Cooperative Multi-Type Multi-Agent Deep Reinforcement Learning for Resource Management in Space-Air-Ground Integrated Networks

Hengxi Zhang
zhanghx20@mails.tsinghua.edu.cn
Tsinghua-Berkeley Shenzhen Institute
Shenzhen, Guangdong, China

Huaze Tang
thz21@mails.tsinghua.edu.cn
Tsinghua-Berkeley Shenzhen Institute
Shenzhen, Guangdong, China

Wenbo Ding*
ding.wenbo@sz.tsinghua.edu.cn
Tsinghua-Berkeley Shenzhen Institute
RISC-V International Open Source Laboratory
Shenzhen, Guangdong, China
Shanghai Artificial Intelligence Laboratory
Shanghai, China

Xiao-Ping Zhang
xiaoping.zhang@ee.ryerson.ca
Tsinghua-Berkeley Shenzhen Institute
RISC-V International Open Source Laboratory
Shenzhen, Guangdong, China
Department of Electrical, Computer and Biomedical
Engineering, Ryerson University
Toronto, Canada

ABSTRACT

The Space-Air-Ground Integrated Network (SAGIN), integrating heterogeneous devices including low earth orbit (LEO) satellites, unmanned aerial vehicles (UAVs), and ground users (GUs), holds significant promise for the advancing smart city applications. However, resource management of the SAGIN is a challenge requiring urgent study in that inappropriate resource management will cause poor data transmission, and hence affect the services in smart cities. In this paper, we develop a comprehensive SAGIN system that encompasses five distinct communication links and propose an efficient cooperative multi-type multi-agent deep reinforcement learning (CMT-MARL) method to address the resource management issue. The experimental results highlight the efficacy of proposed CMT-MARL, as evidenced by key performance indicators such as the overall transmission rate and transmission success rate. These results underscore the potential value and feasibility of future implementation of the SAGIN.

CCS CONCEPTS

• **Computing methodologies** → **Multi-agent reinforcement learning**; • **Networks** → **Wireless access networks**.

KEYWORDS

SAGIN, resource management, multi-agent reinforcement learning

---

*Corresponding author.

1 INTRODUCTION

The advent of 5G technology has ushered in a new era where the Internet of Things (IoT) serves as the backbone of numerous applications and services, such as intelligent transportation systems, home automation, and smart factories [4, 7, 20]. However, the burgeoning computational demands of IoT devices stretch the limits of existing wireless communication networks. Current terrestrial communication paradigms are ill-equipped to cope with this demand. Enter the Space-Air-Ground Integrated Network (SAGIN), a proposed heterogeneous system that amalgamates satellites, unmanned aerial vehicle (UAV)-based air systems, and terrestrial communication systems like base stations [17]. Its objective is to provide versatile network coverage and services.

SAGIN has emerged as a potential solution to meet the mounting computational requirements of IoT services and applications [6, 14, 19]. By integrating terrestrial systems with satellites and unmanned aerial vehicle (UAV)-based air systems, SAGIN offers a flexible and scalable approach that could potentially rise to these challenges [24]. However, the implementation of SAGIN within IoT services is not without its hurdles [11]. The primary obstacle lies in the inherent high mobility of the air system. This mobility results in dynamic and often unpredictable channel conditions and coverage areas. Moreover, the ground connections in current approaches are frequently overlooked. Exceptions are made only for connections between ground users and UAVs, an approach that can lead to inefficiencies and limit the full utilization of the integrated

network [15]. An additional complication arises from the heterogeneity of the SAGIN subsystems. Each subsystem — terrestrial, aerial, and space — uses a unique communication interface, reflecting its particular technological requirements and constraints [17]. Furthermore, channels between these various subsystems possess distinct properties, further complicating communication and integration. Therefore, it is imperative to develop a discerning resource management policy for SAGIN.

The rapid evolution of machine learning techniques offers potential solutions to these challenges [27]. Among these, reinforcement learning (RL) stands out as a reward-centric approach designed to tackle combinatorial optimization problems. An RL agent interacts with and learns from its environment to optimize toward an end-goal, guided by a predefined reward system [23]. In some contexts, there may be multiple agents, or centralized learning methods may prove ineffective due to inherent environmental characteristics [25]. In these situations, the notion of multi-agent RL (MARL) emerges as a decentralized learning method. It has been increasingly adopted by researchers and institutions over recent years [13]. Additionally, there may exist more than one type of agent in the multi-agent system, where different types of agents need to utilize diverse behaviors to coordinate with others [22]. In this work, we develop an ambiguous stochastic optimization (ASO) SAGIN communication model aimed at handling the system's overall resource management problem. We then employ a tailor-made MARL technique, dubbed Cooperative Multi-Type MARL (CMT-MARL), to elaborately study how the multi-agent system will perform in this dynamic SAGIN.

