

ARTICLE

Trustworthy Semantic Knowledge Orchestration for Space-Air-Ground Integrated Edge Computing Using Hierarchical Federated Multi-Agent Deep Reinforcement Learning

Abdullah Alharbi*

Department of Computer Science, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia

*Corresponding Author: Abdullah Alharbi. Email: arharbi@ksu.edu.sa

Received: 14 July 2026; Accepted: 14 August 2026; Published: 22 October 2026

ABSTRACT: Space-Air-Ground Integrated Networks (SAGIN) are networks that combine the use of satellite, High-Altitude Platform Systems (HAPS), Unmanned Aerial Vehicles (UAVs), terrestrial base stations, and Internet of Things (IoT) devices to enable ubiquitous networking and distributed edge intelligence. Efficient semantic edge computing is, however, challenging due to resource heterogeneity, as well as the dynamism of the topology and the evolution of semantic information. This study proposes a novel Hierarchical Graph-Constrained Federated Multi-Agent Deep Reinforcement Learning (HGCF-MARL) framework for trustworthy semantic knowledge orchestration in SAGIN. The framework unifies the placement of semantic knowledge bases, task offloading, communication-computation resource allocation, UAV trajectory planning, and coordination between layers. A Semantic Knowledge Freshness-Mismatch Index (SKFMI) is introduced to quantify temporal staleness, distribution drift, and task mismatch, and trust-aware, Byzantine-robust aggregation and event-triggered policy synchronization mechanisms are introduced to enhance reliability and communication efficiency. HGCF-MARL demonstrated a semantic quality of 0.936 and a knowledge-hit ratio of 94.2% across experiments on EuroSAT, CIFAR-10, AG News, and synthetic SAGIN traces. It shows the average reductions of 13.8% and 56.0% in total energy consumption and federated signaling volume, respectively, compared with the best corresponding baselines, and achieves up to 99.2% model-replacement attack detection rate while maintaining up to 94.1% task-completion ratio in the 10% malicious-participant setting. These results show the effectiveness, robustness, and scalability of HGCF-MARL for dynamic SAGIN edge intelligence.

KEYWORDS: Space-Air-Ground Integrated Networks; semantic edge computing; semantic knowledge orchestration; multi-agent deep reinforcement learning; federated learning; Byzantine robustness

1 Introduction

With the rapid proliferation of intelligent Internet of Things (IoT) applications, emerging 6G networks require ubiquitous connectivity, real-time computation, and autonomous resource management capabilities to support highly dynamic and computation-intensive services [1]. Applications such as remote sensing, autonomous transportation, and disaster monitoring produce delay-sensitive, computation-intensive workloads that cannot be handled by resource-limited devices [2]. To address these challenges, a Space-Air-Ground Integrated Network (SAGIN) is proposed to unify satellites, High-Altitude Platform Systems (HAPS), Unmanned Aerial Vehicles (UAVs), terrestrial networks, and IoT devices. However, due to the highly dynamic topology, heterogeneous computing resources, and complex interactions among network

layers, SAGIN edge computing involves joint decisions on task offloading, resource allocation, mobility management, and service placement, resulting in a high-dimensional and non-convex optimization problem [3]. Most existing work addresses bit-level communication and traditional resource metrics, and solutions can be adapted using Deep Reinforcement Learning (DRL).

The key to semantic communication is to transmit task-related information rather than raw data to reduce transmission overhead; however, this relies on the availability and suitability of semantic knowledge bases [4,5]. In dynamic SAGIN environments, mobility, task variation, and distributional drift may cause these knowledge bases to become outdated or mismatched. However, existing approaches do not jointly optimize semantic knowledge placement, migration, and selection with system-level decisions, limiting their adaptability in highly dynamic SAGIN scenarios [6]. Furthermore, existing Federated Learning (FL) approaches make strong assumptions about the nodes and their regular updates, conditions that are highly impractical in SAGIN, as nodes can be unreliable or malicious and communication bandwidth is limited [7]. Single-timescale optimization is also not very adaptable, and different decisions evolve at various timescales.

Existing studies on semantic communication, graph learning, FL, and Multi-Agent Reinforcement Learning (MARL) in SAGIN generally optimize communication, computation, or learning separately. In this work, we present a new approach called Hierarchical Graph-Constrained Federated Multi-Agent Deep Reinforcement Learning (HGCF-MARL), combining the dynamic graph representation, hierarchical decision-making, trust-aware Byzantine-robust aggregation, and event-triggered synchronization. In addition, the Semantic Knowledge Freshness-Mismatch Index (SKFMI) is formulated as a measure of temporal aging, distribution drift, and task mismatch; the proposed hierarchical policy optimizes jointly semantic knowledge management, task offloading, UAV trajectories, inter-layer association, and communication-computation resources in dynamic SAGIN environments. The main contributions of this work are summarized as follows:

- This study introduces a novel semantically aware SAGIN architecture that enables the dynamic placement, selection, compression, updating, and migration of semantic knowledge bases across satellite, aerial, and terrestrial edge nodes.
- This work proposes a unified optimization problem to orchestrate semantic knowledge, task offloading, UAV trajectory planning, and communication-computation resource allocation subject to latency, energy, reliability, privacy, and semantic quality constraints.
- The proposed SKFMI provides a unified measure of knowledge staleness, data-distribution drift, and task-specific mismatch, thereby enabling adaptive knowledge-base management in dynamic SAGIN environments.
- The MARL framework is proposed by introducing a hierarchical graph constraint between the long timescale of orchestration decisions and the short timescale of task offloading and resource control decisions.
- To achieve better system security, scalability, and communication efficiency, the proposed framework uses trust-aware Byzantine-robust aggregation and event-triggered policy sharing and is extensively evaluated for various dynamic topologies, UAV energy constraints, non-independent and non-identically distributed task arrivals, semantic drift, node failures, and policy-poisoning attacks.

The remainder of this paper is organized as follows. [Section 2](#) reviews related work on SAGIN edge computing, semantic communications, multi-agent DRL, and trustworthy FL. [Section 3](#) describes the system architecture, communication and computation models, semantic knowledge model, and threat assumptions. In [Section 4](#), the problem to be addressed is formulated: how to intertwine semantic orchestration and resource allocation. In [Section 5](#), the proposed HGCF-MARL framework and trust-aware FL mechanism are

presented. The experimental settings and performance evaluation are given in [Section 6](#). [Section 7](#) concludes the paper and provides directions for future research.

2 Related Work

Existing research has progressively shifted from conventional data delivery toward task-oriented and semantic-aware information exchange. In [\[8\]](#), bus trajectories were used to enable street-centric message delivery in urban vehicular networks, while study [\[9\]](#) used autoregressive mobility prediction to optimize vehicular routing. A study in [\[10\]](#) also investigated perception task-oriented information sharing, which is proposed to handle sensing occlusions in Connected and autonomous vehicles. Although these approaches improve information availability, they do not explicitly quantify semantic knowledge aging, distribution drift, or task-specific mismatch.

Other essential issues affecting reliable SAGIN operation include synchronization and positioning accuracy. The influence of time-synchronized spacecraft control and robust GNSS/IMU synchronization has been investigated in [\[11,12\]](#), respectively. A sequence synchronization mechanism was also developed in [\[13\]](#) to reduce unnecessary control exchanges through event-triggered operation. Furthermore, phase-coded signal estimation, multi-constellation ambiguity resolution, and resilient pseudorange prediction were studied in [\[14,15\]](#). Cooperative localization approaches and satellite-constrained positioning mechanisms were presented in [\[16,17\]](#). Although these studies improve synchronization and navigation reliability, they do not consider their integration with semantic knowledge orchestration and federated policy learning in dynamic SAGIN environments.

The semantic representation and multimodal intelligence are increasingly gaining interest. A lightweight question-answering architecture was presented in [\[18\]](#), and hierarchical semantic extraction was used in the detection of malware in [\[19\]](#). The possibilities of semantic data generation with Large Language Models (LLMs), as well as representation refinement for cross-modal data, were explored in [\[20,21\]](#). Moreover, an optimization of collaborative inference for edge large language models was presented in [\[22\]](#). While these works enable efficient semantic extraction and reasoning, they do not consider how semantic knowledge is placed, migrated, and updated, or how to assess its trustworthiness across heterogeneous SAGIN nodes. Edge intelligence and dynamic resource orchestration have also been extensively studied. In [\[23\]](#) an emergency scheduling approach for space-target tracking, based on Auto-ML is proposed. In [\[24\]](#), real-time edge-assisted fault diagnostics have been demonstrated. Collaborative perception and computation offloading was jointly optimized in the 6G air-ground environment in [\[25\]](#), and secure vehicular-edge task offloading was investigated in [\[26\]](#) in combination with low altitude channel characteristics. In general, the above methods focus on optimizing only one of the scheduling, communication, and computing goals while neglecting the whole semantic knowledge lifecycle.

