future internet

MDPI

Systematic Review

Cyber-Resilience of Decentralized Cyber-Physical Systems: A PRISMA-Guided Systematic Mapping, Taxonomy, and Evidence-Based Assessment Framework

Kakim Sagindykov [1], Nazym Mukusheva [1,*], Dinara Tokseit [1], Altyn Urynbassarova [1] and Shyryn Tutkyshbayeva [2]

[1] Department of Information Security, L.N. Gumilyov Eurasian National University, Astana 010000, Kazakhstan; sagindykov_km@enu.kz (K.S.); tokseit1990@gmail.com (D.T.); urynbassarova_azh@enu.kz (A.U.)
[2] School of Software Engineering, Astana IT University, Astana 010000, Kazakhstan; sh.tutkyshbayeva@astanait.edu.kz
* Correspondence: nazymmg@gmail.com

Abstract

Decentralized cyber–physical systems (CPSs) improve scalability and fault tolerance but expose local controllers and neighbor-to-neighbor coordination to integrity, availability, and multi-agent attacks. This study presents a PRISMA-guided systematic mapping of cyber-resilience methods for decentralized CPSs and integrates the evidence into a taxonomy and assessment framework. A Scopus search exported on 14 September 2026 yielded 366 records; 337 primary records were screened, 163 underwent eligibility assessment, 79 were included, and 72 met a six-criterion quality threshold for Core synthesis. Power systems and microgrids represented 55.6% of Core studies, false-data/deception attacks were the largest primary attack family (31.9%), and only 18.1% evaluated explicit multi-vector or hybrid attacks. Simulation-only validation dominated (70.8%), while 29.2% provided strong experimental, real-time, hardware-in-the-loop, laboratory, or co-simulation evidence. The synthesis shows comparatively mature work on resilient control, detection, observer-based compensation, and consensus but limited evidence on recovery, self-healing, safety preservation, cross-domain transfer, and reproducible multi-vector evaluation. The proposed framework connects attack models, topology assumptions, resilience mechanisms, validation strength, and measurable outcomes, providing a traceable basis for comparing methods and designing stronger future evaluations.

Keywords: cyber–physical systems; cyber-resilience; decentralized systems; systematic mapping; PRISMA; resilient control; multi-agent systems; false data injection; denial-of-service; resilience assessment

Check for updates

Academic Editor: Wei Yu

Received: 14 October 2026
Revised: 5 November 2026
Accepted: 10 November 2026
Published: 12 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Cyber–physical systems (CPSs) integrate computation, communication, control and physical processes through feedback loops in which embedded devices and networks monitor, coordinate and actuate physical behavior. This understanding is consistent with the early CPS definition proposed by Lee, where computation and physical processes are tightly coupled, and with the NIST CPS framework, which describes CPSs as engineered systems combining digital, analog, physical, and human components [1,2]. In practical

terms, CPSs underpin smart grids, industrial automation, intelligent transportation, cooperative robotics, medical devices, and other critical services whose disruption can create physical, economic, and societal consequences [3].

The architecture of CPSs is increasingly shifting from centralized control toward distributed and decentralized decision-making. This shift is driven by the scale of modern infrastructures, the need for low-latency edge decisions, autonomy of local subsystems, privacy constraints, and the high cost of maintaining a single trusted control center. Recent work on distributed and resilient industrial CPS architectures also highlights multi-agent systems, service-oriented architectures, virtualization, and autonomous capabilities as enabling features for scalable and resilient industrial deployments [4]. In a decentralized CPS, the system-level objective is achieved through local controllers or agents that exchange information with neighboring agents over a communication graph rather than relying on a central authority. A generic local model can be expressed as follows:

$$x_i(k+1) = f_i(x_i(k), u_i(k), w_i(k)), \qquad u_i(k) = \pi_i(x_i(k), \{y_j(k) : j \in \mathcal{N}_i\}), \tag{1}$$

where agent i updates its physical state x_i using a local control input u_i, uncertainty or disturbance w_i, and neighbor information y_j received from the set of in-neighbors $\mathcal{N}_i$. Equation (1) emphasizes the central difficulty considered in this article: resilience cannot be evaluated only at the level of an isolated device, because the behavior of each benign node depends on both its physical dynamics and the trustworthiness of information received from neighboring nodes.

Decentralization improves scalability, modularity, and robustness against single-point failures, yet it also expands the attack surface. Communication links may be disrupted through denial-of-service, delay, or packet-dropping attacks; sensor and controller data may be manipulated through false data injection, or replay attacks; and individual agents may become malicious, Byzantine, or compromised. A recent survey on security control of CPSs identifies denial-of-service, false data injection, and replay attacks as major attack classes affecting CPS operation and integrity while also emphasizing the role of detection and secure control methods [5]. For decentralized CPSs, these attacks are more difficult to address because no single component has complete situational awareness.

The cyber-resilience perspective is therefore broader than conventional prevention-oriented cybersecurity. NIST cyber-resiliency engineering defines the objective of cyber-resilient systems in terms of anticipating, withstanding, recovering from, and adapting to adverse cyber conditions, stresses, attacks, or compromises [6]. In the CPS domain, this definition must be interpreted jointly with physical performance, safety margins, and service continuity. Thus, a cyber-resilient decentralized CPS is not merely a system that blocks attacks; it is a system that can continue to deliver acceptable mission functionality under attack, degrade gracefully when full performance is impossible, and restore coordinated operation after disruption.

For performance-based reasoning, resilience can be represented through the normalized service or performance trajectory $P(t)/P_0$ during and after an attack. The following simplified metric captures the average retained functionality over a time horizon T, while D expresses the maximum relative degradation:

$$R_T = \frac{1}{T}\int_{t_0}^{t_0+T} \frac{P(t)}{P_0}\, dt, \qquad D = 1 - \frac{P_{\min}}{P_0}. \tag{2}$$

Figure 1 illustrates the conceptual scope adopted in this review, linking decentralized cyber–physical operation with cyber attack exposure, resilience mechanisms, physical performance, and recovery.

The recent literature confirms that resilience is becoming a central research issue for decentralized and multi-agent CPSs. Liao et al. emphasize that the openness and distributed characteristics of CPSs make resilience against malicious attacks a critical problem, and their survey classifies resilient coordination results from the perspectives of physical structure, network topology, and communication mechanism [7]. In parallel, event-triggered resilient distributed optimization has been proposed to reduce communication overhead while tolerating malicious agents and has been validated in multi-microgrid scenarios [8]. These developments indicate a clear movement from isolated security controls toward coordinated, graph-aware, and resource-aware resilience mechanisms.

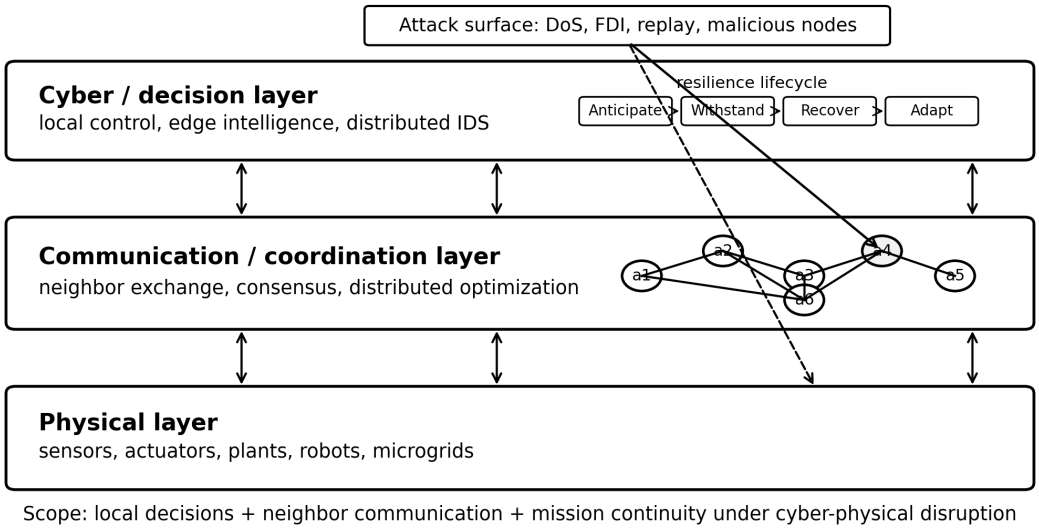


Figure 1. Conceptual scope of cyber-resilience in decentralized cyber–physical systems. Nodes a_1–a_6 denote local agents in the communication/coordination layer; inter-agent lines denote neighbor communication links; vertical bidirectional arrows denote cyber–physical information/control exchange between layers; the solid and dashed arrows from the attack-surface box illustrate representative attack-propagation paths.

However, the current body of work remains fragmented. Some studies focus on attack detection, others on resilient consensus, secure state estimation, distributed optimization, trust management, recovery control, or cyber–physical resilience indicators. A survey on cyber-resilience approaches for CPSs notes that the literature contains a broad variety of techniques and that practical aspects such as metrics, evaluation methods, testing, and validation environments remain open challenges [9]. More recent work on resilience key performance indicators (R-KPIs) for cyber–physical infrastructures also shows that functionality loss, recovery time, and minimum performance level after disturbance are important indicators for resilience modeling [10]. In addition, cyber–physical resilience research increasingly connects technical indicators with organizational and regulatory aspects of critical infrastructure protection [11]. These observations motivate a unified view that connects taxonomy, comparative analysis and assessment metrics specifically for decentralized CPSs. Recent 2025–2026 reviews reinforce both the timeliness of this objective, and the need to delimit the present contribution. General resilience-enhancement surveys synthesize broad CPS techniques, while systematic reviews have examined security modeling, cyber-resilience metrics in industrial IoT, resilience mechanisms in IoT networks, and sustainable/resilient cyber–physical production systems [12–17]. These studies provide valuable domain- or function-specific syntheses; however, they do not jointly quantify how attack family, decentralization/topology assumptions, resilience mechanism, validation strength, and recovery evidence co-vary across decentralized CPS studies. The present mapping is designed around that intersection.

The research gap addressed in this article is therefore not the absence of individual algorithms but the lack of an integrated framework for organizing and evaluating cyber-resilience methods under decentralized assumptions. In particular, existing studies do not always make explicit whether a method requires global information, a trusted coordinator, static topology, synchronous communication, a bounded number of malicious neighbors, or specific graph robustness conditions. These assumptions are decisive for real deployments, because decentralized CPSs such as microgrids, robotic swarms, and industrial Internet-of-Things networks operate under communication constraints, dynamic topology, and heterogeneous device capabilities.

To address this gap, this article develops a taxonomy and assessment framework for cyber-resilience methods in decentralized CPSs. The proposed taxonomy organizes methods by threat model, CPS layer, communication topology, resilience mechanism, and evaluation evidence. The comparative analysis then examines how representative methods differ in terms of attack coverage, information requirements, scalability, communication overhead, recovery capability, and validation setting. Finally, the assessment framework translates cyber-resilience into measurable dimensions, allowing methods to be compared beyond qualitative claims.

A compact assessment score may be represented as a weighted aggregation of normalized indicators, where $\hat{z}_m$ denotes a normalized metric such as detection capability, retained functionality, convergence accuracy, recovery speed, communication efficiency, or service availability, and w_m represents the importance assigned to that metric for a given CPS application:

$$RS = \sum_{m=1}^{M} w_m \hat{z}_m, \qquad \sum_{m=1}^{M} w_m = 1, \quad 0 \le \hat{z}_m \le 1. \tag{3}$$

Accordingly, the objective of this article is to answer the following research question: how can cyber-resilience methods for decentralized CPSs be systematically classified, compared, and assessed under malicious-node and communication attacks? This article focuses on decentralized and multi-agent CPSs where benign agents rely on local computation and neighbor communication. The scope includes distributed attack detection, resilient consensus, trust-based filtering, resilient distributed optimization, reconfiguration, and recovery methods. Purely centralized security architectures and purely cryptographic protocol design are outside the main scope, except where they support decentralized resilience mechanisms.

The remainder of this article is organized as follows. Section 2 summarizes the research contributions, scope, and guiding research questions. Section 3 introduces background concepts and distinguishes centralized, distributed, and decentralized CPS architectures. Section 4 defines the threat model. Section 5 reports the systematic-review methodology. Section 6 develops the taxonomy of cyber-resilience methods, and Section 7 provides the criterion-based comparative analysis. Section 8 presents the systematic evidence synthesis, while Section 9 discusses the findings against the taxonomy and RQ1–RQ4. Section 10 develops the cyber-resilience assessment framework, Section 11 presents the decentralized microgrid case study, Section 12 identifies open challenges and future research directions, and Section 13 concludes this article.

2. Research Contributions and Scope

Cyber-resilience in decentralized CPSs spans security, distributed control, graph-based coordination, and physical performance. Existing work treats attack detection, resilient consensus/optimization, and resilience assessment largely as separate streams [5–8,18]. This article connects them through a common structure linking threat models, topology

and information assumptions, resilience mechanisms, validation strength, and measurable outcomes.

2.1. Contribution 1: Taxonomy of Cyber-Resilience Methods

The first contribution is a taxonomy that organizes cyber-resilience methods according to five dimensions: threat model, CPS layer, decentralization assumption, resilience mechanism, and validation evidence. The taxonomy is designed to distinguish methods that may appear similar at a high level but differ substantially in deployment assumptions. For example, a detection method that requires global state information is not equivalent to a neighbor-based anomaly detector; similarly, a resilient consensus algorithm that assumes a fixed strongly robust graph cannot be directly compared with a method designed for time-varying or directed topologies.

Formally, a cyber-resilience method can be represented as a tuple

$$\mathcal{M} = (\mathcal{A}, \mathcal{L}, \mathcal{I}, \mathcal{G}, \mathcal{R}, \mathcal{E}), \tag{4}$$

where $\mathcal{A}$ denotes the addressed attack class, $\mathcal{L}$ the CPS layer, $\mathcal{I}$ the required information model, $\mathcal{G}$ the graph or topology assumption, $\mathcal{R}$ the resilience mechanism, and $\mathcal{E}$ the evaluation evidence. This representation makes explicit which conditions are necessary for a method to be applicable to a real decentralized CPS.

Unlike a purely attack-oriented classification, the proposed taxonomy is intended to reflect the multi-layered nature of decentralized CPSs. It separates communication attacks from malicious-agent behavior, distinguishes cyber-layer detection from physical-layer recovery, and identifies whether a method provides resistance, absorption, adaptation, or recovery. This structure follows the broader cyber-resiliency perspective promoted by NIST, where systems should be engineered to anticipate, withstand, recover from, and adapt to adverse cyber conditions [6].

2.2. Contribution 2: Comparative Analysis of Existing Methods

The second contribution is a comparative analysis model for evaluating representative methods under a unified set of criteria. The comparison will include preventive techniques, distributed attack detection, resilient consensus, trust-based filtering, resilient distributed optimization, reconfiguration, and self-healing recovery. The objective is to identify not only what each method does but also what it assumes and what it fails to address.

The comparative analysis will use criteria such as attack coverage, information requirement, topology support, tolerance to compromised agents, scalability, communication overhead, convergence or control-performance guarantee, recovery capability, and validation setting. This is important because a method with strong theoretical guarantees may still be impractical if it requires dense communication, synchronous updates, or global knowledge of the network. Conversely, a lightweight method may scale well but provide only partial resilience against coordinated adversaries.

The four contributions are summarized here, and the extended contribution matrix is provided in Supplementary Table S4.

2.3. Contribution 3: Cyber-Resilience Assessment Framework

The third contribution is an assessment framework that translates cyber-resilience into measurable indicators. The framework builds on the idea that resilience is a dynamic property observed before, during, and after disruption. Therefore, a method should be evaluated through its ability to reduce degradation, maintain service continuity, detect abnormal behavior, adapt control or communication strategies, and recover coordinated operation.

Let $\boldsymbol{z}$ be a vector of raw resilience-related indicators:

$$\boldsymbol{z} = \left[T_d,\ D_p,\ T_r,\ E_c,\ \varepsilon_c,\ A_s\right]^\top, \tag{5}$$

where T_d is attack detection time, D_p is physical performance degradation, T_r is recovery time, E_c is communication overhead, ε_c is consensus or coordination error, and A_s is service availability. Because some indicators are minimized and others are maximized, the framework first maps them to normalized indicators $\hat{z}_m \in [0, 1]$ and then applies the weighted aggregation already defined in Equation (3). The weights remain application-specific; the score is therefore a configurable comparative construct rather than a universal resilience metric.

The framework also supports resilience-curve analysis through the retained-functionality measure already defined in Equation (2). This measure captures how much functionality is preserved over the evaluation horizon and complements event-based indicators such as detection and recovery time.

2.4. Contribution 4: PRISMA-Guided Systematic Evidence Synthesis

The fourth contribution is an empirical evidence synthesis that tests the proposed taxonomy and comparative criteria against a systematically screened primary-study set. The executed Scopus stream yielded 366 records. After publication-type filtering, separation of secondary reviews, title/abstract screening, eligibility assessment, and a six-criterion quality appraisal, 72 studies were retained for the Core comparative synthesis, and 7 additional studies were retained as contextual evidence. The extraction matrix records CPS domain, attack family, architecture profile, defense mechanism, validation type, outcome category, and quality score for every Core study. This evidence base is used to quantify which attack models, mechanisms, and validation practices dominate the literature and which parts of the resilience lifecycle remain underrepresented.

2.5. Scope, Boundaries, and Intended Use

The scope of this article is decentralized and multi-agent CPSs in which benign agents rely primarily on local computation and neighbor communication. Representative application domains include decentralized smart grids and microgrids, cooperative robotic systems, distributed industrial automation, vehicular networks, and industrial Internet-of-Things deployments. This article focuses on malicious-node and communication attacks because they directly challenge the assumptions of decentralized coordination.

This article does not aim to provide a complete cryptographic protocol survey or a full catalog of centralized intrusion-detection systems. Such methods are relevant when they support decentralized resilience, but they are not the primary object of analysis. Similarly, purely mathematical stability analysis without cyber disruption is outside the main scope unless it contributes to cyber-resilient control, recovery, or coordination.

The intended use of the proposed structure is threefold. First, researchers can use the taxonomy to position new algorithms and clarify their assumptions. Second, practitioners can use the comparative criteria to select methods suitable for a given decentralized CPS architecture. Third, future studies can use the assessment framework to design reproducible experiments and compare cyber-resilience mechanisms through quantitative metrics rather than qualitative claims alone.

2.6. Research Questions

This article is guided by the following research questions:

- **RQ1:** how can cyber-resilience methods for decentralized CPSs be systematically classified across attack models, CPS layers, topology assumptions, and resilience mechanisms?
- **RQ2:** what are the main strengths and limitations of representative methods such as distributed detection, resilient consensus, trust-based filtering, resilient distributed optimization, and recovery control?
- **RQ3:** which quantitative indicators are most appropriate for assessing cyber-resilience in decentralized CPSs under malicious-node and communication attacks?
- **RQ4:** how can taxonomy, comparative analysis, and assessment metrics be integrated into a reusable framework for future CPS resilience studies?

By answering these questions, this article seeks to provide a bridge between algorithm-oriented research and system-level resilience assessment. The taxonomy and assessment frameworks are not treated only as conceptual proposals: Sections 5–9 test them against a systematically screened evidence base and quantify the distribution of domains, attacks, defense mechanisms, validation practices, and reported outcomes. This bridge is particularly important for decentralized CPSs, where security, control, communication, and physical performance cannot be evaluated independently.

3. Background and Key Concepts

This section establishes the conceptual and mathematical basis used throughout this article. Since the term cyber-resilience is often used together with stability, robustness, security, and fault tolerance, a precise terminology is required before a taxonomy or assessment framework can be formulated. The section first defines cyber–physical systems (CPSs), then distinguishes centralized, distributed, and decentralized architectures, and finally introduces the graph-based and resilience-oriented notation used to analyze decentralized CPSs under malicious-node and communication attacks.

