Computers, Materials & Continua

Tech Science Press

https://doi.org/10.32604/cmc.2026.081224

ARTICLE

Cascading Failure Dynamics and Edge-Intelligent Defense in Space-Air-Ground Integrated Networks for Internet of Things

Peiying Zhang[1,2], Yihong Yu[1,2], Lizhuang Tan[3,4,*], Shuqing He[5], Jian Wang[6] and Ameer El-Sayed[7]

[1]Qingdao Institute of Software, College of Computer Science and Technology, China University of Petroleum (East China), Qingdao, China
[2]Shandong Key Laboratory of Intelligent Oil & Gas Industrial Software, Qingdao, China
[3]Key Laboratory of Computing Power Network and Information Security, Ministry of Education, Shandong Computer Science Center (National Supercomputer Center in Jinan), Qilu University of Technology (Shandong Academy of Sciences), Jinan, China
[4]Shandong Provincial Key Laboratory of Computing Power Internet and Service Computing, Shandong Fundamental Research Center for Computer Science, Jinan, China
[5]School of Information Science and Engineering, Linyi University, Linyi, China
[6]College of Science, China University of Petroleum (East China), Qingdao, China
[7]Department of Information Technology, Faculty of Computers and Informatics, Zagazig University, Zagazig, Egypt

*Corresponding Author: Lizhuang Tan. Email: tanlzh@sdas.org

Received: 26 February 2026; Accepted: 21 April 2026; Published: 15 June 2026

ABSTRACT: As a core information infrastructure in the 6G era, the Space-Air-Ground Integrated Network (SAGIN) integrates space-based, air-based, and ground-based network resources to achieve seamless communication across all domains. However, its characteristics such as heterogeneous node coupling and dynamic topology changes make it prone to cascading failures, severely threatening critical business continuity in Internet of Things (IoT) applications spanning smart cities, healthcare, transportation, and industrial automation. This paper conducts systematic research addressing challenges including modeling difficulties in SAGIN cascading failure propagation, insufficient coordination of defense strategies, and poor resource adaptability. First, a multi-factor coupled dynamic model of cascading failure propagation is established to quantify the synergistic effects of node heterogeneity, link dynamics, and load redistribution. Second, a closed-loop collaborative defense system integrating "early warning-isolation-self-healing" is designed. The system incorporates a lightweight greedy-based self-healing algorithm and uses multi-criteria decision-making (Analytic Hierarchy Process) for resource optimization. These approaches ensure real-time performance and energy efficiency on resource-constrained edge nodes. Third, a joint simulation platform combining NS-3 and MATLAB is built to validate the model and strategies across diverse IoT application scenarios. Experimental results show that the proposed propagation model maintains prediction error within 10%, the defense strategies increase failure recovery rates to 85%–90%, reduce communication interruption duration by over 60%, and lower resource overhead by 20%–25%, providing theoretical support and technical guarantees for stable SAGIN operation in security and resiliency-critical environments.

KEYWORDS: Space-Air-Ground Integrated Network; cascading failure; defense strategy; edge intelligence; Internet of Things; resource allocation and optimization; real-time and energy efficiency; security and resiliency; network reliability

1 Introduction

The Space-Air-Ground Integrated Network (SAGIN) has emerged as a foundational infrastructure for 6th-Generation (6G) mobile communications, enabling global coverage and ubiquitous connectivity [1–3].

By integrating satellite, aerial, and terrestrial networks, SAGIN delivers intelligent services with extensive coverage, high bandwidth, and ultra-low latency [4]. Its convergence with IoT enables smart cities, healthcare, and industrial automation, where service disruptions could lead to catastrophic consequences [5,6]. However, SAGIN's inherent heterogeneity, dynamic topology, and resource constraints introduce significant cascading failure risks that existing reliability research has not adequately addressed—a gap this paper aims to fill.

Analysis of existing literature reveals four critical limitations that hinder effective analysis and defense:

(1) Limited adaptability to structural heterogeneity. Conventional cascading failure models rely on static, isomorphic topology assumptions derived from ground network studies. They fail to capture how differences in physical attributes, resource capacities, and load thresholds across space, air, and ground nodes shape fault propagation paths in SAGIN.

(2) Inadequate modeling of dynamic topology evolution. Frequent reconfigurations from satellite motion, aerial platform repositioning, and mobile terrestrial users render static frameworks ineffective. While dynamic models exist for single-layer networks [7,8], they do not extend to multi-layer SAGIN scenarios.

(3) Absence of coordinated defense mechanisms. Current mitigation strategies target isolated faults or adopt fragmented countermeasures [9]. They lack an integrated framework for early warning, adaptive isolation, and self-healing. Recent satellite network defenses focus on single-layer protection without cross-layer coordination.

(4) Neglect of onboard resource constraints. Satellites and aerial platforms operate under strict resource limits. Most intelligent defense approaches assume unlimited computational resources, leading to excessive complexity and poor deployability.

To address these limitations, this paper proposes a resource-aware adaptive defense paradigm. Building upon foundational load-capacity coupling models [10,11] and extending them to heterogeneous SAGIN environments, this work makes four contributions:

(1) A multi-factor coupled dynamic cascading failure propagation model. Unlike static models, our model defines node load as a composite of computational, transmission, and energy demands, and incorporates link-type dependent propagation probabilities to capture SAGIN-specific heterogeneity.

(2) A hierarchically collaborative closed-loop defense system. Integrating early warning, adaptive isolation, and lightweight self-healing, it formulates resource allocation as a linear programming problem solved via greedy approximation, ensuring deployability on resource-constrained platforms—contrasting with computationally intensive approaches.

(3) A high-fidelity co-simulation platform combining NS-3 and MATLAB to validate the proposed model and strategies.

(4) Systematic experimental validation benchmarking six defense strategies across typical failure scenarios, providing quantitative comparisons and deployment guidelines.

The rest of the paper is structured as follows: Section 2 reviews related work; Section 3 presents the architecture and cascading failure model; Section 4 details the defense mechanism; Section 5 discusses experiments; Section 6 concludes.