Reinforcement Learning (RL) is now one of the main instruments of adaptive decision-making in complex networks. Sample-efficient temporal-difference learning was proposed in [\[27\]](#), and Transformer-based RL was reviewed in [\[28\]](#). In [\[29\]](#), cooperative policy-reward optimization for LLM alignment has been studied. Path planning for safety-critical tasks and distance-sensitive task allocation were analyzed in the works [\[30,31\]](#), respectively. In [\[32,33\]](#), multi-agent task offloading and UAV rescue with Generative AI (GAI) were proposed to ensure fairness. Furthermore, in [\[34\]](#), game-theoretical DRL was utilized for energy-efficient multi-UAV data collection. These approaches achieve a more efficient learning, more safety, more fairness or more control over the trajectory of learning, but tend to use rather flat decision structures without trying to orchestrate jointly both long-term semantic decision and short-term resource allocation.

The graph-based learning approach is a suitable expression for dynamic network relationships. In [35], a graph neural network is used for coordinated multi-cell beamforming. FL is capable of achieving collaborative intelligence without sharing raw data. In [36], a low-coupling vertical FL architecture is proposed, and in [37], the concept of differentially private federated optimization for the Internet of vehicles (IoV) is presented. Secure offloading and hierarchical semantic security mechanisms were investigated in [38]. While these approaches support enhancing privacy and security, they do not fully account for the reliability of policy updates, semantic mismatches, and communication-triggering requirements in dynamic SAGIN. Multi-agent reinforcement learning has been increasingly applied in wireless edge networks to address distributed resource allocation, computation offloading, and coordination problems under dynamic network conditions [39]. Recent studies have investigated dynamic resource scheduling and adaptive resource allocation in edge computing networks, highlighting the importance of temporal awareness and efficient resource utilization in highly dynamic environments [40]. However, those do not offer a consolidated approach for the semantic knowledge freshness evaluation and cross-layer orchestration.

Existing studies have made significant advances in semantic communication, SAGIN optimization, graph neural networks, MARL, FL, and secure edge intelligence. However, most of these solutions focus on addressing these difficulties individually and lack a universal approach to managing the freshness of semantic knowledge, the dynamic network topology, hierarchical decision-making, trust-aware FL, and communication-efficient policy synchronization. In addition, most graph-oriented methods capture only topology representation, most MARL methods focus on network control or resource optimization, and most FL protocols are either mainly geared towards privacy preservation or towards strong aggregation without considering the evolution of the semantics of the trained knowledge or adaptive knowledge lifecycle management. The proposed HGCF-MARL framework aims to overcome these limitations by employing a centralized semantic orchestration architecture that integrates SKFMI, a graph-constrained state representation, hierarchical MARL, trust-aware, Byzantine-robust federated policy aggregation, and event-triggered synchronization. The proposed framework can solve the problems of semantic knowledge management, task offloading, UAV trajectory planning, inter-layer association, and communication/computation resource allocation in a unified and coherent way, and present a semantic-aware, trust-enhanced and hierarchy-driven optimization paradigm for dynamic SAGIN environments.

3 System Model

The system under consideration is a dynamic SAGIN comprising satellites, HAPS, UAVs, terrestrial base stations, and IoT devices, collectively serving as communication and computation facilitators under time-varying conditions arising from mobility, channel variations, and task arrivals. Joint optimization is enabled across heterogeneous layers of the complete node set defined in Eq. (1) in which all entities are unified.

$$\mathcal{N} = \mathcal{S} \cup \mathcal{H} \cup \mathcal{U} \cup \mathcal{B} \cup \mathcal{D}, \quad (1)$$

in this formulation, $\mathcal{N}$ denotes the complete set of network nodes, while $\mathcal{S}$, $\mathcal{H}$, $\mathcal{U}$, $\mathcal{B}$, and $\mathcal{D}$ represent the sets of LEO satellites, HAPS nodes, UAV-mounted edge nodes, terrestrial base stations, and IoT devices, respectively. The system is organized as a graph of nodes and links and structured as a two-timescale system: one macro level with orchestration and one micro level with control. IoT devices have different types of tasks with delay requirements and semantic demands assigned to nodes, and communication requirements at the bit, feature, or semantic level introduce a trade-off between efficiency and quality. They account for mobility, communication, and computation together, with models for UAV dynamics at different levels, SINR-based

links and CPU allocation impacting performance, and task queues capturing congestion. The total task delay combines all of the above delays, transmission, queueing, computation, return, and migration delays as defined by Eq. (2).

$$D_d^{\text{tot}}(t) = D_d^{\text{up}}(t) + D_d^{\text{que}}(t) + D_d^{\text{cmp}}(t) + D_d^{\text{ret}}(t) + D_d^{\text{mig}}(t), \quad (2)$$

where, $D_d^{\text{tot}}(t)$ denotes the total completion delay of task d at time t . $D_d^{\text{up}}(t)$ represents uplink transmission delay, $D_d^{\text{que}}(t)$ denotes queueing delay, $D_d^{\text{cmp}}(t)$ denotes computation delay, $D_d^{\text{ret}}(t)$ represents result-return delay, and $D_d^{\text{mig}}(t)$ denotes semantic knowledge-migration delay. The system stores and manages distributed semantic knowledge bases, with constraints on size, version, embeddings, and update times that can be dynamically placed and migrated. Knowledge can be compacted and transported between nodes at the cost of transportation delay and energy consumption. The SKFMI, introduced in Eq. (3), combines temporal staleness and distribution drift and also incorporates task mismatch.

$$\text{SKFMI}_{d,k,n}(t) = \alpha A_{k,n}(t) + \beta \Delta_{k,n}(t) + \gamma M_{d,k}(t), \quad (3)$$

here, $\text{SKFMI}_{d,k,n}(t)$ denotes the freshness-mismatch index of knowledge object k at node n for task d at time t . $A_{k,n}(t)$ represents the knowledge age (staleness), $\Delta_{k,n}(t)$ denotes the data-distribution drift, and $M_{d,k}(t)$ represents the task-knowledge mismatch. SKFMI measures whether stored semantic knowledge remains suitable for current tasks by jointly considering temporal aging, distribution drift, and task-specific mismatch. It is a unified quality indicator that enables adaptive knowledge updating, migration, and replacement within the HGCF-MARL framework. The coefficients α , β , and γ are weighting factors satisfying $\alpha + \beta + \gamma = 1$.

A lower SKFMI indicates better alignment between knowledge and tasks, while semantic quality depends on channel conditions, compression, retained knowledge, and SKFMI. Event-triggered communication can be used in a federated distributed learning system, where agents learn policies locally but share them via coordinators to minimize communication. A trust-aware mechanism is used to avoid receiving unreliable updates from malicious or faulty nodes, thereby guaranteeing security. Measurements for system performance include latency, power consumption, semantic quality, communication overhead, and robustness. Specifically, the task-completion ratio in Eq. (4) is the percentage of tasks that meet both the delay and semantic-quality constraints.

$$R^{\text{comp}} = \frac{\sum_{t \in \mathcal{T}} \sum_{d \in \mathcal{D}} \mathbb{I}[D_d^{\text{tot}}(t) \leq \tau_d(t), Q_d(t) \geq q_d(t)]}{\sum_{t \in \mathcal{T}} \sum_{d \in \mathcal{D}} a_d(t)}. \quad (4)$$

in which, R^{comp} denotes the task-completion ratio. $\mathcal{T}$ is the set of time slots and $\mathcal{D}$ is the set of tasks. $\mathbb{I}[\cdot]$ is the indicator function, which equals 1 if both conditions are satisfied and 0 otherwise. $D_d^{\text{tot}}(t)$ is the total delay, $\tau_d(t)$ is the delay constraint, $Q_d(t)$ is the achieved semantic quality, $q_d(t)$ is the required semantic-quality threshold, and $a_d(t)$ indicates whether a task arrives at time t . This formulation combines communication, computation, mobility, semantic knowledge management, distributed learning, and security into a single framework, paving the way for efficient and scalable optimization of SAGIN systems in a dynamic and adversarial environment.