3.1. Cyber–Physical Systems

A cyber–physical system is commonly understood as an integration of computation, communication, and physical processes in which embedded computers and networks monitor and control physical dynamics through feedback loops. Lee's early formulation emphasized that CPS design is challenging because physical dynamics, software execution, real-time constraints, and network behavior must be analyzed together rather than as independent components [1]. The NIST CPS Framework similarly treats CPSs as engineered systems in which cyber, physical, and human components interact across functional, timing, trustworthiness, and data dimensions [2]. In modern infrastructure, CPSs appear in smart grids, industrial automation, transportation systems, water networks, autonomous vehicles, healthcare devices, and cooperative robotic systems [4,19,20].

The defining feature of a CPS is the closed-loop coupling between the cyber and physical domains. A simplified discrete-time model for a CPS agent i can be expressed as

$$\begin{aligned} \boldsymbol{x}_i(k+1) &= f_i\big(\boldsymbol{x}_i(k), \boldsymbol{u}_i(k), \boldsymbol{w}_i(k)\big), \\ \boldsymbol{y}_i(k) &= h_i\big(\boldsymbol{x}_i(k), \boldsymbol{v}_i(k)\big), \end{aligned} \tag{6}$$

where $\boldsymbol{x}_i(k)$ denotes the physical state, $\boldsymbol{u}_i(k)$ the control input, $\boldsymbol{y}_i(k)$ the measured output, and $\boldsymbol{w}_i(k)$ and $\boldsymbol{v}_i(k)$ process and measurement disturbances, respectively. In a cyber-resilience context, the same variables may also be affected by intentional cyber disruptions such as false data injection, spoofing, replay, or denial-of-service attacks [3,5,21]. Therefore, CPS security and resilience cannot be reduced to classical information security alone: cyber events may directly alter physical behavior, safety margins, and service availability.

3.2. Decentralized CPSs and Multi-Agent Representation

A decentralized CPS consists of multiple interacting agents that make decisions using local computation and information received from neighboring agents. Each agent typically contains sensors, actuators, embedded computation, communication interfaces, and a local control policy. The collective behavior emerges from local interactions rather than from a single central controller. This makes decentralized CPSs attractive for large-scale systems because decentralization can improve scalability, modularity, and tolerance to single points of failure. At the same time, decentralization creates new resilience challenges because malicious or compromised agents can influence their neighbors and propagate incorrect information through the network [7].

A decentralized CPS can be represented by a graph

$$\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W}), \tag{7}$$

where $\mathcal{V} = \{1, 2, \ldots, N\}$ is the set of agents, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of communication or interaction links, and $\mathcal{W} = [a_{ij}]$ is a weighted adjacency matrix. The neighbor set of agent i is denoted by

$$\mathcal{N}_i = \{j \in \mathcal{V} : (j, i) \in \mathcal{E}\}. \tag{8}$$

In decentralized coordination, agent i updates its decision or control signal using its own state and the information received from $\mathcal{N}_i$. A basic consensus-type update is given by

$$\boldsymbol{x}_i(k+1) = a_{ii}\boldsymbol{x}_i(k) + \sum_{j \in \mathcal{N}_i} a_{ij}\boldsymbol{x}_j(k), \tag{9}$$

where a_{ij} describes the influence of neighbor j on agent i. In benign settings, such mechanisms support synchronization, formation control, state agreement, or distributed optimization. Under attack, however, a malicious node may transmit corrupted values, selectively drop messages, impersonate neighbors, or behave inconsistently toward different agents. Consequently, resilience in decentralized CPSs depends not only on local controller design but also on the topology, information model, and adversarial assumptions of the interaction graph [7,22].

Figure 2 summarizes the agent-level sensing, control, communication, and actuation structure used in the background model.

3.3. Centralized, Distributed, and Decentralized Architectures

The terms centralized, distributed, and decentralized are sometimes used interchangeably, but they imply different assumptions for cyber-resilience. In a centralized CPS, monitoring, decision making, or control depends primarily on a central controller, supervisory server, or cloud platform. This architecture can simplify global optimization and security policy enforcement, but it introduces a potential single point of failure. In a distributed CPS, computation is shared among several nodes, but some coordinating authority or global information exchange may still exist. In a decentralized CPS, by contrast, there is no single trusted center; agents rely on local information and peer-to-peer coordination to achieve system-level objectives.

The architectural distinction used in this article is summarized in Supplementary Table S5; the key criterion is whether attack response can operate primarily from local and neighbor information without relying on a universally trusted center.

In this article, the phrase decentralized CPS refers to systems in which normal operation and resilience mechanisms are primarily based on local agent decisions and neighbor-to-neighbor communication. This includes multi-robot networks, microgrids, distributed

industrial Internet-of-Things systems, vehicular platoons, and other networked control systems where no universally trusted central entity is assumed during attack response.

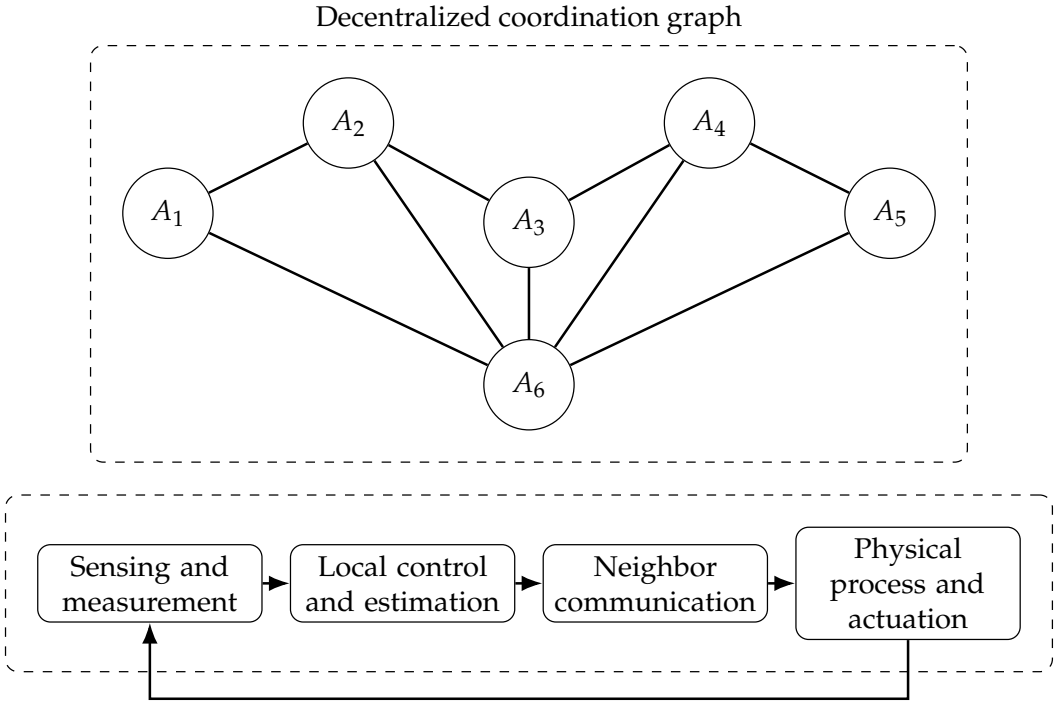


Figure 2. Conceptual representation of a decentralized CPS as a network of interacting agents with local sensing, control, communication, and actuation loops. Agents a_1–a_6 represent local cyber–physical nodes connected by neighbor links; the framed sensing, local-control, communication, and physical-process blocks illustrate the local feedback loop associated with each agent.

3.4. Stability, Robustness, Fault Tolerance, and Cyber-Resilience

Cyber-resilience is broader than stability, robustness, fault tolerance, or cybersecurity alone: it concerns the ability to anticipate, withstand, adapt to, and recover from malicious cyber disruption while preserving mission objectives [6,9]. A method may be stable or fault tolerant yet remain vulnerable to malicious neighbors; conversely, a resilient design may intentionally degrade performance or reconfigure topology to preserve safe operation.

3.5. Cyber-Resilience Lifecycle

For the purposes of this article, cyber-resilience is treated as a lifecycle property. Before an attack, the system should be designed to reduce exposure and prepare redundancy or fallback strategies. During an attack, it should absorb disruption, detect abnormal behavior, and preserve acceptable physical functionality. After an attack, it should recover coordination, restore performance, and adapt policies to reduce future vulnerability. This interpretation is consistent with cyber-resiliency engineering, which places resilience within the system life cycle rather than treating it as a single defensive mechanism [6].

Let $P(t)$ denote a normalized system-performance level, where $P(t) = 1$ corresponds to nominal operation and lower values indicate degradation. A simplified resilience curve can be described by three phases: pre-attack nominal operation, degraded operation under attack, and post-attack recovery. Retained functionality is evaluated using Equation (2); for an interval $[t_0, t_1]$, this corresponds to setting $T = t_1 - t_0$ and, for normalized performance, $P_0 = 1$. A resilient decentralized CPS should maximize retained functionality while minimizing detection time, recovery time, coordination error, and communication overhead. This formulation links the qualitative meaning of cyber-resilience with quantitative assessment metrics that will be developed later in this article.

3.6. Threat-Aware View of Decentralized CPSs

In decentralized CPSs, attacks can target the physical process, the sensing layer, the control logic, or the communication graph. Common examples include denial-of-service attacks that reduce communication availability, false data injection attacks that compromise measurement or control integrity, replay attacks that reuse previously valid data, spoofing attacks that impersonate nodes, and malicious-agent attacks that inject adversarial values into coordination protocols [5,7,21]. These attack types are not independent of the architecture. For example, a communication attack in a centralized CPS may isolate a remote device from the central controller, whereas the same attack in a decentralized CPS may partition the graph and prevent neighboring agents from reaching agreement.

This article therefore adopts a threat-aware interpretation of decentralization. A method is considered relevant to decentralized CPS cyber-resilience if it addresses at least one of the following conditions: limited local information, malicious or compromised agents, unreliable neighbor communication, absence of a fully trusted central monitor, dynamic or uncertain topology, and physical performance degradation caused by cyber disruption. Under these conditions, resilience mechanisms must be evaluated not only by their security function but also by their compatibility with decentralized control and coordination.

3.7. Terminology Adopted in This Article

To avoid ambiguity in the following sections, this article uses the terminology summarized below. A *benign agent* follows the prescribed control or coordination rule. A *malicious agent* intentionally deviates from this rule and may transmit false, inconsistent, or adversarial information. A *compromised agent* is a formerly benign component whose cyber or physical behavior has been altered by an attacker. A *resilience mechanism* is a method that helps the system resist, absorb, adapt to, or recover from a cyber disruption. A *decentralized resilience method* is a resilience mechanism that can operate primarily with local and neighbor-level information rather than global system knowledge.

These definitions provide the basis for the taxonomy in Section 6, where methods will be classified by threat model, CPS layer, information requirement, topology assumption, and resilience function. They also support the comparative analysis and assessment framework, because the effectiveness of a method depends on whether its underlying assumptions match the decentralized CPS environment in which it is deployed.

4. Threat Model for Decentralized Cyber–Physical Systems

A threat model defines what parts of the system may be attacked, what capabilities the adversary is assumed to have, what objectives the adversary pursues, and what assumptions are imposed on the number, location, and behavior of compromised components. For decentralized cyber–physical systems (CPSs), this modeling step is especially important because the absence of a single trusted coordinator changes both the attack surface and the feasibility of defensive mechanisms. Attacks do not only target isolated devices; they can also exploit inter-agent communication, local trust assumptions, graph topology, and the coupling between cyber variables and physical dynamics. Recent CPS security surveys emphasize denial-of-service (DoS), false data injection (FDI), replay, and communication manipulation as recurrent attack classes, while resilient coordination studies show that malicious agents and topology-dependent adversaries are central issues in decentralized CPSs [5,7,21].

This section establishes the threat model used throughout the remainder of this article. It does not assume a single universal attacker; instead, it organizes threats according to the CPS layer, the adversary location, the integrity/availability property being violated, and the resulting physical impact. This is consistent with system security engineering

guidance, where cyber-resilience must be considered across the system life cycle and across interacting cyber and physical components [6,20]. The resulting model provides the basis for the taxonomy of resilience methods in Section 6 and for the assessment framework developed later.

4.1. Decentralized CPS as a Graph-Based Attack Surface

A decentralized CPS can be modeled as a network of N agents connected through a possibly time-varying communication graph

$$\mathcal{G}(k) = (\mathcal{V}, \mathcal{E}(k)), \qquad \mathcal{V} = \{1, 2, \ldots, N\}, \tag{10}$$

where each vertex represents a local cyber–physical agent and each edge $(j, i) \in \mathcal{E}(k)$ denotes that agent i can receive information from agent j at time k. The in-neighbor set of agent i is denoted by $\mathcal{N}_i(k)$. Under nominal conditions, each agent updates its control or coordination decision using local measurements and neighbor information. Under attack, however, the values received by a benign agent may differ from the values generated by their original source.

A compact way to express the attack surface is to distinguish between attacks on sensing, communication, and actuation:

$$\begin{aligned} \tilde{y}_i(k) &= y_i(k) + a_i^s(k), \\ \tilde{z}_{ji}(k) &= z_{ji}(k - \tau_{ji}(k)) + a_{ji}^c(k), \\ \tilde{u}_i(k) &= u_i(k) + a_i^u(k), \end{aligned} \tag{11}$$

where $y_i(k)$ is a local measurement, $z_{ji}(k)$ is information transmitted from neighbor j to agent i, $u_i(k)$ is a control input, and $a_i^s(k)$, $a_{ji}^c(k)$, and $a_i^u(k)$ represent malicious modifications of sensing, communication, and actuation signals, respectively. The delay term $\tau_{ji}(k)$ captures replay, delay, or stale-message attacks. Equation (11) highlights a key difference between conventional cybersecurity and CPS cyber-resilience: an apparently small cyber manipulation may propagate through feedback loops and produce physical performance degradation or unsafe behavior [23,24].

Figure 3 summarizes the threat surface considered in this article. The attacker may operate at the sensor layer, local control layer, communication layer, coordination layer, or physical actuation layer. In a decentralized architecture, these attack vectors can be combined: for example, a compromised agent may inject false local measurements, send inconsistent values to different neighbors, and manipulate distributed coordination at the same time.

4.2. Adversary Objectives and Capabilities

The adversary considered in this article may pursue one or more of four objectives. First, the adversary may attack *availability* by preventing agents from exchanging information or by blocking local computation. DoS, jamming, and packet-dropping attacks belong to this category. Second, the adversary may attack *integrity* by modifying measurements, state estimates, neighbor messages, or control inputs. False data injection, replay, and actuator manipulation are typical examples. Third, the adversary may attack *authenticity* by impersonating a legitimate node or creating false identities. Fourth, the adversary may attack the *physical mission objective*, for example, by increasing frequency deviation in a microgrid, forcing robots away from a formation, destabilizing an industrial process, or preventing consensus.

The adversary capability can be represented by a tuple

$$\mathcal{A} = (\mathcal{C}, \mathcal{I}, \mathcal{T}, \mathcal{O}), \tag{12}$$

where $\mathcal{C}$ denotes the set of compromised components, $\mathcal{I}$ denotes the information available to the attacker, $\mathcal{T}$ denotes the timing pattern of the attack, and $\mathcal{O}$ denotes the attack objective. This representation is intentionally general: a replay attack may have limited actuation capability but strong knowledge of historical data, whereas a Byzantine agent may have broad freedom to transmit arbitrary and inconsistent values to different neighbors.

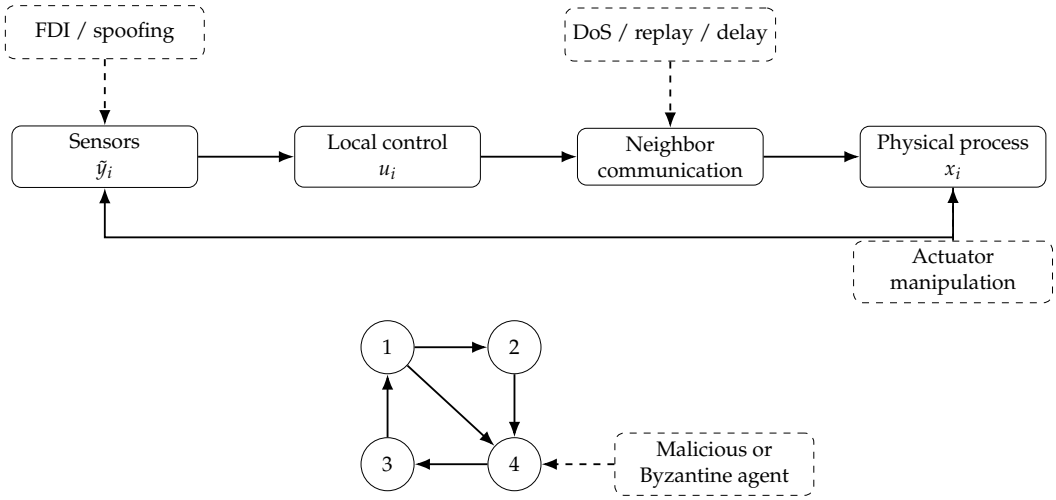


Figure 3. Threat surface of a decentralized CPS. The numbered frames 1–4 denote sensing, local control/decision, communication/coordination, and physical-process/actuation stages, respectively. Attack arrows indicate the targeted path for FDI/spoofing, DoS/replay/delay, malicious or Byzantine-agent behavior, and actuator manipulation.

For decentralized CPSs, the most important distinction is whether malicious components are globally bounded or locally bounded. Let $\mathcal{M} \subseteq \mathcal{V}$ denote the set of malicious or compromised agents. A total adversary model assumes $|\mathcal{M}| \leq f$ in the entire network. An f-local model assumes that every benign agent has at most f malicious in-neighbors:

$$|\mathcal{M} \cap \mathcal{N}_i(k)| \leq f, \qquad \forall i \in \mathcal{V} \setminus \mathcal{M}. \tag{13}$$

The f-local assumption is widely used in resilient consensus because decentralized agents typically cannot know the total number of malicious nodes in the complete network; they only observe values received from neighbors. This assumption also connects threat modeling with graph robustness: a resilience mechanism may tolerate malicious behavior only if the network topology provides sufficient redundant benign paths [7,25].

4.3. Communication-Channel Attacks

Communication-channel attacks target the links through which decentralized agents exchange state estimates, control variables, optimization gradients, alarms, or trust scores. Because decentralized CPSs rely on peer-to-peer communication, link-level disruption can directly affect coordination. The most common availability-oriented attack is DoS, where the adversary blocks or degrades communication by jamming wireless channels, exhausting network resources, or selectively dropping packets. In a centralized CPS, DoS may isolate a remote device from the supervisory controller; in a decentralized CPS, DoS may partition the communication graph and prevent benign agents from reaching agreement.

Integrity-oriented communication attacks include delay, replay, packet manipulation, and man-in-the-middle attacks. In a replay attack, the adversary records previously valid data and retransmits it later, making the receiver operate on stale but plausible information. Scientific Reports research on replay and false data injection in SCADA-oriented power-system contexts notes that both attack types falsify measurements and can disrupt normal operation [26]. In decentralized CPSs, replayed neighbor messages can be particularly difficult to detect because the receiving agent may only have local context and may not be able to compare the message with a global state estimate.

Communication attacks can be modeled using a link availability variable $\gamma_{ji}(k) \in \{0, 1\}$ and an additive or substitutive communication attack signal $a^c_{ji}(k)$:

$$\tilde{z}_{ji}(k) = \gamma_{ji}(k) z_{ji}(k - \tau_{ji}(k)) + a^c_{ji}(k), \tag{14}$$

where $\gamma_{ji}(k) = 0$ represents a blocked or unavailable link. This model captures DoS through link unavailability, replay through non-zero delay $\tau_{ji}(k)$, and message manipulation through $a^c_{ji}(k)$. The model is deliberately abstract because the same mathematical effect may arise from different technical mechanisms, including wireless jamming, routing manipulation, or compromised middleware.

