveys & Tutorials*, 23(4): 2711–2752.
- Chen, B.; Xu, M.; Li, L.; and Zhao, D. 2021. Delay-aware model-based reinforcement learning for continuous control. *Neurocomputing*, 450: 119–128.
- Chen, B.; Xu, M.; Liu, Z.; Li, L.; and Zhao, D. 2020. Delay-aware multi-agent reinforcement learning for cooperative and competitive environments. *arXiv preprint arXiv:2005.05441*.
- Cho, Y. S.; Kim, J.; Yang, W. Y.; and Kang, C. G. 2010. *MIMO-OFDM wireless communications with MATLAB*. John Wiley & Sons.
- Das, A.; Gervet, T.; Romoff, J.; Batra, D.; Parikh, D.; Rabbat, M.; and Pineau, J. 2019. TarMAC: Targeted Multi-Agent Communication. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, 1538–1546. PMLR.
- Ding, Z.; Huang, T.; and Lu, Z. 2020. Learning Individually Inferred Communication for Multi-Agent Cooperation. In Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M.; and Lin, H., eds., *Advances in Neural Information Processing Systems*, volume 33, 22069–22079. Curran Associates, Inc.
- Ellis, B.; Cook, J.; Moalla, S.; Samvelyan, M.; Sun, M.; Mahajan, A.; Foerster, J.; and Whiteson, S. 2023. SMACv2: An Improved Benchmark for Cooperative Multi-Agent Reinforcement Learning. In Oh, A.; Naumann, T.; Globerson, A.; Saenko, K.; Hardt, M.; and Levine, S., eds., *Advances in Neural Information Processing Systems*, volume 36, 37567–37593. Curran Associates, Inc.
- Foerster, J.; Assael, I. A.; De Freitas, N.; and Whiteson, S. 2016. Learning to communicate with deep multi-agent reinforcement learning. In *Advances in neural information processing systems*, volume 29.
- Goldsmith, A. 2005. *Wireless communications*. Cambridge university press.
- Gu, S.; Grudzien Kuba, J.; Chen, Y.; Du, Y.; Yang, L.; Knoll, A.; and Yang, Y. 2023. Safe multi-agent reinforcement learning for multi-robot control. *Artificial Intelligence*, 319: 103905.
- Guo, X.; Shi, D.; and Fan, W. 2023. Scalable Communication for Multi-Agent Reinforcement Learning via Transformer-Based Email Mechanism. In Elkind, E., ed., *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23*, 126–134. International Joint Conferences on Artificial Intelligence Organization.
- Hu, G.; Zhu, Y.; Zhao, D.; Zhao, M.; and Hao, J. 2023. Event-Triggered Communication Network With Limited-Bandwidth Constraint for Multi-Agent Reinforcement Learning. *IEEE Transactions on Neural Networks and Learning Systems*, 34(8): 3966–3978.
- Kuhn, H. W.; and Tucker, A. W. 2014. *Nonlinear Programming*, 247–258. Basel: Springer Basel.
- Li, X.; and Zhang, J. 2024. Context-aware Communication for Multi-agent Reinforcement Learning. In *Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems, AAMAS '24*, 1156–1164. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems. ISBN 9798400704864.
- Liu, Y.; Deng, Y.; Nallanathan, A.; and Yuan, J. 2023. Machine Learning for 6G Enhanced Ultra-Reliable and Low-Latency Services. *IEEE Wireless Communications*, 30(2): 48–54.
- Lowe, R.; Wu, Y.; Tamar, A.; Harb, J.; Abbeel, P.; and Mordatch, I. 2017. Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments. *Neural Information Processing Systems (NIPS)*.
- Mao, H.; Zhang, Z.; Xiao, Z.; Gong, Z.; and Ni, Y. 2020. Learning Agent Communication under Limited Bandwidth by Message Pruning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(04): 5142–5149.
- Meng, Z.; Guo, Y.; Sun, C.; Wang, B.; Sherry, J.; Liu, H. H.; and Xu, M. 2022. Achieving consistent low latency for wireless real-time communications with the shortest control loop. In *Proceedings of the ACM SIGCOMM 2022 Conference, SIGCOMM '22*, 193–206. New York, NY, USA: Association for Computing Machinery. ISBN 9781450394208.
- Oroojlooy, A.; and Hajinezhad, D. 2023. A review of cooperative multi-agent deep reinforcement learning. *Applied Intelligence*, 53(11): 13677–13722.