4 Joint Semantic Knowledge Orchestration and Resource-Allocation Problem Formulation

The system is a dynamic SAGIN with satellites, HAPS, UAVs, base stations, and IoT devices, unified into the global set $\mathcal{N}$ defined in Eq. (1). It adopts a two-timescale decision structure, where macro-level actions (Eq. (5)) determine knowledge placement, migration, UAV trajectory, association, and resource

reservation, while micro-level actions $\mathcal{A}^I(t) = \{\mathbf{X}(t), \mathbf{Z}(t), \boldsymbol{\rho}(t), \mathbf{B}(t), \mathbf{P}(t), \mathbf{F}(t), \boldsymbol{\zeta}(t)\}$ handle real-time offloading and resource allocation.

$$\mathcal{A}^M(m) = \{\mathbf{Y}(m), \mathbf{M}(m), \boldsymbol{\delta}(m), \mathbf{Q}(m), \mathbf{A}^{\text{assoc}}(m), \mathbf{F}^{\text{res}}(m)\}, \quad (5)$$

The overall policy $\Pi = \{\Pi^M, \Pi^I\}$ aims to minimize the long-term cost defined in Eq. (6), where $J(t)$ captures delay, energy, semantic distortion, and system overhead.

$$\min_{\Pi} \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[\sum_{t=1}^T J(t) \right], \quad (6)$$

The total task delay is defined in Eq. (2). Task feasibility is ensured through delay and semantic constraints such as $D_d^{\text{tot}}(t) \leq \tau_d(t)$ and $r_{d,k,n}(t) \text{SKFMI}_{d,k,n}(t) \leq \Gamma_d^{\text{max}}$. Resource constraints include bandwidth $\sum b_{i,j}(t) \leq B_j^{\text{max}}$, computation $\sum f_{d,n}(t) \leq F_n^{\text{max}}$, and UAV mobility $\|\mathbf{v}_u(t)\|_2 \leq V_u^{\text{max}}$. To reduce communication overhead, event-triggered updates are used, as defined in Eq. (7), in which updates are transmitted only when significant changes occur.

$$\zeta_n(t) = \begin{cases} 1, & \Omega_n(t) > \xi_n(t), \\ 0, & \Omega_n(t) \leq \xi_n(t), \end{cases} \quad (7)$$

Finally, the optimization problem in Eq. (8) integrates all objectives and constraints for efficient SAGIN operation. This formulation unifies and scales the use of semantic knowledge, resource allocation, mobility, and FL.

$$\begin{aligned} \mathcal{P}: \quad & \min_{\Pi} \quad \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[\sum_{t=1}^T J(t) \right] \\ & \text{s.t.} \quad \text{C1} - \text{C19}, \end{aligned} \quad (8)$$

5 Proposed HGCF-MARL Framework and Trust-Aware Learning

The proposed HGCF-MARL approach addresses semantic knowledge orchestration, task offloading, resource allocation, UAV control, and FL in dynamic SAGIN environments. The observed network state is encoded as a dynamic graph, processed in a hierarchical MARL, and fine-tuned using trust-aware FL to yield secure and efficient orchestration actions, as shown in Fig. 1.

HGCF-MARL methodological novelty consists in combining SKFMI-based evaluation of knowledge with two-level decision-making, reliability-aware policy aggregation, and event-triggered synchronization. This design relies on a combination of long-term semantic knowledge management and short-term resource control, and makes the system robust to unreliable participants while being light on communication overhead. At each time step, the system constructs a dynamic graph $\mathcal{G}(t) = (\mathcal{V}(t), \mathcal{E}(t), \mathbf{X}(t), \mathbf{Z}(t))$ capturing node states, link conditions, semantic knowledge, and trust information. A graph attention network extracts topology-aware embeddings, where node updates follow a weighted aggregation of neighboring features, and the global representation is obtained as $\mathbf{g}(t) = \text{GAT}(\mathcal{G}(t))$. The control policy is hierarchical, defined as $\Pi = \{\pi^M, \pi^I\}$, where the macro-policy handles long-term orchestration and the micro-policies perform real-time control actions. To enforce constraints, a Lagrangian objective $\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\lambda}) = J(\boldsymbol{\theta}) + \sum_m \lambda_m (C_m(\boldsymbol{\theta}) - C_m^{\text{max}})$ is adopted, balancing performance and constraint satisfaction.

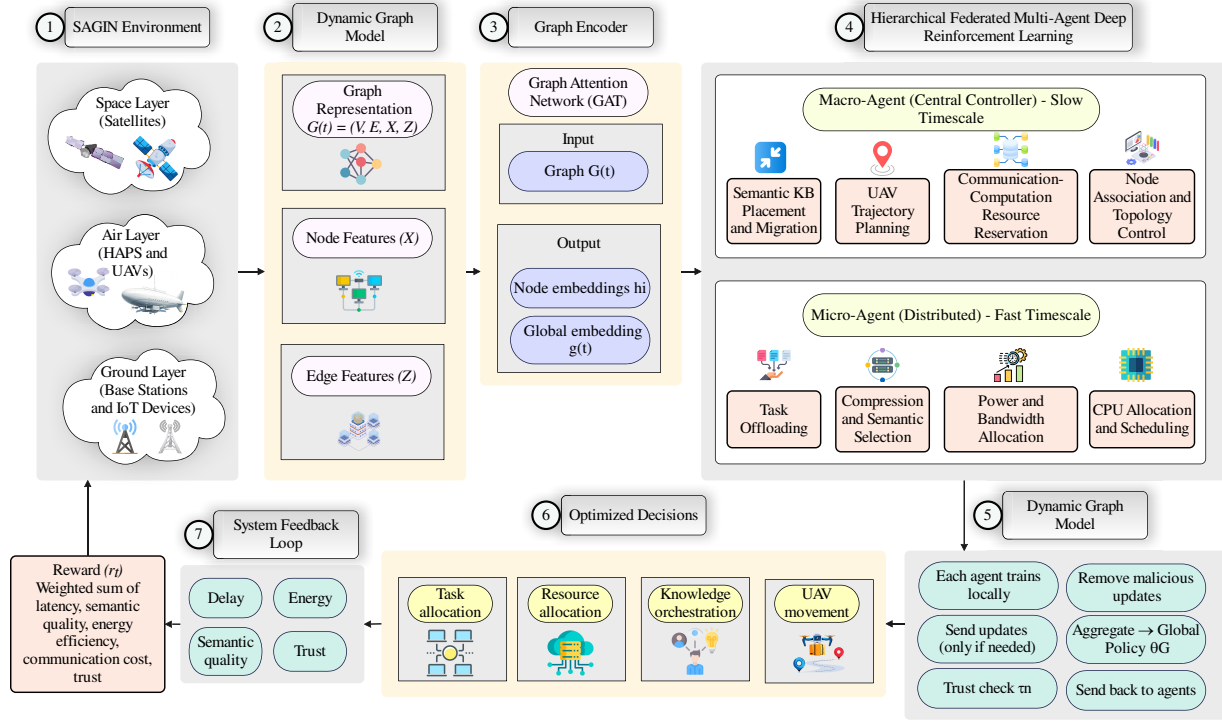


Figure 1: Proposed HGCF-MARL framework for semantic knowledge orchestration in SAGIN, integrating dynamic graph modeling, hierarchical multi-agent decision-making, trust-aware federated learning, and feedback-driven optimization.

During training, each node updates its local policy using gradient descent $\theta_n^{(k+1)} = \theta_n^{(k)} - \eta \nabla_{\theta} L_n(\theta_n^{(k)})$, and only model updates are shared. To ensure robustness, unreliable updates are filtered using deviation and trust criteria, forming a trusted set $\mathcal{H}(t)$, and the global policy is computed using trust-weighted aggregation. The reliability of each participating agent is evaluated before federated policy aggregation using the trust score defined in Eq. (9). The score combines update consistency, historical reliability, and behavioral reliability:

$$\tau_n(t) = \lambda_1 C_n(t) + \lambda_2 H_n(t) + \lambda_3 B_n(t), \quad (9)$$

where $\tau_n(t)$ denotes the trust score of agent n , while $C_n(t)$, $H_n(t)$, and $B_n(t)$ represent update consistency, historical reliability, and behavioral reliability, respectively. The coefficients λ_1 , λ_2 , and λ_3 determine the contribution of each component. Agents with trust scores below $\tau_{\min}$ are excluded or assigned lower aggregation weights to limit the influence of unreliable or malicious updates. The immediate reward used to

train the hierarchical agents is defined in Eq. (10) to balance semantic quality, latency, energy consumption, communication overhead, and trustworthiness:

$$r(t) = w_1 Q_{\text{sem}}(t) - w_2 D(t) - w_3 E(t) - w_4 O(t) + w_5 T(t), \quad (10)$$

where $Q_{\text{sem}}(t)$ denotes semantic knowledge quality, $D(t)$ is the task completion delay, $E(t)$ represents energy consumption, $O(t)$ denotes communication and synchronization overhead, and $T(t)$ is the trust reliability factor. The weights w_1 – w_5 determine the relative importance of these objectives, enabling the agents to learn policies that balance semantic accuracy, service efficiency, communication cost, and robustness. Communication overhead is reduced through event-triggered synchronization, where updates are

transmitted only when a trigger condition $\zeta_n(t)$ exceeds a threshold. The overall computational complexity is dominated by graph processing, given in Eq. (11).

$$\mathcal{O}(|\mathcal{E}|d_h + |\mathcal{V}|d_h^2), \quad (11)$$

The event-triggered synchronization mechanism reduces communication overhead by transmitting policy updates only when the triggering condition is satisfied. Compared with periodic aggregation, whose cost is $\mathcal{O}(N_a d_\theta R)$, HGCF-MARL reduces the effective synchronization frequency to ρR , yielding a communication cost of $\mathcal{O}(N_a d_\theta \rho R)$, where $0 < \rho < 1$. The overall computational cost comprises graph encoding, hierarchical policy optimization, and trust-aware aggregation, with complexities of $\mathcal{O}(|\mathcal{E}|d_h + |\mathcal{V}|d_h^2)$, $\mathcal{O}(N_a T_s C_\pi)$, and $\mathcal{O}(N_a d_\theta)$, respectively. These components form a scalable architecture for dynamic SAGIN environments, and the complete training and execution procedure is summarized in Algorithm 1.