4.4. Node and Agent Attacks

Node and agent attacks are central to decentralized CPSs because each agent participates in local decision making. A compromised agent may still appear as a legitimate member of the network but deviate from the prescribed protocol. Such an agent may inject false values, refuse to update its state, send different values to different neighbors, collude with other compromised nodes, or attempt to bias distributed optimization. In the consensus literature, this is commonly related to Byzantine or malicious-agent models, where the attacker is not restricted to a simple additive disturbance [7,25].

The behavior of a malicious agent $j \in \mathcal{M}$ can be represented as

$$\tilde{z}_{ji}(k) = \phi_{ji}\big(k, \mathcal{H}_j(k), \mathcal{O}\big), \qquad i \in \mathcal{N}^{out}_j(k), \tag{15}$$

where $\mathcal{H}_j(k)$ denotes the information history available to the compromised agent and $\phi_{ji}(\cdot)$ is an arbitrary adversarial message-generation rule. Unlike the additive model in Equation (14), this formulation allows the attacker to transmit inconsistent values to different neighbors. This capability is especially relevant for resilient consensus and distributed optimization because benign agents must decide which neighbor values to trust without a central verifier.

Node attacks also include impersonation and Sybil-like behavior, where an attacker attempts to appear as one or more legitimate agents. In CPSs with resource-constrained devices, weak identity management or poor network segmentation can make these attacks more likely. Operational technology guidance emphasizes authentication, access control, network segmentation, and monitoring because attacks against industrial devices can have direct safety and operational consequences [20]. For decentralized CPSs, identity compromise is not only an access-control problem; it also undermines the assumptions used by graph-based coordination algorithms.

4.5. Sensor and Actuator Attacks

Sensor and actuator attacks directly manipulate the interface between cyber decisions and physical processes. In an FDI attack, the adversary modifies sensor readings or state estimates so that the controller receives plausible but incorrect information. In a spoofing attack, the adversary imitates a legitimate sensor or source. In an actuator attack, the adver-

sary manipulates the control action applied to the physical plant. These attacks are well studied in CPS security because they exploit the feedback structure of control systems and can remain stealthy if the injected signals are consistent with system dynamics [23,24,27].

The impact of a sensor or actuator attack can be expressed through the physical state update

$$x_i(k+1) = f_i(x_i(k), \tilde{u}_i(k), w_i(k)), \qquad \tilde{y}_i(k) = h_i(x_i(k)) + a_i^s(k), \tag{16}$$

where $a_i^s(k)$ affects the measurement available to the cyber layer and $\tilde{u}_i(k)$ affects the control input applied to the physical process. This coupling means that attacks may be judged not only by the amount of data corruption but also by their physical consequence: loss of stability, violation of safety constraints, reduced service quality, or delayed recovery.

In decentralized CPSs, the consequence of a local sensor attack can propagate through the network. If agent i uses a manipulated measurement to compute its local state estimate and then shares this estimate with neighbors, benign agents may indirectly receive corrupted information. Therefore, local attack detection is often insufficient; a resilient system must also limit the spread of corrupted information through coordination and communication mechanisms.

4.6. Coordinated, Stealthy, and Hybrid Attacks

Realistic attacks may combine several vectors. A coordinated attack may compromise multiple agents simultaneously, manipulate communication links, and inject false data into sensors. A stealthy attack may be designed to remain below the threshold of local detectors while gradually degrading physical performance. A hybrid attack may combine DoS with FDI: DoS prevents benign agents from exchanging corrective information, while FDI biases the measurements used by the remaining connected agents. These scenarios are important because defenses optimized for a single attack type may fail when attacks interact.

Hybrid attacks can be represented by the union of compromised sets across CPS layers:

$$\Omega(k) = \mathcal{S}^s(k) \cup \mathcal{S}^c(k) \cup \mathcal{S}^u(k) \cup \mathcal{M}(k), \tag{17}$$

where $\mathcal{S}^s(k)$ is the set of attacked sensors, $\mathcal{S}^c(k)$ is the set of attacked communication links, $\mathcal{S}^u(k)$ is the set of attacked actuators, and $\mathcal{M}(k)$ is the set of malicious agents. The system-level severity of a threat depends not only on the size of $\Omega(k)$ but also on its location in the graph and on the role of the affected components. For example, compromising a bridge node in a sparse communication graph may be more damaging than compromising a peripheral node in a highly redundant graph.

4.7. Threat Classification Adopted in This Article

Table 1 summarizes the threat classes used in this article. The classification is organized around the main target layer, typical attack examples, primary violated property, and expected impact on decentralized CPS operation. This table will be reused conceptually in the taxonomy section: each class of cyber-resilience method should be evaluated according to which threats it can address and under which assumptions.

The threat model adopted in this article makes four assumptions. First, no single fully trusted central monitor is available during attack response. Second, benign agents have access primarily to local measurements and neighbor messages. Third, the adversary may compromise a subset of nodes, links, sensors, or actuators, but resilience analysis should make explicit whether the number of malicious components is globally bounded, locally bounded, or topology-dependent. Fourth, the evaluation of a defense must include both

cyber-level criteria, such as attack detection or communication availability, and physical-level criteria, such as stability, service performance, safety margins, and recovery time.

These assumptions intentionally cover a broad class of decentralized CPSs, including microgrids, cooperative robotic systems, vehicular networks, and industrial Internet-of-Things deployments. The purpose is not to claim that every attack model applies equally to every domain. Rather, the purpose is to provide a common threat-modeling language that allows different cyber-resilience methods to be compared under transparent assumptions. The next section uses this threat model to classify methods according to their resilience function, information requirement, topology assumption, and supported attack class.

Table 1. Threat classes for decentralized cyber–physical systems.

Threat Class	Typical Examples	Violated Property	Main Impact on Decentralized CPS
Communication attacks	DoS, jamming, packet dropping, delay, replay, man-in-the-middle	Availability, integrity, freshness	Graph partition, stale neighbor information, loss of consensus, delayed coordination.
Node and agent attacks	Compromised agent, Byzantine behavior, malicious neighbor, impersonation	Integrity, authenticity, trust	Biased local decisions, inconsistent information exchange, corruption of distributed coordination.
Sensor attacks	False data injection, spoofing, measurement tampering	Integrity, observability	Incorrect state estimation, stealthy degradation, propagation of corrupted measurements.
Actuator attacks	Control-signal manipulation, command injection, actuator blocking	Integrity, safety, availability	Unsafe physical action, reduced stability margin, mission failure, or service interruption.
Coordinated and hybrid attacks	Combined DoS-FDI, multi-agent compromise, stealthy multi-stage attack	Multiple properties	Simultaneous cyber and physical degradation, reduced detectability, harder recovery.

5. Systematic Review Methodology

5.1. Review Design and Reporting Framework

The empirical component of this article was conducted as a systematic mapping and taxonomy-oriented evidence synthesis. Reporting follows the logic of PRISMA 2020 and the PRISMA-S search-reporting extension [28,29]. PRISMA-ScR items were consulted as an applicable reporting crosswalk; this work remains a systematic mapping study and is not represented as a prospectively registered scoping review. After completion of the main review workflow, the protocol was retrospectively registered with INPLASY on 22 October 2026 (registration number: INPLASY2026100097; DOI: https://doi.org/10.37766/inplasy2026.10.0097). This retrospective registration documents the methodology used in the completed review and must not be interpreted as prospective preregistration. A meta-analysis was not attempted because the included studies use heterogeneous CPS domains, attack models, control objectives, topologies, validation platforms, and outcome metrics. Instead, the synthesis combines controlled screening decisions, a six-criterion quality appraisal, and structured categorical extraction.

5.2. Search Strategy and Reproducibility

The working review protocol dated 13 September 2026 specified the following three-block Scopus query (CPS context, decentralization, and cyber resilience). It is reproduced verbatim in Supplementary Table S1. The executed Scopus export is dated 14 September 2026 and contains 366 records. The exported spreadsheet did not preserve the executed query text or the applied interface filters. Consequently, although the protocol

query is documented, its exact identity with the executed Scopus search and the historical result count cannot be independently verified from the available export. The publication window was 2013–2026; search-interface language and document-type filters cannot be independently reconstructed. The publication-type exclusions were applied in the documented screening process.

The primary quantitative identification stream was the executed Scopus export recorded in the review log on 14 September 2026. The export contained 366 records and covered the applied publication window 2013–2026. Publisher platforms, institutional repositories, and DOI records were subsequently used to retrieve full texts, accepted manuscripts, or authoritative extended records for eligibility verification and quality appraisal. These supplementary retrieval sources did not add records to the PRISMA identification count and therefore do not inflate the primary Scopus stream.

5.3. Post Hoc Database-Coverage Sensitivity Check

To address database-coverage uncertainty without retroactively changing the original PRISMA stream, we conducted a dated post hoc coverage sensitivity check on 4 November 2026. The check used the same three conceptual blocks as the protocol query (CPS context, decentralized/distributed architecture, and cyber-resilience/attack/defense concepts) across publicly accessible IEEE Xplore records, public ACM Digital Library indexing queries, and publisher/metadata records explicitly indicating Web of Science indexing. This was a coverage-sensitivity exercise rather than a database-native rerun: authenticated historical result counts and export files were not available for every complementary database, and no such counts are claimed. Candidate records were deduplicated against the 79-study eligible set by DOI and normalized title and then screened using the original eligibility logic.

Nineteen candidate records were documented in the supplementary coverage log. Nine were judged new eligible candidates absent from the original 79-study set, two were duplicates already represented in the eligible corpus, three required full-text confirmation of the decentralization/topology criterion, and five were excluded for publication type, date, or insufficient evidence of decentralized architecture. The nine new eligible candidates all mapped to existing taxonomy dimensions: power systems/microgrids; FDI or hybrid FDI + DoS threats; resilient/secure control, detection, observer compensation, or AI-assisted mitigation; and simulation, experimental, or HIL/real-time validation. Thus, the check confirms a Scopus-only coverage limitation but did not reveal a new taxonomy dimension. Because the exercise was post hoc and not a complete authenticated multi-database rerun, the original PRISMA counts and the 72-study primary synthesis remain unchanged; the additional records are reported only as a coverage-sensitivity layer.

5.4. Eligibility and Screening Procedure

Primary studies were eligible when they addressed a CPS or explicit cyber–physical/networked-control application, used a decentralized, distributed, or multi-agent architecture relevant to decentralized operation, considered a malicious or cybersecurity-relevant threat, proposed or evaluated a resilience/security mechanism, and provided sufficient technical evidence to identify the threat model, response mechanism, and validation basis. Purely centralized methods, accidental-fault-only studies, non-primary publication types, studies without a qualifying malicious threat, and records with insufficient technical evidence were excluded from the Core primary stream.

The 366 exported records were first separated into publication-type exclusions (19 records), secondary reviews retained for background and citation chasing (10 records), and 337 primary candidates. Title/abstract screening excluded 174 primary records and

advanced 163 records to eligibility assessment. Eligibility assessment combined full-text review with authoritative extended records when a full PDF was unavailable; this decision rule is reported explicitly to avoid implying that all 163 PDFs were accessible. A total of 84 records were excluded at eligibility, leaving 79 eligible primary studies.

The normalized eligibility exclusion reasons were centralized or non-decentralized architecture ($n = 65$), insufficient primary technical evidence or unsuitable publication type ($n = 7$), no qualifying malicious cyber threat ($n = 5$), no qualifying resilience/security mechanism ($n = 5$), and wrong CPS/application scope ($n = 2$). These counts sum to the 84 eligibility exclusions.

5.5. Quality Appraisal

All 79 eligible primary studies were scored using a six-dimension 0–2 rubric: Q1, clarity of threat/adversary model; Q2, clarity of decentralization and topology assumptions; Q3, reproducibility of the algorithm or architecture; Q4, strength of validation evidence; Q5, adequacy of baselines, metrics, and repeated/statistical analysis; and Q6, transparency of limitations, implementation details, or code/data availability. The maximum score was 12. Studies scoring at least 7 were retained for the Core comparative synthesis; lower-scoring studies were retained only as contextual evidence. This process yielded 72 Core studies and 7 contextual studies, with no unresolved quality-appraisal records.

5.6. Post Hoc Inter-Rater Reproducibility Audit and Adjudication

To assess the reproducibility of the Q1–Q6 appraisal after completion of the primary review, a post hoc stratified audit sample of 25 of the 79 eligible studies was selected, enriched around the Core/Contextual decision boundary. Two human assessors independently scored the sampled studies while blinded to the original quality scores and to each other's ratings. One sampled record (AUD-14) was available only at abstract level and was therefore excluded from the primary full-text reliability analysis, leaving 24 independently appraised full-text studies. The original assessor ratings were frozen before any disagreement discussion.

Agreement statistics were calculated exclusively from these pre-adjudication ratings. Core/Contextual classification showed exact agreement for 17 of 24 studies (70.8%) with Cohen's $\kappa = 0.30$. Across the six ordinal criteria, exact cell agreement was 88/144 (61.1%). Criterion-level quadratic weighted Cohen's κ values were 0.25 (Q1), 0.41 (Q2), 0.43 (Q3), 0.23 (Q4), 0.37 (Q5), and 0.30 (Q6). Because the audit sample was stratified and intentionally enriched near the decision boundary, these coefficients are reported as reproducibility-audit estimates rather than as agreement statistics for the original screening of all 79 studies.

Adjudication was performed only after the independent ratings had been frozen. Seven studies with Core/Contextual classification disagreement were reviewed first; seven additional same-class studies with the largest score discrepancies were then re-examined; and three residual cases (AUD-17, AUD-19, and AUD-24) received targeted criterion-level checks. This produced a 17-study adjudication log. The clarified rubric required explicit operational attacker assumptions for Q1 = 2, explicit distributed topology and information exchange for Q2 = 2, sufficient algorithmic detail and material parameters for Q3 = 2, ordinarily assigned numerical simulation to Q4 = 1 while reserving Q4 = 2 for HIL, physical, real-time, or genuinely comprehensive validation, and required substantive comparative and quantitative/repeated evidence for Q5 = 2. Consensus scores were retained as a separate sensitivity/adjudication layer and were not substituted back into the frozen independent ratings used for reliability statistics. To avoid retroactively redefining the pre-specified primary synthesis after a post hoc audit, the original 72-Core/7-

Contextual classification was retained for the aggregate evidence synthesis; adjudication is reported separately as a robustness check.

5.7. Data Extraction and Synthesis

For each Core study, the extraction matrix recorded bibliographic metadata, CPS domain, primary and secondary attack tags, architecture profile, defense mechanism, validation type, reported outcome category, and quality score. Attack and defense mechanisms were coded using controlled categories aligned with the taxonomy in Section 6. Defense mechanisms were allowed to receive multiple tags because a single resilience architecture may combine detection, observer compensation, consensus, isolation, and recovery. The evidence synthesis is therefore descriptive and comparative: mutually exclusive primary categories are used for domain, primary attack, validation, and outcome summaries, while mechanism prevalence is reported as multi-label frequency.

The systematic evidence set is used for two purposes. First, it provides empirical support for the taxonomy by showing which attack families and resilience mechanisms are actually represented in the literature. Second, it evaluates evidence maturity by comparing simulation-only studies with hardware-in-the-loop, real-time, laboratory, co-simulation, and other strong validation settings. The resulting evidence synthesis is presented in Section 8, followed by an RQ-oriented discussion in Section 9.

The executed identification stream is intentionally reported as Scopus-based rather than as a multi-database search. A post hoc complementary coverage sensitivity check was added during peer review, but it is reported separately because authenticated database-native exports and historical result counts were not available for every complementary source. The check identified additional eligible candidates within already represented taxonomy strata and therefore confirms the coverage limitation without changing the original PRISMA counts. The inter-rater analysis reported above was likewise a post hoc stratified reproducibility audit rather than prospective duplicate screening of the complete review set.

6. Taxonomy of Cyber-Resilience Methods

This section develops the central taxonomy of this article. While Section 4 described how adversarial actions can affect decentralized CPSs, the present section classifies the methods used to preserve, adapt, and restore the behavior of benign agents under such attacks. The classification is designed for decentralized settings in which each agent has local sensing, local control, and limited neighbor-to-neighbor communication. Consequently, a method is considered cyber-resilient not merely because it blocks an attack but because it supports one or more of the following functions: reducing attack likelihood, detecting abnormal behavior, filtering corrupted information, preserving coordination, isolating compromised components, reconfiguring the communication or control structure, and recovering physical performance.

The proposed taxonomy is defined by the mapping

$$\Theta(\mathcal{M}) = (\mathcal{P}, \mathcal{D}, \mathcal{C}, \mathcal{T}, \mathcal{O}, \mathcal{G}, \mathcal{R}, \mathcal{L}, \mathcal{V}), \tag{18}$$

where $\mathcal{M}$ is a cyber-resilience method, $\mathcal{P}$ denotes preventive controls, $\mathcal{D}$ denotes distributed detection, $\mathcal{C}$ denotes resilient consensus, $\mathcal{T}$ denotes trust or reputation management, $\mathcal{O}$ denotes resilient distributed optimization, $\mathcal{G}$ denotes reconfiguration of the communication or control graph, $\mathcal{R}$ denotes recovery and self-healing, $\mathcal{L}$ denotes AI/ML-enabled learning mechanisms, and $\mathcal{V}$ denotes digital-twin-based validation and decision support. This representation follows the cyber-resilience view that a system should anticipate, withstand, recover from, and adapt to adverse cyber conditions but specializes it to the decentralized

CPS mechanisms discussed in recent work on CPS security control, resilient coordination, distributed optimization, and resilience assessment [5–9,30].

Figure 4 summarizes the taxonomy. The left part contains pre-attack or early-stage mechanisms, the center contains coordination-preserving methods, and the right part contains adaptation, recovery, and validation mechanisms. The taxonomy is not intended to imply a strict sequence. In a real decentralized CPS, several classes must operate simultaneously: for example, authentication and encryption protect communication, neighbor-based detection identifies suspicious data, resilient consensus filters malicious values, trust scores reduce future influence of suspicious nodes, and self-healing control restores physical service.

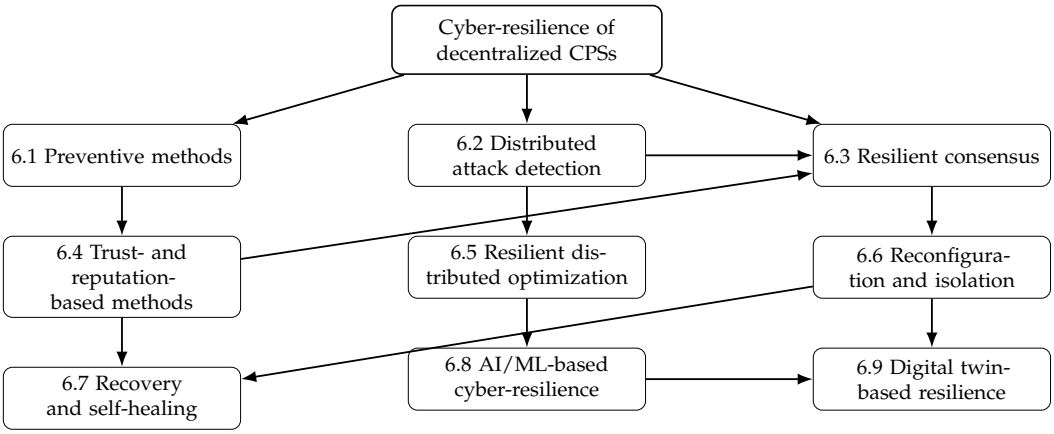


Figure 4. Taxonomy of cyber-resilience methods for decentralized CPSs.