2 HETEROGENEOUS SYSTEM OF WIRELESS COMMUNICATION

In this SAGIN system presented in Fig. 1, we consider four communication categories at three different altitudes, with a bunch of GUs (i.e. vehicles) and a BS on the ground layer, the UAV swarm hovering in the air layer and an LEO satellite in the space layer, and five classes of communication links among these categories. The links started from GUs to BS, UAVs, and satellites are modeled as GU-to-BS (G2B) link, GU-to-UAV (G2U) link, and GU-to-Satellite (G2S) link, respectively. While each UAV can also establish two kinds of links from itself to BS and satellite, expressed similarly as the UAV-to-BS (U2B) link and UAV-to-Satellite (U2S) link.

Both GU and UAV groups continuously generate a series of data packets that need to be transmitted to the remote servers eventually. The BS and satellite are considered as the terminal transmission devices for each GU and UAV to transmit the data packets in the SAGIN, where the UAV in the air layer can be utilized as an intermediary for each GU in the ground to transmit the packet when the G2S or G2B link is temporally crowded or considering the transmit speed of them are relatively low. Specifically, GUs and UAV swarm involve $N_1(= |\mathcal{N}_1|)$ and $N_2(= |\mathcal{N}_2|)$ homogeneous individuals respectively, and can be further treated as heterogeneous system $\mathcal{N}(= \mathcal{N}_1 \cup \mathcal{N}_2)$, where $\mathcal{N}_1 = \{GU_1, GU_2, \ldots, GU_{N_1}\}$ and $\mathcal{N}_2 = \{UAV_1, UAV_2, \ldots, UAV_{N_2}\}$. Each GU can select only one link from G2B, G2U, and G2S links to transmit data, while U2B or U2S are two links for each UAV to choose from. Note that the G2U link can only be generated if the UAV selected by the GU agrees to make this connection.
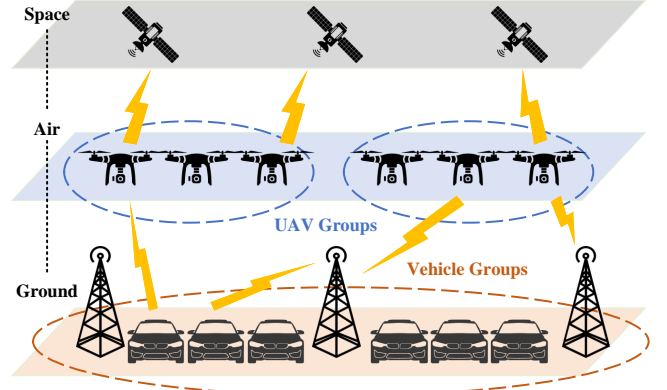

Figure 1: A schematic illustration of SAGIN.

2.1 Wireless Communication Links

The unicast protocol is employed in this work, where each GU or UAV can only select one objective to transmit its message, and the transmission between the same category, such as GU-to-GU and UAV-to-UAV, is prohibited to prevent message congestion, following classic configuration of SAGIN [26]. Since there exist three layers in SAGIN, we allocate two bandwidths W_1 and W_2 that support the low-altitude and high-altitude transmission, respectively, in which low-altitude bandwidth W_1 supports G2B, G2U, and U2B links while G2S and U2S links are on the high-altitude bandwidth W_2.

2.1.1 Low-Altitude Links. The low-altitude links, also known as the terrestrial links, are operated terrestrially, including G2B links, G2U links, and U2B links [3, 9]. The low-altitude links feature relatively lower bandwidth and frequency than high-altitude links and provide communication with lower latency and enhanced reliability due to low pathloss [26]. Taking into account the interfering channels within the sub-band, as discussed in [16], the signal-to-interference-plus-noise ratio (SINR) for low-altitude links operating on the h-th sub-band (where $h \in [H]$) can be formulated as follows,

$$\gamma_i^{LOW}[h] = \frac{p_i^{LOW}[h]g_{i,d_i}[h]}{\sigma^2 + \phi_i^{LOW}[h]}, \tag{1}$$