Algorithm 1: HGCF-MARL training and federated synchronization

Require: Environment, policies θ^M , $\{\theta_n^I\}$, trust threshold $\tau_{\min}$, event thresholds $\{\xi_n\}$

Ensure: Optimized hierarchical and global policies

```

1: Initialize graph encoder, policies, trust scores, and buffers
2: for each episode do
3:   Observe state, build graph  $\mathcal{G}(t)$ , compute embeddings
4:   if macro-slot then
5:     Execute macro-actions (placement, migration, UAV, resources)
6:   end if
7:   for each micro-slot do
8:     Agents select actions and execute task offloading
9:     Observe rewards and update local policies
10:    for each agent  $n$  do
11:      Compute trust  $\tau_n(t)$  and trigger  $\Omega_n(t)$ 
12:      if  $\Omega_n(t) > \xi_n(t)$  then
13:        Upload policy update
14:      end if
15:    end for
16:    Filter unreliable updates and aggregate trusted policies
17:    Broadcast updated global policy
18:  end for
19: end for
20: return Final policies

```

6 Experimental Settings and Performance Evaluation

This section introduces the experimental setup, benchmark methods, evaluation criteria, and experimental protocol for investigating the proposed HGCF-MARL framework. The evaluation is aimed at assessing the efficient orchestration of semantic knowledge, task offloading, communication-computation resource allocation, UAV mobility management, federated-policy efficiency, and resilience to Byzantine participants under varying SAGIN conditions.

Experimental Environment: The SAGIN is deployed over a $20 \times 20 \text{ km}^2$ area with satellites, HAPS, UAVs, base stations, and IoT devices. The default setup includes 100 IoT devices, 4 base stations, 6 UAVs,

2 HAPS, and 4 LEO satellites, with device numbers varying from 40 to 200 for scalability analysis. Tasks are heterogeneous and non-IID, with sizes 0.5–5 MB, computation demand 0.5×10^9 – 5×10^9 CPU cycles, and deadlines 100–800 ms, while semantic quality ranges from 0.80 to 0.95. Nodes have heterogeneous computation and storage, UAVs operate at 100–300 m, and HAPS at 20 km. Satellite links follow orbital visibility, and communication uses probabilistic LoS models with dynamic bandwidth allocation. Detailed parameters are given in [Table 1](#).

Table 1: Key simulation parameters for the SAGIN environment.

Parameter	Value
Service area	$20 \times 20 \text{ km}^2$
IoT devices	100
Base stations/UAVs/HAPS/LEO	4/6/2/4
Micro-slot duration	0.1 s
Task size	0.5–5 MB
Computation demand	$0.5\text{--}5 \times 10^9$ cycles
Task deadline	100–800 ms
Semantic quality	0.80–0.95
UAV altitude/speed	100–300 m/25 m/s
UAV energy budget	300 kJ
Knowledge size	50–250 MB
Retention ratio	0.25–1.00
Malicious ratio	10%
Dirichlet parameter	0.5

Semantic Datasets and Task Construction: The three benchmark datasets, EuroSAT, CIFAR-10, and AG News, are selected to model image-based and text-based semantics tasks. The simulated non-IID conditions are generated using a Dirichlet distribution with $\alpha_{\text{Dir}} = 0.5$, with additional distributional drift and cross-region task mismatch. Accuracy and macro-F1 (for AG News) are used to evaluate performance, both of which are normalized into a unified semantic-quality metric. The details are provided in [Table 2](#).

Semantic Knowledge Catalogue: Each task-specific knowledge object in a semantic knowledge catalog includes knowledge objects trained across different regions and data distributions, along with their embeddings, model parameters, versions, update times, and storage requirements. Different sizes of knowledge can span from 50 to 250 MB, and the compression ratio is between 0.25 and 1.0 for balance of storage, migration cost and accuracy. Bring together knowledge age, data-distribution drift, and task mismatch, with SKFMI used as a means of assessing the suitability of the knowledge. Drift is quantified by Jensen-Shannon divergence, and mismatch is defined as the cosine distance between the task and knowledge embedding. The default weights are applied during experiments and further assessed in a sensitivity evaluation.

Table 2: Semantic datasets and task configuration.

Dataset	Task	Classes	Training Samples	Test Samples	Metric	Input Configuration
EuroSAT	Remote-sensing classification	10	18,900	5,400	Accuracy, Macro-F1	64×64 RGB
CIFAR-10	Image classification	10	45,000	10,000	Accuracy	32×32 RGB
AG News	Text-topic classification	4	108,000	7600	Accuracy, Macro-F1	Max 128 tokens
Synthetic SAGIN traces	Offloading and resource control	4 task types	600,000 tasks	120,000 tasks	Task-success ratio	Non-IID regional arrivals

HGCF-MARL Implementation: The HGCF-MARL two-timescale hierarchy consists of an upper tier that runs once during every 10 micro-slots and distributes knowledge, migrates knowledge, controls the UAVs, and reserves resources, and a bottom tier where agents run within each micro-slot to place and offload tasks, perform transmission control, and perform compression and resource allocation. A three-layer graph attention network (GAT) with 128 hidden units and 8 heads has been used to model the SAGIN topology, capturing node state and link conditions. Standard actor–critic PPO with standard parameters and learning rates is used for both levels. Local policies are updated after 5 epochs, and federated aggregation is performed after every 20 micro-slots or is event-triggered with an adaptive threshold. In every experiment, training episodes run for a total of 3000 and are repeated 10 times, with 200 micro-slots per episode. The average and 95% confidence intervals are shown in [Table 3](#).

Table 3: Key HGCF-MARL training parameters.

Parameter	Value
Training episodes	3000
Micro-slots per episode	200
Independent runs	10
Graph layers/heads	3/8
Hidden dimension	128
Actor–critic layers	256, 128
Learning rates (actor/critic)	$3 \times 10^{-4}/10^{-3}$
Discount factor γ	0.99
PPO clipping	0.20
Batch size	128
Federation interval	20 micro-slots
Event threshold	0.15
Trust threshold	0.60

Byzantine Attack Configuration: The percentage of malicious participants ranges from 0% to 40% and covers sign-flipping attacks, model replacement attacks, random update attacks, stale replay attacks, free-riding attacks, and falsified telemetry. Malicious nodes manipulate updates, reuse outdated models, or exaggerate resource availability. A trust-aware mechanism quantifies changes, validates them, and checks consistency, freshness, and historical behavior before filtering out suspicious updates for aggregation.

Comparison Methods: The proposed HGCF-MARL is compared with the baseline algorithms: LocalExecution, NearestEdgeOffloading, CentralizedPPO, FlatMAPPO, HierarchicalMARL, Graph-MARL, Fed-MARL, and RobustFed-MARL. The ablation studies consist of removing SKFMI, event-triggering, trust mechanisms, and the greedy knowledge placement strategy. The comparison of salient aspects is summarized in Table 4.

Table 4: Characteristics of the compared methods.

Method	KB Orchestration	Graph	Hier.	Fed.	Byzantine	Event
Local Execution	✗	✗	✗	✗	✗	✗
Nearest Edge	✗	✗	✗	✗	✗	✗
Centralized PPO	Limited	✗	✗	✗	✗	✗
Flat MAPPO	Limited	✗	✗	✗	✗	✗
Hierarchical MARL	Limited	✗	✓	✗	✗	✗
Graph-MARL	Limited	✓	✗	✗	✗	✗
Fed-MARL	Limited	✗	✗	✓	✗	✗
Robust Fed-MARL	Limited	✗	✗	✓	✓	✗
Greedy Placement	Popularity-based	✗	✗	✗	✗	✗
HGCF-MARL w/o SKFMI	Age-based	✓	✓	✓	✓	✓
HGCF-MARL w/o Trigger	Dynamic	✓	✓	✓	✓	✗
HGCF-MARL w/o Trust	Dynamic	✓	✓	✓	✗	✓
Proposed HGCF-MARL	SKFMI-based	✓	✓	✓	✓	✓

Note: The symbols “✓” and “✗” indicate whether a method supports the corresponding feature. Specifically, “✓” denotes that the feature is considered or incorporated by the method, whereas “✗” denotes that the feature is not supported or not included. “Limited” indicates partial or restricted support for the corresponding capability.

