

ARTICLE

Dimensionality-Adaptive Reinforcement Learning for Energy-Efficient Connectivity-Maintaining Consensus of Multi-Agent Systems with Hybrid Attacks

Hua-Yu Zhu and Weijie Mao*

State Key Laboratory of Industrial Control Technology, Institute of Cyber-Systems and Control, Zhejiang University, Hangzhou, China

*Corresponding Author: Weijie Mao. Email: wjmao@zju.edu.cn

Received: 08 June 2026; Accepted: 17 August 2026; Published: 22 October 2026

ABSTRACT: Achieving secure and energy-efficient connectivity-maintaining consensus is a quintessential challenge in multi-agent systems (MAS), given the complex coupling between global connectivity constraints, power limitations, and adversarial attacks. This paper introduces a resilient closed-loop framework that synergizes topology maintenance and consensus control across two time-scales, incorporating real-time fault detection and post-attack autonomous survival protocols. To reconstruct secure and energy-efficient topologies, we develop a dimensionality-adaptive reinforcement learning (RL) scheme, which adaptively deploys pre-trained RL topology reconstruction policies tailored to the spatial dimensions of the current operating environment. To obtain these multi-dimensional policies, we propose an efficient topology reconstruction training mechanism. By sequentially assigning values only to the upper triangular elements of the adjacency matrix, this mechanism exploits network symmetry to compress the action search space and strongly penalizes disconnected configurations. The simulations test our proposed method on energy-efficient resilient consensus problems with 10 agents and cyber-physical attacks in 5 s, and we show that the proposed method matches the state of the art on performance while adding targeted-decapitation survivability and dimensional adaptivity.

KEYWORDS: Multi-agent systems; reinforcement learning; consensus; connectivity maintenance; hybrid attacks; energy efficiency

1 Introduction

Connectivity-maintaining consensus refers to achieving consensus while maintaining a connected topology in a graph in the presence of adversarial interference. Maintaining connectivity in Multi-Agent Systems (MASs) is critical for numerous applications, including sensor networks [1], secure communication systems [2], autonomous vehicles [3], and multi-robot systems [4,5].

In recent years, the widespread adoption of cyber-physical systems has brought various security challenges in MASs to the forefront. Notable incidents include the 2009 cyber attack on the energy transmission network [6], the shutdown of Iranian nuclear facilities [7], the uncontrolled operation of a European hydroelectric dam [8], and the 2015 Ukraine power outage [9]. These events have positioned MASs security as a critical research focus. Extensive efforts have been made to address these challenges, yielding significant progress [10–12].

Attacks on MASs can be broadly categorized into two types based on their target: node attacks and inter-node channel attacks. Attacks targeting either nodes or inter-node communication channels in MASs

often result in disruptions to the original communication topology. Such disruptions can render the topology disconnected or impair its spanning tree structure. Note that a topology requires specific properties, such as weak connectivity, to function effectively. Malicious attacks can compromise these properties and undermine system performance. To address this problem, ref. [13] proposed a topology reconfiguration method to restore the original consensus by designing a controller based on the attacked topology. However, this approach is limited to controlling a fixed topology structure without allowing modifications. In cases where attacks cause permanent damage to nodes or communication channels, maintaining the original topology becomes infeasible, jeopardizing the normal operation of the MASs. A natural solution to such failures involves adding new edges to the compromised topology to achieve the desired functionality despite faults [14]. In [15], a method leveraging broadcast mechanisms to introduce redundant information flows was shown to reconstruct the communication topology, enhancing fault tolerance. While this approach effectively utilizes available information in MASs, it increases communication overhead due to more frequent interactions. To mitigate communication costs, ref. [16] introduced a topology reconfiguration algorithm that minimizes the number of connections required to restore connectivity. Additionally, ref. [17] developed a graph Laplacian-based approach to achieve desired formation shapes under network faults, relying solely on relative position measurements between neighbors. Furthermore, ref. [18] investigated persistent formation communication topology reconfiguration under fault constraints. More recently, security in MASs has evolved towards addressing complex, hybrid threat landscapes where cyber-attacks and physical disruptions coexist. For instance, recent works have explored resilient containment control and prescribed-performance secure consensus under multi-layer cyber-physical attacks [19,20]. Despite these advancements, existing control-theoretic methods remain inadequate in addressing the challenge of optimal topology reconstruction for MASs under sudden and irregular node and channel attacks when heterogeneous communication energy profiles are considered. These analytical design approaches typically require precise mathematical models of the attack patterns and struggle to scale when the system topology changes dynamically.

Despite recent advancements, existing methods remain inadequate in addressing the challenge of optimal topology reconstruction for MASs under sudden and irregular node and channel attacks. Real-world systems are inherently dynamic and complex, with malicious attacks potentially targeting different components persistently over time. Moreover, the physical distance between nodes often correlates with communication energy costs, as noted in [21]. However, most studies assume uniform communication energy consumption, overlooking the heterogeneous energy profiles prevalent in real-world MASs. To ensure efficient resource allocation and sustain long-term topology reconstruction efficiency, optimization techniques are essential to streamline controllers, enabling MASs to maintain connectivity, minimize energy consumption, and achieve high efficiency under malicious attacks.

In this paper, we investigate whether a viable approach can be developed to address the connectivity maintenance problem in MASs under simultaneous physical node removal attacks and Denial of Service (DoS) attacks, while minimizing heterogeneous communication costs. In other words, can an MASs autonomously learn to optimize its topology structure to adapt to complex and dynamic environments?

Recent successes in applying machine learning to challenging decision-making domains suggest that such an approach is feasible for MASs topology optimization. In particular, Reinforcement Learning (RL) has proven effective across various applications [22,23]. Lately, several researchers have attempted to leverage Multi-Agent Reinforcement Learning (MARL) to learn optimal communication topologies or edge-activation policies under resource constraints [24,25]. However, these learning-based topology control methods generally assume a fixed action space or rely on rigid heuristic masking when the dimensionality of the workspace or agent network changes. This limits their applicability in highly volatile scenarios where physical attacks alter the operational space. Given the outlined challenges, RL methods are particularly

well-suited for MASs topology optimization for several reasons. First, RL algorithms can autonomously learn and optimize decisions through interaction with the environment, enabling MASs to dynamically adjust their topology to meet evolving demands and challenges. Second, the complex and high-dimensional nature of MASs topologies can be effectively managed by RL, which is capable of identifying globally optimal or near-optimal solutions. Third, RL can address the uncertainties and dynamic changes inherent in MASs environments through policy optimization. Consequently, employing RL to achieve optimal connectivity maintenance under simultaneous physical node attacks and DoS attacks merits investigation. However, the permanent damage caused by physical attacks introduces multidimensional challenges in connectivity maintenance, necessitating tailored interactions and learning across diverse environments. A single RL model is insufficient to address such complexity. Hence, in this paper, we introduce the concept of Dimensionality-Adaptive Reinforcement Learning (DARL) [26], aiming to effectively address the aforementioned challenges.

We develop a DARL-based framework for secure and energy-efficient connectivity-maintaining consensus, and integrate it with a two-time-scale architecture to synergistically optimize topology maintenance and achieve a consensus. Our main contributions are the following:

1. We propose a novel Topology-Consensus Coupled Two-Time-Scale Framework (TCCTTSF) that transcends the traditional decoupled design of connectivity maintenance and consensus control under hybrid malicious attacks in an energy-efficient way. By integrating a detection-isolation-survival protocol into a unified closed-loop architecture, the framework ensures that global consensus is not merely a terminal goal but a resilient process capable of autonomous recovery under unpredictable adversarial interventions.
2. We present a DARL-based approach. This paradigm enables the agents to autonomously learn the optimal trade-off between energy consumption and connectivity constraints.
3. We develop a specialized training pipeline that models the reconstruction process exclusively on the upper triangular elements of the adjacency matrix to address the curse of dimensionality in network optimization. This formulation significantly reduces the state and action spaces, facilitating accelerated convergence and enabling the agent to identify energy-efficient, connected topologies without the need for expert-designed topology-construction heuristics.

The remainder of this paper is organized as follows. [Section 2](#) establishes the theoretical preliminaries, where the adversarial attack and heterogeneous energy consumption models are formulated, followed by a formal definition of the connectivity-preserving consensus problem. [Section 3](#) elaborates on the proposed methodology, detailing the synergistic two-time-scale architecture and the dimensionality-adaptive reinforcement learning approach for topology reconstruction. The efficacy and superiority of the proposed framework are validated through extensive simulations in [Section 4](#). [Section 5](#) gives a discussion. Finally, [Section 6](#) concludes the paper and discusses promising directions for future research.

2 Preliminaries and Problem Formulation

Let $\mathcal{N} = \{1, 2, \dots, N\} \subset \mathbb{Z}^+$ denote the set of node index where $N \geq 2$ and $\mathbb{Z}^+$ is the set of positive integers. Denote by $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ the undirected graph where $\mathcal{V} = \{v_i : i \in \mathcal{N}\}$ is the set of nodes and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges. Note that $\mathbb{R}^N$ is the N -dimensional real vector space. Let $\mathcal{A}(\mathcal{G}) \in \{0, 1\}^{N \times N}$ denote the adjacency matrix of graph $\mathcal{G}$. Let $e_{ij} = (v_i, v_j)$, $i, j \in \mathcal{N}$ denote the undirected edge of $\mathcal{G}$. Denote by $\mathcal{N}_i = \{v_j \in \mathcal{V} : (v_i, v_j) \in \mathcal{E}, i, j \in \mathcal{N}\}$ the set of neighbors of node v_i , $i \in \mathcal{N}$. Let $\mathcal{D}(\mathcal{G})$ denote the degree matrix, then the Laplacian matrix of the graph $\mathcal{G}$ is denoted by $L(\mathcal{G}) = \mathcal{D}(\mathcal{G}) - \mathcal{A}(\mathcal{G})$.

Notations: The $|\cdot|$ symbol denotes the number of elements in a set. $\langle \cdot \rangle$ indicates a sequence. Throughout this paper, $\hat{a}_{ij}$ represents the (i, j) -th entry of the adjacency matrix $\mathcal{A}(\mathcal{G})$, while the notation a denotes the vector of action, and the notation a_k denotes the vector of action at time step k . These symbols correspond to distinct mathematical objects.

Consider a multi-agent system with N nodes and each agent is modeled as

$$\begin{cases} \dot{x}_i(t) = Ax_i(t) + Bu_i(t) \\ y_i(t) = Cx_i(t) \end{cases}, \quad i \in \mathcal{N} \quad (1)$$

where $x_i \in \mathbb{R}^n$ is the state of the agent $i \in \mathcal{N}$, $n \in \mathbb{Z}^+$, $u_i \in \mathbb{R}^m$ is the control input, $m \in \mathbb{Z}^+$, $y_i \in \mathbb{R}$ is the output, and $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{m \times n}$ are corresponding matrices. Furthermore, the physical position of agent i is described by $\mathbf{P}_i(t) \in \mathbb{R}^{n_p}$, $n_p \in \mathbb{Z}^+$, which is included in the state vector $x_i(t)$, e.g., $x_i = [\mathbf{P}_i^\top, \dots]^\top$. $n_p < n$ is the dimension of agent positions. The inter-agent distance is defined as

$$d_{ij} = \|\mathbf{P}_i - \mathbf{P}_j\|, \quad i, j \in \mathcal{N}. \quad (2)$$

The connectivity maintenance problem in multi-agent systems refers to the challenge of ensuring that the communication network among agents remains connected throughout the process of their coordination or movement. Specifically, it requires that there exists at least one spanning tree in graph $\mathcal{G}$ at all times. This paper considers scenarios in which the system is subject to malicious attacks. The presence of such attacks disrupts the original connectivity of the network topology, thereby hindering effective coordination among agents. We specifically examine two distinct models of malicious attacks.

We model physical attacks as a node-removal process. Let $\mathcal{V}_{\text{rem}}$ be the set of physically attacked nodes. Formally, an attacked node $v_i \in \mathcal{V}_{\text{rem}}$ is eliminated from the vertex set $\mathcal{V}$, resulting in a residual graph $\mathcal{G}' = \mathcal{G}(\mathcal{V} \setminus \mathcal{V}_{\text{rem}})$. This removal is irreversible, effectively neutralizing the node's capacity for sensing, communication, and control. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, where Ω is the set of all possible outcomes (attack sequences), $\mathcal{F}$ is a σ -algebra of events, and $\mathbb{P}$ is a probability measure. Let $\{\mathcal{N}(t)\}_{t \geq 0}$ be a homogeneous Poisson process with intensity $\alpha_{\text{int}} > 0$, defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The probability that $\mathcal{K}$ attacks occur within the time interval $[0, t]$ is given by

$$\mathbb{P}(\mathcal{N}(t) = \mathcal{K}) = \frac{(\alpha_{\text{int}} t)^\mathcal{K} e^{-\alpha_{\text{int}} t}}{\mathcal{K}!}, \quad \mathcal{K} = 0, 1, 2, \dots \quad (3)$$

where $\alpha_{\text{int}} \in [0, 1]$ denotes the attack intensity, i.e., the average number of physical attacks per unit time, and e is the Euler's number. Let $\hat{t}_l$ denote the arrival time of the l -th physical attack, defined as $\hat{t}_l := \inf\{t > 0 : \mathcal{N}(t) = l\}$, $l = 1, 2, \dots$. At each attack instant $\hat{t}_l$, the attacked node is selected from the currently active node set $\mathcal{N}(\hat{t}_l^-)$, where $\hat{t}_l^-$ denotes the left-hand limit at $\hat{t}_l$. Define the candidate attack set at time $\hat{t}_l$ as

$$\mathcal{C}(\hat{t}_l) = \operatorname{argmax}_{i \in \mathcal{N}(\hat{t}_l^-)} |\mathcal{N}_i|, \quad (4)$$

where $|\mathcal{N}_i|$ denotes the degree of node i . If $|\mathcal{C}(\hat{t}_l)| = 1$, the unique maximizer is attacked. Otherwise, one node is randomly selected from $\mathcal{C}(\hat{t}_l)$ according to a uniform distribution. The index of attacked failure node is denoted by $b(\hat{t}_l) \in \mathcal{C}(\hat{t}_l)$.