2 Related Work

2.1 Cascading Failure Models for Heterogeneous Networks

Research on cascading failure mechanisms has long been a focal point in complex network reliability. Percolation models rely on network topology to analyze connectivity changes but neglect dynamic factors such as node load. Avalanche models simulate rapid local-to-global failure spread through chain-reaction rules but inadequately account for network heterogeneity. Load-capacity coupling models assume nodes have finite capacity; when load exceeds capacity, nodes fail and redistribute load to neighbors. The seminal work by Motter and Lai [12] established the foundation for this class of models, with subsequent extensions incorporating tunable parameters for load redistribution.

However, these models assume static, homogeneous nodes, failing to capture SAGIN's cross-layer heterogeneity. Recent studies address this but remain single-layer focused: Liu et al. [13] analyzed satellite failures; Zhang and Du [14] developed a low Earth orbit (LEO) satellite dynamic model; Wang et al. [15] investigated Unmanned Aerial Vehicle (UAV) failures due to energy limits. These single-layer approaches cannot capture how failures propagate across space, air, and ground domains—a gap our multi-factor coupled model addresses by integrating computational, transmission, and energy loads across all layers with link-type dependent propagation probabilities.

2.2 Dynamic Topology Evolution Modeling

Frequent reconfigurations from satellite motion and user mobility challenge cascading failure analysis. Valdez et al. [16] reviewed cascading failures in dynamic networks; Xiao et al. [17] emphasized need for time-varying SAGIN models.

Zhang et al. [18] proposed a LEO satellite dynamic model incorporating orbital mechanics. Xu et al. [19] examined cascading effects in UAV swarms. However, these address dynamics within single layers only. While advance single-layer dynamics, they do not extend to multi-layer SAGIN where space, air, and ground dynamics interact and amplify cascading effects—motivating our time-varying topology modeling across all layers.

2.3 Coordinated Defense Mechanisms

In recent years, intelligent methods have been increasingly applied to network fault management. Cui et al. [20] employed multi-agent reinforcement learning for resource allocation in UAV networks, achieving dynamic optimization of communication and computation resources. Chen et al. [21] proposed a node collaborative defense mechanism based on game theory, formulating fault propagation as a non-cooperative game between attackers and defenders. However, these target isolated measures without integrated early warning, isolation, and self-healing. The fragmentation in existing defenses limits system-wide resilience—directly motivating our closed-loop "early warning-isolation-self-healing" framework that coordinates prevention, suppression, and recovery phases.

2.4 Resource-Constrained Defense Optimization

Satellites operate under strict resource limits, often overlooked. Dakic et al. [22] analyzed vehicle-to-everything (V2X) reliability. Fang et al. [23] proposed DRL-driven resource allocation. Al-Zahrani et al. [24] reviewed 5th-Generation (5G) optimization. However, these prioritize performance over resilience, and AI-based methods remain computationally intensive. The impracticality of complex approaches on resource-constrained platforms motivates our lightweight linear programming formulation solved via greedy approximation, ensuring deployability on satellites and aerial platforms.

As summarized in Table 1, existing research lacks a unified approach to SAGIN's heterogeneity, dynamism, and resource constraints. These gaps directly motivate our four contributions: a multi-factor cascading model addressing heterogeneity, a time-varying topology framework addressing dynamics, a closed-loop defense addressing coordination, and a lightweight optimization addressing resource constraints—all building upon and extending prior work to the SAGIN context.

Table 1: Comparison of related work.

Category	Representative Works	Key Limitations	This Paper
Heterogeneous Modeling	Liu et al. [13], Zhang and Du [14], Wang et al. [15]	Single-layer focus, no cross-layer integration	Multi-factor coupled model with cross-layer heterogeneity
Dynamic Topology	Valdez et al. [16], Zhang et al. [18], Xu et al. [19]	Dynamics within single layers only	Time-varying topology modeling across all layers
Coordinated Defense	Cui et al. [20], Chen et al. [21], Zhang et al. [18]	Fragmented measures, no closed-loop integration	"Early warning-isolation-self-healing" closed-loop defense
Resource Constraints	Fang et al. [23], Al-Zahrani et al. [24]	Performance-focused, computationally intensive	Lightweight linear programming with greedy approximation

3 Architecture and Topology

As show in the Fig. 1, the architecture of the SAGIN can be divided into three core levels, each level collaboratively achieving resource scheduling and service delivery.

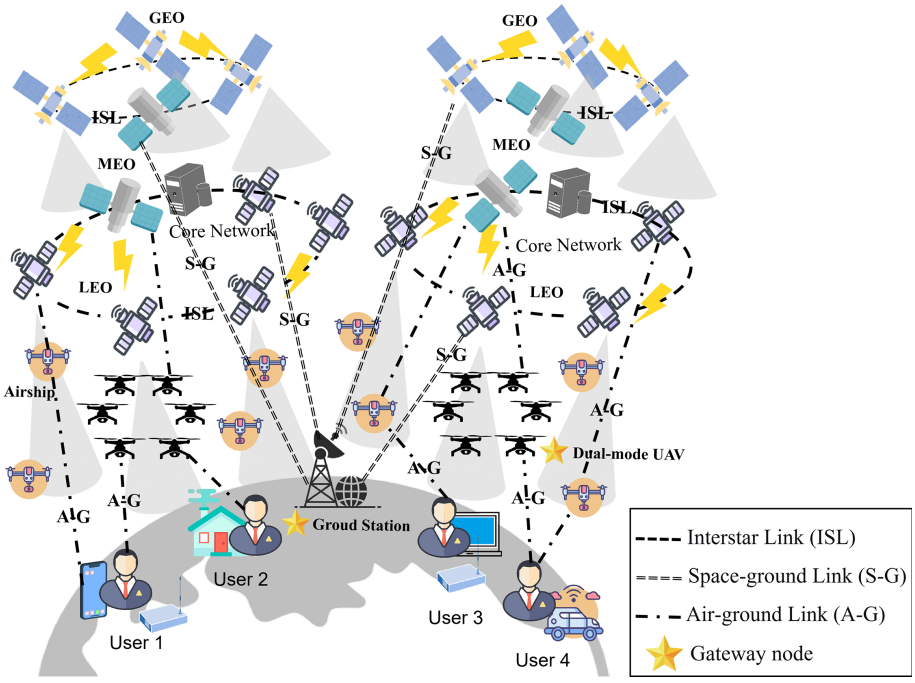

Figure 1: SAGIN hierarchical architecture with heterogeneous node types and link connections.