6.1. Preventive Methods

Preventive methods reduce the probability that an adversary can enter, observe, manipulate, or disrupt the decentralized CPS. In traditional cybersecurity, prevention is often treated as the primary defense layer. In decentralized CPSs, however, prevention is necessary but insufficient: even strong authentication, encryption, and access control cannot eliminate all attacks because field devices may be compromised, communication links may fail, and insider or supply-chain threats may bypass perimeter controls. Therefore, preventive methods should be interpreted as the first layer of a broader cyber-resilience architecture.

The main preventive controls include authentication, encryption, secure communication, access control, redundancy, and secure protocols. Authentication verifies the identity of agents, controllers, and edge devices before accepting data or control commands. Encryption protects confidentiality and integrity of exchanged messages, especially when agents communicate over wireless or shared networks. Secure communication protocols provide message freshness, sequence validation, and replay protection. Access control restricts which nodes can modify set-points, firmware, controller parameters, or network routing. Redundancy increases the availability of sensing, computation, and communication resources. NIST guidance on operational technology security emphasizes the importance of secure architecture, remote access control, segmentation, cryptographic protection, and monitoring for cyber–physical and industrial environments [20]; these principles remain relevant in decentralized CPSs, although their implementation must be adapted to local, resource-constrained agents.

The preventive layer can be expressed through a residual attack probability model:

$$P_{\mathrm{res}}(a) = P_0(a) \prod_{q=1}^{Q} \big(1 - \eta_q(a)\big), \tag{19}$$

where $P_0(a)$ is the baseline probability of attack class a, and $\eta_q(a) \in [0,1]$ is the effectiveness of preventive control q against that attack. Equation (19) is not a universal risk model, but it is useful for illustrating a key point: prevention reduces exposure but rarely drives the residual probability to zero. A decentralized CPS must therefore combine preventive controls with detection, consensus filtering, trust adaptation, and recovery.

Redundancy is especially important because it allows the system to preserve function when one or more components are unavailable or compromised. For example, a sensing task can be supported by multiple neighboring agents, and a communication path can be replaced by an alternative route. In adversarial consensus and Byzantine-resilient coordination, redundancy is also structural: benign nodes need enough reliable neighbors to tolerate a bounded number of malicious values. A simple intuition is that a node that must tolerate f malicious neighbor values needs more than $2f$ informative values before it can discard extremes and still update from benign information. This structural redundancy connects prevention with the graph robustness assumptions discussed in resilient consensus methods.

6.2. Distributed Attack Detection

Distributed attack detection identifies abnormal behavior without relying on a central server. This is critical in decentralized CPSs because global monitoring may be unavailable, delayed, or itself compromised. The detection layer may operate at several levels: local anomaly detection at an individual node, neighbor-based monitoring across adjacent agents, distributed intrusion detection using shared alerts, model-based detection using physical dynamics, and machine learning-based detection using data-driven models. Recent CPS security surveys identify denial-of-service, false data injection, and replay attacks as major attack classes and emphasize attack detection as a key component of CPS security control [5]. Classical model-based CPS attack detection also shows that detectability depends on the relation between attack signals, system dynamics, and available measurements [23,27].

A common local model-based detector compares measured outputs with predicted outputs:

$$r_i(k) = \|y_i(k) - \hat{y}_i(k)\|, \qquad \alpha_i(k) = \begin{cases} 1, & r_i(k) > \tau_i(k), \\ 0, & r_i(k) \le \tau_i(k), \end{cases} \tag{20}$$

where $r_i(k)$ is the residual of agent i, $\hat{y}_i(k)$ is a predicted measurement, and $\tau_i(k)$ is a fixed or adaptive threshold. The binary variable $\alpha_i(k)$ denotes a local alarm. In decentralized environments, however, a single alarm may be unreliable because sensors can be noisy and a compromised node can generate false alarms. Therefore, neighbor-based detection aggregates local evidence from a neighborhood:

$$A_i(k) = \sum_{j \in \mathcal{N}_i \cup \{i\}} \omega_{ij}(k)\alpha_j(k), \qquad \omega_{ij}(k) \ge 0, \quad \sum_j \omega_{ij}(k) = 1. \tag{21}$$

Here, $A_i(k)$ is a local neighborhood alert score. If $A_i(k)$ exceeds a decision threshold, agent i can reduce the influence of suspicious messages, request additional verification, or switch to a fallback controller. This formulation supports distributed intrusion detection while avoiding the assumption that all raw data must be sent to a central monitoring station.

Detection methods can be classified into five subfamilies. First, local anomaly detection uses thresholds, statistical tests, or residuals generated by observers and estimators. Second, neighbor-based monitoring checks consistency between the state or output of an agent and the values reported by its neighbors. Third, distributed intrusion detection exchanges alerts or confidence values among agents. Fourth, model-based detection uses physics, control

models, observers, Kalman filters, or random matrix methods to detect manipulation such as replay or false data injection [26]. Fifth, ML-based detection uses supervised, unsupervised, or self-supervised models to identify abnormal patterns. ML-based detection is powerful when process models are incomplete, but it requires careful validation because training data may not cover all attack scenarios, and the model itself can be targeted by poisoning or adversarial examples.

6.3. Resilient Consensus Methods

Resilient consensus is one of the most important method classes for decentralized CPSs. Many CPS applications require agents to agree on a common value, coordinate a formation, synchronize estimates, or track a shared reference. In the absence of attacks, a standard average consensus update can be written as

$$x_i(k+1) = w_{ii}x_i(k) + \sum_{j\in\mathcal{N}_i} w_{ij}x_j(k), \qquad \sum_{j\in\mathcal{N}_i\cup\{i\}} w_{ij} = 1, \tag{22}$$

where w_{ij} are non-negative weights. Foundational work on consensus and cooperation in networked multi-agent systems shows how graph connectivity and weighting rules determine convergence properties under benign conditions [22]. Under malicious-node or Byzantine behavior, however, Equation (22) can fail because a single adversarial neighbor may inject an extreme or strategically crafted value that pulls the estimate of benign agents away from the true target.

Resilient consensus modifies the update rule so that benign agents can still converge to an acceptable value even when a bounded number of neighbors are malicious. A representative family of algorithms is based on mean-subsequence-reduced (MSR) and weighted MSR (W-MSR) filtering. At each step, agent i sorts the values received from its neighbors, removes up to f largest and up to f smallest suspicious values, and updates using the remaining set $\mathcal{R}_i(k)$:

$$x_i(k+1) = w_{ii}x_i(k) + \sum_{j\in\mathcal{R}_i(k)} w_{ij}x_j(k). \tag{23}$$

This filtering of extreme values is effective when malicious agents attempt to drive benign estimates outside the normal range. Theoretical results on resilient asymptotic consensus show that graph robustness and local malicious-agent assumptions are crucial for guaranteeing convergence [25]. More recent surveys and experimental studies extend this line of work to resilient coordination and dynamic consensus in CPSs, including MSR-type algorithms with input signals and fault-detection/recovery mechanisms [7,31,32].

Resilient consensus methods can be divided into several subgroups. Average consensus methods provide the baseline and are useful for understanding normal coordination behavior. Resilient consensus methods add filtering or robust update rules. Byzantine-resilient consensus considers arbitrary adversarial behavior, including inconsistent information sent to different neighbors. W-MSR algorithms remove extreme values before updating. Trusted-node-based consensus assumes that some nodes are verified and can anchor the network. These subgroups differ in their assumptions: some require only local filtering, while others require graph robustness, trusted nodes, bounds on the number of adversaries, or synchronous updates.

The core idea is that benign agents should reach a correct or acceptable coordinated state even if a subset of nodes behaves maliciously. In practice, the target may not be exact average consensus. For physical CPSs, it may be sufficient that benign agents remain inside a safety interval,

$$x_i(k) \in \left[x_{\min}^{\text{safe}}, x_{\max}^{\text{safe}}\right] \quad \forall i \in \mathcal{V}\setminus\mathcal{F}, \tag{24}$$

where $\mathcal{F}$ is the set of malicious or compromised agents. This safety-oriented view is more realistic for CPSs because physical performance, safety constraints, and service continuity may matter more than exact numerical agreement.

6.4. Trust-Based and Reputation-Based Methods

Trust-based and reputation-based methods allow agents to estimate the reliability of neighboring nodes over time. Instead of treating every received message equally, each agent maintains a trust score for its neighbors. This is particularly useful when malicious behavior is intermittent, stealthy, or not extreme enough to be removed by W-MSR filtering. Trust mechanisms can support confidence-weighted communication, malicious neighbor isolation, and adaptive trust update.

A simple adaptive trust score can be expressed as

$$s_{ij}(k+1) = \lambda s_{ij}(k) + (1-\lambda)\phi_{ij}(k), \qquad 0 \leq \lambda \leq 1, \tag{25}$$

where $s_{ij}(k)$ is the trust score assigned by agent i to neighbor j, and $\phi_{ij}(k)$ is an evidence function derived from consistency checks, residuals, authentication results, packet delivery behavior, or historical reliability. A high value of λ makes trust slow to change, which reduces false isolation but delays response to new attacks. A low value of λ makes trust reactive, but it may overreact to noise or temporary communication failures.

Trust scores can be incorporated into the consensus or optimization weights:

$$\tilde{w}_{ij}(k) = \frac{s_{ij}(k) w_{ij}}{\sum_{\ell \in \mathcal{N}_i} s_{i\ell}(k) w_{i\ell}}, \tag{26}$$

so that suspicious neighbors gradually lose influence even if they are not fully isolated. Reputation mechanisms extend this idea by allowing agents to share trust evidence, while confidence-level methods quantify uncertainty in each neighbor's information. Malicious neighbor isolation occurs when $s_{ij}(k)$ falls below a threshold and the corresponding edge is temporarily removed from the local communication graph. Trust-based methods are attractive because they are adaptive and can handle intermittent attacks, but they also introduce new risks: false accusations, collusion among malicious nodes, reputation manipulation, and delayed response under stealthy attacks.

6.5. Resilient Distributed Optimization

Resilient distributed optimization is essential for decentralized smart grids, microgrids, industrial IoT, and multi-agent CPSs in which agents must jointly optimize a system-level objective using local cost functions. A typical distributed optimization problem is

$$\min_{z \in \mathcal{Z}} F(z) = \sum_{i \in \mathcal{V} \setminus \mathcal{F}} f_i(z), \tag{27}$$

where $f_i(z)$ is the local cost of benign agent i, z is the shared decision variable, and $\mathcal{F}$ is the set of malicious agents. In normal conditions, agents exchange estimates or gradients with neighbors and converge toward a minimizer. Under attack, compromised agents may send false gradients, distorted estimates, or inconsistent values to different neighbors. Such attacks are especially damaging because they can bias the optimizer while appearing as plausible operational data.

Recent work on resilient distributed optimization for CPSs proposes event-triggered methods under adversarial environments, including f-total and f-local attack models. These methods aim to reduce communication overhead while guaranteeing that benign agents converge to a bounded safety interval, and the approach has been validated in multi-

microgrid economic dispatch scenarios [8]. This makes resilient distributed optimization directly relevant to smart grids and microgrids, where distributed economic dispatch must remain stable despite information attacks on individual generation units [3].

A resilient optimization update can be written in generic form as

$$z_i(k+1) = \Pi_{\mathcal{Z}} \left[z_i(k) - \gamma_k \nabla f_i(z_i(k)) + \sum_{j \in \mathcal{R}_i(k)} a_{ij}(k)(z_j(k) - z_i(k)) \right], \tag{28}$$

where $\Pi_{\mathcal{Z}}$ is projection onto the feasible set, γ_k is a step size, and $\mathcal{R}_i(k)$ is the filtered neighbor set after removing suspicious values. Event-triggered versions transmit updates only when a local condition is satisfied:

$$\left\| z_i(k) - z_i(k_i^{\ell}) \right\| \geq \sigma_i \|z_i(k)\| + \epsilon_i, \tag{29}$$

where k_i^{ℓ} is the most recent triggering instant, σ_i regulates sensitivity, and ϵ_i avoids excessive triggering near zero. Event-triggering is important because decentralized CPSs often operate with limited bandwidth, energy, and computation resources. It also reduces unnecessary message exposure, although it may complicate detection because fewer communication samples are available.

Privacy-preserving optimization and federated approaches are also relevant in this class. In industrial CPSs, raw data may be sensitive, proprietary, or too large to transmit. Federated learning and distributed intelligence frameworks process data locally and aggregate model updates rather than centralized raw data; recent work argues that such decentralized processing can improve data privacy and resilience in dynamic CPSs [33]. However, privacy-preserving optimization must be combined with security controls because malicious clients can poison gradients, manipulate local updates, or exploit aggregation rules.

6.6. Reconfiguration and Isolation Methods

Reconfiguration and isolation methods modify the communication, sensing, or control structure after an attack is detected or suspected. Their goal is to prevent local compromise from propagating into system-level failure. In a decentralized CPS, reconfiguration may include isolating compromised nodes, switching topology, rerouting communication, changing controller parameters, activating backup communication links, or moving to fallback control.

Let $\mathcal{G}(k) = (\mathcal{V}, \mathcal{E}(k))$ be the communication graph at time k. If the set of suspicious nodes or edges is denoted by $\mathcal{S}(k)$, a simple graph reconfiguration operation is

$$\mathcal{G}^{+}(k) = \mathcal{G}(k) \setminus \mathcal{S}(k), \qquad \mathcal{S}(k) \subseteq \mathcal{V} \cup \mathcal{E}(k). \tag{30}$$

This expression captures the main idea of malicious node isolation and suspicious link removal. In practice, the operation must be constrained by connectivity, safety, and control requirements. Removing too many edges may protect the system from attack but break the graph conditions required for consensus or optimization. Therefore, isolation must be balanced with maintaining sufficient connectivity among benign agents.

Topology switching and rerouting are useful when DoS, jamming, or packet dropping affects a subset of links. Controller reconfiguration is needed when the physical plant cannot safely operate under the same control law after sensor or actuator compromise. Fallback control provides a conservative local controller that keeps each subsystem within safety bounds until coordination is restored. The main limitation of reconfiguration methods is

that they depend on timely and reliable detection; if detection is wrong, the system may isolate benign nodes or keep malicious nodes active.

6.7. Recovery and Self-Healing Methods

Recovery and self-healing methods restore acceptable functionality after performance degradation has occurred. This class includes recovery control, self-healing mechanisms, backup controllers, state reconstruction, and restoration of network connectivity. Recovery differs from detection and isolation because its focus is not only identifying the attack but restoring mission performance.

A useful recovery metric is the time required for performance to return above an acceptable threshold:

$$T_r = \inf\{t > t_a : P(t) \geq \rho P_0\} - t_a, \qquad 0 < \rho \leq 1, \tag{31}$$

where t_a is the attack or disturbance time, $P(t)$ is the observed performance level, P_0 is nominal performance, and ρ is an acceptable recovery threshold. Resilience assessment studies emphasize that degradation level, recovery time, and minimum performance level are key indicators for cyber–physical infrastructures [10,11,18,30].

Recovery control may use observers to reconstruct states corrupted by false data injection or safe-mode controllers to keep physical variables within constraints. Self-healing mechanisms may rejoin isolated nodes after validation, rebuild routing tables, restore communication links, or redistribute tasks among remaining benign agents. Backup controllers are useful for critical processes where service cannot be interrupted. Restoration of network connectivity is particularly important in decentralized CPSs because coordination methods require a sufficiently connected graph. The main challenge is that recovery must be cyber–physical: restoring the cyber network without considering physical transients may create unsafe control actions, while restoring physical operation without verifying cyber integrity may reintroduce the attacker.

6.8. AI/ML-Based Cyber-Resilience Methods

AI/ML-based methods are increasingly used for anomaly detection, adaptive defense, predictive maintenance, and autonomous response in CPSs. Their relevance to decentralized CPSs is growing because edge devices and local controllers can process data near the physical process, reducing latency, and dependence on a central cloud. The main AI/ML-based approaches include anomaly detection using machine learning, federated learning, graph neural networks, reinforcement learning for adaptive defense, and edge intelligence.

ML-based anomaly detection can learn normal operating patterns from sensor and communication data. Federated learning allows multiple agents to train a shared model while keeping raw data local, which supports privacy-preserving resilience in CPSs [33]. Graph neural networks are attractive because decentralized CPSs are naturally graph-structured: nodes represent agents, sensors, or controllers, and edges represent communication or physical coupling. Reinforcement learning can support adaptive defense by learning response policies under changing attack conditions, while edge intelligence enables fast local decisions without centralized inference.

However, AI/ML-based resilience introduces its own attack surface. Adversaries can poison training data, manipulate features, craft adversarial examples, compromise model aggregation, or exploit distribution shifts. Surveys on adversarial examples in deep learning-enabled cybersecurity systems show that machine learning models used for cyber defense can themselves be vulnerable to adversarial attacks and require dedicated defenses [34]. Therefore, AI/ML methods should not be treated as replacements for model-based resilience, secure control, or formal graph-theoretic guarantees. Instead, they should

be integrated as complementary layers that improve detection, prediction, and adaptation while being protected by robust training, validation, and monitoring.

6.9. Digital Twin-Based Resilience

Digital twin-based resilience is an emerging direction for decentralized CPSs. A digital twin is a dynamic virtual representation of a physical system that can be used for monitoring, simulation, attack analysis, degradation prediction, strategy validation, and decision support. In the resilience context, a digital twin can compare expected and observed behavior, simulate attack scenarios before applying mitigation actions, estimate the effect of isolating nodes or links, and test recovery strategies without endangering the real system.

A generic digital twin state can be represented as

$$\mathcal{D}(t) = \Psi\big(\mathcal{X}(t), \mathcal{C}(t), \Theta_p(t), \mathcal{H}(t)\big), \tag{32}$$

where $\mathcal{X}(t)$ denotes physical-state data, $\mathcal{C}(t)$ denotes cyber and communication state, $\Theta_p(t)$ denotes model parameters, and $\mathcal{H}(t)$ denotes historical observations. The twin can be used to compute predicted behavior and compare it with actual observations. If the deviation becomes significant, the system may trigger additional diagnosis or recovery actions.

Digital twins are especially useful for CPS cybersecurity because they make it possible to evaluate attack propagation, test response policies, and support operator decision-making. Recent work on digital twins and cybersecurity highlights their value for analysis, optimization, and cyber–physical evaluation while also noting that digital twins introduce security risks because unauthorized access to the twin may reveal sensitive system information [35,36]. Therefore, digital twin-based resilience must include access control, data integrity protection, synchronization validation, and secure interfaces between the twin and the physical CPS.

6.10. Summary of the Taxonomy

The nine method classes and their distinguishing assumptions are summarized in Supplementary Table S7. In brief, prevention reduces exposure, detection identifies abnormal behavior, consensus and optimization preserve coordination, trust and reconfiguration adapt information flow, recovery restores service, and AI/ML or digital-twin methods add predictive and adaptive capabilities.

A lifecycle mapping of the method classes is provided in Supplementary Table S8 and is used to structure the comparison in the next section.

The main implication of this taxonomy is that no single method class is sufficient for decentralized CPS cyber-resilience. Prevention without recovery is brittle; detection without isolation does not stop propagation; resilient consensus without detection may tolerate only bounded attacks; AI/ML without adversarial robustness may become a new vulnerability; and digital twins without secure synchronization may mislead decision-making. A rigorous assessment should therefore compare methods as complementary components of a multi-layer cyber-resilience architecture rather than as isolated security tools.

7. Comparative Analysis of Existing Methods

The taxonomy in Section 6 identifies the main classes of cyber-resilience methods for decentralized CPSs. However, a taxonomy alone does not show whether two methods are comparable, which assumptions they require, or under which attack and topology conditions they are likely to be effective. For this reason, this section develops a criterion-based comparative analysis. The objective is not only to describe existing methods but to compare them through a common analytical lens that captures attack coverage, architectural assump-

tions, information requirements, resilience mechanism, topology support, communication cost, scalability, recovery capability, and validation evidence.