where $p_i^{LOW}[h]$ and $g_{i,d}[h]$ refer to the transmit power and the interfering channel from i-th ($i \in \mathcal{N}$) device (GU or UAV) to its designated communication endpoint denoted as d_i (base station or another device) over h-th sub-band. $\phi_i^{LOW}[h]$ denotes the interference of low-altitude links and can be expressed as, $\phi_i^{LOW}[h] = \sum_{i' \in \mathcal{N}_i} x_{(i',d_{i'})}[h]p_{i'}^{LOW}[h]g_{i',d_{i'}}[h]$, where $p_{i'}^{LOW}[h]$ and $g_{i',d_{i'}}[h]$ similarly represent the transmit power and the interfering channel from i'-th device to the communication destination $d_{i'}$. $x_{(i',d_{i'})}(\cdot)$ is the indicator function equal to 1 if the h-th sub-band is occupied and 0 otherwise.

2.1.2 High-Altitude Links. The high-altitude links, also known as the non-terrestrial links, encompass the connections between terrestrial devices [8]. Specifically, the high-altitude links consist of G2U links and U2S links. These links enable communication and data exchange between the terrestrial devices (GUs and UAVs) and the satellite. Since the satellite can be utilized to support long-range transmission due to its wide horizon, the high-altitude link in the

proposed SAGIN model can be treated as a relatively smaller BS with a relatively wide bandwidth for receiving a series of messages from both GUs and UAVs [12]. Therefore, we express the SINR of high-altitude link over h-th ($h \in [H]$) sub-band as,

$$\gamma_i^{HIGH}[h] = \frac{p_i^{HIGH}[h]g_{i,S}[h]}{\sigma^2 + \phi_i^{HIGH}[h]}, \tag{2}$$

where $p_i^{HIGH}[h]$ and $g_{i,S}[h]$, similar to the high-altitude link, are the transmit power and the interfering channel from i-th terrestrial device to the satellite over h-th sub-band. Considering there are only G2S and U2S links on the high-altitude bandwidth W_2, we use $\phi_i^{HIGH}[h] = \sum_{i' \in \mathcal{N}^i} x_{i',S}[h]p_{i'}^{HIGH}[h]g_{i',S}[h]$ to indicate the interference power of high-altitude link from i-th device to the satellite over h-th sub-band.

2.1.3 Transmit Rate and Latency. Since transmit rate and latency are critical performance metrics that can directly affect the overall quality and reliability of heterogeneous network communication, we mainly focus on optimizing them in this work. Based on all types of SINRs above, the transmit rates of different links over the same h-th sub-band can be hence established as [21],

$$R^a[h] = W_b \log(1 + \gamma^a[h]), \tag{3}$$

where $a \in \{LOW, HIGH\}$ represents different types of links and $b \in \{1, 2\}$ is the bandwidth occupied for transmission, where W_1 and W_2 denote the bandwidth for low- and high-altitude links.

3 OPTIMIZATION OBJECTIVE

We introduce the ambiguous stochastic optimization (ASO) model in this section. ASO refers to the optimization of a system or process under conditions of uncertainty, where the probabilities of different outcomes or events are not precisely known [18].

Given an object-power-channel decision profile $(\mathbf{x}, \mathbf{p}, \mathbf{h})$ of this heterogeneous communication system, the ambiguous transmission rate $R : \mathcal{X} \times \mathcal{P} \times \mathcal{H} \times \Xi \to \mathbb{R}$, and a set of plausible distributions Ω over the uncertain parameters, the goal is to find the optimal decision profile $(\mathbf{x}^*, \mathbf{p}^*, \mathbf{h}^*)$ that maximizes the worst-case expected transmission rate of R over the set of plausible distributions,

$$\sup_{\mathbf{x},\mathbf{p},\mathbf{h}} \inf_{M \in \Omega} \mathbb{E}_M \left[R^a(\mathbf{x}, \mathbf{p}, \mathbf{h}; \xi) \right]$$

$$\text{s.t. } (a) \sum_j x_{i,j} + x_{i,B} + x_{i,S} = 1, \forall \, GU_i, i \in \mathcal{N}_1,$$

$$(b) \, x_{j,B} + x_{j,S} = 1, \forall \, UAV_j, j \in \mathcal{N}_2, \tag{4}$$

$$(c) \, x_{i,j}, x_{i,B}, x_{i,S}, x_{j,B}, x_{j,S} \in \{0, 1\},$$

$$(d) \, p_{i,j}, p_{i,B}, p_{i,S}, p_{j,B}, p_{j,S} \geq 0,$$

where ξ some random variables representing the uncertain parameters, and $\mathbb{E}_M$ denotes the expected value with respect to the distribution M such that $i \in \mathcal{N}_1$ and $j \in \mathcal{N}_2$. This formulation seeks to find a transmit decision that minimizes the maximum expected value of the objective function over all plausible distributions, accounting for the ambiguity and uncertainty in this SAGIN system.