3.1 SAGIN Resilience Modeling

The space layer comprises LEO, medium Earth orbit (MEO), and high Earth orbit (HEO) satellites delivering wide-area coverage and long-distance communications. LEO satellites offer low latency and high bandwidth but require constellation networking for global coverage; MEO satellites balance coverage and latency for regional needs; HEO satellites provide extensive regional coverage with higher latency and limited bandwidth. Satellite nodes face challenges including frequent link handovers and power constraints due to orbital mechanics.

The aerial layer consists of drones and high-altitude airships acting as relay hubs between space and ground layers, enhancing regional communication and supporting edge computing. Drones offer flexible deployment and rapid repositioning, while airships provide extended endurance and wide coverage. Both face reliability fluctuations due to weather and energy constraints. Inter-layer communication is enabled by gateway nodes such as satellite ground stations and multi-mode UAVs, explicitly modeled in our simulation. The ground layer comprises 5G base stations, core network equipment, and end-user devices. Base stations provide ample computing power, stable links, and low latency; core network equipment manages resource allocation and data routing as the operational hub [25], end-user devices—IoT devices, smartphones, and vehicles—act as service initiators and recipients. Due to limited ground coverage, wide-area coordination depends on the space-air layer.

3.2 Topological Dynamism and Heterogeneity

Topological Dynamism: Satellite motion, aerial platform repositioning, and mobile ground users cause frequent link handovers and topology reconfigurations, creating a highly dynamic SAGIN topology characterized by time-varying connectivity [26].

Topological Heterogeneity: Space, air, and ground nodes differ significantly in physical characteristics, resource capabilities [27], and communication protocols. Satellite links suffer from free-space loss and rain fade, causing fluctuations in delay, bandwidth, and loss rate; air links are affected by weather and platform vibrations; ground links operate in stable environments with ample bandwidth and low latency. These disparities across link types increase the complexity of network coordination and fault propagation analysis. The topology is a hybrid of three-layer hierarchical architecture with intra-layer mesh connectivity.

3.3 Improved Cascading Fault Model Based on Load Balancing

Building upon foundational load-capacity coupling models, we extend the classical framework to accommodate SAGIN heterogeneity. The classical model defines node failure when load exceeds capacity, triggering load redistribution to neighbors. However, this model assumes static topology and homogeneous nodes, limiting its applicability to SAGIN. Our improved model incorporates multi-dimensional load definitions, link-type dependent propagation probabilities, and dynamic load evolution to capture SAGIN-specific characteristics. Node load is defined as the weighted sum of computational task load, data transmission load, and energy consumption load, calculated as

$$L_i = w_1 L_{i,comp} + w_2 L_{i,trans} + w_3 L_{i,energy} \tag{1}$$

based on SAGIN node resource characteristics. The coefficients w_1, w_2, and w_3 are determined by node type and application scenarios using the Analytic Hierarchy Process (AHP), a multi-criteria decision-making method that systematically combines expert knowledge and network operational objectives. For instance, satellite nodes, with limited power budgets, are assigned higher energy consumption load coefficients ($w_3 > w_1, w_2$), while ground base stations, with stable power supply, prioritize computational and data

transmission loads ($w_1, w_2 > w_3$). This approach ensures that weight selection is grounded in domain expertise rather than arbitrary assignment. Node capacity, representing the maximum load a node can handle, is defined as the minimum of computational capacity, transmission capacity, and energy capacity,

$$C_i = \min(C_{i,comp}, C_{i,trans}, C_{i,energy}) \tag{2}$$

$C_{i,comp}$ denotes the computational capacity of node i, measured by its maximum supported CPU utilization; $C_{i,trans}$ represents the transmission capacity of node i, quantified by its maximum supported bandwidth occupancy; $C_{i,energy}$ indicates the energy capacity of node i, defined by the load threshold corresponding to the minimum remaining energy ratio.

The initial failure node is selected based on a node failure probability model. The failure probability $P_{fail,i}$ of a node depends on its operational time t, resource consumption rate r_i, and external environmental interference intensity s_i, defined as

$$P_{fail,i} = 1 - e^{-\alpha t - \beta r_i - \gamma s_i} \tag{3}$$

where α, β, and γ are scaling factors. Through random sampling, the initial failure node set is determined according to each node's failure probability. When node i fails, its load L_i migrates to adjacent healthy nodes following these rules: First, adjacent nodes are ranked by link priority coefficient $p_{i,j}$ (which correlates with link bandwidth, transmission delay, and reliability). Then, the load is evenly distributed to adjacent healthy nodes in descending order of priority, with the allocation ratio to node j calculated as

$$p_{i,j} / \sum_{j \in N_i} p_{i,j} \tag{4}$$

If the remaining capacity of adjacent healthy nodes is insufficient to handle the migrated load, the remaining load continues to migrate to the next-level adjacent nodes.

To account for inter-layer link attenuation, the fault propagation probability $P_{i,j}$ is defined as the probability of a node i's fault propagating to node j, given by

$$P_{i,j} = \frac{L_{j,received}}{C_j} \cdot \eta_{i,j} \tag{5}$$

where, $L_{j,\text{received}}$ represents the received migration load at node j, while $\eta_{i,j}$ is the link attenuation coefficient, which varies with the link type. The attenuation coefficients decrease progressively from inter-satellite links to space-ground links, with terrestrial links exhibiting the lowest attenuation.

To describe the dynamic changes of load over time (reflecting business growth or sudden traffic surges), the load dynamic evolution formula is introduced:

$$L_i(t+1) = L_i(t) + \Delta L_{i,\text{external}}(t) + \sum_{j \in N_i} \Delta L_{j \to i}(t) \tag{6}$$

where $DeltaL_{i,\text{external}}$ denotes the externally added load, and $sum_{j \in N_i} \Delta L_{j \to i}$ represents the migrated load from neighboring node j, load transfer during fault propagation.

The recovery process of faulty nodes can be described by a time and resource-based recovery probability model:

$$P_{\text{recover},i}(t) = 1 - \exp(-\lambda_i t) \tag{7}$$

where λ_i is the recovery rate coefficient positively correlated with node resources (e.g., computing power, energy), and t is the fault duration.

Early warning threshold formula provides a quantitative early warning trigger condition for fault prediction based on multi-source information fusion.

$$\text{Alert if: } \hat{L}_i(t+\Delta t) > \theta \cdot C_i \quad \text{or} \quad \hat{Q}_{i,j}(t+\Delta t) < Q_{th} \tag{8}$$

where $\hat{L}_i$ is the predicted load, θ is the safety factor, $\hat{Q}_{i,j}$ is the predicted link quality, and Q_{th} is the quality threshold. A summary of the key variables and their notations is provided in Table 2.