For clarity, we provide an example to demonstrate the scenario. Fig. 1 illustrates an example of a physical attack on a multi-agent system with four nodes. Fig. 1a shows the topology of the network before the attack, while Fig. 1b depicts the attacker launching a physical attack on Node 1, which has the most neighbors. Fig. 1c

illustrates the communication connectivity among the nodes after the attack. As observed, Node 1 is no longer operational, and communication remains only between Nodes 2 and 3.

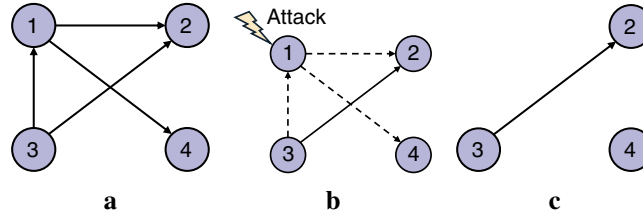


Figure 1: Illustration of physically attacked nodes: (a) Communication topology before the attack; (b) Location of the targeted node under attack; (c) Communication topology after the attack.

In this study, we consider intermittent DoS attacks that intermittently disrupt the data exchange within the network. Unlike permanent physical failures, these attacks are characterized by their temporal sparsity and resource-constrained nature. Let $\mathcal{I}_\ell := [t_\ell, t_\ell + v_\ell]$ denote the ℓ -th DoS interval, where $t_\ell \geq 0$ represents the onset of the ℓ -th attack and $v_\ell > 0$ denotes its corresponding duration. The index $\ell \in \mathbb{N}_{\geq 1}$ identifies the sequence of attack events. For a given observation interval $[\varsigma, t]$, the set of time instants during which the communication is blocked (active DoS) $\Xi(\varsigma, t)$ and the set of time instants during which the network remains functional (inactive DoS) $\Theta(\varsigma, t)$ are respectively defined as:

$$\begin{aligned} \Xi(\varsigma, t) &:= \left(\bigcup_{\ell=1}^{N_{\text{att}}} \mathcal{I}_\ell \right) \cap [\varsigma, t], \\ \Theta(\varsigma, t) &:= [\varsigma, t] \setminus \Xi(\varsigma, t), \end{aligned} \quad (5)$$

where N_{att} is the total number of DoS attack intervals occurring up to time t . For any node $v_i \in \mathcal{V}$, the actual output signal $\hat{y}_i(t)$ subject to DoS attacks is modeled as a switching-like mechanism: $\hat{y}_i(t) = \begin{cases} 0, & t \in \Xi(0, T) \\ y_i(t), & t \in \Theta(0, T) \end{cases}$, where $y_i(t)$ denotes the nominal (unperturbed) output.

Assumption 1 (DoS Frequency [27]): Let $n_b(\varsigma, t)$ denote the number of DoS off-to-on transitions in the interval $[\varsigma, t)$. There exist constants $\xi \geq 0$ and $T_f > 0$ such that: $n_b(\varsigma, t) \leq \xi + \frac{t-\varsigma}{T_f}$, $\forall t > \varsigma \geq 0$.

Assumption 2 (DoS Duration [27]): Let $|\Xi(\varsigma, t)|$ denote the Lebesgue measure of the set $\Xi(\varsigma, t)$. There exist constants $o \geq 0$ and $T_d > 1$ such that: $|\Xi(\varsigma, t)| \leq o + \frac{t-\varsigma}{T_d}$, $\forall t > \varsigma \geq 0$.

Remark 1: Assumptions 1 and 2 encapsulate the constrained adversarial resources. Specifically, T_f and T_d represent the average dwell-time and the inverse of the duty cycle of the attacks, respectively. This framework encompasses a broad class of attack patterns, including periodic, aperiodic, and stochastic pulse-width modulated jamming.

The DoS attack location is formally defined by a subset of compromised nodes $\mathcal{V}_{\text{DoS}}(t) \subset \mathcal{V}$. At any instant t , the set of inaccessible nodes is given by: $\mathcal{V}_{\text{DoS}}(t) = \{v_i \in \mathcal{V} \mid \mathbb{I}_{\text{DoS}}(t \in \Xi^i) = 1\}$, where $\mathbb{I}_{\text{DoS}}(\cdot)$ is the indicator function. The physical impact of the DoS attack at location v_i is the instantaneous nullification of all incident edges. Accordingly, the set of active communication links at time t is defined by

$$\mathcal{E}(t) = \{(v_j, v_i) \in \mathcal{E} \mid v_i \notin \mathcal{V}_{\text{DoS}}(t) \text{ and } v_j \notin \mathcal{V}_{\text{DoS}}(t)\}. \quad (6)$$

In addition to security-related factors such as malicious attacks, this work incorporates heterogeneous communication energy consumption as a critical constraint in connectivity maintenance.

Definition 1: Let $N(t)$ be the number of currently active nodes at time t , and $\mathcal{N}(t) = \{1, 2, \dots, N(t)\} \subset \mathbb{Z}^+$ be the set of currently active node indices, which excludes the physically attacked nodes and DoS attacked nodes. The communication quality between agents $i \in \mathcal{N}(t)$ and $j \in \mathcal{N}(t)$ is modeled as $g_e(d_{ij}) = \frac{1}{1 + \exp(\hat{\alpha}(d_{ij} - \rho))}$, where

d_{ij} is defined by (2), $\hat{\alpha} = \frac{2}{\rho_{2i} - \rho_{1i}} \log \frac{1-\epsilon}{\epsilon}$, $\rho = \frac{\rho_{1i} + \rho_{2i}}{2}$, $\epsilon \in (0, 0.5)$, and ρ_{1i}, ρ_{2i} define the effective communication range of the i -th agent. A higher $g_e(d_{ij})$ indicates better communication quality.

Assumption 3: The system is sampled at intervals of Δt . During each sampling interval $[k\Delta t, (k+1)\Delta t)$, the agent positions $\mathbf{P}_i$ and the communication topology remain constant.

Definition 2 (Heterogeneous communication energy): The total communication energy of the multi-agent system at time t is $E(t) = \sum_{i=1}^{N(t)} \sum_{j \in \mathcal{N}_i} P_{ij}(t)$, where the instantaneous power consumption P_{ij} is defined as $P_{ij}(t) = \gamma_i^e d_{ij}^\beta(t) \max\{0, \varphi - g_e(d_{ij}(t))\}$. Here, $\gamma_i^e > 0$ is the heterogeneous energy coefficient of agent $i \in \mathcal{N}(t)$, $\mathcal{N}(t)$ is the set of active node indices at time t , $\beta \geq 2$ is the path-loss exponent, $\varphi \in (0, 1)$ is a signal energy parameter, and $d_{ij}(t) = \|\mathbf{P}_i(t) - \mathbf{P}_j(t)\|$, $i, j \in \mathcal{N}(t)$.

Remark 2: Maximizing $g_e(d_{ij})$ corresponds to minimizing communication energy consumption. The heterogeneous factor γ_i^e accounts for agent-specific energy differences in practical systems.

To maintain full-node communication connectivity in a multi-agent system, we aim to ensure that the topology formed by the remaining functional agents remains connected (i.e., $\lambda_2(\mathcal{G}(t)) > 0$) under potential malicious attacks, while simultaneously minimizing the total communication energy consumption J . Here, $\lambda_2(\mathcal{G}(t))$ is the second smallest eigenvalue of the Laplacian matrix $L(\mathcal{G}(t))$ at time t . Meanwhile, to avoid the curse of dimensionality and excessive computational overhead, the topology reconstruction problem is formulated at a slow time-scale, independent of the instantaneous continuous-time agent dynamics. The problem is formulated as a combinatorial optimization problem:

$$\min_{\mathcal{A}(t)} J = \int_{t=0}^T E(t; \mathcal{A}(t)) \quad (7)$$

$$\text{s.t. } \mathcal{A}(t) \in \{0, 1\}^{|\hat{\mathcal{V}}| \times |\hat{\mathcal{V}}|}, \quad \mathcal{A}(t) = \mathcal{A}^\top(t), \quad \forall t \quad (8)$$

$$\lambda_2(\mathcal{G}(t)) > 0, \quad \forall t \quad (9)$$

$$a_{ij}(t) = 1 \implies d_{ij}(t) \leq \max(\eta_i, \eta_j), \quad \forall i \neq j, \forall t \quad (10)$$

where T is the prescribed time horizon, $\mathcal{A}(t)$ is the adjacency matrix at sampling instant t , representing the network connectivity configuration; $\hat{\mathcal{V}}$ denotes the set of agents that remain functional after attacks; $\hat{a}_{ij}(t)$ is the (i, j) -th element of $\mathcal{A}(t)$ and $i, j \in \mathcal{N}(t)$; $\lambda_2(\mathcal{G}(t))$ is the second smallest eigenvalue of the Laplacian matrix $L(\mathcal{G}(t))$; $\eta_i = \rho_{2i} + \epsilon_{\text{dis}}$ is the maximum communication range of agent $i \in \mathcal{N}(t)$; ρ_{2i} is the ρ_2 parameter of the agent i ; $\epsilon_{\text{dis}} > 0$ is a small constant; $E(t; \mathcal{A}(t))$ is the total communication energy consumption at time t given the adjacency matrix. It is notable that Constraint (9) ensures that $\lambda_2(\mathcal{G}(t)) > 0$. This graph-theoretic condition serves as the necessary and sufficient guarantee for the feasibility of the subsequent low-level consensus dynamics.

Remark 3: According to the lemma in [28], $\lambda_2(\mathcal{G}(t)) > 0$ implies that the graph $\mathcal{G}(t)$ is connected and contains a spanning tree, i.e., there exists at least one path for information to propagate between any pair of nodes.

Definition 3 (Consensus [29]): The multi-agent system described in (1) achieves consensus if $\lim_{t \rightarrow \infty} \|y_i(t) - y_j(t)\| = 0$, $\forall i, j \in \mathcal{N}(t)$, where $y_i(t) \in \mathbb{R}$ denotes the output of agent i .

Remark 4: Consensus is guaranteed if the underlying communication graph $\mathcal{G}(t)$ remains connected for all $t \geq 0$.

This paper aims to develop a general framework for multi-agent systems subject to both physical and cyber attacks. The framework is designed to maintain an optimal network topology that minimizes heterogeneous communication energy consumption, preserves connectivity, and enables the system to achieve consensus. Specifically, under the malicious attack models (3)–(6) and the constraints of the optimization problem (7)–(10), the proposed framework ensures that the system (1) satisfies the consensus conditions defined in Definition 3. The main challenge addressed in this study arises from the stochastic and heterogeneous nature of malicious attacks, combined with the need to minimize communication energy consumption. Ensuring in real time that the multi-agent system maintains a connected topology with a spanning tree, while satisfying multiple operational constraints and achieving consensus, is a nontrivial problem.

3 Methodology

3.1 Framework

In this section, we design the TCCTTSF to address the secure and energy-efficient connectivity-maintaining consensus of MASs under malicious attacks. As illustrated in Fig. 2, the framework integrates a Resilient Connectivity Maintenance Module (RCMM) with a Consensus Control Module (CCM), forming a synergistic closed-loop architecture. The core philosophy of this framework lies in its Attack-Adaptive Event-Triggered Two-Time-Scale Mechanism (AETS): the RCMM operates at a discrete, adaptive-time-scale to mitigate topology-level perturbations (e.g., DoS attacks), while the CCM executes at a continuous, fast-time-scale to rapidly drive the agents towards a synchronized state.

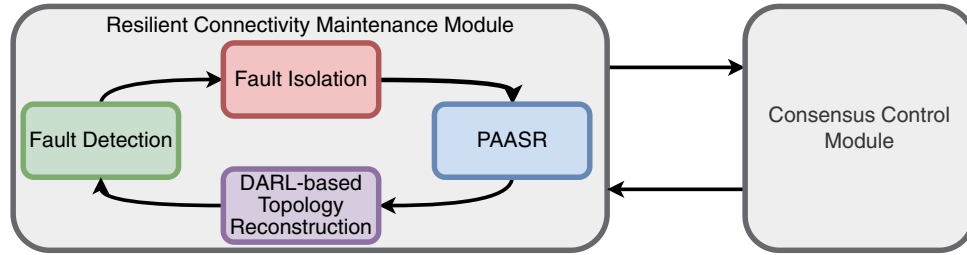


Figure 2: Topology-consensus coupled two-time-scale framework.