4 MARKOV DECISION PROCESS OF SAGIN

Considering the aforementioned SAGIN optimization objectives are non-convex functions and difficult to figure out with conventional
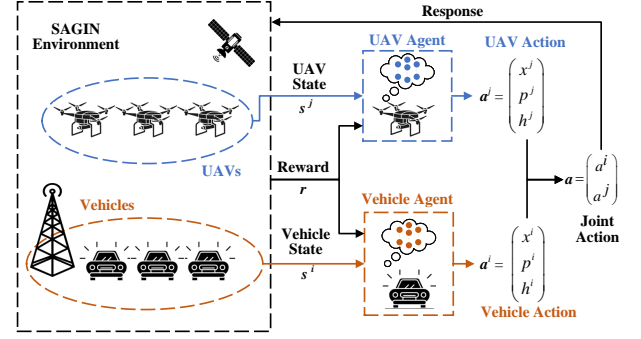

Figure 2: The workflow of Markov decision process in SAGIN.

programming methods, we model it as a Markov decision process (MDP), given as $\langle \mathcal{S}, \mathcal{A}, P, r, \beta \rangle$ for embedding the framework of MARL method, where $\mathcal{S} \triangleq \{\mathcal{S}_1, \mathcal{S}_2\}$ and $\mathcal{A} \triangleq \{\mathcal{A}_1, \mathcal{A}_2\}$ are the joint observation and action spaces of GU and UAV agent, respectively. P indicates the transition probability, r refers to the reward given by the SAGIN environment and β represents the factor that determines the importance of future rewards.

4.1 Observation Space

In this SAGIN, we model GUs and UAVs as two types of agents that cooperatively explore the uncertain and dynamic communication environment. As discussed in Section 2.1, SINR, on the one hand, is a measure of the quality of the received signal relative to the interference and noise. Agents can use the SINR to estimate the power of the received signal relative to the interference and noise, which can be used to make decisions about scheduling data transmissions. On the other hand, interference power is a measure of the power of the interfering signal caused by other data transmissions. This metric can be utilized by agents to estimate the impact of other data transmissions on the quality of their own data transmissions. In addition, the data packet size is the size of the data packet that needs to be transmitted. Agents can leverage this information to estimate the time required to transmit the data packet and make decisions about scheduling data transmissions. Based on the above information, the observation space of agents can be defined as a vector of features that includes SINR γ, interference power ϕ, and data packet size B. We hence establish the observation of GU agent i as, $s^i = Y_{GU}^i(\gamma, \phi, B) \in \mathcal{S}_1, \forall \, GU_i, i \in \mathcal{N}_1$, nd similarly, the observation of UAV agent j can be expressed as, $s^j = Y_{UAV}^j(\gamma, \phi, B) \in \mathcal{S}_2, \forall \, UAV_j, j \in \mathcal{N}_2$, where $Y_{GU}(\cdot)$ and $Y_{UAV}(\cdot)$ represent the observation functions of GU and UAV agents, respectively. In addition, the SINR $\gamma = \{\gamma^{LOW}, \gamma^{HIGH}\}$ and the interference power $\phi = \{\phi^{LOW}, \phi^{HIGH}\}$, both of which are modeled in Section 2.1.

4.2 Action Space

The action space represents the set of feasible actions that can be taken by the GU or UAV agent in the SAGIN environment. The goal of each individual GU or UAV is to develop a collaborative policy that optimizes their collective long-term reward by selecting actions from their own designated action space. For each GU agent,

it needs to select a receiver to transmit the data package at first, i.e. the communication link $x^i \in \mathcal{X}_1$, and then the appropriate communication channel $h^i \in \mathcal{H}_1$, and transmit power $p^i \in \mathcal{P}_1$ are supposed to be chosen for acceptable transmission efficiency. Therefore, each GU agent makes an object-power-channel decision profile, $a^i = (x^i, p^i, h^i) \in (\mathcal{X}_1, \mathcal{P}_1, \mathcal{H}_1) \subseteq \mathcal{A}_1, \forall GU_i$, Similarly, the decision profile that each UAV agent at each time step is expressed as, $a^j = (x^j, p^j, h^j) \in (\mathcal{X}_2, \mathcal{P}_2, \mathcal{H}_2) \subseteq \mathcal{A}_2, \forall UAV_j$.