A criterion-based comparison is particularly important for decentralized CPSs because the same method may perform differently depending on graph connectivity, adversary distribution, availability of trusted nodes, local observability, and physical-process constraints. For example, a residual-based detector may be effective when the local model is accurate, but it may miss stealthy false data injection attacks; a W-MSR consensus algorithm may tolerate malicious values under robust graph conditions, but it does not by itself reconstruct damaged physical states; and an event-triggered resilient optimization method may reduce communication load while introducing additional assumptions about triggering thresholds and bounded adversaries [8,23,25,37]. The comparison therefore treats each method as a combination of assumptions, mechanisms, and evidence rather than as an isolated algorithmic label.

7.1. Comparison Criteria

The criteria used in this article are summarized in Table 2. They are selected to support a fair comparison across the heterogeneous method classes introduced in Section 6. The first four criteria describe the problem setting: attack type, architecture, required information, and resilience mechanism. The next three criteria describe operational feasibility: topology support, communication overhead, and scalability. The last two criteria describe practical usefulness: recovery capability and validation type.

Table 2. Criteria for comparative analysis of cyber-resilience methods in decentralized CPSs.

Criterion	Meaning in the Context of Decentralized CPSs
Type of attack	Attack class addressed by the method, such as denial-of-service (DoS), false data injection (FDI), replay, spoofing, malicious node, Byzantine behavior, or compromised actuator.
CPS architecture	Architectural setting assumed by the method: centralized, distributed, decentralized, or hybrid. For this article, decentralized and multi-agent settings are the primary focus.
Required information	Scope of information needed by each agent: purely local measurements, neighbor-based messages, partial network information, trusted-node information, or global state/topology information.
Resilience mechanism	Main mechanism used to preserve or restore functionality, such as detection, filtering, isolation, resilient consensus, reconfiguration, optimization, recovery, or self-healing.
Topology support	Network assumption under which the method operates: fixed graph, switching graph, dynamic topology, directed topology, undirected topology, or time-varying communication.
Communication overhead	Additional communication introduced by the method, classified qualitatively as low, medium, or high depending on message exchange, alert propagation, and coordination requirements.
Scalability	Expected ability to scale with the number of agents, classified as low, medium, or high depending on computational and communication complexity.
Recovery capability	Whether the method explicitly supports post-attack recovery: yes, partial, or no. Detection-only methods are usually partial unless linked to isolation or recovery control.
Validation type	Evidence used in the literature: theory, numerical simulation, hardware-in-the-loop experiment, laboratory testbed, or real-system validation.

For a structured comparison, let $\mathcal{M} = \{M_1, \ldots, M_R\}$ denote a set of representative method classes and let $\mathcal{K} = \{K_1, \ldots, K_Q\}$ denote the set of comparison criteria. The comparative matrix can be written as

$$\mathbf{C} = [c_{rq}] \in \mathcal{V}^{R \times Q}, \qquad c_{rq} = \Gamma(M_r, K_q), \tag{33}$$

where $\Gamma(\cdot)$ maps method M_r to a qualitative or quantitative value under criterion K_q, and $\mathcal{V}$ denotes the set of possible values such as *local* , *neighbor-based*, *global*, *low*, *medium*, *high*, *yes*, *partial*, and *no*. In later assessment, these qualitative values can be converted into normalized scores only after the application context and safety priorities are specified.

7.2. Comparative Matrix of Representative Method Classes

The full method-class comparative matrix is moved to Supplementary Table S9. The main text focuses on the cross-cutting differences that affect deployment: attack coverage, information requirements, topology assumptions, overhead, scalability, validation strength, and recovery capability.

The matrix shows that the main trade-off is not simply between secure and insecure methods but between local autonomy, information availability, communication cost, and recovery depth. Preventive controls are necessary but insufficient for resilience because they mainly reduce exposure before an attack. Detection methods provide situational awareness but usually require isolation, consensus filtering, or recovery control to affect physical outcomes. Resilient consensus and distributed optimization are well aligned with decentralized CPSs because they use neighbor-level information; however, their guarantees depend strongly on graph robustness, adversary bounds, and convergence assumptions. Recovery and digital twin-based approaches provide stronger post-attack support, but they introduce additional modeling, synchronization, and validation requirements.

7.3. Evaluation Logic and Scoring of Method Suitability

For rigorous comparative analysis, the comparison should not stop at a qualitative table. A transparent scoring logic can be introduced to evaluate the suitability of a method for a given CPS application. Let $s_{rq} \in [0, 1]$ be the normalized suitability of method M_r under criterion K_q, and let w_q be the importance weight assigned by the application context. A generic method suitability score is

$$S(M_r) = \sum_{q=1}^{Q} w_q s_{rq}, \qquad \sum_{q=1}^{Q} w_q = 1, \tag{34}$$

where high values indicate stronger suitability for the selected scenario. The scoring direction must be defined carefully. For example, high scalability and strong recovery capability increase s_{rq}, whereas high communication overhead decreases it. If $o_r \in \{1, 2, 3\}$ denotes low, medium, and high overhead, a simple normalized overhead score may be written as

$$s_{r,\text{overhead}} = 1 - \frac{o_r - 1}{2}, \tag{35}$$

so that low overhead receives score 1, medium overhead receives score 0.5, and high overhead receives score 0. Similar ordinal mappings can be used for scalability, topology flexibility, and recovery depth, provided that the assumptions are explicitly stated.

This scoring logic is not intended to replace experimental validation. Instead, it provides a transparent bridge between the qualitative taxonomy and the quantitative resilience assessment framework developed later in this article. In a smart-grid scenario, for example, detection delay, FDI tolerance, and recovery time may receive higher weights. In a multi-

robot scenario, topology flexibility, consensus accuracy, and communication overhead may be more important. In industrial CPSs, validation type and safe fallback behavior may dominate the evaluation because unsafe transients can directly affect equipment and operators.

7.4. Scenario-Oriented Interpretation

The scenario interpretation indicates that a practical decentralized CPS should rarely rely on a single method class. A resilient smart grid may combine authentication, PMU-based FDI detection, event-triggered distributed optimization, and digital-twin-supported recovery testing. A cooperative robot network may combine W-MSR consensus, neighbor trust estimation, and reconfiguration after suspicious behavior. An industrial CPS may combine secure protocols, distributed detection, fallback control, and post-incident recovery procedures. The comparative analysis therefore supports a layered interpretation of cyber-resilience, where methods are selected according to threat assumptions and operational constraints.

7.5. Key Findings from the Comparative Analysis

The comparison leads to five main findings. First, decentralized CPS resilience is strongly constrained by information locality. Methods that require global information may be analytically attractive but are less suitable for large-scale decentralized systems where agents can only observe local or neighbor states. Second, graph topology is a central hidden assumption. Consensus and distributed optimization methods typically require sufficient redundancy, robustness, or connectivity; otherwise, malicious nodes can dominate local neighborhoods. Third, detection-only approaches are insufficient for resilience unless they are connected to isolation, reconfiguration, or recovery control. Fourth, communication overhead is a practical limiting factor: high-frequency alert exchange, model synchronization, or digital twin updates can reduce scalability and may even create new attack surfaces. Fifth, validation depth remains uneven across the literature. Many methods are supported by theory and simulation, while fewer are validated on hardware-in-the-loop platforms, laboratory CPS testbeds, or real operational systems.

These comparative findings define the hypotheses that are tested against the systematic evidence base in the next section. The evidence synthesis quantifies whether the literature is in fact dominated by particular CPS domains, attack families, resilience mechanisms, and validation practices. Section 9 then interprets those empirical patterns against the taxonomy and research questions before Section 10 translates them into measurable indicators such as detection time, performance degradation, convergence error, communication overhead, compromised-node tolerance, recovery time, and service availability.

8. Results: Systematic Evidence Synthesis

8.1. Study Selection and Quality Profile

Figure 5 summarizes the completed selection process. The executed Scopus export contained 366 records. After excluding 19 unsuitable publication types and separating 10 secondary reviews, 337 primary records entered title/abstract screening. Of these, 174 were excluded and 163 advanced to eligibility assessment. A total of 84 were excluded at eligibility, leaving 79 eligible primary studies. Quality appraisal retained 72 studies for the Core comparative synthesis and 7 as contextual evidence. Thus, the empirical results below are based on the 72 Core studies, whereas the contextual set is used only to illustrate emerging or weakly validated directions. Table 3 summarizes the study-selection and quality-appraisal counts.

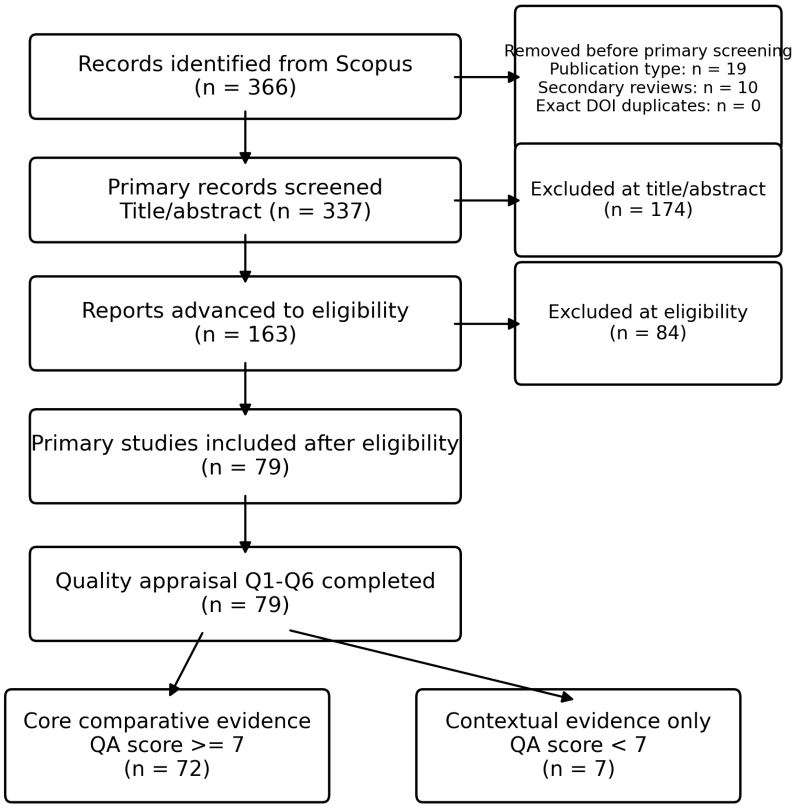


Figure 5. PRISMA-style flow of the executed Scopus screening, eligibility, and quality-appraisal process. Publisher platforms and repositories were used for retrieval and verification but were not counted as additional identification sources.

Table 3. Study selection and quality-appraisal summary.

Stage	Count	Share/Note
Records identified from Scopus	366	Primary export
Publication-type exclusions	19	Pre-screen exclusion
Secondary reviews separated	10	Background/citation chasing
Primary records screened	337	Title/abstract stage
Records excluded at title/abstract	174	51.6% of primary screened
Reports assessed for eligibility	163	Include + unclear after TA screen
Reports excluded at eligibility	84	51.5% of eligibility set
Eligible primary studies	79	Entered Q1–Q6 appraisal
Core comparative synthesis	72	QA score ≥ 7
Contextual evidence only	7	QA score < 7

8.2. Post Hoc Database-Coverage Sensitivity

The complementary coverage check documented 19 public candidate records. Nine were new eligible candidates not present in the original 79-study set, two were duplicates of already included studies, three remained uncertain because the decentralized/topological criterion could not be confirmed from the public record, and five were excluded. All nine new eligible candidates fell within existing categories, predominantly power/microgrid applications and FDI-oriented resilient-control or detection mechanisms; one explicitly addressed hybrid FDI + DoS attacks, and several used experimental, OPAL-RT, or HIL-style validation. No new CPS-domain, attack-family, defense-mechanism, or validation category

was required. The result therefore indicates that database coverage affects corpus breadth and may influence category prevalence, but the sampled complementary evidence does not alter the structure of the proposed taxonomy. Detailed records, DOI/title deduplication status, screening rationale, and source URLs are provided in the supplementary coverage log.

8.2.1. Inter-Rater Reproducibility Audit

The post hoc reproducibility audit contained 25 sampled records, of which 24 had full text available to both assessors; AUD-14 remained abstract-only and was excluded from the primary reliability calculation. On the 24 full-text studies, Core/Contextual classification agreed in 17 cases (70.8%), yielding Cohen's $\kappa = 0.30$. The mean absolute difference between assessor total scores was 2.25 points on the 0–12 scale. Across all 144 paired Q1–Q6 cells, exact agreement was 61.1% (88/144). The lowest criterion-level agreement occurred for validation strength (Q4; exact agreement 45.8%, weighted $\kappa = 0.23$) and threat-model clarity (Q1; weighted $\kappa = 0.25$), which motivated clarification of the operational scoring boundaries.

After the independent ratings were frozen, 17 studies underwent documented adjudication: seven classification disagreements, seven large within-class discrepancies, and three targeted residual checks. Relative to the original quality-appraisal classification, the adjudicated scores would move 3 studies from Contextual to Core and 2 from Core to Contextual, yielding 73 Core and 6 Contextual in the post hoc sensitivity classification. Because this was a post hoc reproducibility audit, the primary aggregate synthesis retains the original pre-specified classifications; the consensus decisions are reported as a sensitivity layer and were not used to recalculate the pre-adjudication agreement coefficients.

8.2.2. Post Hoc Classification Sensitivity

To test whether these boundary reclassifications materially affected the descriptive synthesis, the principal distributions were recomputed using the 73-study consensus Core set while retaining the original 72-study set as the primary analysis. The substantive pattern was stable. Power systems and microgrids remained the dominant CPS domain (40/72, 55.6%, versus 40/73, 54.8%); FDI/deception/integrity remained the largest primary attack family (23/72, 31.9%, versus 24/73, 32.9%); simulation/numerical studies remained the predominant validation mode (51/72, 70.8%, versus 52/73, 71.2%); and resilient/secure control remained present in all Core studies. The only notable rank change was within the middle attack categories: DoS/DDoS increased from 12/72 (16.7%) to 14/73 (19.2%) and moved above multi-vector/hybrid attacks, which remained 13 studies (17.8% under the 73-study denominator). Secondary defense ranks changed only at the margin, with event-triggered resilience and AI/ML-assisted defense exchanging adjacent positions. Thus, the post hoc consensus classification altered boundary composition but did not change the leading domain, leading attack family, dominant validation mode, or principal defense pattern; the main taxonomy and descriptive conclusions were therefore robust to the adjudicated reclassification.

The complete study-level evidence table for all 72 Core studies is provided as Supplementary Table S3. The main text retains only aggregate distributions needed to answer the research questions and support the taxonomy. Supplementary Table S2 preserves the seven contextual studies from the primary appraisal, while the separate adjudication log documents the post hoc sensitivity reclassifications.

The Core evidence base comprised 55 journal articles (76.4%) and 17 conference papers (23.6%) published between 2013 and 2026. The mean quality score was 8.28/12 and the median was 8/12; 32 studies scored at least 9/12. This distribution indicates that the

Core set is not composed only of minimal-threshold papers, but the quality profile is still heterogeneous, especially in validation depth, statistical evidence, and transparency.

8.3. CPS Domains

The evidence base was strongly concentrated in energy applications. Power systems and microgrids accounted for 40 of 72 Core studies (55.6%), followed by connected transport and vehicles (18.1%) and generic multi-agent/networked CPS models (12.5%). Industrial CPS/IIoT, UAV/robotic systems, and water/critical-infrastructure applications were comparatively rare. The domain distribution in Table 4 therefore shows that general claims about decentralized CPS resilience are often extrapolated from an evidence base dominated by power-electronic and microgrid settings.

Table 4. CPS domains represented in the 72-study Core evidence base.

CPS Domain	Studies	Percent
Power systems and microgrids	40	55.6%
Connected transport and vehicles	13	18.1%
Generic MAS/networked CPS	9	12.5%
General CPS	4	5.6%
Industrial CPS/IIoT/ICS	3	4.2%
UAV/robotics/swarm	2	2.8%
Water/critical infrastructure	1	1.4%

8.4. Attack Models

False-data, deception, and integrity attacks were the largest mutually exclusive primary attack family, representing 23 studies (31.9%). DoS/DDoS was the primary family in 12 studies (16.7%). A total of 13 studies (18.1%) explicitly evaluated multi-vector or hybrid attacks, while 20 studies (27.8%) used a broader or application-specific adversarial model that could not be reduced to a single standard attack family. Byzantine/malicious-node attacks and stand-alone jamming were much less common as primary categories. Table 5 highlights the central gap: most studies still evaluate one dominant attack mechanism rather than interacting integrity, availability, and communication-layer attacks.

Table 5. Primary attack families in the Core evidence base.

Primary Attack Family	Studies	Percent
FDI/deception/integrity	23	31.9%
Generic/other cyberattack	20	27.8%
Explicit multi-vector/hybrid	13	18.1%
DoS/DDoS	12	16.7%
Byzantine/malicious nodes	3	4.2%
Jamming	1	1.4%

Multi-label coding provides additional context: 34 studies contained an FDI/deception/integrity component and 20 contained a DoS/DDoS component. Only isolated studies explicitly addressed spoofing/MITM, replay, or jamming. Thus, although the literature recognizes diverse threats conceptually, the experimental evidence remains centered on integrity manipulation and availability disruption.

8.5. Defense Mechanisms

All Core studies contained some form of resilient or secure control because this was an eligibility requirement for the primary synthesis. Beyond that common layer, detection/IDS appeared in 40 studies (55.6%), observer or estimator compensation in 22

(30.6%), and resilient consensus/coordination in 18 (25.0%). Event-triggered resilience and AI/ML-assisted defense each appeared in 10 studies. Recovery/self-healing and isolation/reconfiguration were found in only six studies each, while digital-twin and safety/barrier-function approaches were rare. Because mechanisms are multi-label, the counts in Table 6 intentionally exceed 72 when summed.

The mechanism distribution reveals a lifecycle imbalance. Detection and attack-tolerant control are relatively mature, but explicit recovery, self-healing, safe reconfiguration, and formal safety enforcement are much less represented. This finding supports the lifecycle-oriented taxonomy in Section 6: cyber-resilience is frequently operationalized as "continue operating under attack" rather than as the complete sequence of detect, absorb, adapt, recover, and learn.

Table 6. Prevalence of resilience mechanisms in the Core evidence base (multi-label coding).

Mechanism	Studies	Percent of Core
Resilient/secure control	72	100.0%
Detection/IDS	40	55.6%
Observer/estimator compensation	22	30.6%
Resilient consensus/coordination	18	25.0%
Event-triggered resilience	10	13.9%
AI/ML-assisted defense	10	13.9%
Trust/reputation	8	11.1%
Isolation/reconfiguration	6	8.3%
Recovery/self-healing	6	8.3%
Authentication/cryptography	3	4.2%
Digital twin	2	2.8%
Safety/barrier-function control	1	1.4%

8.6. Validation Strength

Validation is the clearest weakness of the evidence base. Fifty-one studies (70.8%) relied primarily on simulation or numerical studies. Only 21 studies (29.2%) received the strongest Q4 validation score, corresponding to hardware-in-the-loop, real-time, physical/laboratory, co-simulation, or other source-described strong experimental evidence. Within the mutually exclusive validation coding, six studies used HIL/real-time testbeds, six used physical/laboratory experiments, one used explicit co-simulation/cyber–physical testbed validation, and eight had other strong experimental validation documented by the source. Only five studies (6.9%) obtained the highest Q5 score for repeated/statistical evidence. Table 7 summarizes the mutually exclusive primary validation categories.

Table 7. Primary validation approaches in the Core evidence base.