Under the two-time-scale temporal framework, RCMM generates connected graph structures as inputs for CCM. In turn, CCM provides RCMM with instantaneous state updates, enabling the latter to derive near-optimal topologies informed by real-time system dynamics. Let t_{rec} be the adaptive update interval of topology reconstruction, and t_{con} be the update interval of fast consensus control. The AETS is defined as

$$\begin{cases} \mathcal{G}' = f_{\text{rec}}(\mathbf{x}, \mathcal{G}), & \text{with Time Interval } t_{\text{rec}} \\ \mathbf{u} = f_{\text{con}}(\mathbf{x}, \mathbf{y}), & \text{with Time Interval } t_{\text{con}} \end{cases} \quad (11)$$

where $\mathcal{G}'$ denotes the updated topology graph for the next time step, f_{rec} is the adaptive function for topology reconstruction, f_{con} is the function for fast consensus, and $\mathbf{x} = [x_1^T, \dots, x_N^T]^T$ is the single agent state vector. Typically, to achieve rapid consensus, we set $t_{\text{con}} = \varpi$, where ϖ is a fixed, relatively small constant.

Furthermore, the adaptive time interval of topology reconstruction t_{rec} is defined as $t_{\text{rec}}(\alpha_{\text{int}}) = \tau_{\min} + (\tau_{\max} - \tau_{\min})e^{-k_{\text{sen}}\alpha_{\text{int}}}$ where $\tau_{\min}$ and $\tau_{\max}$ denote the lower and upper bounds of the inner-execution time, respectively, while k_{sen} is a design parameter regulating the sensitivity to the attack intensity α_{int} . The exponential term ensures that the triggering interval decreases monotonically with respect to α_{int} , enabling

more frequent topology reconstruction under stronger attacks. Note that there exists a strictly positive constant $\tau_{\min} > 0$ such that Zeno behavior is excluded.

3.2 Resilient Connectivity Maintenance

We use a multi-stage pipeline (Fig. 2) to resiliently maintain topological connectivity, ensuring MASs remain functional under malicious attacks. We begin by detecting fault nodes with an attack-adaptive mechanism that dynamically adjusts detection frequency based on real-time attack activity. This provides a timely and resource-efficient estimation of the compromised node set, enabling rapid response under high-intensity attacks while avoiding unnecessary monitoring overhead in benign conditions. Next, we isolate the nodes flagged as failed during the previous detection stage. This adjusts the effective network topology by removing unreliable agents and reconfiguring local interaction patterns, thereby preventing the propagation of corrupted information and preserving system integrity. Then, we deploy a post-attack autonomous survival and reconnection mechanism to maintain the essential operational continuity of the remaining healthy sub-networks. Finally, we reconstruct the network topology according to its current state using a DARL-based policy, which facilitates multi-agent topological reconfiguration under varying node conditions.

3.2.1 Fault Detection

For the first stage of the connectivity maintenance pipeline, we aim to reliably detect agents that become completely unavailable due to DoS attacks or abrupt failures. In such scenarios, methods relying on state comparisons or residual checks are no longer applicable, as no information can be received from compromised agents. Therefore, we develop a timeout-based fault detection mechanism that identifies inactive agents based on communication absence.

For each agent $i \in \mathcal{N}$, define t_i^{last} as the most recent time instant at which agent i successfully receives information from its neighbors. We introduce a local timer for each agent, and define the inter-communication interval as $\Upsilon_i(t) = t - t_i^{\text{last}}$. Let $\hat{\mathcal{F}}(t)$ be the compromised node set at time t . A node is declared as faulty if its communication inactivity exceeds a prescribed threshold, i.e., $i \in \hat{\mathcal{F}}(t)$ if $\Upsilon_i(t) > \mu_i$, where $\mu_i > 0$ is a design parameter representing the maximum allowable communication interval. To account for time-varying attack intensity, the threshold μ_i can be adaptively adjusted as $\mu_i(t) = \frac{\mu_0}{1 + \alpha_{\text{int}}(t)}$, where $\mu_0 > 0$ is a nominal threshold and $\alpha_{\text{int}}(t)$ characterizes the estimated attack intensity. This detection mechanism provides a distributed and model-free way to identify communication-loss faults, ensuring timely detection under severe DoS conditions while avoiding false alarms during normal intermittent communication.

3.2.2 Fault Isolation

The second stage of the connectivity maintenance pipeline aims at mitigating the impact of detected faulty agents by excluding them from the interaction network. This step ensures that the system dynamics evolve only over reliable communication links, thereby preventing structural degradation caused by inactive nodes.

Define the effective node set as $\mathcal{V}_{\text{eff}}(t) = \mathcal{V} \setminus \hat{\mathcal{F}}(t)$, and construct the induced subgraph $\mathcal{G}_{\text{eff}}(t) = (\mathcal{V}_{\text{eff}}(t), \mathcal{E}_{\text{eff}}(t))$, where $\mathcal{E}_{\text{eff}}(t) = \{(i, j) \in \mathcal{E}(t) \mid i, j \in \mathcal{V}_{\text{eff}}(t)\}$. At the matrix level, the adjacency matrix $\mathcal{A}(t)$ is updated as $\tilde{\mathcal{A}}_{ij}(t) = \begin{cases} \mathcal{A}_{ij}(t) & i, j \in \mathcal{E}_{\text{eff}}(t), \\ 0, & \text{otherwise} \end{cases}$. Consequently, all communication links associated with faulty agents are removed, and the system dynamics are restricted to the subgraph induced by $\mathcal{V}_{\text{eff}}(t)$. This isolation mechanism effectively redefines the network topology by pruning inactive nodes,

preserving meaningful interactions among operational agents and preparing the system for subsequent topology reconstruction.

3.2.3 Post-Attack Autonomous Reconnection

The third stage of the connectivity maintenance pipeline aims to bridge the gap between instantaneous fault isolation and global network recovery, ensuring that the remaining agents maintain physical integrity while autonomously restoring a cohesive communication fabric. We present Post-Attack Autonomous Survival and Reconnection Mechanism (PAASR). This mechanism is structured into two concurrent phases: local safety preservation and stochastic neighbor discovery.

Upon the exclusion of compromised nodes, the surviving agents immediately transition into a local sensor-based safety mode. In this state, the coupling between distributed coordination and local control is temporarily suspended to prioritize fundamental safety constraints. By leveraging onboard heterogeneous sensors (e.g., LiDAR or IMU), each agent activates a standby sensor-based fallback controller. This ensures that even in the absence of external setpoints, the agent maintains collision-free maneuvers within its local safe region, effectively preventing catastrophic failures during the transient phase of network fragmentation.

To mitigate the isolation caused by attacks, an active-passive hybrid re-pairing strategy is invoked. Fragmented agents initiate an omnidirectional broadcast scan to identify $\mathcal{K}$ active peers within their communication radius η . To enhance the robustness of the recovering graph, a safe connection redundancy protocol is implemented, where nodes concurrently maintain a passive listening state to accept incoming handshake requests from authenticated neighbors. This dual-link establishment enhances connection density and structural redundancy, facilitating the rapid transition of disjoint clusters into a well-connected mesh topology. Once the distributed reconnection and redundancy injection phase satisfies the basic connectivity requirement, i.e., the reconstructed graph $\mathcal{G}_{\text{rec}}$ contains a spanning tree, the MASs immediately reinstates the consensus control protocol. Although the initial topology formed through stochastic reconnection may be sub-optimal in terms of energy cost, it provides the necessary algebraic connectivity to pull the agents back from their local safety modes towards a common objective.

Notably, upon the cessation of DoS attacks, the previously compromised nodes invoke the distributed reconnection and redundancy injection mechanism. This facilitates the rapid restoration of inter-agent communication and strengthens topological robustness, thereby ensuring a seamless transition from the post-attack survival phase to the global consensus recovery.

3.2.4 DARL-Based Topology Reconstruction

The final stage of the connectivity maintenance pipeline focuses on reconstructing the compromised topology by DARL. DARL tackles high-dimensional state or action spaces by dynamically adjusting the complexity of the learning model to match the task's intrinsic dimension, improving sample efficiency and scalability. To maintain the topological connectivity of the MASs under malicious attacks, we present a resilient connectivity maintenance method based on DARL, utilizing a library of pre-trained RL policies.

In the DARL-based topology reconstruction scheme, a virtual computational leader $\mathcal{L}_1$ is dynamically elected based on residual computational capacity and topological centrality to serve as the execution kernel.

To eliminate the single point of failure and counter targeted decapitation attacks (e.g., degree-based attacks in Eq. (4) that inherently target high-centrality nodes like $\mathcal{L}_1$), we incorporate a pre-designated k -tier leader succession protocol inspired by hardware warm-standby mechanisms.

During the dynamic election phase (executed over the physically re-connected prototype graph $\mathcal{G}_{\text{rec}}$ formed by PAASR), agents compute a ranked sequence of candidate kernels $\mathcal{L}_{\text{leader}}(t) =$

$\{\mathcal{L}_1, \mathcal{L}_2, \dots, \mathcal{L}_k\}$, $k \geq 2$. Here, $\mathcal{L}_1$ is selected based on top computational capacity and topological centrality to maximize inference efficiency. To prevent an adversary from predicting and sequentially targeting successors via degree-centric profiles (4), the succession ranks $\{\mathcal{L}_2, \dots, \mathcal{L}_k\}$ are selected from nodes meeting a minimum compute threshold using a ranking criterion *decorrelated* from topological degree (e.g., inverse-degree weighting combined with a pseudo-random seed shared during consensus election). This ensures the succession order stops being the attack order.

The primary policy execution and failover logic operates as follows. The pre-trained DARL policy network is fully replicated and onboarded across all operational agents prior to deployment, enabling any agent $\mathcal{L}_m \in \mathcal{L}_{\text{leader}}$ to assume the execution role immediately. Upon election, $\mathcal{L}_1$ transmits a low-overhead heartbeat to $\mathcal{L}_2$ at interval $\Delta\tau$. To support seamless continuation without high signaling overhead, each heartbeat carries a lightweight telemetry digest: $H = \langle k, \rho(k) \rangle$, containing the current MDP sub-step index k and the state/action pointer $\rho(k)$ (occupying only a few bytes).

To handle transient DoS attacks where blocked heartbeats could trigger a false promotion and cause a split-brain condition (two active primaries): (1) Pursuant to Assumption 2, τ_{failover} is explicitly calibrated such that $\tau_{\text{failover}} > \tau_{\text{DoS, max}}$, ensuring transient blackouts clear before timeout expiration. (2) As a fail-safe, $\mathcal{L}_1$ monitors its outbound transmission status; if an isolated communication blackout exceeds τ_{failover} , $\mathcal{L}_1$ automatically abdicates to a standby state. Upon timeout, $\mathcal{L}_2$ broadcasts a single-hop promotion signal and assumes the primary kernel role, continuing the reconstruction directly from sub-step k via $\rho(k)$. If all pre-designated successors in $\mathcal{L}_{\text{leader}}$ are sequentially targeted before reconstruction completes, the system gracefully falls back to the baseline P2P topology $\mathcal{G}_{\text{rec}}$ to execute consensus control until attack intensity subsides.

Remark 5: *The integration of the multi-tier succession protocol fundamentally resolves the nested single point of failure vulnerability. While selecting $\mathcal{L}_1$ based on centrality optimizes inference performance (at the cost of aligning with the adversary's degree-centric targeting profile in Eq. (4)), the succession order for $\{\mathcal{L}_2, \dots, \mathcal{L}_k\}$ is deliberately decorrelated from topological degree. Consequently, an adversary executing degree-based decapitation cannot predict or walk down the succession list in sequence. Combined with minimal telemetry heartbeats and self-demotion guards against DoS-induced split-brain execution, this pre-designed handoff guarantees seamless recovery continuity even under targeted attacks.*

The architecture of the DARL-based topology reconstruction (Fig. 3) consists of three main modules: (1) *Encoder*. Given the observed topology, the encoder performs dimension alignment. Specifically, the remaining active nodes are reindexed as $(1, 2, \dots, N(t))$, where $N(t)$ denotes the number of agents at current time t . This ensures compatibility with the corresponding pre-trained RL model. (2) *Hybrid DeepTR Models*. This module contains $N - 2$ pre-trained policies, covering agent counts from 3 to N . Each policy is trained independently under a fixed-dimension setting. Based on $N(t)$, the appropriate policy is selected and applied to the encoded state. (3) *Decoder*. The decoder restores the original node indices and maps the generated connectivity decisions back to the physical node identities. A new topology is thus reconstructed for execution.

The rationale behind the DARL-based topology reconstruction scheme is practical rather than theoretical. Although variable-size neural architectures exist, training a single policy to handle all possible agent counts would significantly increase model complexity and training instability. Instead, we decompose the global problem into multiple fixed-dimension subproblems, which improves training stability and policy specialization. Therefore, the DARL-based topology reconstruction scheme can be viewed as a dimension-consistent policy selection framework, enabling robust connectivity maintenance under dynamic agent loss.