4.3 Scenario Reward

Since the objective is to maximize the transmission rate of the SA-GIN, we can naturally utilize the rate to design the reward that encourages agents to take actions that maximize the overall rate while still ensuring the stability and robustness of the network. Then, we stimulate coordination between agents. The reward should encourage agents to coordinate their actions to maximize the rate. Third, the reward encourages exploration of the environment to discover new strategies that improve the rate. For example, agents can be rewarded for taking actions that have not been tried before, or for taking actions that improve the rate in a novel way. Notably, the robustness and performance need to be balanced.

Based on these considerations, we specifically design the reward function for a cooperating MARL architecture in the aforementioned SAGIN as,

$$r = \alpha_1 \cdot r_L + \alpha_2 \cdot r_{coo} + \alpha_3 \cdot r_{exp}, \qquad (5)$$

where r is the common reward that all agents share, known as a cooperative reward. $\{\alpha_n\}_{n=1}^4$ are constant hyper-parameters that balance different reward components. r_L stands for the latency of the network. To incentivize agents to minimize the latency of the SAGIN, we choose to use the negative of the latency as the reward. The higher the negative value of the latency, the greater the reward,

$$r_L = -(\kappa_1 \sum_{i \in \mathcal{N}_1} L_i + \kappa_2 \sum_{j \in \mathcal{N}_2} L_j), \qquad (6)$$

which ensures that agents will prioritize reducing the latency of the network. κ_1 and κ_1 are two hyper-parameters that balance the latencies of different communication devices. And r_{coo} is a reward for actions that improve coordination between agents. For encouraging collaboration and coordination between Gus and UAVs to minimize latency, we give a reward to agents who take actions contributing to a coordinated strategy that reduces latency,

$$r_{coo} = \sum_{i \in \mathcal{N}_1} w_i \delta_i + \sum_{j \in \mathcal{N}_2} w_j \delta_j, \qquad (7)$$

where w_i and w_j refer to the weights assigned to GU i and UAV j indicating the importance of each agent's role in the network. δ_i and δ_j are binary indicators of whether their actions contributed to a coordinated strategy that reduced latency. To measure the reduction in latency resulting from a coordinated strategy, we utilize a metric such as the difference between the latency before and after the strategy is implemented. If the coordinated strategy results in a lower latency compared to the latency without the strategy, then we can consider that the action of GU i or UAV j contributes to a coordinated strategy that reduces the average latency of the SAGIN

Table 1: Simulation settings [1, 2, 5, 10]

SAGIN parameter	Value
Number of vehicle	$\{2, 5, 10, 20, 30\}$
Number of UAV	$\{1, 2, 5, 10, 15\}$
Number of BS	1
Number of satellite	1
Environment area	$200 \times 100 \ \mathrm{m}^2$
Height of vehicle	1.5 m
Height of UAV	25 m
Height of BS	100 m
Height of satellite	600 km
Angle of satellite	30°
Carrier frequency of terrestrial links	2 GHz
Bandwidth of terrestrial links	1 MHz
Carrier frequency of non-terrestrial links	30 GHz
Bandwidth of non-terrestrial links	100 MHz
GU & UAV transmit power	$\{23, 10, 5, -100\}$ dBm
BS antenna gain	8 dBi
UAV & GU antenna gain	3 dBi
GU & UAV receiver noise figure	9 dB
BS receiver noise figure	5 dB
Noise power	-114 dBm
Time slot for package delivery	100 ms

system $\bar{L}$, designed as,

$$\delta = \begin{cases} 1, & \text{if } \bar{L}_t - \bar{L}_{t+1} > 0, \\ 0, & \text{otherwise.} \end{cases} \qquad (8)$$

Further, r_{exp} indicates a reward for novel actions that improve latency. To motivate agents to explore the environment and try new strategies, we can use an exploration bonus that rewards agents for taking actions that have not been tried before or that improve latency in a novel way. One way to do this is to give a reward to agents that take actions that increase the diversity of strategies used to reduce latency,

$$r_{exp} = \zeta_1 \sum_{i \in \mathcal{N}_1} \mathcal{H}^i(\pi^i) + \zeta_2 \sum_{j \in \mathcal{N}_2} \mathcal{H}^j(\pi^j), \qquad (9)$$

where ζ_1 and ζ_2 are two weighting factors that control the influence of the exploration bonus, and $\mathcal{H}(\pi)$ is the entropy of the probability distribution π over the set of strategies used to reduce latency. We in this article utilize the Shannon entropy as a measure of the uncertainty or randomness of a probability distribution, given as $\mathcal{H}(\pi) = -\sum_{a \in \mathcal{A}} \pi(a) log_2 \pi(a)$, in which $\pi(a)$ is the probability of selecting action a over its action space $\mathcal{A}$. This encourages agents to explore different strategies and avoid relying on a single strategy that may not be effective in all situations.