Validation Approach	Studies	Percent
Simulation/numerical study	51	70.8%
Experimental/strong validation (source-described)	8	11.1%
HIL/real-time testbed	6	8.3%
Physical/laboratory experiment	6	8.3%
Co-simulation/cyber–physical testbed	1	1.4%

This imbalance matters because graph-theoretic guarantees and simulation performance do not automatically transfer to real deployments with packet loss, timing jitter, model mismatch, hardware limits, and simultaneous cyber–physical disturbances. The evidence therefore supports stronger use of HIL, cyber ranges, digital twins, and reproducible multi-seed campaigns rather than one-off deterministic simulations.

8.7. Reported Outcomes

The reported outcomes were concentrated in two categories: consensus/coordination/tracking (28 studies, 38.9%) and voltage/frequency/power regulation (27 studies, 37.5%). Detection/isolation/localization was the primary outcome in eight studies, while seven emphasized stability or resilient operation. Only two studies placed safety or constraint preservation at the center of the reported result. No mutually exclusive primary category was dominated by recovery time or service restoration, reinforcing the underrepresentation of post-attack recovery. Table 8 summarizes these primary outcome categories.

Taken together, the results show a technically substantial but uneven evidence base. Resilient control under FDI and DoS is well represented, especially in microgrids, whereas explicit multi-vector attack interaction, non-energy domains, lifecycle recovery, safety guarantees, and strong repeated experimental validation remain comparatively sparse.

Table 8. Primary reported outcome categories in the Core evidence base.

Primary Outcome Category	Studies	Percent
Consensus/coordination/tracking	28	38.9%
Voltage/frequency/power regulation	27	37.5%
Detection/isolation/localization	8	11.1%
Stability/resilient operation	7	9.7%
Safety/constraint preservation	2	2.8%

9. Discussion

The systematic evidence synthesis changes the role of the taxonomy from a purely conceptual classification into an empirically tested organizing structure. The 72-study Core set is large enough to reveal recurring patterns, but the distribution of evidence is highly uneven across domains, attacks, mechanisms, and validation depth. This section interprets those patterns against RQ1–RQ4 and separates evidence-supported conclusions from broader future research directions, which are reserved for Section 12.

9.1. RQ1: Classification of Cyber-Resilience Methods

The evidence supports a multi-dimensional taxonomy rather than a single list of algorithms. FDI/deception and DoS-like disruption dominate the attack dimension, while power systems and microgrids account for 55.6% of the Core evidence. Resilient control, detection, observer compensation, and consensus dominate the mechanism dimension, whereas recovery, self-healing, explicit safety preservation, and digital-twin-supported resilience are uncommon. These concentrations mean that an algorithm label alone is not an adequate descriptor: the same "resilient control" label can hide different attacker capabilities, topology assumptions, information requirements, and validation conditions.

The eligibility process also confirms that "distributed" and "decentralized" should not be treated as synonyms. A substantial number of otherwise relevant studies were excluded because their operation still depended on a supervisory or centralized element. Architecture and topology must therefore remain explicit taxonomy dimensions. This distinction is consistent with recent reviews of distributed industrial CPS architectures and security modeling, which emphasize architecture-level assumptions and multi-layer attack paths as determinants of resilience [4,13].

9.2. RQ2: Strengths and Limitations of Representative Methods

The strongest evidence exists for model-based resilient control combined with detection, observer-based compensation, or resilient coordination. These approaches benefit from mature control-theoretic tools: stability or boundedness can often be stated under explicit

attack and graph assumptions, and performance can be evaluated through tracking, regulation, or convergence errors. This helps explain why consensus/coordination/tracking and voltage/frequency/power regulation account for most primary outcomes.

However, attack tolerance should not be equated with full cyber-resilience. Detection without isolation does not stop propagation; observer compensation can preserve performance without identifying the adversary; resilient consensus may depend on graph-robustness and bounded-neighbor assumptions; and event-triggered designs reduce communication burden while adding timing and threshold assumptions. The evidence is especially thin after the "withstand" phase: recovery/self-healing and isolation/reconfiguration each occur in only a small minority of Core studies. Recent resilience reviews similarly emphasize that rapid response, restoration, and adaptation remain distinct from conventional prevention and secure control [12,15].

AI/ML-assisted defense is present but not yet dominant. Its strongest role in the Core set is anomaly detection, classification, and adaptive estimation rather than formally guaranteed closed-loop recovery. The practical limitation is not only data dependence; learned components also introduce distribution shift, poisoning, and adversarial-evasion risks. For decentralized CPSs, a defensible research direction is therefore hybrid: use learning to improve perception or detection while retaining explicit safety, control, and fallback constraints around the learned component.

9.3. RQ3: Quantitative Indicators for Decentralized CPS Resilience

The evidence shows that the field measures control performance more consistently than resilience lifecycle performance. Convergence error, tracking error, voltage/frequency regulation, stability, and detection accuracy are common, but recovery time, retained functionality, service availability, safety-margin violation, and repeated statistical uncertainty are reported much less consistently. Only 6.9% of Core studies received the highest Q5 score for repeated or statistical evidence.

This supports the assessment framework in Section 10. A method should be evaluated with a vector of cyber and physical indicators rather than a single accuracy or convergence metric. At minimum, experiments should report attack-detection/isolation delay, physical degradation during the attack, post-attack recovery time, service availability, communication/computational overhead, and the attack/topology conditions under which the guarantee is claimed. Metric-oriented IIoT reviews reach a compatible conclusion: cyber-resilience assessment is useful only when metrics are tied to explicit assets, objectives, and operational context [14].

9.4. RQ4: Integration into a Reusable Assessment Framework

The combined taxonomy and synthesis support a six-step evaluation workflow: define the attacker and compromised components; state decentralization and topology assumptions; identify the resilience mechanism; measure cyber-layer detection and physical degradation; quantify adaptation/recovery; and report validation strength and reproducibility. The evidence shows why all six are necessary. Two studies can both claim FDI resilience while differing fundamentally in whether they require global information, tolerate unbounded injections, preserve physical safety, assume a fixed graph, or rely only on numerical validation.

The weighted resilience score proposed later in this article should therefore be interpreted as application-specific, not as a universal ranking. The underlying indicators and weights must remain visible. This is particularly important because the literature is domain-imbalanced: a weighting scheme suitable for microgrid secondary control cannot be transferred unchanged to UAV swarms, vehicle platoons or industrial process control.

9.5. Limitations of the Evidence Base and Review Process

Several limitations constrain the strength of the conclusions. First, the reproducible PRISMA identification stream is based on an executed Scopus export rather than a completed authenticated multi-database search. A post hoc public coverage sensitivity check identified nine additional eligible candidates and therefore confirms that the original stream does not exhaust the literature. However, those candidates mapped to existing taxonomy dimensions, so the check supports structural robustness of the taxonomy while cautioning against interpreting the original category percentages as database-independent prevalence estimates. Second, authoritative publisher or repository records were used for a subset of eligibility decisions when open PDFs were unavailable; the audit trail distinguishes these cases, but they are not equivalent to direct full-text access. Third, the formal inter-rater analysis was conducted post hoc on a stratified 25-study audit sample rather than prospectively across the complete screening set; one abstract-only record was excluded, leaving 24 full-text studies for the primary reliability estimate. Because the sample intentionally oversampled boundary cases, the observed Cohen's $\kappa = 0.30$ should be interpreted as an audit-sample reproducibility estimate rather than as a prevalence-neutral measure of agreement for all 79 studies. Fourth, heterogeneous outcomes prevent a defensible pooled meta-analysis. Finally, the strong concentration in power and microgrid studies limits external validity for other decentralized CPS domains.

These limitations are material but transparent. They imply that the present work should be read as a PRISMA-guided systematic mapping and evidence synthesis rather than as an exhaustive multi-database meta-review. The main contribution is the traceable link between a controlled evidence set and the taxonomy/assessment framework; future updates should extend database coverage and prospectively formalize duplicate screening and agreement assessment rather than changing the already reported Scopus counts retrospectively.

10. Cyber-Resilience Assessment Framework

The taxonomy in Section 6 and the comparative analysis in Section 7 identify what classes of methods exist and how they differ in architectural assumptions, information requirements, threat coverage, and validation depth. However, an evidence-grounded contribution requires a further step: the methods must be mapped to measurable cyber-resilience indicators. The purpose of this section is therefore to propose an assessment framework that connects decentralized CPS behavior, adversarial disturbances, resilience dimensions, and quantitative metrics into a single evaluation logic. This framework is not intended to replace domain-specific safety or reliability standards. Rather, it provides a common structure for comparing decentralized CPS resilience methods across smart grids, microgrids, industrial IoT, cooperative robotics, and other multi-agent CPS applications.

Recent resilience assessment studies emphasize that cyber-resilience should be evaluated dynamically rather than only through static security controls. NIST frames cyber-resilience in terms of anticipating, withstanding, recovering from, and adapting to adverse cyber conditions [6]. Quantitative CPS-oriented frameworks model resilience through performance degradation and recovery trajectories [30], while R-KPI studies highlight functionality loss, recovery time, and minimum performance level after disturbances as important indicators for cyber–physical infrastructures [10]. In addition, recent cyber-resilience quantification research uses resilience curves and area-under-curve reasoning to capture dynamic behavior over attack and recovery phases [38]. These ideas motivate the framework developed below.

Figure 6 summarizes the assessment logic. A threat scenario first affects cyber or physical components of the decentralized CPS. A resilience method then modifies the

system response through prevention, detection, filtering, isolation, reconfiguration, or recovery. The resulting performance trajectory is measured through indicators such as detection time, degradation level, convergence error, communication overhead, and service availability. Finally, normalized metrics are aggregated into an interpretable resilience score.

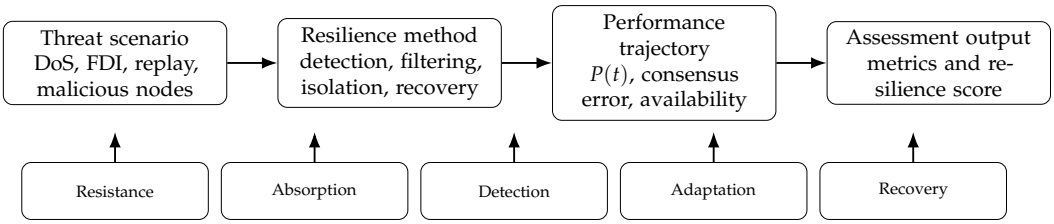


Figure 6. Proposed assessment logic for cyber-resilience in decentralized CPSs.

10.1. Assessment Dimensions

Cyber-resilience in decentralized CPSs should be assessed across multiple dimensions because no single metric can fully capture the behavior of a networked cyber–physical system under attack. A method may detect attacks quickly but cause high communication overhead; another method may preserve consensus but recover slowly; a third method may provide strong prevention but weak post-compromise adaptation. Therefore, this article proposes five assessment dimensions: resistance, absorption, detection, adaptation, and recovery. They represent the main stages through which a decentralized CPS experiences, mitigates, and recovers from malicious disruption. Table 9 summarizes their meanings and typical indicators.

Table 9. Assessment dimensions for cyber-resilience in decentralized CPSs.

Dimension	Meaning	Typical Indicators
Resistance	Ability to reduce the probability or initial impact of attack penetration before harmful propagation occurs.	Authentication success rate, encryption coverage, access-control effectiveness, attack success probability, redundant-path availability.
Absorption	Ability to maintain acceptable service and physical performance while the attack is active.	Minimum retained functionality, bounded state deviation, degraded-mode service level, load served, safety-margin preservation.
Detection	Ability to identify abnormal or malicious behavior rapidly and accurately without relying on a central monitor.	Attack detection time, false-positive rate, false-negative rate, residual threshold crossing, neighbor-alert consistency.
Adaptation	Ability to change communication, control, or decision-making behavior in response to detected compromise.	Topology switching time, trust-score update speed, fallback mode activation, event-triggered adjustment, policy reconfiguration.
Recovery	Ability to restore coordinated operation, performance, and connectivity after attack isolation or disturbance termination.	Recovery time, restored connectivity, convergence error after recovery, service availability, performance restoration ratio.

Let the assessment state of a decentralized CPS under a given attack scenario be represented by the normalized vector

$$\boldsymbol{\rho} = \left[\rho_{\text{res}}, \rho_{\text{abs}}, \rho_{\text{det}}, \rho_{\text{adp}}, \rho_{\text{rec}}\right]^{\top} \in [0,1]^5, \tag{36}$$

where the entries denote resistance, absorption, detection, adaptation, and recovery capabilities, respectively. This vector makes it possible to compare methods that emphasize different parts of the resilience lifecycle. For example, encryption and authentication primarily increase ρ_{res}, W-MSR consensus primarily improves ρ_{abs} and ρ_{rec} under bounded malicious neighbors, and digital-twin-supported response validation can improve ρ_{adp} and ρ_{rec}.

The assessment dimensions should be interpreted relative to a baseline operating condition. Let P_0 denote the nominal performance level and $P(t)$ the observed mission performance during a horizon $[t_0, t_0 + T]$. A normalized performance trajectory can be defined as

$$\hat{P}(t) = \frac{P(t)}{P_0}, \qquad 0 \le \hat{P}(t) \le 1. \tag{37}$$

This trajectory is central for assessing absorption and recovery because it captures how much mission functionality is retained during and after the attack. If $P(t)$ cannot be measured directly, it can be approximated using domain-specific indicators such as load served in a smart grid, production rate in an industrial CPS, consensus accuracy in a multi-agent network, or task completion ratio in a robot swarm.

10.2. Proposed Metrics

The framework uses measurable indicators derived from simulation, HIL/laboratory tests, or operational logs, including detection time, degradation, recovery time, coordination error, service availability, safety violations, and communication overhead. Definitions and interpretation are provided in Supplementary Table S10 [10,18,30,38].

For decentralized CPSs, these metrics must be defined at both local and network levels. Local metrics describe how an individual agent reacts to attacks, while network-level metrics describe whether the benign subnetwork can preserve coordination. For a benign-agent set $\mathcal{B}$, a consensus-oriented convergence error can be written as

$$E_c(k) = \frac{1}{|\mathcal{B}|} \sum_{i \in \mathcal{B}} \|x_i(k) - x^{\star}(k)\|, \tag{38}$$

where $x_i(k)$ is the state of benign agent i and $x^{\star}(k)$ is the desired agreement, reference, or optimal trajectory. For distributed optimization, $x^{\star}(k)$ may be replaced with the optimizer of the global or constrained aggregate objective when available.

Service availability can be measured using a threshold-based indicator. Let $P_{\min}$ be the minimum acceptable performance level for safe or useful operation. Then

$$A_S = \frac{1}{T} \int_{t_0}^{t_0+T} \mathbb{I}\left(\hat{P}(t) \ge \frac{P_{\min}}{P_0}\right) dt, \tag{39}$$

where $\mathbb{I}(\cdot)$ equals one when the condition is true and zero otherwise. This metric is especially useful for critical infrastructures because it separates complete restoration from minimal viable service. A system may not fully recover immediately, but it may still remain available if it delivers an acceptable degraded service level.

Communication overhead should be included because many decentralized resilience methods depend on extra neighbor messages, alert exchange, trust propagation, or digital twin synchronization. A normalized overhead metric can be defined as

$$O_C = \frac{N_{\text{res}} - N_0}{N_0}, \tag{40}$$

where N_0 is the number of messages in normal operation and N_{res} is the number of messages required when the resilience mechanism is active. Since high overhead can reduce scalability and increase exposure to DoS attacks, communication efficiency can be expressed as

$$C = \frac{1}{1 + O_C}. \tag{41}$$

Thus, a method with no additional overhead has $C = 1$, while methods with large communication costs receive lower scores.

The performance trajectory is summarized using the retained-functionality integral in Equation (2), after normalizing or clipping the trajectory to the intended $[0, 1]$ scale where appropriate. This complements the maximum degradation G_P because it captures both severity and duration of performance loss without introducing a second equivalent definition.

10.3. General Resilience Score

The final step of the assessment framework is to aggregate normalized indicators into a general resilience score. The score should be transparent, adjustable, and scenario-dependent. The baseline aggregation is the weighted form already defined in Equation (3). For the five-component implementation used here, the normalized indicator vector is $[D, A, R, C, S]$, where D is detection capability, A is absorption capability, R is recovery capability, C is communication efficiency, and S is system stability or service availability; the corresponding weights $w_1, \ldots, w_5$ sum to one. The weights reflect the priorities of a specific CPS application. For example, w_1 and w_5 may be high in smart-grid state estimation, w_3 may dominate in self-healing microgrids, and w_4 may be important in low-bandwidth IIoT or robot-swarm networks.

Each component used in Equation (3) should be normalized so that higher values indicate better resilience. A possible normalization is

$$D = 1 - \min\left(\frac{T_D}{T_D^{\max}}, 1\right), \tag{42}$$

$$A = 1 - G_P, \tag{43}$$

$$R = 1 - \min\left(\frac{T_R}{T_R^{\max}}, 1\right), \tag{44}$$

$$C = \frac{1}{1 + O_C}, \tag{45}$$

$$S = A_S, \tag{46}$$

where $T_D^{\max}$ and $T_R^{\max}$ are application-defined maximum acceptable detection and recovery times. This formulation penalizes slow detection, high degradation, slow recovery, and high communication overhead while rewarding continuous service availability.

The score can be extended to include resistance and adaptation explicitly when the application requires more detail. For instance, a seven-component score can be written as

$$RS^+ = \alpha_1\rho_{\text{res}} + \alpha_2\rho_{\text{abs}} + \alpha_3\rho_{\text{det}} + \alpha_4\rho_{\text{adp}} + \alpha_5\rho_{\text{rec}} + \alpha_6 C + \alpha_7 S, \tag{47}$$

with $\sum_{j=1}^{7} \alpha_j = 1$. Equation (47) is useful when prevention, adaptation, and recovery must be evaluated separately, while Equation (3) provides the compact baseline form for comparative reporting.

The resilience score is not universal by default; it must be configured according to the CPS mission. A high score in one scenario does not necessarily imply superiority in another unless the same weights, threat assumptions, and validation conditions are used.

10.4. Assessment Procedure

The proposed framework can be applied through a five-step procedure. First, define the decentralized CPS model, including agents, communication topology, physical performance indicator, and benign/malicious node assumptions. Second, specify the attack scenario using the threat model from Section 4; this includes attack type, target layer, start time, duration, and adversary capability. Third, select the cyber-resilience method or method combination to be assessed. Fourth, measure or simulate the metrics defined in Supplementary Table S10. Fifth, normalize the metrics and compute the resilience score using Equation (3) or Equation (47).

This procedure provides a bridge between the qualitative and quantitative parts of this article. The taxonomy identifies candidate methods, the comparative analysis clarifies their assumptions, and the assessment framework determines how their resilience should be measured. It also supports reproducibility because future studies can report threat assumptions, metric definitions, weights, and validation type explicitly. In this sense, the proposed assessment framework complements the comparative matrix in Section 7 and prepares the methodological basis for a case study or simulation scenario in the next section.

11. Reviewer-Driven Simulation Case Study

This case study was refined in response to Reviewer 1 to test three issues that are easily obscured in consensus-only illustrations: graph assumptions for W-MSR, cyber-to-physical effects, and the distinction between attack onset and detector alarm time. The analysis remains a reduced-order synthetic simulation and is not presented as hardware-in-the-loop or field validation.

11.1. Graph and Adversary Assumptions

The evaluated network contains ten agents, with agents 3 and 7 treated as adversarial and the W-MSR filtering parameter fixed at $f = 1$. The original communication configuration violates the checked F-local assumption because one benign agent has two adversarial neighbors; it is therefore treated only as an assumption-violation stress case. A separately constructed 36-edge topology (Valid36) satisfies the checked static F-local placement condition, while an edge-count-matched 36-edge topology (Stress36) provides a stress comparison with a different adversarial-neighborhood pattern. These static checks do not establish uniform robustness of the time-varying graph during DoS-driven link interruptions.