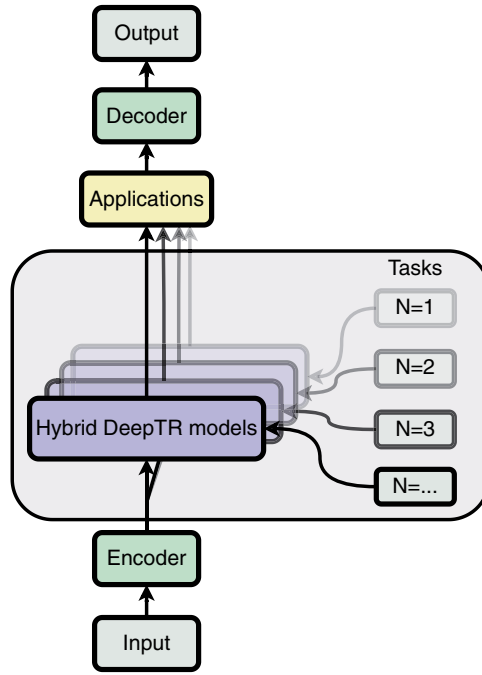


Figure 3: Architecture of the DARL-based topology reconstruction.

The DeepTR policy is designed for topology reconstruction in networks with a fixed number of nodes. In the DARL-based topology reconstruction, each pre-trained RL model, corresponding to a specific number of nodes, is trained by DeepTR. The specific design details of DeepTR are as follows.

Markov Formulation. The connectivity of an undirected graph $\mathcal{G}$ is completely determined by its adjacency matrix $\mathcal{A}(\mathcal{G}) \in \{0, 1\}^{N \times N}$, which is symmetric with zero diagonal elements. Since $\mathcal{A}(\mathcal{G})$ is symmetric, its independent elements are contained in the strictly upper triangular part. Let the strictly upper triangular elements of $\mathcal{A}(\mathcal{G})$ be reordered into a sequence

$$f_{\text{trans}}(\mathcal{A}) \in \{0, 1\}^M, \quad M = \frac{N(N-1)}{2} \in \mathbb{R} \quad (12)$$

where $f_{\text{trans}}(\cdot)$ is the function to transform the strictly upper triangular elements of $\mathcal{A}$ into a sequence. The ordering is defined row-wise from top to bottom, and within each row from left to right, i.e., $(1, 2), (1, 3), \dots, (1, N), (2, 3), \dots, (N-1, N)$. For example, for $N = 4$, the strictly upper triangular elements are $f_{\text{trans}}(\mathcal{A}) = \langle \hat{a}_{12}, \hat{a}_{13}, \hat{a}_{14}, \hat{a}_{23}, \hat{a}_{24}, \hat{a}_{34} \rangle$. Thus, the topology reconstruction problem is equivalent to determining the binary sequence $f_{\text{trans}}(\mathcal{A}) \in \{0, 1\}^M$.

In this paper, the sequential construction of the upper-triangular elements $f_{\text{trans}}(\mathcal{A}) \in \{0, 1\}^M$ is modeled as a discrete-time Markov Decision Process (MDP) $\{s_k, a_k\}_{k=1}^M$ for offline policy training. The transition satisfies

$$\mathbb{P}(s_{k+1} = s' \mid s_k = s, a_k = a, s_{k-1}, \dots) = \mathbb{P}(s_{k+1} = s' \mid s_k = s, a_k = a). \quad (13)$$

where $k \in \{1, \dots, M\}$ represents the sub-step index for reconstructing the $\rho(k)$ -th upper-triangular element, $\rho(k) \in \{1, \dots, M\}$ is the sequential pointer indicating the index of the upper-triangular element currently being decided at sub-step k , s is the current RL state, s' is the RL state of the next decision-making time step, s_k is the RL state at step k , a_k is the RL action at step k , and a is the current RL action.

State Space. In the RL framework, the state represents all information required to characterize the system at a decision instant. Let the decision process evolve in discrete time with sampling instants k . The system state at time step k is defined by

$$s_k = \left(f_{\text{trans}}(\mathcal{A}(k)), \rho(k), \{\gamma_i^e\}_{i \in \mathcal{N}(t)}, \beta, \varphi, \{\eta_i\}_{i \in \mathcal{N}(t)}, \mathbf{x}(t) \right), \quad (14)$$

where $\mathcal{A}(k) \in \{0, 1\}^{N \times N}$ is the evolving adjacency matrix at sub-step k . Specifically, $\mathcal{A}(1)$ is initialized with all in-range candidate links connected ($\mathbb{I}(\|d_{ij}\| < \min(\rho_{2i}, \rho_{2j}))$). At sub-step k , the status of the first $k - 1$ candidate upper-triangular links has been finalized ($a(m) \in \{0, 1\}$ for $m < k$), while the remaining candidate elements ($m \geq k$) retain their initial in-range connectivity; γ_i^e is the energy coefficient of agent $i \in \mathcal{N}(t)$; $\mathcal{N}(t)$ is the set of active node indices at time t ; $\beta \geq 2$ is the path-loss exponent; $\varphi > 0$ is a signal energy parameter; η_i is the max communication range of agent $i \in \mathcal{N}(t)$; and $\mathbf{x}(t)$ is the stacked system state vector of all agents. Thus, the state space is defined as $\mathcal{S} = \{(f_{\text{trans}}(\mathcal{A}), \rho, \{\gamma_i^e\}_{i \in \mathcal{N}}, \beta, \varphi, \{\eta_i\}_{i \in \mathcal{N}}, \mathbf{x})\}$. The state transition satisfies the Markov property $\mathbb{P}(s_{k+1} | s_k, a_k, \dots) = \mathbb{P}(s_{k+1} | s_k, a_k)$ where a_k denotes the action taken at RL time instant k .

Remark 6: The dimensionality of the state space directly affects learning efficiency and convergence. In practice, redundant or slowly varying variables can be excluded to reduce complexity, provided that the Markov property is preserved.

Action Space. At each decision step k , the RL agent selects a topology reconstruction action. Since the communication graph is undirected, its adjacency matrix $\mathcal{A}(k)$ is symmetric with zero diagonal elements. Therefore, the independent decision variables are the strictly upper triangular elements. The action at time step k is defined by

$$a_k \in \{0, 1\} \quad (15)$$

where $a_k = 1$ assigns an edge to the $\rho(k)$ -th strictly upper-triangular entry of $\mathcal{A}$, corresponding to the node pair $(i, j) = f_{\text{trans}}^{-1}(\rho(k))$, and $a_k = 0$ indicates no edge. Thus, the action space is $\mathcal{A} = \{0, 1\}$.

To determine the action, a stochastic policy $\pi(a|s)$ is employed, which defines the probability of selecting action a under state s . In this work, the policy is parameterized by Deep Neural Network (DNN), which serves as a function approximator for topology optimization.

Remark 7: Due to graph undirectedness and symmetry, the full adjacency matrix $\mathcal{A} \in \{0, 1\}^{N \times N}$ is uniquely reconstructed from the completed upper-triangular sequence $f_{\text{trans}}(\mathcal{A}) \in \{0, 1\}^M$ at the final decision step $k = M$.

Remark 8: By decomposing the topology generation into an M -step sequential decision process with a binary scalar action space $\mathcal{A} = \{0, 1\}$, we avoid the exponential action space 2^M of one-shot topology selection. This significantly improves training stability and computational efficiency.

Reward Setup. The reward function $r(s_k, a_k)$ is designed to guide the agent to minimize communication energy consumption while maintaining network connectivity. At time step k , the immediate reward is defined as

$$r_k = -\zeta_1 \eta_{\text{cst}} E(k) + \zeta_2 \eta_{\text{cst}} r_{\lambda_2}(k) - \zeta_3 \eta_{\text{cst}} \mathbf{1}_{\text{invalid}}(k) - \zeta_4 \eta_{\text{cst}} r_{\text{shaped}}(k) - \zeta_5 \eta_{\text{cst}} \mathbf{1}_{\text{disconnect}}(k) + \zeta_6 \eta_{\text{cst}} \mathbf{1}_{\text{secure}}(k) \quad (16)$$

where $E(k) = \sum_{i=1}^{N(k)} \sum_{j \in \mathcal{N}_i} P_{ij}(k)$ denotes the instantaneous communication energy consumption; $r_{\lambda_2}(k)$ is a constant reward that increases if $\lambda_2 > 0$ at decision step k ; $\mathbf{1}_{\text{invalid}}(k)$ is an indicator function that equals 1 if any constructed edge violates the communication range constraint, and 0 otherwise; $r_{\text{shaped}}(k) = \Phi(k) - \Phi(k-1)$ is a potential-based shaping reward; $\Phi(k) = \|\mathcal{E}(k)\| - (N + \alpha_{\text{mar}})\|_2$; $\alpha_{\text{mar}} \in \mathbb{Z}^+$ is a connectivity

margin. Note that the shaping potential $r_{\text{shaped}}(k)$ is evaluated at each sub-step; $\mathbf{1}_{\text{disconnect}}(k)$ is an indicator function that equals 1 if $\lambda_2 \leq 0$, and 0 otherwise. $\mathbf{1}_{\text{secure}}(k)$ is an indicator function that equals 1 if $|\mathcal{E}(k)| = N + \alpha_{\text{mar}}$ when $\lambda_2 > 0$, and 0 otherwise; $\zeta_1, \zeta_2, \zeta_3, \zeta_4, \zeta_5, \zeta_6 > 0$ are weighting parameters, and η_{cst} is a constant reward scaling factor. In our reward setting, the first term penalizes energy consumption, the second term encourages connectivity, the third term imposes a penalty for physically invalid topologies (violating communication range constraints), the fourth term encourages an appropriate number of edges, the fifth term penalizes disconnected topologies, and the sixth term impose a large reward for final secure connection. During offline training, the algebraic connectivity $\lambda_2(L(k))$ is evaluated at every sub-step k on the intermediate adjacency matrix $\mathcal{A}(k)$ to verify network connectivity and determine if $\mathbf{1}_{\text{disconnect}}$ is triggered. It is worth noting that $\mathbf{1}_{\text{secure}}$ effectively pins the final target edge count to $N + \alpha_{\text{mar}}$. Consequently, the energy minimization objective operates within this cardinality constraint by selecting which $N + \alpha_{\text{mar}}$ edges minimize energy, rather than choosing the total edge count freely. The episode terminates only when the maximum decision horizon is reached, or the constructed topology becomes infeasible ($\mathbf{1}_{\text{invalid}} = 1$ or $\mathbf{1}_{\text{disconnect}} = 1$), or $\mathbf{1}_{\text{secure}} = 1$. Notably, these three rules are hard episode-ending rules.

Remark 9: In practical MASs, the achievable vertex connectivity $k_{\text{con}} \in \mathbb{N}$ is fundamentally upper-bounded by the agents' physical communication range η and their spatial distribution density. Specifically, a k_{con} -connected graph may be physically unfeasible if the number of neighbors within the disk of radius η is strictly less than k_{con} for any agent. In this paper, rather than mandating a high-order k_{con} -connectivity during the training of the RL agent, the reward function is designed to prioritize global connectivity ($\lambda_2 > 0$) with a sparse edge density to minimize communication overhead. Although the resulting topology might be susceptible to targeted attacks, the proposed PAASR provides a robust safety layer. By integrating local safety protocols with distributed reconnection, the system exhibits operational redundancy that compensates for the topological fragility during active attack phases, while preserving energy efficiency in nominal or post-attack states.

Recall that the DeepTR algorithm is formulated within a Markov decision process framework. At each decision step k , the agent observes the system state (14) and the action space is defined as $a_k \in \mathcal{A} = \{0, 1\}$. Specifically, each action a_k represents the binary decision for adding or omitting a single candidate edge. The policy network $\pi_\theta(a|s)$ outputs a probability distribution over $\mathcal{A}$. The value network $V_\phi(s)$ estimates the state value function. Both networks are parameterized by DNNs. θ and ϕ are corresponding network parameters, respectively. The objective of RL is to maximize the expected discounted return $J^{\text{RL}}(\theta) = \mathbb{E}_{\pi_\theta} [\sum_{k=0}^K \gamma^k r_k]$, where $\gamma \in (0, 1)$ is the discount factor, and K is the prescribed sampling time horizon. To improve training stability under high-dimensional binary action spaces, DeepTR adopts the Proximal Policy Optimization (PPO) [30] algorithm. The clipped surrogate objective is defined as

$$\mathcal{L}^{\text{PPO}}(\theta) = \mathbb{E} \left[\min \left(\hat{\rho}_k(\theta) \hat{A}_k, \text{clip}(\hat{\rho}_k(\theta), 1 - \varepsilon, 1 + \varepsilon) \hat{A}_k \right) \right], \quad (17)$$

where $\hat{\rho}_k(\theta) = \frac{\pi_\theta(a_k|s_k)}{\pi_{\theta_{\text{old}}}(a_k|s_k)}$. The Temporal-Difference (TD) error is $\delta_k = r_k + \gamma V_\phi(s_{k+1}) - V_\phi(s_k)$, and the advantage function is computed using Generalized Advantage Estimation (GAE): $\hat{A}_k = \sum_{l=0}^{K-k} (\gamma \lambda^{\text{GAE}})^l \delta_{k+l}$, where $\lambda^{\text{GAE}} \in [0, 1]$ controls the bias-variance tradeoff. The complete training procedure is summarized in Algorithm 1.