Table 2: Summary of key variables and notations.

Variable	Definition	Physical Meaning	Variable	Definition	Physical Meaning
L_i	Node load	Weighted sum of comp/trans/energy load of node i	$L_{i,comp}$	Computational load	CPU utilization of node i
$L_{i,trans}$	Transmission load	Bandwidth occupancy of node i	$L_{i,energy}$	Energy load	Reciprocal of remaining energy ratio of node i
w_1, w_2, w_3	Load weights	Coefficients for node type/scenario	C_i	Node capacity	Max load node i can handle (min of comp/trans/energy)
$C_{i,comp}$	Computational capacity	Max supported CPU utilization of node i	$C_{i,trans}$	Transmission capacity	Max supported bandwidth occupancy of node i
$C_{i,energy}$	Energy capacity	Load threshold for min remaining energy	$P_{fail,i}$	Node failure probability	Probability node i fails (time/resource/env)
α, β, γ	Scaling factors	Parameters scaling failure probability impact	$p_{i,j}$	Link priority coefficient	Priority of link (i, j) (bandwidth/delay/reliability)
$P_{i,j}$	Fault propagation probability	Probability failure of i propagates to j	$L_{j,recovered}$	Received migration load	Load received by j from failed nodes
$\eta_{i,j}$	Link attenuation coefficient	Attenuation factor for link (i, j) (sat/air/terr)	$\Delta L_{i,external}$	External load addition	Load increase from business/traffic surge
$\Delta L_{j\to i}$	Migrated load	Load transferred from j to i	$P_{recover,i}(t)$	Recovery probability	Probability i recovers within time t
λ_i	Recovery rate coefficient	Correlated with node resources (comp/energy)	$\hat{L}_i$	Predicted load	Load forecast for node i
θ	Safety factor	Threshold coefficient for early warning	$\hat{Q}_{i,j}$	Predicted link quality	Quality forecast for link (i, j)
Q_{th}	Quality threshold	Minimum acceptable link quality			

4 Cascade Fault Defense and Self-Healing Algorithm Design

The proposed defense strategy establishes a prevention-inhibition-recovery framework. Prevention optimizes resource allocation and protects critical nodes to balance load and minimize initial failures. Upon fault detection, rapid detection, load diversion, and link isolation contain faults locally. Recovery restores failed nodes via fast localization, topology reconstruction, and redundant resource scheduling, minimizing service interruption. Resource allocation is optimized throughout to reduce overhead, ensuring engineering feasibility on resource-constrained platforms.

4.1 Layered Cooperative Defense Mechanism

The prevention stage optimizes resource allocation according to the importance of the node, evaluated by means of a composite metric that combines centrality of the between, centrality of the degree and functional dependency. Blockchain-enabled collaborative frameworks have shown promise in securing resource coordination across distributed logistics networks [28], inspiring our trusted resource allocation design. Betweenness centrality reflects routing significance, degree centrality indicates connectivity density, and functional dependency quantifies inter-node functional reliance. The node importance index $I_i = aB_i + bD_i + cF_i$ uses weights a, b, and c determined by the AHP, a multi-criteria decision method that quantifies the relative importance of different centrality metrics without introducing prohibitive computational overhead. Resources are allocated proportionally to these indices. High-importance hub nodes, including high-orbit satellites and core ground base stations, receive priority allocation of bandwidth, computing power, and energy to enhance capacity and fault tolerance, alongside dedicated redundant backups such as standby satellites and redundant links. Regular nodes have their resource ratios dynamically adjusted based on service demand and current load to maintain network load balancing.

The suppression stage implements proactive load diversion triggered by fault precursor detection. A multi-source fusion prediction model collects real-time node status, link performance, and topology evolution data. Long Short-Term Memory time-series analysis identifies precursor features, predicting potential faults and locating at-risk nodes and links ahead of occurrence. Upon detecting precursors, dynamic load diversion is activated. Partial loads from at-risk nodes are preemptively migrated to adjacent healthy nodes, with paths selected based on node capacity, link quality, and service priority. Idle unmanned aerial vehicles are deployed to fault-prone areas to establish temporary links, facilitating load offloading and coverage maintenance. A link priority mechanism prioritizes core services for bandwidth allocation; non-core services may be temporarily suspended to contain fault propagation and protect critical network functionality.

4.2 Lightweight Self-Healing Algorithm Implementation

The system achieves rapid fault identification via link quality monitoring and node heartbeat feedback. Link reliability is measured by transmission latency, bandwidth, and packet loss rate. Node failures are detected through periodic heartbeat exchanges with multi-node collaborative verification. Algorithm 1 formalizes this process.

Algorithm 1: Fault detection algorithm

Input: Let N represent the cardinality of the node set, and E denote the link set, T_{out} represent the timeout threshold;
Output: The faulty node set F_n and faulty link set F_l are output;

```
F_n ← ∅, F_l ← ∅
for i ∈ N do
    Send heartbeat signal to node i
    if no response received within T_out then
        F_n ← F_n ∪ {i}
        for (i, j) ∈ E do
            F_l ← F_l ∪ {(i, j)}
        end for
    end if
end for
```

(Continued)

Algorithm 1 (continued)

for $(i, j) \in E$ **do**
  Measure delay d_{ij} and packet loss rate p_{ij}
  if $d_{ij} > d_{\text{threshold}}$ OR $p_{ij} > p_{\text{threshold}}$ **then**
    $F_l \leftarrow F_l \cup \{(i, j)\}$
  end if
end for
return F_n, F_l

Upon fault detection, the system reconstructs topology through a greedy strategy, prioritizing nearby healthy nodes with sufficient residual capacity. For satellite layers, pre-planned reconnection procedures based on orbital predictions are applied. Core link failures trigger redundant link activation with dynamic bandwidth adjustment. Algorithm 2 details the reconstruction process.