11.2. Simulation Design, Detection, and Resilience Score

The consensus layer is coupled to a reduced-order synthetic microgrid response model that outputs frequency deviation (Hz), voltage deviation (p.u.), and active-power-sharing error (p.u.). The electrical coupling is represented by

$$P_i = \sum_j B_{ij} V_i V_j \sin(\delta_i - \delta_j), \tag{48}$$

$$Q_i = \sum_j B_{ij} \Big(V_i^2 - V_i V_j \cos(\delta_i - \delta_j) \Big), \tag{49}$$

and the reduced-order dynamics are

$$M\dot{f}_i = -D_f f_i + u_{f,i} - P_i - \ell_i, \tag{50}$$

$$\dot{\delta}_i = 2\pi f_i, \tag{51}$$

$$\tau_V \dot{V}_i = -(V_i - 1) - K_Q Q_i + u_{v,i}. \tag{52}$$

The principal simulation parameters were $M = 0.80$, $D_f = 1.60$, $\tau_V = 0.35$ s, and $K_Q = 0.10$. These values were declared simulation parameters rather than identified hardware parameters. The model was not calibrated to an IEEE benchmark feeder, field measurements, or an HIL platform; the results therefore quantify within-model behavior only. The main comparison uses Mean aggregation and W-MSR under FDI, DoS, and combined FDI + DoS. For the refined comparison, 50 paired random seeds (9000–9049) were used per topology/attack/method combination with numerical integration step $\Delta t = 0.0125$ s. Smaller seed sets were used for parameter-sensitivity sweeps and are not pooled with the main 50-seed results.

Attack onset is treated as a programmed intervention, not a detection event. A separately calibrated one-step innovation/dropout detector produces an alarm time and detection delay. Detector thresholds were frozen from clean calibration seeds 7000–7019 before held-out attack evaluation: the innovation thresholds were $h_f = 0.016436$ and $h_V = 0.00491068$, and four consecutive missing messages triggered the dropout alarm. Clean-run false-alarm evaluation used separate seeds and is reported separately from calibration; zero observed false alarms in a finite test set is not interpreted as a zero population false-positive rate.

For each run, a five-component simulation score summarizes detection (D), absorption/operational continuity (A), recovery (R), communication delivery reliability (C_{rel}), and service stability (S):

$$RS_5 = w_D D + w_A A + w_R R + w_C C_{\mathrm{rel}} + w_S S, \qquad w_D + w_A + w_R + w_C + w_S = 1. \tag{53}$$

For the reported Stage-7 values, equal weights $w_D = w_A = w_R = w_C = w_S = 0.2$ were used, with $T_D^{\max} = 2$ s and $T_R^{\max} = 10$ s. The communication term differs deliberately from the generic overhead-oriented C in Equation (41); here, it measures delivery preservation relative to the matched clean run,

$$C_{\mathrm{rel}} = \min\left(\frac{\mathrm{delivery}_{\mathrm{attack}}}{\mathrm{delivery}_{\mathrm{clean,matched}}}, 1\right). \tag{54}$$

The normalization anchors and weights are application dependent. Accordingly, RS_5 is interpreted together with the underlying physical and communication indicators rather than as a universally validated resilience metric. Means and 95% bootstrap confidence intervals were used for simulation outputs, with paired seed-level comparisons for Mean versus W-MSR; sensitivity checks varied physical parameters, FDI strength, packet delivery, integration step, score weights, and normalization caps.

11.3. Refined Stage-7 Results

Table 10 summarizes the refined 50-seed FDI + DoS comparison for W-MSR. Valid36 produced markedly smaller simulated frequency and voltage deviations than Stress36 under the evaluated attack and parameter distribution. This difference is consistent with the evaluated neighborhood structure but does not establish an unconditional topology guarantee.

Table 10. Refined simulation results for W-MSR under FDI + DoS ($\Delta t = 0.0125$ s).

Topology	n	Mean Peak $\lvert\Delta f\rvert$ (Hz)	Mean Peak $\lvert\Delta V\rvert$ (p.u.)	Mean RS_5
Valid36	50	0.0151	0.0023	0.9644
Stress36	50	0.0742	0.0123	0.8464

The observed detector fraction was 50/50 for both configurations under this test distribution. In 100 clean detector runs per relevant configuration, no false alarms were observed; the corresponding upper 95% binomial bound is approximately 3.7%, so the result must not be described as a zero false-positive probability. The observed 100% detection fraction likewise applies only to the evaluated attacks and settings.

For Valid36 under FDI + DoS, the paired comparison gave mean $RS_5 \approx 0.9644$ for W-MSR and 0.8451 for Mean aggregation. Under DoS alone, the two methods were nearly indistinguishable ($RS_5 \approx 0.93524$), which rules out a claim that W-MSR is uniformly superior across attack types. Sensitivity analyses further showed that effect magnitudes and score rankings depend on physical parameters, FDI strength, packet delivery, score weights, and normalization caps.

11.4. Interpretation and Limitations

The case study supports three conditional conclusions. First, empirical convergence under a graph that violates the checked F-local assumption does not validate the theorem assumptions, while violation of a sufficient condition does not imply failure in every run. Second, the reduced-order physical extension makes cyber-to-physical consequences visible, but it is not an experimentally validated microgrid model. Third, separating attack onset from detector alarm time prevents artificial zero-delay "detection" claims.

The analysis has six principal limitations: synthetic reduced-order electrical dynamics; static rather than uniform temporal graph certification under DoS; fixed adversarial placements and a limited FDI/DoS family; sample-specific detector performance; application-dependent RS_5 weights/normalization and recovery-censoring rules; and a finite sensitivity grid with unequal seed counts between main and auxiliary runs. The results therefore support simulation-based, conditional claims only. External benchmark, HIL, and field validation remain future work.

12. Open Challenges and Future Research Directions

The evidence base identifies a focused research agenda rather than an unrestricted list of emerging technologies. Table 11 summarizes the seven gaps that are directly supported by the Core-study frequencies. The subsections that follow explain their implications for decentralized CPS research.

12.1. From Single Attacks to Adaptive Multi-Vector Adversaries

Most studies still isolate one attack family or apply multiple attacks without modeling their interaction. Real attackers can combine availability disruption with false-data injection, spoofing, compromised insiders, or delayed replay to exploit the defender's transition between normal, degraded, and recovery modes. Security-modeling research in 2026 likewise identifies simplistic treatment of dynamic, multi-layer, and multi-path attacks as a persistent limitation [13]. Future experiments should therefore predefine attacker budgets, knowledge, timing, overlap, compromised-node placement, and adaptation rules and should report whether guarantees hold under simultaneous rather than sequential disruptions.

Table 11. Evidence-based research gaps and recommended research directions.

Research Gap	Evidence Signal	Recommended Direction
Multi-vector attack realism	Only 13/72 Core studies (18.1%) explicitly evaluate multi-vector or hybrid attacks.	Evaluate interacting integrity, availability, and communication attacks with controlled timing, overlap, and attacker budgets rather than isolated single-attack cases.
Domain concentration	40/72 studies (55.6%) concern power systems or microgrids.	Validate transfer to UAV/robot swarms, transportation, water, industrial processes, and heterogeneous CPSs and report which assumptions survive domain transfer.
Simulation-heavy validation	51/72 studies (70.8%) use simulation/numerical validation as the strongest verified mode.	Increase HIL, real-time, cyber-range, digital-twin, and physical-testbed validation including network and hardware non-idealities.
Weak repeated/statistical evidence	Only 5/72 studies (6.9%) receive the highest Q5 score.	Use multi-seed campaigns, confidence intervals, effect sizes, sensitivity analysis, and attack-parameter sweeps.
Limited reproducibility/transparency	Only 29/72 studies receive any Q6 credit.	Release code, parameter files, datasets, or synthetic generators and report assumptions and failure cases explicitly.
Recovery/self-healing underrepresentation	Recovery/self-healing and isolation/reconfiguration each appear in only 6/72 studies.	Couple detection with isolation, degraded-mode control, reconfiguration, safe recovery, and controlled reintegration.
Safety guarantees remain rare	Safety/barrier-function mechanisms appear in 1/72 studies; safety/constraint preservation is the primary outcome in only 2/72.	Integrate resilient control with invariant sets, barrier functions, or other formal safety constraints during attack, detection delay, and recovery.

12.2. Cross-Domain Transfer and Heterogeneous Decentralized Architectures

Power systems provide mature models and testbeds, but their dominance risks overfitting the field's assumptions to comparatively structured electrical networks. Robotics, autonomous vehicles, UAV swarms, water systems, and industrial process networks introduce mobility, stronger timing constraints, intermittent links, heterogeneous sensing, and different safety envelopes. Recent systematic reviews of cyber–physical production systems and IoT resilience show parallel demand for adaptive and domain-aware resilience mechanisms [15,16]. A method should therefore report which parts of its guarantee are graph-theoretic and transferable and which depend on domain-specific plant models or communication properties.

12.3. Validation, Statistical Evidence, and Reproducibility

The clearest maturity gap is experimental. Simulation is valuable for mechanism development but does not capture all effects of jitter, asynchronous updates, packet loss, device saturation, implementation latency, and model mismatch. Stronger studies should progress from deterministic simulation to multi-seed campaigns, co-simulation or network emulation, HIL/real-time execution, and, where feasible, physical testbeds. Statistical reporting should include uncertainty, effect size, and sensitivity rather than only a single representative trajectory. Supplementary packages should contain attack schedules, random seeds, controller parameters, and metric definitions. This is particularly important for review-to-practice transfer because reproducibility determines whether comparative claims can be independently tested.

12.4. Recovery, Safe Degradation, and Self-Healing

The evidence base is substantially stronger for detection and attack-tolerant control than for restoration. A resilient CPS should specify what happens after a malicious node is identified: whether it is isolated, whether topology remains connected, which degraded mode is safe, how state estimates are reconstructed, and under what conditions the node can rejoin. Recovery logic also needs explicit safety constraints. Without such constraints, an aggressive reconfiguration or controller switch can itself create unsafe transients. Formal tools such as invariant-set methods and control barrier functions are therefore promising complements to resilient consensus and observer-based control, especially when detection and communication delays are non-negligible.

12.5. Trustworthy AI and Digital-Twin-Supported Resilience

AI and digital twins are promising but should not be treated as resilience guarantees by themselves. Learning-based detectors require robustness to drift, poisoning, and adversarial manipulation, while digital twins require trustworthy synchronization and explicit handling of model discrepancy. A practical architecture should bind learned or twin-based components to verifiable decision boundaries, fallback controllers, and audit logs. Recent broad CPS resilience surveys and IoT-resilience frameworks emphasize the same transition from prevention-only security toward adaptive response and recovery [12,15]. The key research question is therefore not whether AI or digital twins improve detection in isolation but whether they improve closed-loop mission resilience under bounded computational and communication resources.

12.6. Benchmarking and Reporting Standards

The field would benefit from a minimum reporting profile for decentralized CPS resilience experiments. Each study should state the following: graph/topology assumptions; malicious-node model; attacker knowledge; attack schedule and budget; physical and cyber metrics; baselines; repeated-run protocol; computational/communication cost; failure conditions; and code/data availability. A common profile would not force one universal resilience score, but it would make heterogeneous studies comparable and support future meta-analysis where domain-specific metrics converge.

12.7. Synthesis

The most consequential future work will integrate rather than merely add technologies. Detection must connect to isolation and safe control; resilient coordination must expose its topology and adversary assumptions; learning must operate inside assurance boundaries; and recovery must be measured as explicitly as attack detection. The evidence suggests that progress will be judged increasingly by reproducibility and end-to-end validation rather than by the novelty of an isolated algorithm. For decentralized CPSs, this means demonstrating that local decisions preserve mission-relevant physical performance under realistic network and attacker conditions and that the system can return safely to coordinated operation after disruption.

13. Conclusions

This study examined cyber-resilience in decentralized CPSs through a PRISMA-guided systematic mapping, a multi-dimensional taxonomy, comparative analysis, and an evidence-based assessment framework. The executed Scopus stream yielded 366 records; 337 primary records were screened, 163 underwent eligibility assessment, 79 were included, and 72 met the quality threshold for Core synthesis. The evidence is technically substantial but uneven: 55.6% of Core studies concern power systems or microgrids, only 18.1% explicitly evaluate

multi-vector or hybrid attacks, and 70.8% rely on simulation-only validation. Strong experimental or real-time evidence is present in 29.2% of studies, while the highest level of repeated/statistical evidence is rare.

Across the evidence base, resilient control, distributed detection, observer-based compensation, and resilient consensus are the most mature mechanisms. In contrast, recovery, self-healing, explicit safety preservation, cross-domain transfer, and reproducible multi-vector evaluation remain comparatively weak. These findings support a taxonomy that keeps attack family, architecture/topology, information requirements, resilience mechanism, and validation evidence visible rather than collapsing them into a single method label.

The proposed assessment framework translates those dimensions into measurable resistance, absorption, detection, adaptation, and recovery indicators. Its purpose is not to produce a universal ranking but to make assumptions, trade-offs, and evidence strength explicit for a given CPS mission. The microgrid case study illustrates how such metrics can be operationalized, while the systematic synthesis shows where stronger validation is still needed.

This review is limited by its reproducible Scopus-only primary identification stream, the use of authoritative extended records for a subset of eligibility decisions, a post hoc rather than prospectively designed inter-rater audit, and heterogeneous outcomes that preclude a meaningful global meta-analysis. A separate post hoc public database-coverage sensitivity check identified nine additional eligible candidates absent from the original corpus; all mapped to existing taxonomy dimensions, indicating that coverage affects corpus breadth but did not require structural revision of the taxonomy. The audit nonetheless provides a transparent reproducibility estimate (17/24 Core/Contextual agreement; Cohen's $\kappa = 0.30$) while preserving the frozen pre-adjudication ratings. A separate post hoc sensitivity analysis using the 73-study consensus Core classification produced only small distributional shifts and preserved the leading domain, attack family, validation mode, and defense pattern, supporting the robustness of the principal descriptive conclusions to boundary reclassification. These limitations define the next methodological steps: complementary database searches, prospectively defined duplicate screening and agreement assessment, open extraction artifacts, and stronger cross-domain testbed evidence. Within those boundaries, this study provides a traceable foundation for comparing decentralized CPS cyber-resilience methods and for designing future evaluations that connect cyber attack handling to physical service continuity and safe recovery.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/fi18110624/s1. Table S1: Protocol-defined Scopus search query; Table S2: Seven contextual studies (not in the 72-study Core synthesis); Table S3: Complete 72-Core-study evidence table; Table S4: Extended contribution matrix; Table S5: Architectural assumptions relevant to CPS cyber-resilience; Table S6: Distinction between related resilience concepts; Table S7: Detailed taxonomy of cyber-resilience method classes; Table S8: Method classes mapped to the resilience lifecycle; Table S9: Comparative matrix of method classes; Table S10: Proposed assessment metrics. The bibliographic sources supporting the supplementary tables are included in the main reference list [6–10,18,20,23–27,31–37,39–111].

Author Contributions: Conceptualization, K.S. and N.M.; Methodology, K.S., N.M. and D.T.; Formal analysis, K.S. and N.M.; Investigation, N.M., D.T., A.U. and S.T.; Data curation, N.M. and A.U.; Validation, D.T., A.U. and S.T.; Visualization, N.M. and S.T.; Writing—original draft preparation, N.M.; Writing—review and editing, K.S., N.M., D.T., A.U. and S.T.; Supervision, K.S.; Project administration, K.S. All authors have read and agreed to the published version of the manuscript

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The systematic-review protocol details, screening and quality-appraisal summaries, supplementary evidence tables, adjudication log, database-coverage sensitivity log, and Stage-7 reproducibility materials are documented in the accompanying supplementary/reproducibility files. Publisher-copyrighted full texts are not redistributed. Additional working records may be requested from the corresponding author.

Acknowledgments: During the preparation of this manuscript, the authors used ChatGPT-5.6 (OpenAI) solely for language assistance, including translation from Kazakh and improvement of English style and readability. The authors reviewed and edited all AI-assisted text and take full responsibility for the content of the publication.

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

References

1. Lee, E.A. Cyber Physical Systems: Design Challenges. In Proceedings of the 11th IEEE International Symposium on Object and Component-Oriented Real-Time Distributed Computing (ISORC), Orlando, FL, USA, 5–7 May 2008; pp. 363–369. https://doi.org/10.1109/ISORC.2008.25.
2. NIST Cyber-Physical Systems Public Working Group. *Framework for Cyber-Physical Systems: Volume 1, Overview*; Technical Report NIST SP 1500-201; National Institute of Standards and Technology: Gaithersburg, MD, USA, 2017. https://doi.org/10.6028/NIST.SP.1500-201.
3. Rouhani, S.H.; Su, C.L. Future Cyber-Resilient Renewable and Sustainable Smart Grids: A Critical Review from Power System Researchers' Perspective on Emerging Threats and Solutions. *Renew. Sustain. Energy Rev.* **2025**, *226*, 116299. https://doi.org/10.1016/j.rser.2025.116299.
4. Alves, J.; Sousa, P.; Cruz, T.; Mendes, J. A Review of Architecture Features for Distributed and Resilient Industrial Cyber-Physical Systems. *J. Manuf. Syst.* **2025**, *82*, 1069–1090. https://doi.org/10.1016/j.jmsy.2025.07.012.
5. Xing, W.; Shen, J. Security Control of Cyber-Physical Systems under Cyber Attacks: A Survey. *Sensors* **2024**, *24*, 3815. https://doi.org/10.3390/s24123815.
6. Ross, R.; Pillitteri, V.; Graubart, R.; Bodeau, D.; McQuaid, R. *Developing Cyber-Resilient Systems: A Systems Security Engineering Approach*; Technical Report NIST SP 800-160 Vol. 2 Rev. 1; National Institute of Standards and Technology: Gaithersburg, MD, USA 2021. https://doi.org/10.6028/NIST.SP.800-160v2r1.
7. Liao, Z.; Shi, J.; Wang, S.; Zhang, Y.; Mu, R.; Sun, Z. A Survey of Resilient Coordination for Cyber-Physical Systems Against Malicious Attacks. *IEEE Internet Things J.* **2025**, *12*, 46500–46525. https://doi.org/10.1109/JIOT.2025.3612764.
8. Liao, Z.; Wang, S.; Shi, J.; Li, M.; Zhang, Y.; Sun, Z. Resilient Distributed Optimization for Cyber-Physical Systems under Adversarial Environments: An Event-Based Method. *ISA Trans.* **2024**, *149*, 1–15. https://doi.org/10.1016/j.isatra.2024.04.015.
9. Segovia-Ferreira, M.; Rubio-Hernan, J.; Cavalli, A.R.; Garcia-Alfaro, J. A Survey on Cyber-Resilience Approaches for Cyber-Physical Systems. *ACM Comput. Surv.* **2024**, *56*, 1–37. https://doi.org/10.1145/3652953.
10. Aghazadeh Ardebili, A.; Boscolo, M.; Longo, A.; Pourmadadkar, M.; Ficarella, A.; Padoano, E. Resilience in Cyber-Physical Infrastructures: R-KPI Prioritization, Framework Development, and Case Study Insights. *J. Saf. Sci. Resil.* **2025**, *6*, 100194. https://doi.org/10.1016/j.jnlssr.2024.12.005. Erratum in *J. Saf. Sci. Resil.* **2026**, *7*, 100373.
11. Longo, A.; Aghazadeh Ardebili, A.; Lazari, A.; Ficarella, A. Cyber-Physical Resilience: Evolution of Concept, Indicators, and Legal Frameworks. *Electronics* **2025**, *14*, 1684. https://doi.org/10.3390/electronics14081684.
12. Gomez-Cabrera, A.; Escamilla-Ambrosio, P.J.; Al Faruque, M.A. A Review of Resilience Enhancement Techniques for Cyber-Physical Systems. *IEEE Access* **2025**, *13*, 138061–138082. https://doi.org/10.1109/ACCESS.2025.3593257.
13. Huang, S.; Poskitt, C.M.; Shar, L.K. Security Modelling for Cyber-Physical Systems: A Systematic Literature Review. *ACM Trans. -Cyber-Phys. Syst.* **2026**, *10*, 1–29. https://doi.org/10.1145/3776549.
14. Lezzi, M.; Corallo, A.; Lazoi, M.; Nimis, A. Measuring Cyber Resilience in Industrial IoT: A Systematic Literature Review. *Manag. Rev. Q.* **2026**, *76*, 1107–1161. https://doi.org/10.1007/s11301-025-00495-8.
15. Darbandeh, F.; Asghar, M.R.; Chen, L. Cyber Resilience in IoT Networks Revisited: Frameworks, Mechanisms, and Future Directions. *ACM Comput. Surv.* **2026**, *58*, 1–45. https://doi.org/10.1145/3818682.
16. Barrero-Arciniegas, H.A.; Bataleblu, A.A.; Rauch, E.; Matt, D.T. Sustainable and Resilient Cyber-Physical Production Systems: A Systematic Literature Review. *J. Comput. Des. Eng.* **2026**, *13*, 215–249. https://doi.org/10.1093/jcde/qwag043.
17. Tian, E.; Wu, Z.; Huang, Y.; Chen, H. A Survey of Cyber-Security in Cyber-Physical Systems: Attack Strategies and Protection Mechanisms. *IEEE Trans. Ind.-Cyber-Phys. Syst.* **2026**, *4*, 392–404. https://doi.org/10.1109/TICPS.2026.3696239.