- Palanisamy, P. 2020. Multi-Agent Connected Autonomous Driving using Deep Reinforcement Learning. In *2020 International Joint Conference on Neural Networks (IJCNN)*, 1–7.
- Peng, P.; Wen, Y.; Yang, Y.; Yuan, Q.; Tang, Z.; Long, H.; and Wang, J. 2017. Multiagent bidirectionally-coordinated nets: Emergence of human-level coordination in learning to play starcraft combat games. *arXiv preprint arXiv:1703.10069*.
- She, C.; Yang, C.; and Quek, T. Q. S. 2018. Cross-Layer Optimization for Ultra-Reliable and Low-Latency Radio Access Networks. *IEEE Transactions on Wireless Communications*, 17(1): 127–141.
- Singh, A.; Jain, T.; and Sukhbaatar, S. 2019. Individualized Controlled Continuous Communication Model for Multiagent Cooperative and Competitive Tasks. In *International Conference on Learning Representations*.

- Song, S.; Lin, Y.; Han, S.; Yao, C.; Wu, H.; Wang, S.; and Lv, K. 2025. CoDe: Communication Delay-Tolerant Multi-Agent Collaboration via Dual Alignment of Intent and Timeliness. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(22): 23304–23312.
- Sukhbaatar, S.; Fergus, R.; et al. 2016. Learning multiagent communication with backpropagation. In *Advances in neural information processing systems*, volume 29.
- Sun, C.; She, C.; Yang, C.; Quek, T. Q. S.; Li, Y.; and Vucetic, B. 2019. Optimizing Resource Allocation in the Short Blocklength Regime for Ultra-Reliable and Low-Latency Communications. *IEEE Transactions on Wireless Communications*, 18(1): 402–415.
- Sun, C.; Zang, Z.; Li, J.; Li, J.; Xu, X.; Wang, R.; and Zheng, C. 2024. T2MAC: Targeted and Trusted Multi-Agent Communication through Selective Engagement and Evidence-Driven Integration. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 15154–15163.
- Wang, R.; He, X.; Yu, R.; Qiu, W.; An, B.; and Rabinovich, Z. 2020a. Learning Efficient Multi-agent Communication: An Information Bottleneck Approach. In III, H. D.; and Singh, A., eds., *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, 9908–9918. PMLR.
- Wang, T.; Wang, J.; Zheng, C.; and Zhang, C. 2020b. Learning Nearly Decomposable Value Functions Via Communication Minimization. In *International Conference on Learning Representations*.
- Wang, Y.; fangwei zhong; Xu, J.; and Wang, Y. 2022. ToM2C: Target-oriented Multi-agent Communication and Cooperation with Theory of Mind. In *International Conference on Learning Representations*.
- Wei, H.; Xu, N.; Zhang, H.; Zheng, G.; Zang, X.; Chen, C.; Zhang, W.; Zhu, Y.; Xu, K.; and Li, Z. 2019. Colight: Learning network-level cooperation for traffic signal control. In *Proceedings of the 28th ACM international conference on information and knowledge management*, 1913–1922.
- Yu, C.; Velu, A.; Vinitzky, E.; Gao, J.; Wang, Y.; Bayen, A.; and WU, Y. 2022. The Surprising Effectiveness of PPO in Cooperative Multi-Agent Games. In Koyejo, S.; Mohamed, S.; Agarwal, A.; Belgrave, D.; Cho, K.; and Oh, A., eds., *Advances in Neural Information Processing Systems*, volume 35, 24611–24624. Curran Associates, Inc.
- Yuan, L.; Wang, J.; Zhang, F.; Wang, C.; Zhang, Z.; Yu, Y.; and Zhang, C. 2022. Multi-Agent Incentive Communication via Decentralized Teammate Modeling. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(9): 9466–9474.
- Yuan, T.; Chung, H.-M.; Yuan, J.; and Fu, X. 2023. DA-COM: Learning Delay-Aware Communication for Multi-Agent Reinforcement Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(10): 11763–11771.
- Zhang, S. Q.; Zhang, Q.; and Lin, J. 2019. Efficient Communication in Multi-Agent Reinforcement Learning via Variance Based Control. In Wallach, H.; Larochelle, H.; Beygelzimer, A.; d'Alché-Buc, F.; Fox, E.; and Garnett, R., eds., *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.
- Zhu, C.; Dastani, M.; and Wang, S. 2022. A survey of multi-agent reinforcement learning with communication. *arXiv preprint arXiv:2203.08975*, 1.