Algorithm 1: DeepTR training procedure

```

1 Initialize: policy parameters  $\theta$ , value parameters  $\phi$ ;
2 Set: total edges  $M = \frac{N(N-1)}{2}$ , discount factor  $\gamma$ , GAE parameter  $\lambda^{\text{GAE}}$ , clipping coefficient  $\varepsilon$ ;
3 Set: update epochs  $\mathcal{E}$ , minibatch size  $\mathcal{B}$ ;
4 for each episode do
5   Randomly initialize multi-agent physical configuration (positions, ranges);
6   Initialize partially constructed adjacency matrix  $\mathcal{A}(1) \leftarrow \left[ \mathbb{I}(\|d_{ij}\| < \min(\rho_{2i}, \rho_{2j})) \right]_{N \times N}$  where
        $\mathbb{I}(\text{Condition}) = 1$  if the condition is true, 0 otherwise;
7   Observe initial state  $s(1)$  with edge pointer  $\rho(1) = 1$ ;
8   for  $k = 1$  to  $M$  do
9     Sample binary topology action  $a(k) \sim \pi_\theta(\cdot|s(k))$ , where  $a(k) \in \{0, 1\}$ ;
10    Determine the status of the  $\rho(k)$ -th candidate link in  $\mathcal{A}(k)$  via  $a(k) \in \{0, 1\}$  (where
         $a(k) = 1$  retains the edge and  $a(k) = 0$  prunes it) to obtain  $\mathcal{A}(k+1)$ ;
11    if  $k = M$  or constraints violated ( $\mathbf{l}_{\text{invalid}} = 1$  or  $\mathbf{l}_{\text{disconnect}} = 1$  or  $\mathbf{l}_{\text{secure}}$ ) then
12      Compute terminal/immediate reward  $r(k)$  and set done = True;
13      Get next state  $s(k+1)$  and store transition  $(s(k), a(k), r(k), s(k+1), \text{done})$ ;
14      break;
15    else
16      Compute intermediate step reward  $r(k)$  and set done = False;
17      Update pointer  $\rho(k+1) \leftarrow \rho(k) + 1$ ;
18      Get next state  $s(k+1)$  and store transition  $(s(k), a(k), r(k), s(k+1), \text{done})$ ;
19    Compute TD errors  $\delta_k$ ;
20    Compute advantages  $\hat{A}_k$  via GAE;
21    for  $c = 1$  to  $\mathcal{E}$  do
22      Sample minibatches of size  $\mathcal{B}$ ;
23      Update  $\theta$  by maximizing  $\mathcal{L}^{\text{PPO}}(\theta)$ ;
24      Update  $\phi$  by minimizing value loss;
25    Update old policy:  $\theta_{\text{old}} \leftarrow \theta$ ;

```

In this paper, the sequential link determination over upper-triangular candidate elements $f_{\text{trans}}(\mathcal{A}) \in \{0, 1\}^M$ is formulated as a discrete-time MDP $\{s_k, a_k\}_{k=1}^M$ for offline policy training. During training, the agent sequentially decides $a(k) \in \{0, 1\}$ to prune or retain candidate edges starting from the maximal in-range topology $\mathcal{A}(1)$. Once training is complete, the optimal topology matrix $\mathcal{A}^*$ is inferred and applied directly in a single step during online steady-state operation.

Remark 10: *The proposed topology reconstruction strategy operates as a two-stage resilience mechanism under DoS attacks: (1) Attack-Active Stage: The system activates the maximal in-range candidate topology $\mathcal{A}(1)$ to maximize algebraic connectivity $\lambda_2(L)$ and structural redundancy, ensuring network integrity against DoS disruption. (2) Attack-Free Stage: Once the attack ceases, the pre-trained RL policy directly evaluates the candidate topology and deploys the energy-optimized sparse network in a single step (one-shot reconfiguration), rather than dynamically pruning edges online in real time. It is worth emphasizing that the sequential link pruning process (deciding $a(k) \in \{0, 1\}$ at sub-step k) serves exclusively as the offline MDP training formulation, which allows the agent to efficiently learn structural topological features.*

3.3 Cooperative Consensus

To fully demonstrate the efficiency of the proposed framework, we consider a general class of MASs. The CCM is designed to achieve consensus among all operational agents by leveraging the resilient topologies generated by the RCMM.

Given the MASs (1), the objective of the CCM is to leverage a standard distributed control law $u_i(t)$ such that the outputs of all agents satisfy $\lim_{t \rightarrow \infty} \|y_i(t) - y_j(t)\| = 0$ for all $i, j \in \mathcal{V}_{\text{eff}}$. Based on the attack-adaptive topology $\mathcal{G}_{\text{eff}}(t)$ provided by the upper-level RCMM, we adopt a generic consensus protocol of the following form:

$$\hat{u}_i(t) = \sum_{j \in \mathcal{N}_i(t)} \tilde{a}_{ij}(t) (\hat{x}_j(t) - \hat{x}_i(t)) \quad (18)$$

where $\hat{u}_i \in \mathbb{R}$ is the independent input component of agent i , $\hat{x}_i \in \mathbb{R}$ is the independent state component excluding the position state. Here, we assume that each independent state component excluding the position state $\hat{x}_i$ is controlled by an independent input $\hat{u}_i$. Crucially, the control law (18) is executed at the fast-time-scale t_{con} . Since the topology update interval satisfies $t_{\text{rec}} > t_{\text{con}}$, the adjacency matrix $\tilde{\mathcal{A}}(t)$ remains constant within each interval $[t_k, t_{k+1})$. This temporal separation allows the CCM to utilize the optimal connectivity margins established by the DARL-based topology reconstruction, ensuring that the consensus performance is not degraded by the topological perturbations caused by malicious attacks.

3.4 Two-Time-Scale Modeling

Consider a singularly perturbed cascading architecture. Let $t_{\text{rec}} \in \mathbb{R}_{\geq 0}$ denote the slow continuous-time scale governing the high-level policy updating loop, and let $t_{\text{con}} \in \mathbb{R}_{\geq 0}$ represent the fast continuous-time scale governing the evolution of the low-level MASs. The two distinct time scales are coupled via a small perturbation parameter ϵ , where $\frac{dt_{\text{rec}}}{dt_{\text{con}}} = \epsilon$, $0 < \epsilon \ll 1$. The high-level driving subsystem modeling the parameter update error of the PPO algorithm on the slow scale t_{rec} is defined as:

$$\frac{d\mathbf{E}(t_{\text{rec}})}{dt_{\text{rec}}} = f_{\text{rec}}(\mathbf{E}(t_{\text{rec}})) \quad (19)$$

where $\mathbf{E}(t_{\text{rec}}) = \theta(t_{\text{rec}}) - \theta^* \in \mathbb{R}^m$ represents the policy weight deviation from the ideal optimal parameter θ^* , $f_{\text{rec}} : \mathbb{R}^m \rightarrow \mathbb{R}^m$ denotes the evolution dynamics governing the high-level policy parameter adaptation on the slow time scale t_{rec} . Inherently, $f_{\text{rec}}(\mathbf{E}(t_{\text{rec}}))$ encapsulates the expected parameter update mechanism of the PPO algorithm under the topological constraints. Specifically, it can be conceptualized as $f_{\text{rec}}(\mathbf{E}(t_{\text{rec}})) = \alpha \cdot \text{Proj}_{\Omega}(\nabla_{\theta} \mathcal{L}^{\text{PPO}}(\pi_{\theta}))$ where $\nabla_{\theta} \mathcal{L}^{\text{PPO}}$ represents the gradient of the clipped surrogate objective in PPO, Proj_{Ω} denotes the projection operator keeping the parameters on the compact manifold Ω defined by $\lambda_2(L(\theta)) \geq \underline{\lambda}$, and $\alpha > 0$ is the learning rate. To ensure the well-posedness of the high-level subsystem, we assume that $f_{\text{rec}}(\cdot)$ is locally Lipschitz continuous on the domain of interest. Furthermore, $\mathbf{E} = 0$ is an isolated equilibrium point of the error system (19), meaning that $f_{\text{rec}}(0) = 0$. Concurrently, the low-level driven subsystem delineating the consensus tracking error dynamics on the fast scale t_{con} is formulated as:

$$\frac{d\xi(t_{\text{con}})}{dt_{\text{con}}} = -L(\theta(t_{\text{rec}}))\xi(t_{\text{con}}) \quad (20)$$

where $\xi(t) \in \mathbb{R}^{N(t)-1}$ denotes the consensus error vector. Eqs. (19) and (20) fundamentally constitute a single-directional, discrete-continuous cascading singularly perturbed system.

3.5 Main Results

Assumption 4: For the nominal (attack-free) closed-loop topology generated by the high-level policy π_θ , the algebraic connectivity of the corresponding graph Laplacian $L(\theta)$ satisfies $\lambda_2(L(\theta)) \geq \underline{\lambda} > 0$. (Note: In implementation, this is encouraged via a strong negative penalty ζ_5 in the reward function rather than an explicit hard projection onto the parameter space.)

Assumption 5: During the training phase on the compact policy space, the evolution of the policy parameter error $\tilde{\theta}(\tau) = \theta(\tau) - \theta^*$ under the high-level update scheme satisfies the local stability condition: $\dot{V}_1(\tilde{\theta}(\tau)) = (\nabla_{\tilde{\theta}} V_1)^\top \mathbf{u}_\theta(\tilde{\theta}) < 0$ for all $\tilde{\theta} \neq \mathbf{0}$ within a compact neighborhood of the optimal policy parameter θ^* , where $V_1(\tilde{\theta}) = D_{KL}(\pi_{\theta^*} \parallel \pi_\theta)$.

Theorem 1: Consider the cascaded singularly perturbed system consisting of the high-level topology decision framework (19) and the low-level multi-agent consensus dynamics (20) in the absence of adversarial attacks. Under Assumptions 4 and 5, there exists a critical threshold $\epsilon^* > 0$ such that for all $\epsilon \in (0, \epsilon^*)$, the joint nominal error origin $(\tilde{\theta}, \xi) = (\mathbf{0}, \mathbf{0})$ is Locally Asymptotically Stable (LAS).

Proof: The analysis employs singular perturbation and Input-to-State Stability (ISS) cascading techniques across two time scales for the attack-free scenario.

Step 1 (Slow Driving Subsystem): Consider the candidate Lyapunov function $V_1(\tilde{\theta}(\tau)) = D_{KL}(\pi_{\theta^*} \parallel \pi_{\theta(\tau)})$. By Assumption 5, $\dot{V}_1(\tilde{\theta}(\tau)) < 0$ holds locally for $\tilde{\theta} \neq \mathbf{0}$, establishing that the slow parameter driving subsystem (19) is Locally Asymptotically Stable at $\tilde{\theta} = \mathbf{0}$.

Step 2 (Fast Boundary-Layer Subsystem): On the fast time scale t , treating $\theta(\tau)$ as a quasi-steady-state variable ($\epsilon \rightarrow 0$), define $V_2(\xi(t)) = \frac{1}{2} \xi^\top(t) \xi(t)$. Under Assumption 4, evaluating the derivative yields:

$$\frac{dV_2(\xi(t))}{dt} = -\xi^\top(t) L(\theta(\tau)) \xi(t) \leq -\lambda_2(L(\theta(\tau))) \|\xi\|^2 \leq -2\underline{\lambda} V_2(\xi(t)) \quad (21)$$

which proves that the nominal boundary-layer system is Globally Exponentially Stable (GES).

Step 3 (Composite System Stability): Re-writing the fast error dynamics with parameter perturbation as $\dot{\xi} = -L(\theta^*) \xi + \Delta(\tilde{\theta}, \xi)$, where $\|\Delta(\tilde{\theta}, \xi)\| \leq L_g \|\tilde{\theta}\| \|\xi\|$. Thus, the driven fast subsystem is ISS with respect to input $\tilde{\theta}$. By two-time-scale cascading stability theory, there exists $\epsilon^* > 0$ such that for all $0 < \epsilon < \epsilon^*$, the joint origin $(\tilde{\theta}, \xi) = (\mathbf{0}, \mathbf{0})$ is locally asymptotically stable. $\square$

Remark 11: When subjected to DoS attacks satisfying Assumptions 1 and 2 (bounded attack frequency and duration), the fast system switches between nominal exponential decay ($\lambda_2 \geq \underline{\lambda}$) and bounded growth intervals ($\lambda_2 = 0$). By applying Average Dwell Time (ADT) switching stability analysis, exponential stability of the fast subsystem is preserved provided the attack duration does not exceed the critical ratio determined by the decay rate $\underline{\lambda}$ and the growth bound during attack intervals. In contrast, topology attacks involving physical node removals represent discrete, irreversible events that alter the state dimension of ξ .