Evaluation Metrics: The evaluation metrics include normalized accuracy (or macro-F1) for semantic quality, task delay, task completion ratio, and total energy consumption for system performance. Knowledge efficiency is assessed using the knowledge-hit ratio and migration traffic, while system overhead is evaluated based on federated signaling. Security performance is evaluated using detection accuracy and false positive rate, whereas learning performance is measured using episodic return and convergence behavior.

Convergence Evaluation: Convergence speed and stability of HGCF-MARL compared with baseline methods are evaluated using the average episodic return over 3000 episodes, as shown in Fig. 2. The comparison with Graph-MARL and Hierarchical MARL demonstrates the benefits of graph-based state representation and hierarchical decision-making. In addition, the contributions of individual HGCF-MARL components are investigated through ablation experiments by removing specific modules, including SKFMI, trust mechanisms, and event-triggered synchronization. Robustness is further evaluated under Byzantine attacks introduced at episode 1500, where recovery performance reflects the effectiveness of trust-aware aggregation and event-triggered updates.

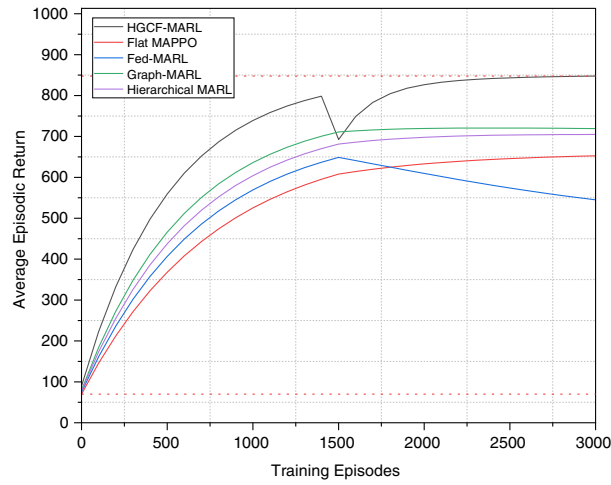


Figure 2: Convergence behavior of average episodic return across training episodes, including the effect of Byzantine disturbances.

Task Latency and Completion Performance: In Fig. 3, the average task delay is plotted as the number of IoT devices increases, contributing to greater contention, queueing, and migration overhead. A comparison of HGCF-MARL with baseline techniques, especially for high-load conditions, is presented. The corresponding completion ratio is shown in Fig. 4, which is defined as successful only if a task meets both the latency and the inputs-on-output semantic-quality requirements.

Energy Consumption: Total energy consumption under varying task-arrival rates is illustrated in Fig. 5, including communication, computation, UAV propulsion, and knowledge-migration components, with a detailed breakdown provided in Table 5. While baseline methods incur higher computational or communication costs, HGCF-MARL reduces overall energy consumption through the joint optimization of compression, power allocation, task placement, and UAV mobility.

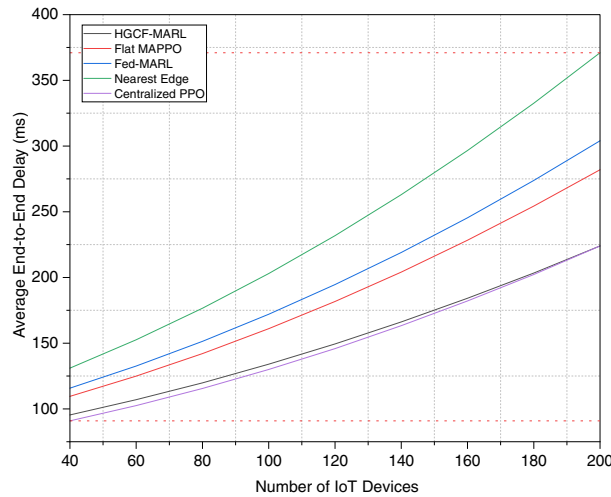


Figure 3: Average end-to-end task delay under increasing numbers of IoT devices.

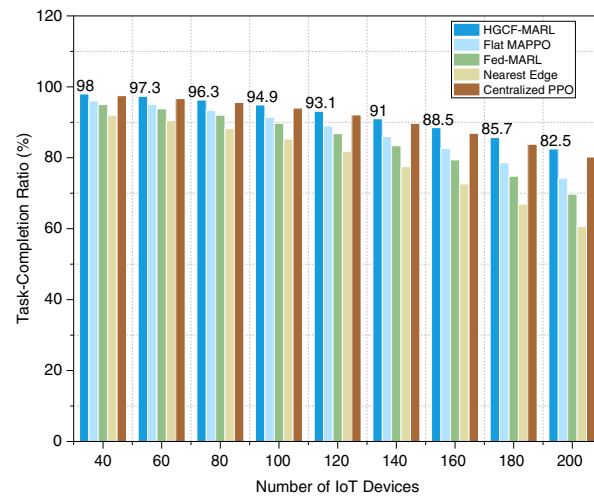


Figure 4: Task-completion ratio under increasing IoT device density, reflecting system reliability.

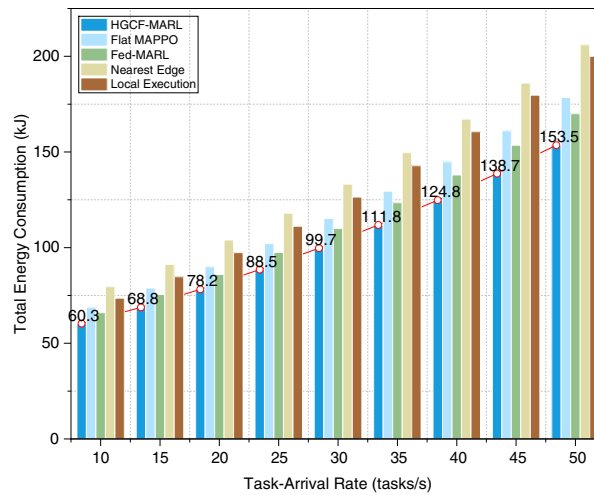


Figure 5: Total network energy consumption under different task-arrival rates.

Table 5: Component-wise energy consumption under default load.

Method	IoT Comm. (kJ)	Comp. (kJ)	UAV Prop. (kJ)	KB Mig. (kJ)	Total (kJ)
Local Execution	9.8	118.6	0.0	0.0	128.4
Nearest Edge	42.7	65.3	51.8	0.0	159.8
Centralized PPO	31.5	57.4	45.6	8.9	143.4
Flat MAPPO	29.8	54.7	43.2	8.1	135.8
Fed-MARL	28.4	52.9	41.7	7.8	130.8
Proposed HGCF-MARL	23.7	46.1	36.4	6.5	112.7

Semantic Quality and SKFMI Analysis: As demonstrated in Fig. 6, semantic quality decreases as distribution drift increases; however, HGCF-MARL maintains superior performance compared with age-based and static approaches by leveraging SKFMI for adaptive knowledge selection. Fig. 7 illustrates the relationship between semantic quality and task delay across different compression ratios, demonstrating effective adaptation for both image and text tasks.

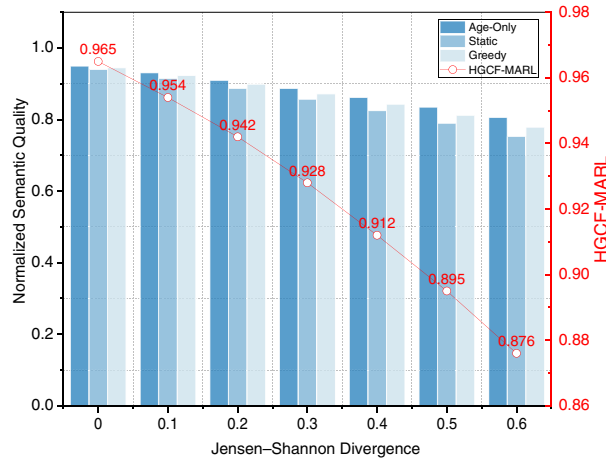


Figure 6: Normalized semantic quality under increasing data-distribution drift.

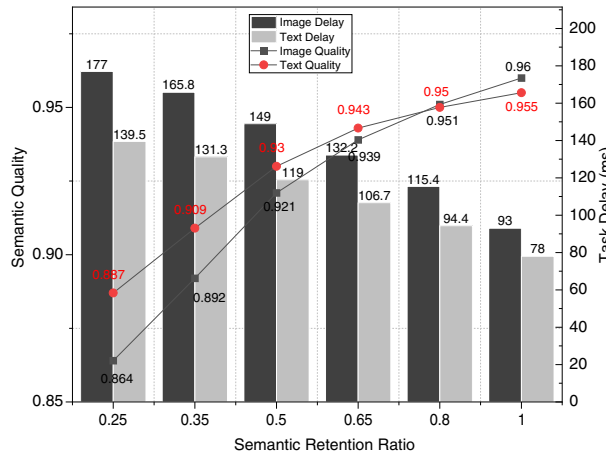


Figure 7: Trade-off between semantic quality and task delay under different retention ratios.