5 EXPERIMENTS AND RESULTS

We have novelly designed a SAGIN scenario to implement the proposed CMT-MARL in this part. Elaborately, the G2B channel is modeled using the WINNER II channel model [10], the U2B channel is modeled according to the definition provided in 3GPP TR 36.777 Rel. 15 [1], the G2U channel follows the definition in [5], and the non-terrestrial links (G2S and U2S) are modeled based on
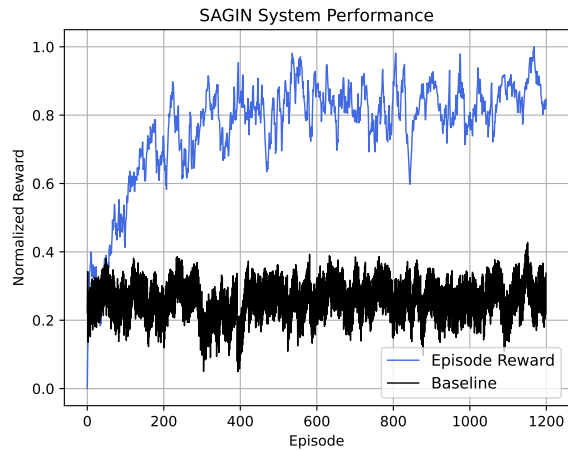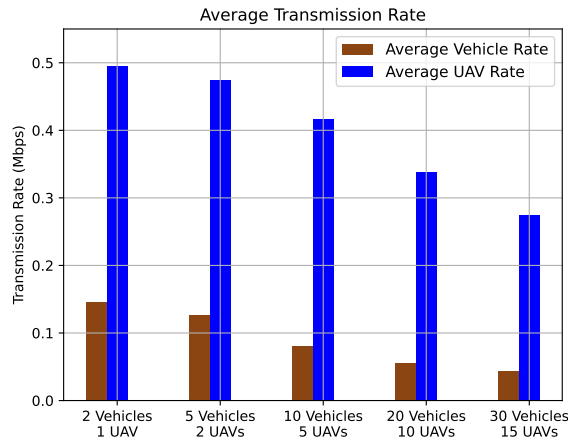
Figure 3: SAGIN system performance.

Figure 4: Average transmission rates with different numbers of vehicles and UAVs.

the definition in 3GPP TR 38.811 Rel. 15 [2]. The specific simulation experiment settings can be found in Table 1.

To evaluate the performance of the proposed CMT-MARL method in the training phase, we first employ 2 vehicles and 1 UAV and exhibit the accumulated reward of the SAGIN system, in which the reward is normalized within the range of [0, 1] to effectively showcase the capabilities of the multi-type multi-agent system. Each episode comprises 100 steps, corresponding to a duration of 100 ms, during which the vehicle and UAV agents transmit packages. The convergence of system performance can be observed in Figure 3, where it becomes apparent that the performance stabilizes after approximately 500 episodes. The baseline represents that the vehicle and UAV agents randomly select actions at any time.

During the test phase, we assess the effectiveness of the CMT-MARL approach by varying the quantities of vehicle-UAV combinations, while maintaining a constant number of one base station (BS) and one satellite throughout the evaluations, as listed in Table 1. The test area spans 200×100 m^2. Figure 4 illustrates the average transmission rates achieved by the vehicle and UAV groups. The results show that the vehicle group attains an average transmission rate of approximately 0.15 Mbps, while the UAV group achieves around 0.50 Mbps. Notably, even when the number of agents increases to
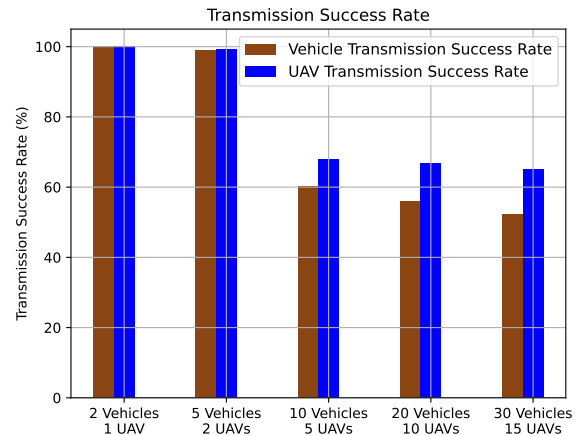

Figure 5: Transmission success rates with different numbers of vehicles and UAVs.