4 Evaluation

4.1 Experimental Setup and Parameters

To demonstrate the performance of the proposed connectivity-maintaining topology reconstruction scheme under malicious attacks, we adopt a linear state-space dynamical system for evaluation. While highly complex, non-linear dynamics with persistent disturbances or parametric uncertainties could be employed, such intricate coupling tends to introduce extraneous fluctuations. This makes it challenging to quantitatively isolate whether the observed tracking metrics (e.g., consensus error) stem from the topology reconstruction mechanism or the unmodeled physical dynamics themselves. Therefore, using a controllable

state-space formulation allows us to transparently benchmark the topological response and attack resilience. Notably, this continuous-time dynamical model directly abstracts real-world engineering systems, such as secondary frequency and voltage coordination in islanded micro-grids at fixed operating points. Specifically, we consider a multi-agent system composed of $N = 10$ agents with the following dynamics.

$$\begin{cases} \dot{x}_i(t) = Ax_i(t) + Bu_i(t) \\ y_i(t) = Cx_i(t) \end{cases}, \quad i \in \mathcal{N}(t) \quad (22)$$

where $x_i(t) = [x_{i,1}(t), x_{i,2}(t), x_{i,3}(t)]^\top \in \mathbb{R}^3$, $A = \mathbf{0}_{3 \times 3} \in \mathbb{R}^{3 \times 3}$, $B = [0, 0, 1]^\top \in \mathbb{R}^{3 \times 1}$, and $C = [0, 0, 1] \in \mathbb{R}^{1 \times 3}$.

The initial state components of the MASs along the x and y axes are independently generated from a truncated normal distribution within the interval $[0, 8]$. To ensure reproducibility, the random seeds for the respective axes are fixed at 18 and 22. The initial communication graph $\mathcal{G}_0$ (Fig. 4) is chosen as a ring topology to ensure basic connectivity.

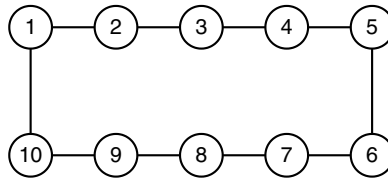


Figure 4: Initial communication topology $\mathcal{G}_0$.

To simulate a harsh networked environment, we implement a hybrid attack model. The intensity of the targeted node removal attack is set as $\alpha_{\text{int}} = 1$. The random seed is set to 42. The attacker identifies and removes the agent with the node with the most neighbors using (4), to maximize the disruption of information flow. The DoS attack follows an average dwell-time constraint with $\tau_{\text{ADT}} = 0.6$ s and an attack frequency $f_D = 1.2$ Hz, targeting the backbone communication links.

The parameterization of the communication quality and energy consumption models is configured as follows. We set the scaling factor $\epsilon = 0.1$, the path loss exponent $\beta = 3$, and the efficiency coefficient $\varphi = 0.8$. To account for heterogeneity across N entities, the vector $\boldsymbol{\rho}_1 = [\rho_{1,1}, \dots, \rho_{1,N}]^\top \in \mathbb{R}^N$ is drawn from a normal distribution within the range $[2, 4]$ with a random seed of 12, while $\boldsymbol{\rho}_2 = \rho_{\text{cst}} \cdot \mathbf{1}_N$ and $\rho_{\text{cst}} = 4$ is a constant. Similarly, the vector of energy coefficient $\boldsymbol{\gamma}^e = [\gamma_1^e, \dots, \gamma_N^e]^\top \in \mathbb{R}^N$ is generated following a normal distribution in the interval $[1, 1.5]$ with the random seed fixed at 36. The remaining key simulation parameters are summarized in Table 1.

Table 1: Other simulation parameters.

Parameters	Value	Parameters	Value	Parameters	Value	Parameters	Value
ϵ_{dis}	0.1	γ	0.99	α_{mar}	3	ζ_3	100
τ_{min}	0.1	λ^{GAE}	0.95	η_{cst}	$N/6$	ζ_4	5
τ_{max}	1.5	ϵ	0.2	ζ_1	2	ζ_5	100
k_{sen}	1	t_{con}	0.01	ζ_2	0.1	ζ_6	100
μ_0	3	$\mathcal{K}$	3				

The performance is evaluated using the following metrics: (i) Resilience Index (RI), (ii) Total Energy Depletion (TED), (iii) Success Rate of Reconnection (SRR), and (iv) Total Connection Time Ratio (TCT).

Specifically, the RI is defined as the weighted accumulation of errors over time, defined by $RI = \int_0^T \|\chi(t)\| dt$ where T is the total running time, and $\chi(t) = \frac{1}{|\mathcal{N}(t)|} \sum_{i \in \mathcal{N}(t)} \left(x_{i,3} - \sum_{j \in \mathcal{N}(t)} \frac{1}{|\mathcal{N}(t)|} x_{j,3} \right)^2 \in \mathbb{R}$ is the consensus mean squared error. The metric TED aims to measure the total energy consumption during the whole simulation process, defined by $TED = \int_0^T E(t) dt$. The SRR indicator characterizes the recovery efficiency of a compromised system's connectivity. Specifically, it is defined as the ratio between the frequency of trials in which the system regains global connectivity from a partitioned state within a given duration and the total count of adversarial perturbations. The SRR is defined by $SRR = \frac{1}{\hat{M}} \sum_{i=1}^{\hat{M}} \mathbb{I}(\lambda_2(L_{\text{recon}}^{(i)}) > 0)$, where $\hat{M}$ is the number of total attacks, $\mathbb{I}(\cdot)$ is the indicator function which equals 1 if the condition is satisfied, otherwise the value is 0. $\lambda_2(L_{\text{recon}}^{(i)})$ is the algebraic connectivity of the Laplacian matrix of the i -th topology reconstruction. $\lambda_2 > 0$ represents the success of the reconnection. The TCT ratio is defined as the proportion of the cumulative duration in which the system maintains global connectivity, characterized by the algebraic connectivity condition $\lambda_2(L) > 0$, relative to the entire simulation horizon. The formula is given by $TCT = \frac{1}{T} \int_0^T \mathbb{I}(\lambda_2(t) > 0) dt$.

4.2 RL Training and Convergence Analysis

To evaluate the effectiveness of the proposed DeepTR in maintaining secure and energy-efficient connectivity for topology reconstruction, we compare the PPO algorithm employed by DeepTR with Asynchronous Advantage Actor-Critic (A2C) [31] and Deep Q-Network (DQN) [32]. As foundational model-free deep reinforcement learning algorithm, A2C and DQN serve as representative baselines to benchmark the optimization performance in dynamic topology synthesis. In addition, we conduct a sensitivity analysis of the PPO method with respect to different hyperparameters, i.e., the clipped range and learning rate. Fig. 5 illustrates the training performance on topology reconstruction environments. From Fig. 5a,b, we see that the PPO employed by DeepTR outperforms the previous methods on the topology reconstruction environment. From Fig. 5c,d, we see that the PPO employed by DeepTR has the best performance when the clipped range is 0.3 and the learning rate is set to 0.001.

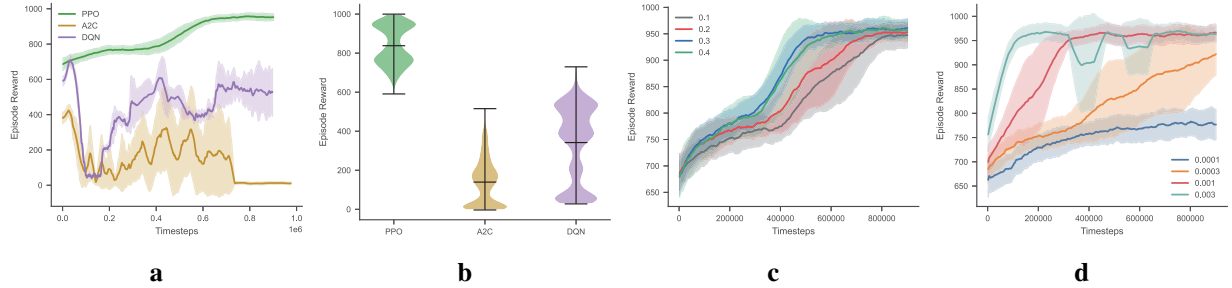


Figure 5: Training performance on topology reconstruction environments, training for one million timesteps. (a) Learning curves of PPO, A2C, and DQN. Solid lines represent smoothed mean rewards, and shaded areas indicate standard deviation across 20 independent runs. (b) Violin plots of PPO, A2C, and DQN. The horizontal line segments within each violin indicate the maximum, minimum, and mean values of the distribution, respectively. The width of the shaded area represents the frequency of the reward values. (c) Learning curves of PPO with clipped ranges 0.1, 0.2, 0.3, and 0.4. (d) Learning curve of PPO with learning rates 0.0001, 0.0003, 0.001, and 0.003.

4.3 Performance under Hybrid Attacks

To evaluate the resilience and efficiency of the TCCTTSF in adversarial environments, we conducted experiments on a ring-based multi-agent system subjected to simultaneous multi-layer attacks. The evaluation focuses on the system's ability to maintain consensus while optimizing energy consumption throughout a 5 s duration, under 20 independent stochastic realizations (different random seeds) to ensure result reliability.

The real-time response of the system under a combination of DoS and node-removal attacks is depicted in Fig. 6. In Fig. 6a, we observe the convergence of the consensus error. During the initial attack-free period ($t < 0.6$ s), all strategies initiate error reduction. However, the onset of DoS attacks (red shaded regions) and intermittent node removals (grey impulses) triggers significant disturbances. The proposed TCCTTSF (Ours) demonstrates better resilience; despite the attacks, it maintains the fastest convergence rate and achieves the lowest residual error. In contrast, the Baseline and Random methods exhibit substantial performance degradation, with their error trajectories lingering at higher values after each attack pulse. The underlying mechanism for this resilience is revealed by the algebraic connectivity (λ_2) in Fig. 6b. Our method actively reconfigures the topology to maintain a significantly higher connectivity margin (stabilizing near $\lambda_2 \approx 0.35$ during recovery) compared to the baseline, which fluctuates near the edge of disconnection ($\lambda_2 \approx 0.12$). This confirms that the TCCTTSF successfully identifies critical nodes and links to preserve network integrity under active interference.

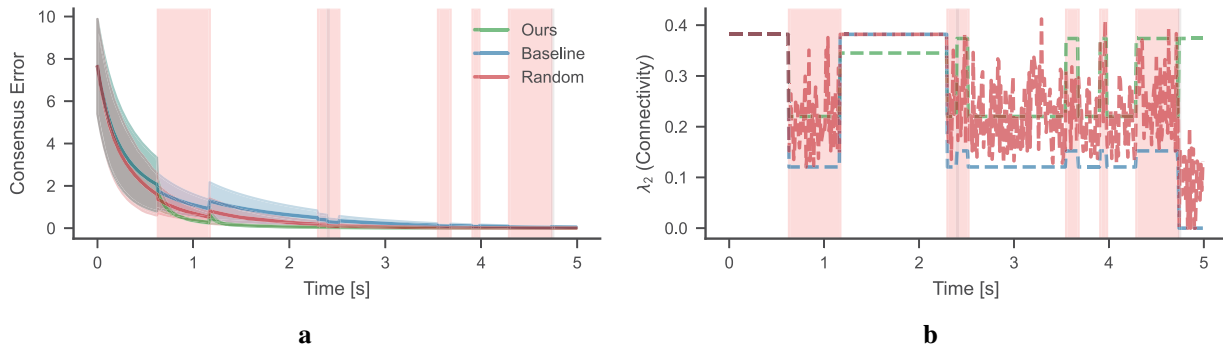


Figure 6: Performance of our proposed method, baseline, and random protocol on energy-efficient connectivity-maintaining consensus problems with hybrid attacks, communicating using the initial topology $\mathcal{G}_0$ from Fig. 4. (a) Trajectories of consensus errors across different strategies, with red shaded areas indicating DoS attack intervals and grey impulses representing node removal events. (b) Evolution of the algebraic connectivity (λ_2) reflecting the topological robustness maintained by the different policies.

To quantify the long-term benefits of the proposed strategy, we analyzed four key performance metrics over the full mission profile, as summarized in Fig. 7. The proposed method achieves the lowest Cumulative Error (RI = 2.91), outperforming the baseline (4.40) by approximately 34%. Simultaneously, it exhibits the highest energy efficiency with a Total Energy Depletion (TED = 6338.17), which is nearly 14% lower than the random reconfiguration strategy. This demonstrates that the TCCTTSF agent effectively balances the trade-off between control performance and energy conservation. Our approach demonstrates exceptional reliability, reaching a SRR of 1.00, indicating that every topology breach was successfully mended. The TCT for our method is maintained at 1.00, ensuring continuous network-wide communication. Conversely, both the Baseline and Random strategies suffer from lower restoration success (≈ 0.86) and fragmented connectivity, highlighting the risk of total system failure without the intelligent dimensionality-adaptive policy. Statistical consistency is evidenced by the error bars representing the standard deviation across 20 seeds.

The minimal variance in our method's SRR and TCT underscores its high deterministic performance in stochastic adversarial environments.