Knowledge Placement and Migration Performance: Table 6 outlines and compares knowledge-hit ratio, storage utilization, and migration traffic across different methods. While Fig. 8 displays the effect of user mobility and higher mismatch with static placement, and high overhead due to periodic synchronization. HGCF-MARL, however, adapts, balancing semantic performance and communication cost.

Event-Triggered Federated Communication: As seen in Fig. 9, periodic Fed-MARL incurs high communication overhead, whereas the proposed event-triggered mechanism transmits less information when there is no significant change in the information. Table 7 compares the cost of signaling and the performance to observe the balance between signaling cost and timely policy adaptation with adaptive thresholds.

Table 6: Semantic knowledge placement and migration performance.

Method	Hit Ratio (%)	Avg. SKFMI	Storage Util. (%)	Migration (GB)	Sem. Quality
Static Placement	68.4	0.421	61.7	5.8	0.814
Greedy Placement	76.9	0.354	78.5	14.3	0.847
Periodic Synchronization	91.6	0.188	92.4	38.7	0.913
Age-Only Selection	83.7	0.273	81.6	21.2	0.878
HGCF-MARL w/o SKFMI	86.5	0.241	83.1	18.9	0.892
Proposed HGCF-MARL	94.2	0.146	84.8	15.6	0.936

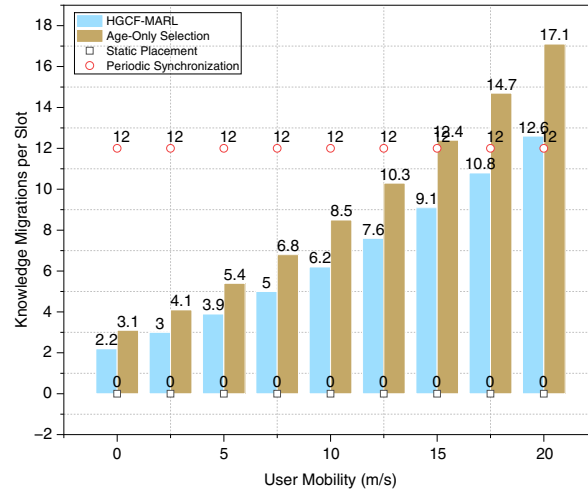
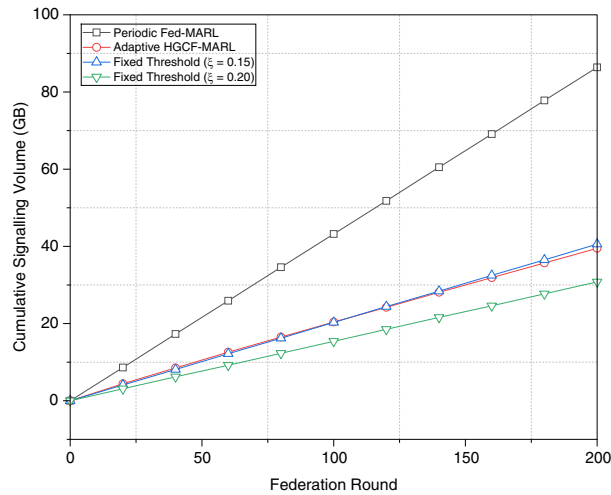
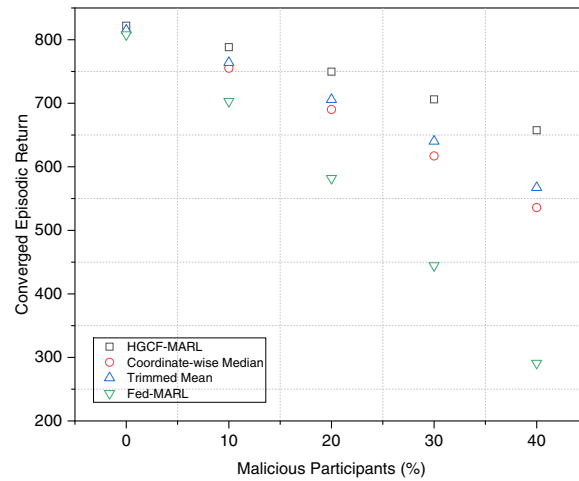
**Figure 8:** Semantic knowledge migration frequency under varying user mobility levels.**Figure 9:** Cumulative federated signaling overhead across federation rounds for different synchronization strategies.

Table 7: Effect of event-triggering threshold on federated policy sharing.

Strategy	Threshold	Uploads	Volume (GB)	Reward	Comp. Ratio (%)	Recovery
Periodic Federation	–	18,000	86.4	824.7	94.8	47
Fixed	0.05	14,760	70.8	823.9	94.7	45
Fixed	0.10	11,340	54.4	821.8	94.4	41
Fixed	0.15	8460	40.6	818.9	94.1	38
Fixed	0.20	6390	30.7	809.6	92.8	51
Fixed	0.30	4050	19.4	785.3	89.9	74
Adaptive Threshold	Dynamic	7920	38.0	821.6	94.3	34

Byzantine Resilience: Performance under the increase of malicious participants is shown in Fig. 10 where the HGCF-MARL is better than standard and robust baselines. Detection accuracy is summarized in Table 8 and the accuracy is higher for stronger attacks than subtle attacks. Fig. 11 also shows that the trust-aware mechanism applies to falsified telemetry behavior to keep the delay and reliability stable under adversarial conditions.

**Figure 10:** Converged episodic return under different malicious-participant ratios.**Table 8:** Byzantine-attack detection and robustness results.

Attack Type	Malicious (%)	Detection (%)	False Pos. (%)	Reward Ret. (%)	Comp. Ratio (%)	Attack Succ. (%)
Sign Flipping	10	98.7	1.8	97.3	93.8	1.3
Sign Flipping	30	96.1	2.7	92.4	90.6	3.9
Model Replacement	10	99.2	1.5	97.8	94.1	0.8
Model Replacement	30	96.8	2.9	91.7	89.8	3.2
Random Update	20	97.4	2.1	94.9	92.2	2.6
Stale Replay	20	92.6	3.4	92.1	90.5	7.4
Free Riding	20	89.8	4.1	90.6	89.2	10.2
Falsified Telemetry	20	93.7	3.8	91.9	90.1	6.3
Mixed Attacks	40	88.4	5.6	84.7	83.5	11.6

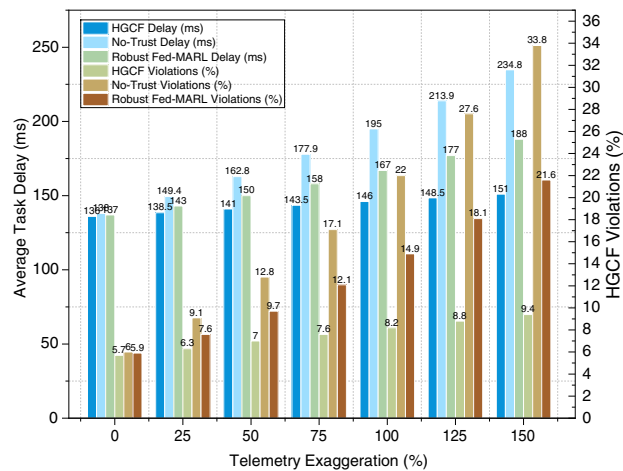


Figure 11: Impact of falsified telemetry on task delay and deadline-violation ratio.

UAV Mobility and Satellite Dynamics: As illustrated in Fig. 12, the more energy that is available to the UAV, the more tasks it is able to complete due to better positioning, and as predicted, this effect diminishes with increased energy. While Fig. 13 compares the satellite visibility, the HGCF-MARL proactively moves tasks and knowledge away from the satellite to minimize handover disruption.

Scalability: Fig. 14 presents decision latency and memory consumption per network size. However, a centralized PPO has issues with world optimization, and a Flat MAPPO is highly complex. By contrast, HGCF-MARL with hierarchical control and local graph-based decisions has low latency and high scalability.

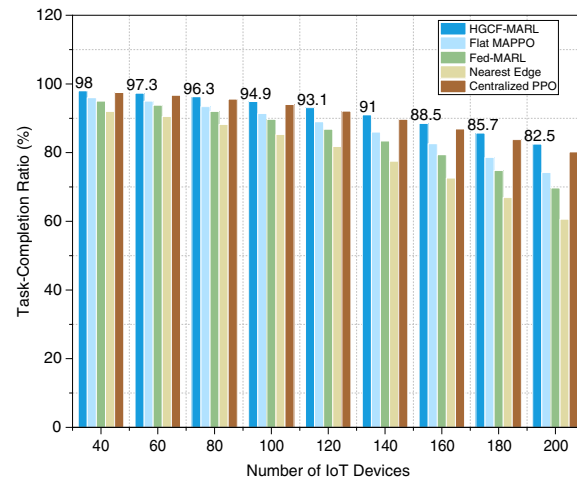


Figure 12: Task-completion performance and UAV propulsion energy under different energy budgets.