30 vehicles and 15 UAVs, indicating a crowded communication environment, the average transmission rates remain relatively stable at approximately 0.05 Mbps for the vehicle group and 0.28 Mbps for the UAV group.

Furthermore, in Figure 5, we demonstrate the transmission efficacy as reflected by the success rates across varying quantities of vehicle-UAV combinations. As the agent amount rises, the escalating interference within the SAGIN system poses hurdles to the transmissions. Consequently, agents must endeavor to discern and implement more efficient transmission policies to overcome the intensified interference and successfully accomplish their package transmission tasks. Hereto, The CMT-MARL technique allows the SAGIN system to achieve a nearly 100% success rate with a relatively smaller number of agents, i.e., {2, 1} and {5, 2} vehicle-UAV combinations. As the agent number expands from {10, 5} to {30, 15}, the success rates for both vehicle and UAV agents remain consistently high, hovering around 55% to 60% for the former and approximately 65% for the latter. This attests to the robustness and effectiveness of the CMT-MARL technique in facilitating reliable and successful transmissions within the SAGIN environment.

6 CONCLUSION

In this study, to elaborately investigate the resource management problem within the SAGIN system, we have innovatively constructed a SAGIN system incorporating five distinct communication links and proposed a specific MARL technique called CMT-MARL to solve it. This method is tested using different amounts of vehicle-UAV agent combinations and showcases its reliability and robustness in the SAGIN, providing the hope of implementing it over the physical SAGIN environment. In future works, distributed decision-making, as well as the scalability and efficiency issues, will further be investigated using mean field and MARL methods in consideration that each vehicle or UAV agent can make decisions based on local observations and interactions with their immediate environment, without requiring centralized control. This distributed approach allows the network to scale effectively as the number of agents increases.

ACKNOWLEDGMENTS

REFERENCES

[1] 2017. *3rd Generation Partnership Project; Technical Specification Group Radio Access Network;Study on Enhanced LTE Support for Aerial Vehicles: (Release 15)* (Standard 3GPP TR, Dec 2017).

[2] 2020. *3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Study on New Radio (NR) to support non-terrestrial networks : (Release 15)* (Standard 3GPP TR, Sep 2020).

[3] Mamta Agiwal, Abhishek Roy, and Navrati Saxena. 2016. Next generation 5G wireless networks: A comprehensive survey. *IEEE Communications Surveys & Tutorials* 18, 3 (2016), 1617–1655.

[4] Ala Al-Fuqaha, Mohsen Guizani, Mehdi Mohammadi, Mohammed Aledhari, and Moussa Ayyash. 2015. Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications. *IEEE Communications Surveys & Tutorials* 17, 4 (2015), 2347–2376.

[5] Akram Al-Hourani, Sithamparanathan Kandeepan, and Simon Lardner. 2014. Optimal LAP Altitude for Maximum Coverage. *IEEE Wireless Communications Letters* 3, 6 (2014), 569–572.

[6] Marcelo Albuquerque, Arun Ayyagari, Michael A Dorsett, and Michael S Foster. 2007. Global information grid (GIG) edge network interface architecture. In *MILCOM 2007-IEEE Military Communications Conference.* IEEE, 1–7.

[7] Jeffrey G Andrews, Stefano Buzzi, Wan Choi, Stephen V Hanly, Angel Lozano, Anthony CK Soong, and Jianzhong Charlie Zhang. 2014. What will 5G be? *IEEE Journal on selected areas in communications* 32, 6 (2014), 1065–1082.

[8] Maurizio Casoni, Carlo Augusto Grazia, Martin Klapez, Natale Patriciello, Angelos Amditis, and Evangelos Sdongos. 2015. Integration of satellite and LTE for disaster recovery. *IEEE Communications Magazine* 53, 3 (2015), 47–53.