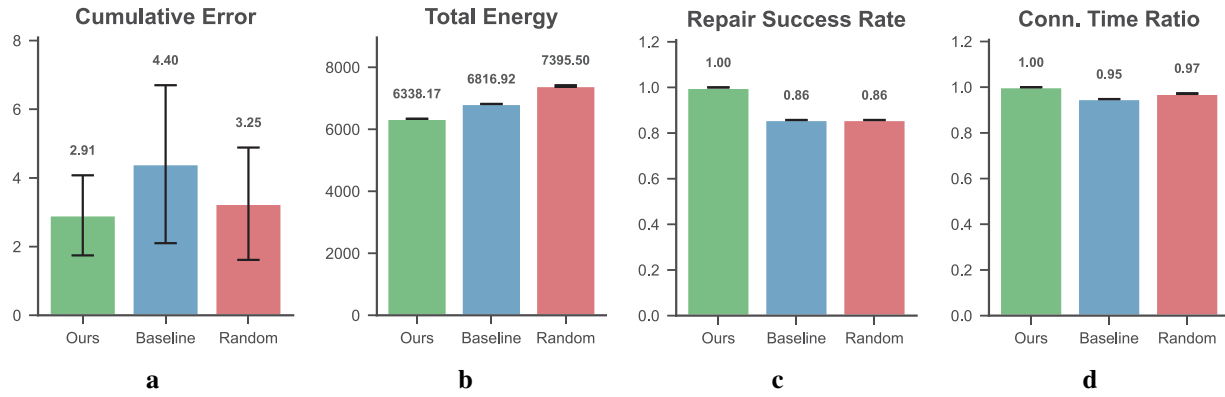


Figure 7: Comparative performance evaluation of connectivity-maintaining consensus strategies under hybrid attacks with topology $\mathcal{G}_0$. (a) Cumulative consensus error (RI); (b) Total energy dissipation (TED); (c) Topology repair success rate (SRR); (d) Connectivity time ratio (TCT).

4.4 Comparative Analysis

Next, we compare our proposed method to several methods from the literature, which are considered effective for resilient multi-agent consensus problems. We compare against tuned implementations of the following algorithms: graph-theoretic structural approach [33], event-trigger-based strategy [34], meta-heuristic search adapted from Enhanced Binary Particle Swarm Optimization (EBPSO) [35], and self-adaptive method [36]. These methods are applicable to scenarios involving connection maintenance or energy conservation in multi-agent systems. Therefore, they are suitable baselines. For our method, we use the hyperparameters from the previous section. To ensure a fair comparison, all experiments are conducted under identical multi-layer attack profiles over a 5 s operational horizon. We execute 10 independent trials with randomized initial states $x_{i,3}(0)$, while $x_{i,1}$, $x_{i,2}$, the attack sequences, and agent trajectories are fixed to maintain consistency across comparisons. The metric of Average Final Error (AFE) is defined as $AFE = \xi(T)$. The metric of Connected Mass (CM) is defined as $CM = \frac{1}{T} \int_0^T \frac{|C_{max}(t)|}{N(0)} dt$, where $|C_{max}(t)|$ is the largest connected component at time t .

Table 2 illustrates the performance of the proposed method and the other methods on the energy-efficient resilient consensus problems. From Table 2, we see that our proposed method outperforms the other compared methods on the energy-efficient resilient consensus environment with respect to metrics AFE, RI, TED, SRR, TCT, and CM. As indicated by the AFE and RI, our method has better error suppression throughout the entire attack duration. TED indicates that our method identifies more sparse yet effective topological structures that fulfill consensus requirements with minimal actuation effort. Furthermore, our method achieves a 100% SRR and a 100% TCT, ensuring that the network remains fully integrated despite adversarial interferences. In terms of CM, our method achieves the highest score of 0.897 alongside SOTA, outperforming other baseline approaches. As shown in Table 2, compared with the state-of-the-art baselines, the proposed method achieves comparable consensus accuracy at slightly lower communication energy consumption. Notably, for the TED metric, the graph-based method, event-based method, SOTA, and our proposed method achieve a standard deviation of zero due to the deterministic deployment of fixed agent positions across evaluation runs in the experimental setup.

Table 2: Comparison of the proposed method and other algorithms on the energy-efficient resilient consensus problems. Data are shown as mean $\pm$ standard deviation across 10 independent runs. **Bold** indicates the best performance.

Algorithm	AFE ($\downarrow$)	RI ($\downarrow$)	TED ($\downarrow$)	SRR ($\uparrow$)/TCT ($\uparrow$)/CM ($\uparrow$)
Graph-based method [33]	$5.22 \pm 4.06 \times 10^{-3}$	3.38 ± 2.19	6697.22 ± 0.00	85.71%/94.80%/0.614
Event-based method [34]	$4.06 \pm 3.37 \times 10^{-3}$	3.13 ± 2.10	6721.15 ± 0.00	85.71%/94.80%/0.628
Adapted from EBPSO [35]	$0.87 \pm 0.60 \times 10^{-3}$	3.02 ± 1.50	6415.70 ± 14.13	100%/98.4%/0.812
SOTA [36]	$0.78 \pm 0.47 \times 10^{-3}$	2.97 ± 1.67	6369.39 ± 0.00	100%/100%/0.897
Ours	$0.66 \pm 0.38 \times 10^{-3}$	2.91 ± 1.42	6338.17 ± 0.00	100%/100%/0.897

4.5 Ablation Study

To investigate the individual contribution of the AETS and the PAASR mechanisms, we conduct an ablation study. The mechanisms we evaluate are (1) only AETS, (2) only PAASR, and (3) the integrated AETS + PAASR framework. Table 3 illustrates the performance of the three mechanisms on the energy-efficient resilient consensus problems. From Table 3, we see that the mechanism utilizing only AETS achieves the highest energy efficiency. However, this efficiency comes at the cost of reduced reliability. As shown in the safety metrics, while it maintains a 100% SRR, its TCT drops to 98.8%. This indicates that while AETS effectively minimizes operational costs, it lacks the necessary restorative strength to counter sustained adversarial interferences, leading to transient disconnections. Notably, the only PAASR mechanism achieves a remarkably low AFE of 0.52×10^{-8} . This precision is a deterministic physical consequence of unconstrained topology reinforcement rather than a random statistical artifact. Without the energy-conserving regulation of AETS, the PAASR mechanism restores damaged links using maximum power, forming densely connected, highly redundant graph topologies. In spectral graph theory, the consensus convergence rate is lower-bounded by the algebraic connectivity $\lambda_2(L)$, with the tracking error decaying as $\mathcal{O}(e^{-\lambda_2 t})$. The unconstrained redundancy under the only PAASR mechanism generated an exceptionally large λ_2 , driving the system to near-instantaneous state alignment and pushing the terminal error down to numerical precision limits (10^{-8}) within 5 s. However, this precision exposes a performance trade-off between control accuracy and resource sustainability. By integrating AETS with PAASR, the proposed framework established a trade-off operating point: it preserves the security guarantees of PAASR (100% SRR and 100% TCT) and maintains high operational precision (0.66×10^{-3}), while reducing total energy consumption by 70.7%. In practical networked engineering systems, a tracking error on the order of 10^{-3} is already well within operational tolerances for consensus maintenance. This explicit trade-off validates that the synergy between both mechanisms is essential for deploying long-term resilient multi-agent systems under battery or transmission power constraints.

Table 3: Comparison of only AETS mechanism, only PAASR mechanism, and AETS + PAASR mechanism on the energy-efficient resilient consensus problems. Data are shown as mean $\pm$ standard deviation across 10 independent runs. **Bold** indicates the best performance.

Mechanism	AFE ($\downarrow$)	RI ($\downarrow$)	TED ($\downarrow$)	SRR ($\uparrow$)/TCT ($\uparrow$)
Only AETS	$1.13 \pm 0.48 \times 10^{-3}$	3.84 ± 1.88	6189.29	100%/98.8%
Only PAASR	$0.52 \pm 0.36 \times 10^{-8}$	1.07 ± 0.60	21662.66	100%/100%
AETS + PAASR	$0.66 \pm 0.38 \times 10^{-3}$	2.91 ± 1.42	6338.17	100%/100%

4.6 Robustness Analysis

To evaluate the resilience of the TCCTTSF against varying levels of adversarial pressure, we conducted a robustness analysis by modulating the intensity of the node-removal attacks. We defined three intensity levels, characterized by the scaling parameter $\alpha_{\text{int}} \in \{1, 1.5, 2\}$, representing a progressive escalation in the impact of malicious interference. Table 4 illustrates the performance of the previous attack intensities on the energy-efficient resilient consensus problems. From Table 4, we see that as the intensity α_{int} increases from 1 to 2, the RI remains stable at approximately 2.89, with negligible fluctuations in its standard deviation. Intriguingly, the AFE exhibits a counterintuitive improvement, decreasing from 0.66×10^{-3} to 0.26×10^{-3} . This phenomenon is primarily attributed to the reduction in the active agent population caused by the intensified node-removal attacks. Under higher α_{int} , the removal of a larger number of agents effectively simplifies the network's collective dynamics and increases $\lambda_2(\mathcal{G})$, because the energy constraint limits the number of links, and fewer agent nodes leads to a bigger $\lambda_2(\mathcal{G})$. With fewer remaining agents to synchronize, the residual sub-network attains a lower-dimensional state space, which facilitates more rapid state alignment and a more precise consensus in the terminal phase. Thus, the diminished final error reflects the inherent convergence advantages of a reduced-scale system, even though it operates under more severe structural depletion. As theoretically expected for sparse communication topologies, smaller surviving network scales yield larger algebraic connectivity λ_2 . In our attack simulation, as the surviving node count decreases from 10 to 8, 6, and 4 under high attack intensity, the average non-zero λ_2 increases from 0.62, 0.92, 1.70, 4, respectively. The system's energy expenditure reflects a rational and adaptive response to the adversarial environment. The TED exhibits a controlled linear-like increase, rising from 6338.17 ($\alpha_{\text{int}} = 1$) to 8140.30 ($\alpha_{\text{int}} = 2$). This is due to the PAASR mechanism. This increase of approximately 28% is a necessary investment for maintaining network integrity under intensified removal attempts. Despite the doubled attack intensity, the SRR and TCT remain invariant at a perfect 100%. Additionally, the CM exhibits a decreasing trend from 0.897 to 0.462 as α_{int} scales from 1 to 2. This reduction is attributed to the substantial loss of active agent nodes (from 8 down to 4), which inevitably shrinks the network's structural redundancy and topological consensus margin, even though the remaining sub-network achieves higher state alignment precision.

Table 4: Performance of different attack intensities on the energy-efficient resilient consensus problems. Data are shown as mean $\pm$ standard deviation across 10 independent runs.

Intensity	AFE ($\downarrow$)	RI ($\downarrow$)	TED ($\downarrow$)	SRR ($\uparrow$)/TCT ($\uparrow$)/CM ($\uparrow$)
$\alpha_{\text{int}} = 1$, (8 nodes left)	$0.66 \pm 0.38 \times 10^{-3}$	2.91 ± 1.42	6338.17	100% /100%0.897
$\alpha_{\text{int}} = 1.5$, (6 nodes left)	$0.35 \pm 0.28 \times 10^{-3}$	2.89 ± 1.46	7693.27	100%/100%0.671
$\alpha_{\text{int}} = 2$, (4 nodes left)	$0.26 \pm 0.39 \times 10^{-3}$	2.89 ± 1.52	8140.30	100%/100%0.462

4.7 Scalability Analysis

To evaluate the scalability of our proposed method, we evaluate DARL on the energy-efficient resilient consensus problem across varying numbers of agents N . It is worth emphasizing that for every distinct agent scale N , a dedicated, per-dimension policy is pre-trained and executed in accordance with the DARL framework. The metric Average Connected Time (ACT) denotes the average value for topological connection inference time. The scalability results are presented in Table 5. As shown in Table 5, when using a compact hidden architecture of 128×128 , our method achieves robust performance for $N \leq 10$. However, when a dedicated 128×128 policy is trained for $N = 11$, a distinct performance cliff is observed: the AFE and RI increase, SRR decreases by 24.29%, and TCT decreases by 5.20%. This performance drop does not stem from policy misallocation or zero-shot extrapolation; rather, it reveals a *neural network capacity bottleneck*—an

architecture of 128×128 lacks the representational expressiveness needed to capture the exponential growth in state-action interactions for $N \geq 11$. By scaling the policy network capacity to $256 \times 256 \times 128$, DARL successfully recovers optimal connectivity and consensus precision for $N \in [11, 23]$, before encountering another representational limit at $N = 24$. Similarly, further expanding the network capacity to $512 \times 512 \times 256$ maintains system connectivity up to $N = 32$, with $N = 33$ starting to stress this capacity tier. These empirical observations provide a key design insight for scalable MARL: maintaining resilient consensus under hybrid attacks requires non-linearly scaling the policy network capacity alongside the agent dimension N . In addition, as the number of agents increases, the metric ACT increases accordingly, indicating that our method may introduce inference delays when deployed in very large-scale systems. Meanwhile, TED indicates that deploying more agents incurs higher total communication energy, reflecting the inherent physical overhead of scale.

Table 5: Performance of different numbers of agents on the energy-efficient resilient consensus problems. Data are shown as mean $\pm$ standard deviation across 10 independent runs.