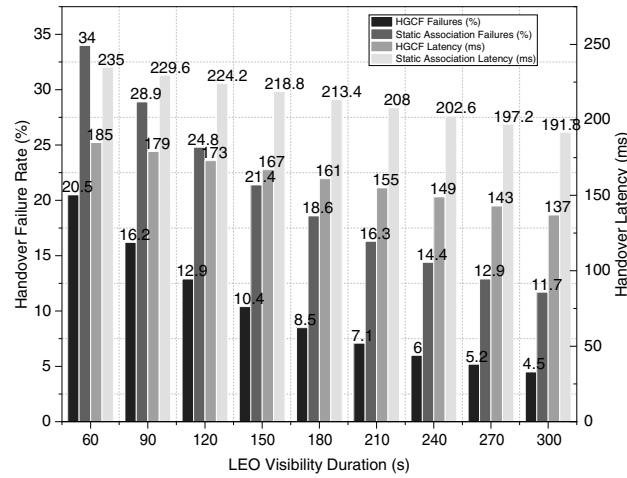


Figure 13: Handover failures and latency under varying LEO satellite visibility durations.

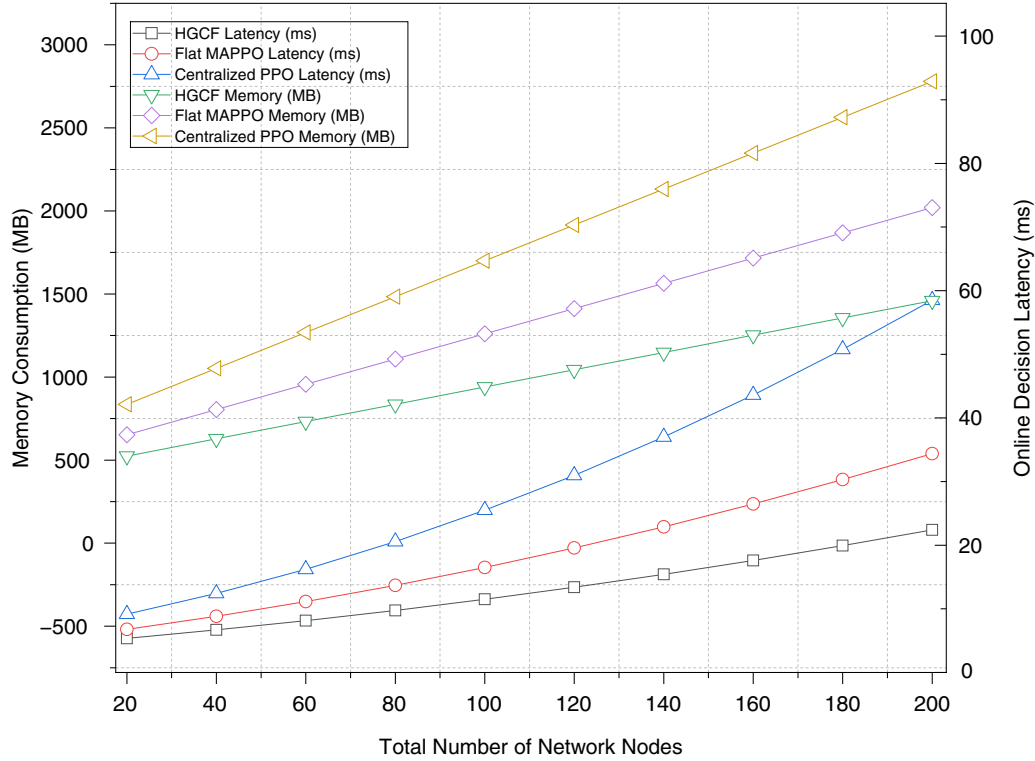


Figure 14: Scalability analysis in terms of online latency and memory consumption under increasing network size.

Ablation Study: The effect of each important component of HGCF-MARL is examined and reported in Table 9, in which the full model is compared with versions that omit graph attention, hierarchy, SKFMI, trust filtering, event-triggering, or dynamic migration. Many indicators have demonstrated that improvements in performance are not attributable to any single component in this specific set but to the integrated design as a whole.

Table 9: Ablation analysis of HGCF-MARL.

Variant	Reward	Delay (ms)	Energy (kJ)	Sem. Quality	Comp. Ratio (%)	Migration (GB)	FL Signal (GB)
Without Graph Attention	751.8	164.7	125.3	0.902	88.6	18.1	40.4
Without Hierarchy	724.6	178.9	131.2	0.896	86.9	19.7	42.8
Without SKFMI	769.3	151.8	120.5	0.892	89.7	18.9	39.6
Without Trust Filtering	681.4	193.5	137.1	0.861	82.8	21.3	38.2
Without Event Triggering	824.7	139.1	114.2	0.937	94.8	15.4	86.4
Without Dynamic KB Migration	738.9	171.6	126.8	0.883	87.5	0.0	36.9
Complete HGCF-MARL	821.6	136.4	112.7	0.936	94.3	15.6	38.0

Reproducibility Protocol: To avoid interference from different experimental environments, all methods are tested using the same parameters: seeds, task traces, channels, mobility, and topology, and their parameters are determined by the validation process. Results are averaged over 10 runs and shown with a 95% confidence interval. The full installation has been provided, including datasets, traces, attacks, etc., and figures are generated directly from the logs to ensure accuracy and reproducibility.

The evaluation is focused on HGCF-MARL with respect to orchestration efficiency, semantic effectiveness, communication efficiency, security resilience, and scalability. Delays, completion ratio, energy and semantic quality, migration traffic, signaling overhead, attack robustness, and decision latency are the metrics used to validate the system's performance in dynamic SAGIN settings.

7 Conclusion and Future Work

In this work, HGCF-MARL was proposed for semantic knowledge orchestration in the SAGIN edge computing, which simultaneously optimized the knowledge management, task offloading, resource allocation, UAV mobility, and FL. The framework aims to enhance semantic quality, task completion, latency, energy efficiency, convergence, and robustness against unreliable participants through SKFMI-based knowledge evaluation, graph-constrained hierarchical control, trust-aware aggregation, and event-triggered synchronization. In future studies, realistic large-scale SAGIN validation, more compelling comparisons with the latest advances in semantic communication, federated RL, graph-based orchestration approaches, lightweight policies, distributed graph processing, hierarchical aggregation, and adaptive synchronization for scalability will be directions for future work. Additional extensions can include 6G semantic communication, digital twins, multimodal and continual learning, LLM-assisted reasoning, privacy-preserving mechanisms, and theoretical guarantees of convergence and robustness.

Acknowledgement: The authors would like to thank King Saud University, Riyadh, Saudi Arabia, for its support and encouragement.

Funding Statement: This work was funded by the Ongoing Research Funding Program (ORF-2026-444), King Saud University, Riyadh, Saudi Arabia.

Availability of Data and Materials: Data will be available on reasonable request from the authors.

Ethics Approval: Not applicable. This study did not involve human participants, human data, animals, or biological materials.

Conflicts of Interest: The author declares that there are no conflicts of interest.