Algorithm 2: Path reconstruction algorithm

Input: Faulty node set F_n, Healthy node set H, Affected service flows S,
Residual capacity $C_{\text{res}}[j]$ for each $j \in H$, Load requirement load_s for each s
Output: Reconstructed routing paths

Sort H by $C_{\text{res}}[j]$ in descending order
for $f \in F_n$ **do**
  Identify service flows $S_f \subseteq S$ handled by node f
  for $s \in S_f$ **do**
    for $j \in$ sorted H **do**
      if $C_{\text{res}}[j] \geq \text{load}_s$ **then**
        Reroute flow s through node j
        $C_{\text{res}}[j] \leftarrow C_{\text{res}}[j] - \text{load}_s$
        Mark flow s as successfully rerouted
        Break
      end if
    end for
  end for
end if
for $(i, j) \in F_l$ (critical links) **do**
  Activate highest-priority redundant backup link
  Adjust bandwidth to meet real-time demands
end for
return reconstructed routing paths

This reconstruction problem can be viewed as a linear program: maximize successfully rerouted flows subject to node capacity constraints. The greedy algorithm provides an approximate solution with significantly lower overhead than exact solvers.

Algorithm 1 has time complexity $O(N + E)$; Algorithm 2 achieves $O(N \log N)$. Both have space complexity $O(N + E)$, meeting real-time requirements on resource-constrained platforms.

5 Experimental Result

This section validates the proposed model and defense strategies through four experiments: cascading failure propagation analysis examining topology density and load threshold impacts, model prediction accuracy comparing our multi-factor model with the Cascading Failure Model (CASCADE), defense performance evaluation quantifying advantages of our closed-loop strategy over traditional approaches, and strategy differentiation analyzing performance differences among six specific defense schemes. The experimental environment settings, including simulation tools, network scale, node/link parameters, and fault scenarios, are detailed first. These six defense strategies are S1 (Proposed), a closed-loop approach integrating early warning, isolation, and self-healing; S2, static redundancy backup; S3, static load balancing; S4, early warning only; S5, adaptive isolation only; and S6, rule-based self-healing. This comparative assessment isolates the contribution of each individual component and substantiates the superiority of their integrated operation in the proposed S1 strategy. The following presents the detailed experimental process and results.

5.1 Experimental Design and Results Verification

This experiment employs NS-3 as the primary simulation tool for network-level modeling, supporting heterogeneous node configuration, mobility patterns, and link-level performance simulation. MATLAB is integrated for offline data processing, model solving, and visualization. The co-simulation operates iteratively: NS-3 generates trace files containing link delays, packet loss rates, and throughput at each time step; MATLAB reads these traces, computes node loads and failure probabilities using Eqs. (1)–(8), and determines defense actions; updated configuration parameters are then fed back to NS-3 for the next simulation cycle. Table 3 shows the network scale settings. This modular approach requires no custom API-level interfacing, using standard file I/O for data exchange, ensuring ease of replicability. The combination of NS-3 and MATLAB ensures both authenticity and flexibility while meeting the experimental verification requirements.

Table 3: Network topology scale configuration (based on typical SAGIN testbed deployments).

Network Layer	Node Type	Node Quantity
Satellite Layer	Low Earth Orbit (LEO) Satellite	10
	Medium Earth Orbit (MEO) Satellite	5
	Geostationary Earth Orbit (GEO) Satellite	2
Air Layer	UAV	20
	High-Altitude Airship	5
Ground Layer	Ground Base Station	30
	Terminal User Node	100

The node attribute parameter settings are shown in Table 4.

The link parameter settings are shown in Table 5.

All experiments are based on three defined failure scenarios: LEO satellite node failure (simulated by disabling network interface), ground base station overload (CPU utilization exceeds threshold), and UAV swarm disconnection (introducing high path loss for air-ground links). These are implemented in NS-3 through custom modules that modify node status or environmental conditions at specified times. We conducted validation experiments on cascading failure propagation characteristics to analyze the

relationship between topology density, load threshold, and both the propagation speed and impact scope of cascading failures. Different topology densities and load thresholds were configured, and multiple simulation experiments were performed across three typical failure scenarios. The time required for failure propagation to reach a stable state and the proportion of failed nodes were recorded.

Table 4: Node attribute parameter configuration (referencing operational data from Starlink and UAV networks [15]).

Node Type	CPU Utilization	Bandwidth Utilization	Minimum Remaining Energy Ratio
LEO Satellite	80%	90%	20%
MEO Satellite	70%	85%	25%
GEO Satellite	60%	80%	30%
UAV	50%	70%	15%
High-Altitude Airship	60%	75%	20%
Ground Base Station	90%	95%	Unrestricted (Stable Power Supply)

Table 5: Link parameter configuration (derived from 3GPP NTN standards and prior simulation studies [14]).

Link Type	Transmission Delay	Bandwidth	Packet Loss Rate
Inter-Satellite Link	50–100 ms	100–500 Mbps	0.1%–0.5%
Satellite-Ground Link	100–200 ms	50–200 Mbps	0.5%–1.0%
Air-Ground Link	50–100 ms	30–100 Mbps	1.0%–2.0%
Ground Link	1–10 ms	100–1000 Mbps	0.01%–0.1%

For the density analysis in Figs. 2 and 3, we scaled this baseline by factors of 0.5, 1.0, and 1.5 to generate low (density = 5), medium (density = 10), and high (density = 15) densities. This variable-controlling approach allows us to isolate the effect of network scale on fault propagation while preserving SAGIN's hierarchical structure.

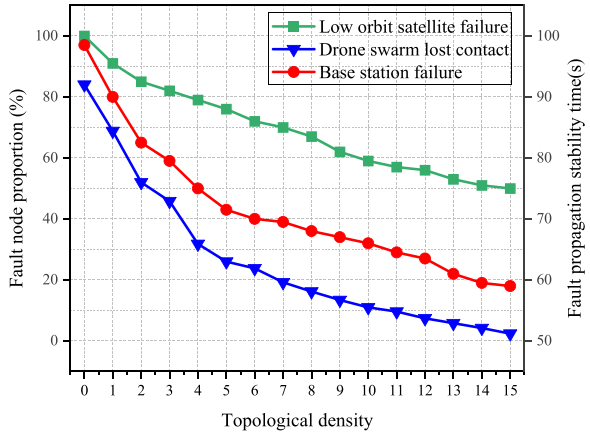

Figure 2: Impact of topology density on fault propagation. Density values (5, 10, 15) correspond to 0.5×, 1.0×, and 1.5× scaling of baseline topology.