Scalability (Hidden Layer)	AFE ($\downarrow$)	ACT ($\downarrow$)	RI ($\downarrow$)	TED ($\downarrow$)	SRR ($\uparrow$)/TCT ($\uparrow$)
$N = 10$ (128×128)	$0.66 \pm 0.38 \times 10^{-3}$	$2.03 \pm 0.28 \times 10^{-3}$	2.91 ± 1.42	6338.17	100%/100%
$N = 11$ (128×128)	$4.87 \pm 4.43 \times 10^{-3}$	$2.07 \pm 0.46 \times 10^{-3}$	3.22 ± 2.11	6417.60	85.71%/94.80%
$N = 23$ ($256 \times 256 \times 128$)	$1.47 \pm 1.12 \times 10^{-3}$	$3.17 \pm 0.53 \times 10^{-2}$	6.96 ± 3.57	14493.68	100%/100%
$N = 24$ ($256 \times 256 \times 128$)	$10.83 \pm 8.60 \times 10^{-3}$	$3.22 \pm 0.61 \times 10^{-2}$	9.35 ± 4.12	14378.61	71.43%/77.40%
$N = 32$ ($512 \times 512 \times 256$)	$5.43 \pm 2.31 \times 10^{-3}$	$9.24 \pm 1.48 \times 10^{-2}$	25.24 ± 12.95	28595.80	100%/100%
$N = 33$ ($512 \times 512 \times 256$)	$51.14 \pm 39.86 \times 10^{-3}$	$9.87 \pm 1.63 \times 10^{-2}$	66.24 ± 39.74	24976.14	57.14%/55.60%

5 Discussion

While the numerical effectiveness of the proposed TCCTTSF has been established, implementing the algorithm in real-world cyber-physical systems environments (e.g., islanded microgrids, industrial autonomous fleets) requires careful consideration of physical hardware and networking constraints. During operation, the agent only executes a single forward pass through the trained policy network. Execution profiling on an Apple M2 acting as a desktop-class ARM surrogate benchmarking system indicates an average inference time of 2 ms per control step for 10 agents. In practical wireless networks, transmission delays and packet losses can corrupt state feedback. The proposed AETS mechanism enforces a minimum dwell-time constraint ($\tau_{\min} = 0.1$ s), which inherently acts as a low-pass filter to reject transient channel delays and prevent high-frequency control chatter under network jitter. The topology reconstruction logic requires time-stamped state information to evaluate algebraic connectivity λ_2 . Standard industrial synchronization protocols, such as IEEE 1588 Precision Time Protocol (PTP), provide sub-millisecond clock alignment, which is more than sufficient for the proposed event-triggered sampling horizons. The complete policy library consisting of 31 pre-trained models (for $N = 3$ to 33) occupies a total memory footprint of 15.3 MB, enabling execution on modern edge platforms. To rigorously isolate the impact of cyber-attack intensity and control topology optimization, all evaluation scenarios utilized a fixed initial spatial configuration for the agents. This deterministic setup results in zero standard deviation across runs for energy metrics such as TED. Generalizing the dynamic policy evaluation across randomized initial spatial topologies remains an important target for future work.

6 Conclusion

In this study, we addressed the core challenge of balancing global connectivity, energy efficiency, and adversarial resilience in multi-agent systems by proposing a novel, two-time-scale TCCTTSF. This unified

closed-loop architecture integrates a detection-isolation-survival protocol to ensure resilient consensus under unpredictable, hybrid malicious attacks. Crucially, to circumvent the curse of dimensionality in network optimization, we developed a DARL-based paradigm that models the topological reconstruction process exclusively on the upper triangular elements of the adjacency matrix. This formulation significantly compresses the state and action spaces, enabling agents to autonomously discover energy-efficient, connected topologies without expert-designed topology-construction heuristics. Empirical evaluations under hybrid attacks—including periodic DoS interferences and node removals—demonstrate that our framework consistently defines the Pareto frontier, achieving a 100% TCT and 100% SRR while cutting energy consumption compared to state-of-the-art baselines. These results confirm that secure consensus is a direct consequence of intelligent topological adaptation.

Acknowledgement: None.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: The authors confirm contribution to the paper as follows: Conceptualization, Hua-Yu Zhu and Weijie Mao; methodology, Hua-Yu Zhu; software, Hua-Yu Zhu; validation, Hua-Yu Zhu and Weijie Mao; formal analysis, Hua-Yu Zhu; investigation, Hua-Yu Zhu; resources, Weijie Mao; writing—original draft preparation, Hua-Yu Zhu; writing—review and editing, Hua-Yu Zhu and Weijie Mao; visualization, Hua-Yu Zhu; supervision, Weijie Mao; project administration, Weijie Mao. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: All reviewed materials are publicly available and cited, primarily sourced from major academic repositories and publishers, including IEEE Xplore, ACM Digital Library, ScienceDirect (Elsevier), SpringerLink, AAAI, Nature, and the arXiv repository. The simulation datasets and custom software code generated and analyzed during the current study (including the experimental metrics in [Figs. 5–7](#) and [Tables 1–5](#)) are not publicly available due to privacy and institutional restrictions, but are available from the corresponding author upon reasonable request.

Ethics Approval: Not applicable.

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

References

1. Cofta P. Simplifying the operation of hybrid ad-hoc sensor networks with neural networks as the sole data reconstruction method. *Expert Syst Appl.* 2026;2025(1–3):128981. doi:10.1016/j.eswa.2025.128981.
2. Zhang Y, Wu H, Cao J. Prescribed-time synchronization for stochastic reaction-diffusion networks with delayed impulse and application in secure communication. *Expert Syst Appl.* 2026;301(3):130308. doi:10.1016/j.eswa.2025.130308.
3. Wang C, Jiang Y, Zhu S, Song LW, Du X, Guan X. Collision-free connectivity topology switching and maintenance for autonomous underwater vehicles. *IEEE Trans Control Syst Technol.* 2025;33(6):2021–36. doi:10.1109/tcst.2025.3570095.
4. Ou J, Song G, Guo J, Cao Y, Wang Y. GPU-enabled decentralized, multi-robot path planning based on global evolutionary dynamic programming and local particle swarm optimization. *Expert Syst Appl.* 2026;321(1):132321. doi:10.1016/j.eswa.2026.132321.
5. Wang S, Sun J, Fang W, Palade V. RDPSO-based multi-robot cooperation method for target search in unknown environments. *Expert Syst Appl.* 2025;302(10):130471. doi:10.1016/j.eswa.2025.130471.
6. Gorman S. Electricity grid in US penetrated by spies. *Wall Street J.* 2009;8(8):1–3.
7. Rid T. Cyber war will not take place. *J Strategic Stud.* 2012;35(1):5–32. doi:10.1080/01402390.2011.608939.
8. Cherdantseva Y, Burnap P, Blyth A, Eden P, Jones K, Soulsby H, et al. A review of cyber security risk assessment methods for SCADA systems. *Comput Secur.* 2016;56(9):1–27. doi:10.1016/j.cose.2015.09.009.

9. Liang G, Weller SR, Zhao J, Luo F, Dong ZY. The 2015 Ukraine blackout: implications for false data injection attacks. *IEEE Trans Power Syst.* 2016;32(4):3317–8.
10. Wang H, Chen F, Wang Y. Co-design of distributed dynamic event-triggered scheme and extended dissipative consensus control for singular Markov jumping multi-agent systems under periodic denial-of-service jamming attacks. *Expert Syst Appl.* 2024;251(3):124091. doi:10.1016/j.eswa.2024.124091.
11. Zhang W, Wang Z, Liu Y, Ding D, Alsaadi FE. Sampled-data consensus of nonlinear multiagent systems subject to cyber attacks. *Int J Robust Nonlinear Control.* 2018;28(1):53–67. doi:10.1002/rnc.3855.
12. Xu W, Hu G, Ho DW, Feng Z. Distributed secure cooperative control under denial-of-service attacks from multiple adversaries. *IEEE Trans Cybern.* 2019;50(8):3458–67. doi:10.1109/tcyb.2019.2896160.
13. Zhang P, Shan Q, Li T, Chen CP. Multi-consensus recovery of multi-agent system with cyber-attack based on topology reconstruction controller. *Int J Robust Nonlinear Control.* 2022;32(18):9882–901. doi:10.1002/rnc.6361.
14. Yang H, Han QL, Ge X, Ding L, Xu Y, Jiang B, et al. Fault-tolerant cooperative control of multiagent systems: a survey of trends and methodologies. *IEEE Trans Ind Inform.* 2019;16(1):4–17.
15. Ma HJ, Yang GH. Adaptive fault tolerant control of cooperative heterogeneous systems with actuator faults and unreliable interconnections. *IEEE Trans Autom Control.* 2015;61(11):3240–55. doi:10.1109/tac.2015.2507864.
16. Yang HE, Jiang B, Zhang Y. Fault-tolerant shortest connection topology design for formation control. *Int J Control Autom Syst.* 2014;12(1):29–36. doi:10.1007/s12555-012-9315-x.
17. Lin Z, Wang L, Han Z, Fu M. A graph Laplacian approach to coordinate-free formation stabilization for directed networks. *IEEE Trans Autom Control.* 2015;61(5):1269–80. doi:10.1109/tac.2015.2454711.
18. Wang G, Luo H, Hu X. Communication topology reconstruction for a three-dimensional persistent formation with fault constraint. *IEEE Trans Netw Sci Eng.* 2024;11(6):6574–88. doi:10.1109/tNSE.2024.3457489.
19. Xu Y, Zhang W, Zhang Y, Wu ZG. Distributed resilient control of cyber-physical multi-agent systems under attacks. *IEEE Trans Ind Cyber Phys Syst.* 2025;3:241–50. doi:10.1109/ticps.2025.3543291.
20. Anand A, Guha D, Purwar S. Observer-aided adaptive consensus control of cyber-physical multi-agent systems with hybrid attacked topology. *IEEE Trans Autom Sci Eng.* 2026;23:5250–61. doi:10.1109/tase.2026.3663295.
21. Dai R, Mesbahi M. Optimal topology design for dynamic networks. In: *Proceedings of the 2011 50th IEEE Conference on Decision and Control and European Control Conference*; 2011 Dec 12–15; Orlando, FL, USA. p. 1280–5.
22. Perolat J, De Vylder B, Hennes D, Tarassov E, Strub F, de Boer V, et al. Mastering the game of stratego with model-free multiagent reinforcement learning. *Science.* 2022;378(6623):990–6. doi:10.1126/science.add4679.
23. Degraeve J, Felici F, Buchli J, Neunert M, Tracey B, Carpanese F, et al. Magnetic control of tokamak plasmas through deep reinforcement learning. *Nature.* 2022;602(7897):414–9. doi:10.1038/s41586-021-04301-9.
24. de Mol B, Barbieri D, Viebahn J, Grossi D. Centrally coordinated multi-agent reinforcement learning for power grid topology control. In: *Proceedings of the 16th ACM International Conference on Future and Sustainable Energy Systems*; 2025 Jul 14–17; Bangalore, India. p. 460–75.
25. Li X, Wang X, Bai C, Zhang J. Exponential topology-enabled scalable communication in multi-agent reinforcement learning. *arXiv:2502.19717.* 2025.
26. Nakada R, Imaizumi M. Adaptive approximation and generalization of deep neural network with intrinsic dimensionality. *J Mach Learn Res.* 2020;21(174):1–38.
27. Liu S, Jiang B, Mao Z, Zhang Y. Neural-network-based adaptive fault-tolerant cooperative control of heterogeneous multiagent systems with multiple faults and DoS attacks. *IEEE Trans Neural Netw Learn Syst.* 2023;35(5):6273–85. doi:10.1109/tnnls.2023.3282234.
28. Fiedler M. Algebraic connectivity of graphs. *Czechoslovak Math J.* 1973;23(2):298–305. doi:10.21136/cmj.1973.101168.
29. Olfati-Saber R, Murray RM. Consensus problems in networks of agents with switching topology and time-delays. *IEEE Trans Autom Control.* 2004;49(9):1520–33. doi:10.1109/tac.2004.834113.
30. Schulman J, Wolski F, Dhariwal PW, Radford A, Klimov O. Proximal policy optimization algorithms. *arXiv:1707.06347.* 2017.

31. Mnih V, Badia AP, Mirza M, Graves A, Lillicrap T, Harley T, et al. Asynchronous methods for deep reinforcement learning. *Proc Mach Learn Res.* 2016;48:1928–37.
32. Mnih V, Kavukcuoglu Kr, Silver D, Graves A, Antonoglou I, Wierstra D, et al. Playing atari with deep reinforcement learning. *arXiv:1312.5602.* 2013.
33. Zou L, Liu X, Zhang X, Su H. Resilient consensus for multi-agent systems under distributed denial-of-service attacks: a graph-based approach. *J Control Decis.* 2024;69(4):1–13. doi:10.1080/23307706.2024.2442464.
34. Zhang Y, Wu ZG, Kwok KW, Huang T, Chakrabarti P, Zhou L. Secure event-based consensus control for multi-agent systems under DoS attacks and input saturation. *IEEE Trans Syst Man Cybern Syst.* 2025;55(7):4816–27. doi:10.1109/tsmc.2025.3558286.
35. Tijjani S, Ab Wahab MN, Noor MHM. An enhanced particle swarm optimization with position update for optimal feature selection. *Expert Syst Appl.* 2024;247(7):123337. doi:10.1016/j.eswa.2024.123337.
36. Huang J, Chen M, Liu Y. Achieving resilient and self-adaptive topology configuration in 3D UAV networks. *ACM Trans Internet Technol.* 2026;26(1):1–30. doi:10.1145/3747350.