References

1. Wang L, Fan M, Yang N, Ma X, Liang Y, Zhang H. Toward intelligent space-air-ground integrated network: architecture, challenges, and emerging directions. *J Commun Inf Netw*. 2025;10(2):87–102. doi:10.23919/jcin.2025.11083695.
2. Ibrahim RW, Rodrigues TK, Kato N, Ariyoshi M, Hasegawa Y. A survey on AI-empowered high-frequency and FSO-enabled mobile edge computing in non-terrestrial networks for 6G. *IEEE Trans Cogn Commun Netw*. 2026;12:9504–28. doi:10.1109/tccn.2026.3706515.
3. Sun G, Wang Y, Yu H, Guizani M. Proportional fairness-aware task scheduling in space-air-ground integrated networks. *IEEE Trans Serv Comput*. 2024;17(6):4125–37. doi:10.1109/tsc.2024.3478730.
4. Huang C, Chen X, Chen G, Xiao P, Li GY, Huang W. Deep reinforcement learning-based resource allocation for hybrid bit and generative semantic communications in space-air-ground integrated networks. *IEEE J Sel Areas Commun*. 2025;43(12):3942–54. doi:10.1109/jsac.2025.3623157.
5. Betalo ML, Ullah I, Tesema FB, Wu Z, Li J, Bai X. Generative AI-driven multi-agent DRL for task allocation in UAV-assisted EMPD within 6G-enabled SAGIN networks. *IEEE Internet Things J*. 2025;12(17):35890–907. doi:10.1109/jiot.2025.3579780.
6. Zhang L, Zhao H, Xu B, Zhu H, Wang X. Agentic AI for SAGIN resource management: semantic awareness, orchestration, and optimization. *IEEE Netw*. 2026. doi:10.1109/mnet.2026.3695611.
7. Zhang P, Wang G, Chen S, Yu Y, Chen L. The application of mobile edge computing in the space-air-ground integrated network. *J Intell Comput Netw*. 2025;1(1):71–97. doi:10.64509/jicn.11.13.
8. Sun G, Zhang Y, Liao D, Yu H, Du X, Guizani M. Bus-trajectory-based street-centric routing for message delivery in urban vehicular ad hoc networks. *IEEE Trans Veh Technol*. 2018;67(8):7550–63. doi:10.1109/tvt.2018.2828651.
9. Sun G, Song L, Yu H, Chang V, Du X, Guizani M. V2V routing in a VANET based on the autoregressive integrated moving average model. *IEEE Trans Veh Technol*. 2019;68(1):908–22. doi:10.1109/tvt.2018.2884525.
10. Xiao Z, Shu J, Jiang H, Min G, Chen H, Han Z. Overcoming occlusions: perception task-oriented information sharing in connected and autonomous vehicles. *IEEE Netw*. 2023;37(4):224–9. doi:10.1109/mnet.018.2300125.
11. Li D, Tong S, Yang H, Hu Q. Time-synchronized control for spacecraft reorientation with time-varying constraints. *IEEE/ASME Trans Mechatron*. 2025;30(3):2073–83. doi:10.1109/tmech.2024.3430953.
12. Li J, Sun R, Wang Y, Ochieng WY. A robust time synchronization algorithm for GNSS/IMU integrated navigation in urban environments. *Meas Sci Technol*. 2025;36(3):036302. doi:10.1088/1361-6501/ada976.
13. Cao H, Hu Q, Jiang H, Ge SS, Li D. Predefined-sequential-synchronized control with exploration of event-triggered mechanism. *Automatica*. 2026;183:112645.
14. Cheng Q, Chen W, Wang J, Mi X, Yang Y, Sun R. Multi-constellation instantaneous single difference triple-carrier ambiguity resolution in urban environments. *GPS Solut*. 2025;29(4):186. doi:10.1007/s10291-025-01946-1.
15. Sun R, Sheng Q, Cheng Q, Shang X, Ochieng WY. 3-D grid-based resilient pseudorange error prediction for adaptive GNSS/IMU integrated navigation in urban areas. *IEEE Internet Things J*. 2025;12(12):19264–79. doi:10.1109/jiot.2025.3533077.
16. Liang Z, Liu L, Yang J, Yu J, Zhang S, Zhang X, et al. AUV cooperative localization method based on M–H resampling and state covariance matrix optimized particle filter. *IEEE Trans Instrum Meas*. 2025;74(10):1–14. doi:10.1109/tim.2025.3609326.
17. Chen J, Fang Z, Gao M, Xu Y. Robust cooperative GNSS positioning under satellite-constrained conditions: accounting for physical correlations in stochastic modeling. *Measurement*. 2026;287(9):122499. doi:10.1016/j.measurement.2026.122499.
18. Yin L, Wang L, Cai Z, Lu S, Wang R, AlSanad A, et al. DPAL-BERT: a faster and lighter question answering model. *Comput Model Eng Sci*. 2024;141(1):771–86. doi:10.32604/cmesci.2024.052622.
19. Wang T, Liu M, Li H, Zhao L, Jiang C, Xia C, et al. ArchSentry: enhanced Android malware detection via hierarchical semantic extraction. *IEEE Trans Netw Serv Manag*. 2025;22(3):2822–37. doi:10.1109/tnsm.2025.3559255.
20. Huang Z, Li T, Li Y. SMAUG: semantic-enhanced mobile app usage data generation with LLM. *IEEE Trans Serv Comput*. 2026;19(1):671–84. doi:10.1109/tsc.2025.3648444.

21. Wang K, Hu Y, Liu H, Shao J, Nie L. Cross-modal representation shift refinement for point-supervised video moment retrieval. *ACM Trans Inf Syst.* 2026;44(3):54. doi:10.1145/3786606.
22. Jin L, Zhang Y, Li Y, Wang S, Yang HH, Wu J, et al. MoE2: optimizing collaborative inference for edge large language models. *IEEE Trans Netw.* 2026;34(120):4637–51. doi:10.1109/ton.2026.3677371.
23. Long X, Chen J, Yang L, Huang H. An emergency scheduling method based on AutoML for space maneuver objective tracking. *Expert Syst Appl.* 2026;298(5):129759. doi:10.1016/j.eswa.2025.129759.
24. Wan A, Zhu Z, Al-Bukhaiti K, Cheng X, Jiang J, Ji X, et al. Real-time aero-engine fault diagnosis using 5G edge computing and deep learning. *Measurement.* 2026;257(2):118784. doi:10.1016/j.measurement.2025.118784.
25. Jiang H, Guo J, Xiao Z, Yang K, Li T, Li B. Collaborative perception and computing offloading in 6G air-ground integrated networks. *IEEE J Sel Areas Commun.* 2026;44:2821–38. doi:10.1109/jsac.2025.3645021.
26. Lei T, Mao K, Zhu Q, Zhang R, Ma Z, Hua B, et al. Ultra-wideband channel modeling and characterization under low-altitude scenarios. *Chin J Aeronaut.* 2026;39(5):104077. doi:10.1016/j.cja.2026.104077.
27. Liu Q, Chen P, Lin K, Zhao K, Ding J, Li YJ. Sample-efficient backtrack temporal difference deep reinforcement learning. *Knowl Based Syst.* 2025;330(3):114613. doi:10.1016/j.knosys.2025.114613.
28. Yuan W, Chen J, Chen S, Feng D, Hu Z, Li P, et al. Transformer in reinforcement learning for decision-making: a survey. *Front Inf Technol Electron Eng.* 2024;25(6):763–90. doi:10.1631/fitee.2300548.
29. Liu Q, Song Z, Liang Y, Xie Z, Zhang S, Zhang J, et al. CoRLHF: reinforcement learning from human feedback with cooperative policy-reward optimization for LLMs. *Expert Syst Appl.* 2026;301:130113. doi:10.1016/j.eswa.2025.130113.
30. Xu Z, Wang K, Mu C, Qiu T. Safety-critical path planning for obstacle avoidance based on reinforcement learning and control barrier functions. *IEEE Internet Things J.* 2025;12(23):51410–21. doi:10.1109/jiot.2025.3614857.
31. Zhang J, Chen P, Yang X, Wu H, Li W. Distance-sensitive task allocation for mobile crowdsensing: a reinforcement learning-based mechanism. *IEEE Trans Netw Sci Eng.* 2026;13:9500–17. doi:10.1109/tnse.2026.3691000.
32. Peng X, Qin C, Zhang X, Xu L, Zhang X, Jiang L. DRUDM-CFG: a fairness-aware multi-agent DRL algorithm for AMEC-assisted task offloading in post-disaster scenarios. *IEEE Trans Mob Comput.* 2026;25(9):14135–50. doi:10.1109/tmc.2026.3676811.
33. Tang X, Chen Q, Weng W, Jin C, Liu Z, Wang J, et al. Task assignment and exploration optimization for low-altitude UAV rescue via generative AI-enhanced multi-agent reinforcement learning. *IEEE Trans Mob Comput.* 2026;25(1):627–43. doi:10.1109/tmc.2025.3594188.
34. Qi N, Jiang H, Xiao S, Wu D, Zhou F, Li C, et al. Robust and energy-efficient multi-UAV trajectory planning for data collection: a game-theoretic and deep reinforcement learning approach. *IEEE Trans Green Commun Netw.* 2026;10:3086–101. doi:10.1109/tgcn.2026.3695901.
35. Liu X, Fischione C. Coordinated beamforming for multi-cell ISAC using graph neural networks. *IEEE Trans Wirel Commun.* 2026;25:5876–89. doi:10.1109/twc.2025.3621437.
36. Dai W, Cui T, Zhang T, Chen B. VFL+: low-coupling vertical federated learning with privileged information paradigm. *IEEE Trans Emerg Top Comput Intell.* 2025;9(5):3533–47. doi:10.1109/tetci.2025.3543769.
37. Li Y, Wang S, Chang T. Differentially private federated stochastic primal-dual learning for Internet of Vehicles. *IEEE Internet Things J.* 2025;12(11):17034–50. doi:10.1109/jiot.2025.3534982.
38. Xiao Z, Chen Y, Li T, Tian S, Min G, Tang Z, et al. Toward secure task offloading in vehicular edge computing networks. *IEEE Trans Dependable Secure Comput.* 2026. doi:10.1109/tdsc.2026.3697937.
39. Zhao N, Ye Z, Pei Y, Liang YC, Niyato D. Multi-agent deep reinforcement learning for task offloading in UAV-assisted mobile edge computing. *IEEE Trans Wirel Commun.* 2022;21(9):6949–60. doi:10.1109/twc.2022.3153316.
40. Zhang H, Zhao H, Liu R, Gao X, Xu S. Dynamic user association and computation offloading in satellite edge computing networks via deep reinforcement learning. *IEEE Trans Green Commun Netw.* 2024;8(4):1888–901. doi:10.1109/tgcn.2024.3357813.