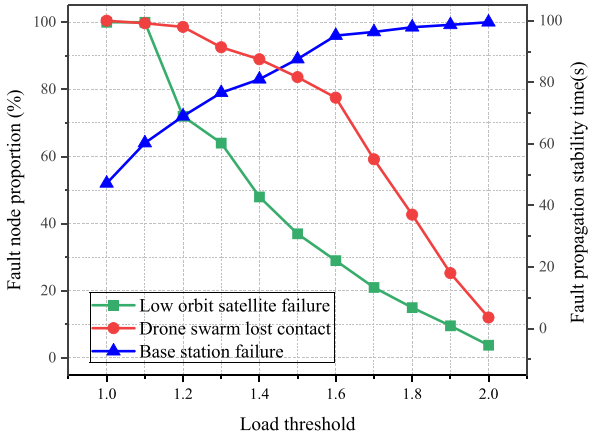

Figure 3: Impact of load threshold on fault propagation. Load thresholds delay propagation and reduce failure scope under fixed topology density.

As illustrated in Fig. 2, under identical load thresholds, higher topology density accelerates failure propagation and expands its scope. For ground base station failure at load threshold 1.6, increasing density from 5 to 10 to 15 reduces stabilization time from 18 to 12 to 8 time units, while the failed node proportion rises from 35% to 55% to 70%. Tighter connectivity creates more propagation paths and accelerates load migration, facilitating rapid failure diffusion.

Conversely, as shown in Fig. 3, under fixed topology density, higher load thresholds slow failure propagation and reduce its scope. For LEO satellite failure at density 10, raising the threshold from 1.2 to 1.6 to 2.0 increases stabilization time from 10 to 15 to 22 time units, while the failed node proportion drops from 65% to 45% to 25%. Higher thresholds enhance node bearing capacity, enabling nodes to absorb more migrated load and reduce secondary failures.

We validated our model against actual failure scenarios using LEO satellite failure. Our model achieved prediction errors of 5%–8% for stabilization time and 4%–7% for failed node proportion, compared to 20%–25% and 15%–20% for CASCADE. For an actual stabilization time of 15 time units, our model predicted 14–16 units vs. CASCADE's 11–19 units; for an actual failed node proportion of 45%, our model predicted 42%–48% vs. CASCADE's 36%–54%.

We further conducted comparative analyses between this model and CASCADE as well as Zhang and Du LEO-specific model [14], covering three scenarios (LEO satellites, ground base stations, and drone swarms). Fig. 4 presents a quantitative comparison of prediction errors among the three models. Building on the load-capacity framework, our model incorporates multi-dimensional loads and cross-layer propagation, capturing interdependencies across space, air, and ground. For LEO satellite failure, our model achieves errors of 5%–8% (stabilization time) and 4%–7% (failed node proportion), outperforming CASCADE (20%–25%, 15%–20%) and Zhang and Du [14] (12%–15%, 10%–12%). Similar improvements across all scenarios demonstrate generalizability.

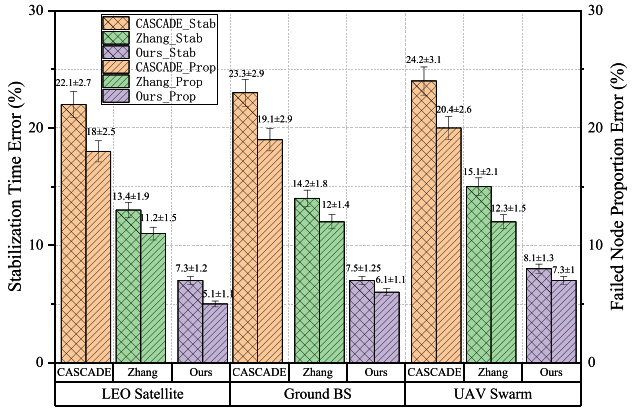

Figure 4: Benchmark comparison of prediction errors across three failure scenarios.

We evaluated the performance of the Early Warning-Isolation-Self-Healing closed-loop collaborative defense strategy from three dimensions: failure recovery rate, communication interruption duration and resource overhead. Two experimental groups were set up for comparison, one adopting the proposed defense strategy and the other using traditional defense strategies of static redundancy backup and static load balancing. Simulation experiments were carried out across three typical failure scenarios, with all relevant performance metrics recorded for analysis.

As shown in Figs. 5 and 6, our strategy achieved failure recovery rates of 85%–90%, compared to 60%–65% for traditional strategies. For example, in the UAV cluster disconnection scenario, our strategy

recovered 88% of failed nodes and links, whereas the traditional strategy recovered only 62%. Furthermore, as shown in Fig. 7, our strategy achieved average communication interruption durations of 5–8 time units, significantly outperforming the 15–20 time units observed with traditional strategies. In telemedicine applications, for instance, our strategy constrained communication interruptions to within 6 time units, satisfying the stringent latency requirements for remote surgical guidance, whereas traditional strategies' prolonged interruptions risked compromising healthcare service continuity. Regarding resource overhead, our strategy reduced bandwidth occupation and energy consumption by 20%–25% compared to traditional approaches. Specifically, in satellite node failure scenarios, our strategy decreased satellite energy consumption by 22%, effectively enhancing satellite node endurance. The resource utilization rate are shown in Fig. 8.

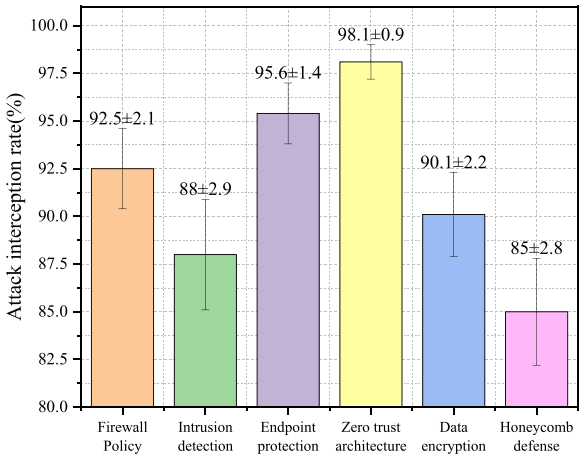

Figure 5: Failure recovery rate under three SAGIN failure scenarios.