[9] Sathyanarayanan Chandrasekharan, Karina Gomez, Akram Al-Hourani, Sithamparanathan Kandeepan, Tinku Rasheed, Leonardo Goratti, Laurent Reynaud, David Grace, Isabelle Bucaille, Thomas Wirth, et al. 2016. Designing and implementing future aerial communication networks. *IEEE Communications Magazine* 54, 5 (2016), 26–34.

[10] Martin Döttling, Werner Mohr, and Afif Osseiran. 2010. *WINNER II Channel Models.* 39–92.

[11] Manal El Tanab and Walaa Hamouda. 2016. Resource allocation for underlay cognitive radio networks: A survey. *IEEE Communications Surveys & Tutorials* 19, 2 (2016), 1249–1276.

[12] John Farserotu and Ramjee Prasad. 2000. A survey of future broadband multimedia satellite systems, issues and trends. *IEEE Communications Magazine* 38, 6 (2000), 128–133.

[13] Jakob Foerster, Ioannis Alexandros Assael, Nando De Freitas, and Shimon Whiteson. 2016. Learning to communicate with deep multi-agent reinforcement learning. *Advances in neural information processing systems* 29 (2016).

[14] Victor P Hubenko, Richard A Raines, Robert F Mills, Rusty O Baldwin, Barry E Mullins, and Michael R Grimaila. 2006. Improving the global information grid's performance through satellite communications layer enhancements. *IEEE Communications Magazine* 44, 11 (2006), 66–72.

[15] Nei Kato, Zubair Md Fadlullah, Fengxiao Tang, Bomin Mao, Shigenori Tani, Atsushi Okamura, and Jiajia Liu. 2019. Optimizing space-air-ground integrated networks by artificial intelligence. *IEEE Wireless Communications* 26, 4 (2019), 140–147.

[16] Le Liang, Hao Ye, and Geoffrey Ye Li. 2019. Spectrum sharing in vehicular networks based on multi-agent reinforcement learning. *IEEE Journal on Selected Areas in Communications* 37, 10 (2019), 2282–2292.

[17] Jiajia Liu, Yongpeng Shi, Zubair Md Fadlullah, and Nei Kato. 2018. Space-air-ground integrated network: A survey. *IEEE Communications Surveys & Tutorials* 20, 4 (2018), 2714–2741.

[18] Krzysztof Postek, Aharon Ben-Tal, Dick Den Hertog, and Bertrand Melenberg. 2018. Robust optimization with ambiguous stochastic constraints under mean and dispersion information. *Operations Research* 66, 3 (2018), 814–833.

[19] Tereza Pultarova. 2015. Telecommunications-space tycoons go head to head over mega satellite network [news briefing]. *Engineering & Technology* 10, 2 (2015), 20–20.

[20] Bodong Shang, Yang Yi, and Lingjia Liu. 2021. Computing over space-air-ground integrated networks: Challenges and opportunities. *IEEE Network* 35, 4 (2021), 302–309.

[21] Claude E Shannon. 1948. A mathematical theory of communication. *The Bell system technical journal* 27, 3 (1948), 379–423.

[22] Sriram Ganapathi Subramanian, Pascal Poupart, Matthew E Taylor, and Nidhi Hegde. 2020. Multi type mean field reinforcement learning. *arXiv preprint arXiv:2002.02513* (2020).

[23] Richard S Sutton and Andrew G Barto. 2018. *Reinforcement learning: An introduction.* MIT press.

[24] Gang Xiong, Fenghua Zhu, Xisong Dong, Haisheng Fan, Bin Hu, Qingjie Kong, Wenwen Kang, and Teng Teng. 2016. A kind of novel ITS based on space-air-ground big-data. *IEEE intelligent transportation systems magazine* 8, 1 (2016), 10–22.

[25] Kaiqing Zhang, Zhuoran Yang, and Tamer Başar. 2021. Multi-agent reinforcement learning: A selective overview of theories and algorithms. *Handbook of reinforcement learning and control* (2021), 321–384.

[26] Ning Zhang, Shan Zhang, Peng Yang, Omar Alhussein, Weihua Zhuang, and Xuemin Sherman Shen. 2017. Software defined space-air-ground integrated vehicular networks: Challenges and solutions. *IEEE Communications Magazine* 55, 7 (2017), 101–109.

[27] Conghao Zhou, Wen Wu, Hongli He, Peng Yang, Feng Lyu, Nan Cheng, and Xuemin Shen. 2020. Deep reinforcement learning for delay-oriented IoT task scheduling in SAGIN. *IEEE Transactions on Wireless Communications* 20, 2 (2020), 911–925.