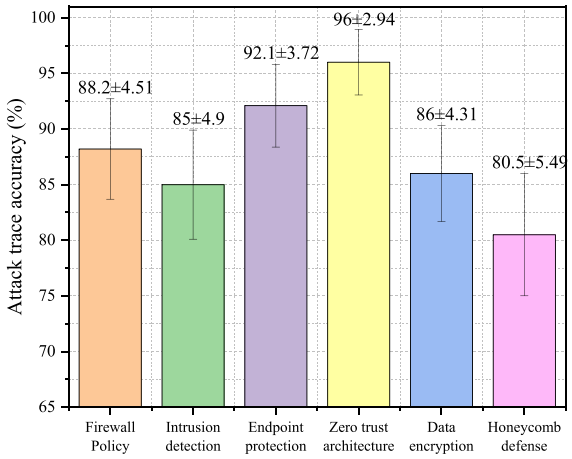

Figure 6: Defense success rate across LEO, ground, and UAV failure scenarios.

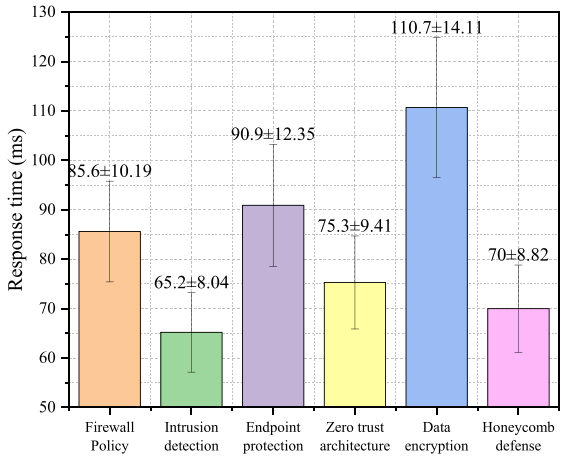

Figure 7: Communication interruption duration in three failure scenarios.

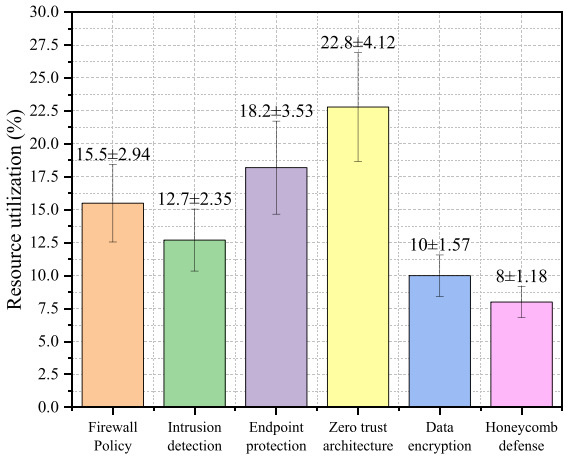

Figure 8: Resource utilization efficiency under different failure scenarios.

These experimental results demonstrate that our designed defense and self-healing algorithms significantly outperform traditional strategies in terms of failure recovery capability, communication continuity assurance, and resource utilization efficiency, substantiating their marked performance advantages.

6 Conclusion

This study investigates cascading fault propagation mechanisms and defense strategies in integrated space-air-ground networks. We innovatively developed a multi-factor coupled dynamic model that breaks through the static isomorphism assumption of traditional models by incorporating key factors like node heterogeneity and dynamic load shifting. Experimental validation shows the model reduces prediction errors for fault propagation stability time and faulty node proportion to 5%–8% and 4%–7%, respectively, significantly outperforming the classic CASCADE model with substantially improved fitting accuracy. Through multi-scenario simulations, we revealed the regulatory patterns of topology density and load thresholds: increasing topology density under the same load threshold accelerates fault propagation and expands impact scope, while raising load thresholds under the same topology density delays propagation and narrows impact range. The "early warning-isolation-self-healing" closed-loop defense achieves 85%–90% recovery rates while reducing resource consumption by 20%–25%, demonstrating practical deployability on resource-constrained platforms.

To further enhance the reliability and intelligence of integrated space-air-ground networks, future research will adopt a multi-directional approach. Key priorities include: expanding multi-fault concurrent scenarios by optimizing models and defense coordination mechanisms through consideration of initial fault nodes' distribution, types, and temporal differences; calibrating models using operational data to refine parameters and thresholds for practical engineering applications; integrating AI with network control theory to develop adaptive defense strategies and multi-agent coordination mechanisms, thereby improving adaptability to dynamic environments and complex fault scenarios. Additionally, research outcomes will be extended to emerging fields like space internet and 6G communications, optimizing strategy designs to drive continuous advancements in reliability theory and technology.

Acknowledgement: None.

Funding Statement: This work is supported by the National Natural Science Foundation of China under Grants 62471493 and 62402257, and partially supported by the Natural Science Foundation of Shandong Province under Grants ZR2023LZH017, ZR2024MF066, and 2023QF025, and partially supported by the Open Foundation of Key Laboratory of Computing Power Network and Information Security, Ministry of Education, Qilu University of Technology (Shandong Academy of Sciences) under Grant 2023ZD010.

Author Contributions: The authors confirm their contribution to the paper as follows: Conceptualization and Design: Peiying Zhang, Yihong Yu; Methodology: Peiying Zhang; Software: Peiying Zhang, Lizhuang Tan, Shuqing He, Jian Wang, Ameer El-Sayed; Investigation: Yihong Yu; Data Curation: Yihong Yu; Funding Acquisition: Peiying Zhang; Project Administration: Peiying Zhang, Ameer El-Sayed, Lizhuang Tan; Writing—Original Draft: Yihong Yu, Peiying Zhang; Writing—Review & Editing: Yihong Yu, Shuqing He; Supervision: Peiying Zhang, Lizhuang Tan, Jian Wang. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The general created dataset is available upon request.

Ethics Approval: This study did not involve any human or animal subjects, and therefore, ethical approval was not required.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Cui H, Zhang J, Geng Y, Xiao Z, Sun T, Zhang N, et al. Space-air-ground integrated network (SAGIN) for 6G: requirements, architecture and challenges. China Commun. 2022;19(2):90–108.
2. Haider ZA, Ullah I, Shareha AA, Nasimov R, Memon SA. Artificial Intelligence (AI)-Enabled Unmanned Aerial Vehicle (UAV) systems for optimizing user connectivity in sixth-generation (6G) ubiquitous networks. Comput Mater Contin. 2026;86(1):1–16. doi:10.32604/cmc.2025.071042.
3. Cheng N, He JC, Yin ZS, Zhou CH, Wu HQ, Lyu F, et al. 6G service-oriented space-air-ground integrated network: a survey. Chin J Aeronaut. 2022;35(9):1–18. doi:10.1016/j.cja.2021.12.013.
4. Zhang Y, Wang X, Gang Y, Wang J, Wu S, Zhang P, et al. 6G SAGIN information transmission model. IEEE Commun Magaz. 2025;63(6):98–105. doi:10.1109/mcom.001.2400351.
5. Jia Z, Jin F, Xie J, He Y. Recurrent MAPPO for joint UAV trajectory and traffic offloading in space-air-ground integrated networks. Comput Mater Contin. 2026;86(1):1–15. doi:10.32604/cmc.2025.069128.
6. He S, Cheng B, Wang H, Huang Y, Chen J. Proactive personalized services through fog-cloud computing in large-scale IoT-based healthcare application. China Commun. 2017;14(11):1–16. doi:10.1109/cc.2017.8233646.
7. Zhao G, Kang Z, Huang Y, Wu S. A routing optimization method for LEO satellite networks with stochastic link failure. Aerospace. 2022;9(6):322. doi:10.3390/aerospace9060322.
8. Zhang L, Du Y. Cascading failure model and resilience enhancement scheme of space information networks. Reliab Eng Syst Safety. 2023;237(3):109379. doi:10.1016/j.ress.2023.109379.
9. Xing L. Cascading failures in Internet of Things: review and perspectives on reliability and resilience. IEEE Internet Things J. 2020;8(1):44–64.
10. Guo C, Gong C, Guo J, Wei Z, Han Y, Khan SZ. Software-defined space-air-ground integrated network architecture with the multi-layer satellite backbone network. Comput Mater Contin. 2020;64(1):527–40. doi:10.32604/cmc.2020.09788.
11. Zhang H, Yao X, Xu K, Wu Z, Li W, Lu Y, et al. Trustworthiness evaluation toward 6G support of space-air-ground integrated network. IEEE Wirel Commun. 2025;32(2):34–40. doi:10.1109/mwc.001.2400273.
12. Motter AE, Lai YC. Cascade-based attacks on complex networks. Phys Rev E. 2002;66(6):065102. doi:10.1103/physreve.66.065102.
13. Liu Z, Han J, Wang Y, Li X, Chen S. Performance analysis of routing algorithms in satellite network under node failure scenarios. In: Proceedings of the 2014 IEEE Global Communications Conference; 2014 Dec 8–12; Austin, TX, USA. p. 2838–43.
14. Zhang L, Du Y. A dynamic cascading failure model for LEO satellite networks. IEEE Trans Netw Service Manag. 2023;21(2):1672–89. doi:10.1109/tnsm.2023.3343357.
15. Wang H, Cai Z, Liao C, Li B. Energy-constrained UAV network topology recovery based on graph convolutional networks. In: Advanced intelligent computing technology and applications. Singapor: Springer; 2025. p. 74–85.
16. Valdez LD, Shekhtman L, La Rocca CE, Zhang X, Buldyrev SV, Trunfio PA, et al. Cascading failures in complex networks. J Complex Netw. 2020;8(2):cnaa013. doi:10.1093/comnet/cnaa022.
17. Xiao Y, Ye Z, Wu M, Li H, Xiao M, Alouini MS, et al. Space-air-ground integrated wireless networks for 6G: basics, key technologies and future trends. IEEE J Selected Areas Commun. 2024;42(12):3327–54.
18. Zhang L, Du Y, Li A. Rapid cascading risk assessment and vulnerable satellite identification schemes for LEO satellite networks. Reliab Eng Syst Safety. 2025;256(4):110699. doi:10.1016/j.ress.2024.110699.
19. Xu B, Bai G, Zhang Y, Fang Y, Tao J. Failure analysis of unmanned autonomous swarm considering cascading effects. J Syst Eng Electron. 2022;33(3):759–70. doi:10.23919/jsee.2022.000069.
20. Cui J, Liu Y, Nallanathan A. Multi-agent reinforcement learning-based resource allocation for UAV networks. IEEE Trans Wirel Commun. 2019;19(2):729–43. doi:10.1109/twc.2019.2935201.
21. Chen K, Zhang L, Zhong J. Space-air-ground integrated network (SAGIN) in disaster management: a survey. IEEE Trans Netw Service Manag. 2025;22(5):4021–49. doi:10.1109/tnsm.2025.3580965.
22. Dakić A, Rainer B, Priller P, Nan G, Momić A, Ye X, et al. Wireless V2X communication testbed for connected, cooperative and automated mobility. In: Proceedings of the 2024 IEEE Vehicular Networking Conference (VNC); 2024 May 29–31; Kobe, Japan. p. 9–16.

23. Fang C, Hu Z, Meng X, Tu S, Wang Z, Zeng D, et al. DRL-driven joint task offloading and resource allocation for energy-efficient content delivery in cloud-edge cooperation networks. IEEE Trans Veh Technol. 2023;72(12):16195–207. doi:10.1109/tvt.2023.3297362.
24. Al-Zahrani FA, Khan I, Zareei M, Zeb A, Waheed A. Resource allocation and optimization in device-to-device communication 5G networks. Comput Mater Contin. 2021;69(1):1201–14. doi:10.32604/cmc.2021.018386.
25. Fang C, Xu H, Yang Y, Hu Z, Tu S, Ota K, et al. Deep-reinforcement-learning-based resource allocation for content distribution in fog radio access networks. IEEE Internet Things J. 2022;9(18):16874–83. doi:10.1109/jiot.2022.3146239.
26. Lin P, Zhang Z, Liu L. Research on space-air-ground integrated network application. In: Proceedings of the 2024 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB); 2024 Jun 19–21; Toronto, ON, Canada. p. 1–6.
27. Wang X, Shen T, Zhang Y, Chen X. An efficient topology emulation technology for the space-air-ground integrated network. In: Proceedings of the IEEE INFOCOM 2023-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS); 2023 May 20; Hoboken, NJ, USA. p. 1–8.
28. Fu D, Hu S, Zhang L, He S, Qiu J. An intelligent cloud computing of trunk logistics alliance based on blockchain and big data. J Supercomput. 2021;77(12):13863–78. doi:10.1007/s11227-021-03800-w.