18. Cassottana, B.; Roomi, M.M.; Mashima, D.; Sansavini, G. Resilience Analysis of Cyber-Physical Systems: A Review of Models and Methods. *Risk Anal.* **2023**, *43*, 2359–2379. https://doi.org/10.1111/risa.14089.
19. Rajkumar, R.; Lee, I.; Sha, L.; Stankovic, J. Cyber-Physical Systems: The Next Computing Revolution. In Proceedings of the 47th Design Automation Conference, Anaheim, CA, USA, 13–18 June 2010; pp. 731–736. https://doi.org/10.1145/1837274.1837461.
20. Stouffer, K.; Pease, M.; Tang, C.; Zimmerman, T.; Pillitteri, V.; Lightman, S.; Hahn, A.; Saravia, S.; Sherule, A.; Thompson, M. *Guide to Operational Technology (OT) Security*; Technical Report NIST SP 800-82 Rev. 3; National Institute of Standards and Technology: Gaithersburg, MD, USA, 2023. https://doi.org/10.6028/NIST.SP.800-82r3.
21. Humayed, A.; Lin, J.; Li, F.; Luo, B. Cyber-Physical Systems Security—A Survey. *IEEE Internet Things J.* **2017**, *4*, 1802–1831. https://doi.org/10.1109/JIOT.2017.2703172.
22. Olfati-Saber, R.; Fax, J.A.; Murray, R.M. Consensus and Cooperation in Networked Multi-Agent Systems. *Proc. IEEE* **2007**, *95*, 215–233. https://doi.org/10.1109/JPROC.2006.887293.
23. Pasqualetti, F.; Dörfler, F.; Bullo, F. Attack Detection and Identification in Cyber-Physical Systems. *IEEE Trans. Autom. Control* **2013**, *58*, 2715–2729. https://doi.org/10.1109/TAC.2013.2266831.
24. Dibaji, S.M.; Pirani, M.; Flamholz, D.B.; Annaswamy, A.M.; Johansson, K.H.; Chakrabortty, A. A Systems and Control Perspective of CPS Security. *Annu. Rev. Control* **2019**, *47*, 394–411. https://doi.org/10.1016/j.arcontrol.2019.04.011.
25. LeBlanc, H.J.; Zhang, H.; Koutsoukos, X.; Sundaram, S. Resilient Asymptotic Consensus in Robust Networks. *IEEE J. Sel. Areas Commun.* **2013**, *31*, 766–781. https://doi.org/10.1109/JSAC.2013.130413.
26. Shen, Y.; Qin, Z. Detection, Differentiation and Localization of Replay Attack and False Data Injection Attack Based on Random Matrix. *Sci. Rep.* **2024**, *14*, 2758. https://doi.org/10.1038/s41598-024-52954-z.
27. Teixeira, A.; Shames, I.; Sandberg, H.; Johansson, K.H. A Secure Control Framework for Resource-Limited Adversaries. *Automatica* **2015**, *51*, 135–148. https://doi.org/10.1016/j.automatica.2014.10.067.
28. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. *BMJ* **2021**, *372*, n71. https://doi.org/10.1136/bmj.n71.
29. Rethlefsen, M.L.; Kirtley, S.; Waffenschmidt, S.; Ayala, A.P.; Moher, D.; Page, M.J.; Koffel, J.B.; PRISMA-S Group. PRISMA-S: An Extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. *Syst. Rev.* **2021**, *10*, 39. https://doi.org/10.1186/s13643-020-01542-z.
30. Cao, Z.; Zhao, H.; Wang, Y.; He, C.; Zhou, D.; Han, X. A Resilience Quantitative Assessment Framework for Cyber-Physical Systems: Mathematical Modeling and Simulation. *Appl. Sci.* **2025**, *15*, 8285. https://doi.org/10.3390/app15158285.
31. Wang, J.; Deng, X.; Guo, J.; Zeng, Z. Resilient Consensus Control for Multi-Agent Systems: A Comparative Survey. *Sensors* **2023**, *23*, 2904. https://doi.org/10.3390/s23062904.
32. Manfredi, S.; Angeli, D.; Tortora, C. A Resilient Consensus Algorithm with Inputs for the Distributed Monitoring of Cyber-Physical Systems. *Control Eng. Pract.* **2025**, *154*, 106166. https://doi.org/10.1016/j.conengprac.2024.106166.
33. Azeri, N.; Hioual, O.; Hioual, O. A Distributed Intelligence Framework for Enhancing Resilience and Data Privacy in Dynamic Cyber-Physical Systems. *Clust. Comput.* **2024**, *27*, 6289–6304. https://doi.org/10.1007/s10586-024-04349-y.
34. Macas, M.; Wu, C.; Fuertes, W. Adversarial Examples: A Survey of Attacks and Defenses in Deep Learning-Enabled Cybersecurity Systems. *Expert Syst. Appl.* **2024**, *238*, 122223. https://doi.org/10.1016/j.eswa.2023.122223.
35. Homaei, M.; Mogollón-Gutiérrez, Ó.; Sancho, J.C.; Ávila, M.; Caro, A. A Review of Digital Twins and Their Application in Cybersecurity Based on Artificial Intelligence. *Artif. Intell. Rev.* **2024**, *57*, 201. https://doi.org/10.1007/s10462-024-10805-3.
36. Jiang, Y.; Wang, W.; Ding, J.; Lu, X.; Jing, Y. Leveraging Digital Twin Technology for Enhanced Cybersecurity in Cyber-Physical Production Systems. *Future Internet* **2024**, *16*, 134. https://doi.org/10.3390/fi16040134.
37. Paudel, S.P.S.; Shaaban, M. An Evaluation of Methods for Detecting False Data Injection Attacks in the Smart Grid. *Front. Comput. Sci.* **2024**, *6*, 1504548. https://doi.org/10.3389/fcomp.2024.1504548.
38. Alhidaifi, S.M.; Asghar, M.R.; Ansari, I.S. Cyber Resilience Quantification: A Probabilistic Estimation Model for IT Infrastructure. *Reliab. Eng. Syst. Saf.* **2026**, *265*, 111473. https://doi.org/10.1016/j.ress.2025.111473.
39. Lian, Z.; Zhu, Y.; Guo, F.; Peng, C.; Zhou, Q. Distributed Cyber Resilient Control Strategy for Remote DC Microgrids Under Integrated Satellite Terrestrial Networks. *IEEE Trans. Ind. Inform.* **2025**, *21*, 2363–2372. https://doi.org/10.1109/TII.2024.3507174.
40. Wang, Y.; Rajabinezhad, M.; Zhang, Y.; Zuo, S. Cyber-Physical Defense for Heterogeneous Multi-Agent Systems Against Exponentially Unbounded Attacks on Signed Digraphs. *IEEE Trans. Netw. Sci. Eng.* **2025**, *12*, 2165–2179. https://doi.org/10.1109/TNSE.2025.3545280.
41. Yang, J.; Zhao, Y.; Wang, Z.; Wang, S. Terminal-Based Adaptive Cooperative Control and Protection Strategy in Cyber-Physical System Under Hybrid Cyberattacks. *IEEE Internet Things J.* **2026**, *13*, 7228–7239. https://doi.org/10.1109/JIOT.2025.3638426.
42. Al Maruf, A.; Niu, L.; Clark, A.; Mertoguno, J.; Poovendran, R. A Compositional Approach to Safety-Critical Resilient Control for Systems with Coupled Dynamics. In Proceedings of the IEEE Conference on Decision and Control, Cancún, Mexico, 6–9 December 2022. https://doi.org/10.1109/CDC51059.2022.9992810.

43. Tomić, I.; Breza, M.; McCann, J. Jamming-resilient control and communication framework for cyber physical systems. In Proceedings of the IET Conference Publications, London, UK, 1–2 May 2019. https://doi.org/10.1049/cp.2019.0132.
44. Zuo, S.; Yue, D. Resilient Containment of Multigroup Systems against Unknown Unbounded FDI Attacks. *IEEE Trans. Ind. Electron.* **2022**, *69*, 2864–2873. https://doi.org/10.1109/TIE.2021.3066941.
45. Almalaq, Y.; Kchaou, M.; Alateeq, A.A.; Abbassi, R.; Jerbi, H. Secure event-triggered decentralised control for large-scale system under deception attacks: A dynamic output approach. *Int. J. Syst. Sci.* **2026**, 1–18. https://doi.org/10.1080/00207721.2026.2687893.
46. Roy, S.; Chaudhary, A.; Guha, D.; Negi, R. Adaptive Event-Triggered Finite-Time Control of Affine Cyber-Physical Systems Under Denial-of-Service (DoS) and Deception Attacks. *Int. J. Robust Nonlinear Control* **2026**. https://doi.org/10.1002/rnc.7848.
47. Niu, Y.; He, L.; He, C.; Han, X.; Cao, Z.; Zhou, D. Privacy-preserving resilience control for cyber-physical systems based on the dynamic heterogeneous redundancy architecture. *Sci. Sin. Inf.* **2026**, *56*, 164. https://doi.org/10.1360/SSI-2025-0010.
48. Kachhwaha, M.; Modi, H.; Nehra, M.; Fulwani, D. Resilient Control of DC Microgrids Against Cyber Attacks: A Functional Observer Based Approach. *IEEE Trans. Power Electron.* **2024**, *39*, 459–468. https://doi.org/10.1109/TPEL.2023.3326308.
49. Cao, G.; Jia, R.; Dang, J. Distributed Resilient Mitigation Strategy for False Data Injection Attack in Cyber-Physical Microgrids. *Front. Energy Res.* **2022**, *10*. https://doi.org/10.3389/fenrg.2022.845341.
50. Ning, X.; Jiang, J. Defense-in-depth against insider attacks in cyber-physical systems. *Internet Things -Cyber-Phys. Syst.* **2022**, *2*, 203–211. https://doi.org/10.1016/j.iotcps.2022.12.001.
51. Ali, Z.; Terriche, Y.; Su, C.L.; Parise, G.; Ahmad, S.; Elsisi, M. Distributed Resilient Control for Inverter Based Resources in Shore-Ship Cyber-Physical Microgrids. In Proceedings of the Conference Record - Industrial and Commercial Power Systems Technical Conference, Montreal, QC, Canada, 11–14 May 2025. https://doi.org/10.1109/ICPS64254.2025.11030366.
52. Ahmed, Z.; Lu, Y.; Nasir, M.; Iqbal, R. Digital Twin-Based Resilient Consensus Control of Unmanned Aerial Vehicles under Attack. *IEEE Internet Things J.* **2026**, *13*, 45427–45442. https://doi.org/10.1109/JIOT.2026.3718582.
53. Tadepalli, P.; Pullaguram, D.; Alam, M. Cyber-Resilient Strategy for DC Microgrids Against Concurrent FDI and DoS Attacks. *IEEE Trans. Ind. Inform.* **2025**, *21*, 4756–4767. https://doi.org/10.1109/TII.2025.3545102.
54. Franze, G.; Famularo, D.; Lucia, W.; Tedesco, F. A resilient control strategy for cyber-physical systems subject to denial of service attacks: A leader-follower set-theoretic approach. *IEEE/CAA J. Autom. Sin.* **2020**, *7*, 1204–1214. https://doi.org/10.1109/JAS.2020.1003189.
55. Ma, K.; Cao, Q.; Zhao, P.; Dong, Y.; Hua, C.; Li, L.; Yang, J. Distributed Resilient Control for Frequency/Voltage Recovery and Power Sharing in Seaport Microgrids with All-Electric Ships. *IEEE Trans. Transp. Electrif.* **2025**, *11*, 4070–4082. https://doi.org/10.1109/TTE.2024.3453316.
56. Dai, J.; Chen, W.; Zhou, X. Fuzzy high order differentiator observer based resilient control for distributed battery energy storage systems against unbounded FDI attacks. *IET Control Theory Appl.* **2024**, *18*, 2180–2191. https://doi.org/10.1049/cth2.12622.
57. Zeng, R.; Li, Y.; Hu, S.; Shao, X.; Xu, Y.; Yang, X.; Cao, Y. Edge Resilient Control for Feeder Automation System under Hybrid Attacks: A Hybrid Physical Model-Driven and Data-Driven Approach. *IEEE Trans. Smart Grid* **2025**, *16*, 4184–4196. https://doi.org/10.1109/TSG.2025.3581370.
58. Kchaou, M.; Ali, M.; Abbassi, R.; Jerbi, H. Memory Event-Triggered-Based Adaptive Dynamic Output Decentralized Sliding-Mode Control for Large-Scale Systems Under Deception Attacks. *IEEE Internet Things J.* **2026**, *13*, 38847–38859. https://doi.org/10.1109/JIOT.2026.3704408.
59. Tahani, V.; Zarif, M.; Narm, H.; Amerian, M. Event-Triggered Control and Internal-Feedbacks for Detecting and Mitigating False Data Injection Attacks in Islanded DC Microgrids. *IET Gener. Transm. Distrib.* **2026**, *20*, e70339. https://doi.org/10.1049/gtd2.70339.
60. Deng, C. Distributed Resilient Control for Cyber-Physical Systems under Denial-of-Service Attacks. In Proceedings of the 2019 23rd International Conference on Mechatronics Technology, ICMT 2019, Salerno, Italy, 23–26 October 2019. https://doi.org/10.1109/ICMECT.2019.8932108.
61. Deng, C.; Wang, Y.; Wen, C.; Xu, Y.; Lin, P. Distributed Resilient Control for Energy Storage Systems in Cyber-Physical Microgrids. *IEEE Trans. Ind. Inform.* **2021**, *17*, 1331–1341. https://doi.org/10.1109/TII.2020.2981549.
62. Hou, B.; Zeng, S.; Xi, B.; Jia, S.; Guo, Q.; Xu, L.; Sun, H. Performance of Neighborhood-Watch-Based Resilient Distributed Energy Management Algorithm under Different Types of Cyberattacks. In Proceedings of the Proceedings of 2021 IEEE 4th International Electrical and Energy Conference, CIEEC 2021, Wuhan, China, 28–30 May 2021. https://doi.org/10.1109/CIEEC50170.2021.9510236.
63. Liu, Y.; Dong, X.; Ren, Z.; Liu, J. Secure formation control for resilient multi-agent cyber-physical systems under DoS attacks and faults. *Int. J. Robust Nonlinear Control* **2023**, *33*, 3607–3626. https://doi.org/10.1002/rnc.6602.
64. Tang, Q.; Deng, C.; Wang, Y.; Guo, F.; Fan, S. Iterative Observer-Based Resilient Control for Energy Storage Systems in Microgrids under FDI Attacks. *IEEE Trans. Smart Grid* **2024**, *15*, 4744–4753. https://doi.org/10.1109/TSG.2024.3382125.

65. Eslami, A.; Khorasani, K. Event-Based Privacy Preserving Resilient Leader-Follower Consensus Control against Cyber-Attacks in Multi-Agent Cyber-Physical Systems. *ACM Trans. -Cyber-Phys. Syst.* **2026**, *10*, 1–25. https://doi.org/10.1145/3776544.
66. Jeyaraj, P.; Rajan Samuel Nadar, E.; Mihet-Popa, L. Deep-Block Network for Cyberattack Mitigation and Assessment in Smart Grid Power System with Resilience Indices. *Electr. Power Components Syst.* **2024**, *52*, 2367–2383. https://doi.org/10.1080/15325008.2023.2268073.
67. Mokari, H.; Firouzmand, E.; Sharifi, I.; Doustmohammadi, A. Resilient control strategy and attack detection on platooning of smart vehicles under DoS attack. *ISA Trans.* **2024**, *144*, 51–60. https://doi.org/10.1016/j.isatra.2023.11.019.
68. Khoshnevisan, L.; Liu, X. Resilient neural network-based control of nonlinear heterogeneous multi-agent systems: A cyber-physical system approach. *Nonlinear Dyn.* **2023**, *111*, 19171–19185. https://doi.org/10.1007/s11071-023-08840-w.
69. Lian, Z.; Wen, C.; Guo, F.; Deng, C. Distributed Resilient Secondary Voltage Control for AC Microgrids Under DoS Attacks. In Proceedings of the IEEE International Conference on Control and Automation, ICCA, Naples, Italy, 27–30 June 2022. https://doi.org/10.1109/ICCA54724.2022.9831846.
70. Wang, Y.; Mondal, S.; Deng, C.; Satpathi, K.; Xu, Y.; Dasgupta, S. Cyber-Resilient Cooperative Control of Bidirectional Interlinking Converters in Networked AC/DC Microgrids. *IEEE Trans. Ind. Electron.* **2021**, *68*, 9707–9718. https://doi.org/10.1109/TIE.2020.3020033.
71. Severson, T.; Croteau, B.; Rodríguez-Seda, E.; Kiriakidis, K.; Robucci, R.; Patel, C. A resilient framework for sensor-based attacks on cyber–physical systems using trust-based consensus and self-triggered control. *Control Eng. Pract.* **2020**, *101*, 104509. https://doi.org/10.1016/j.conengprac.2020.104509.
72. Alhalali, S.; Nielsen, C.; El-Shatshat, R. Mitigation of cyber-physical attacks in multi-area automatic generation control. *Int. J. Electr. Power Energy Syst.* **2019**, *112*, 362–369. https://doi.org/10.1016/j.ijepes.2019.05.014.
73. Mokari, H.; Firouzmand, E.; Sharifi, I.; Doustmohammadi, A. Deception Attack Detection and Resilient Control in Platoon of Smart Vehicles. In Proceedings of the 2022 30th International Conference on Electrical Engineering, ICEE 2022, Tehran, Iran, 17–19 May 2022. https://doi.org/10.1109/ICEE55646.2022.9827376.
74. Gong, X.; Cao, Y.; Li, X.; Lin, H.; Shu, Z.; Wen, G. Resilient Output Containment of Heterogeneous Multi-Agent Systems Against Byzantine Attacks. *IEEE Trans. Signal Inf. Process. Over Netw.* **2025**, *11*, 938–951. https://doi.org/10.1109/TSIPN.2025.3592314.
75. Xu, Y.; Zhang, W.; Zhang, Y.; Wu, Z.G. Distributed Resilient Control of Cyber-Physical Multi-Agent Systems Under Attacks. *IEEE Trans. Ind.-Cyber-Phys. Syst.* **2025**, *3*, 241–250. https://doi.org/10.1109/TICPS.2025.3543291.
76. Sahoo, S.; Dragicevic, T.; Blaabjerg, F. An Event-Driven Resilient Control Strategy for DC Microgrids. *IEEE Trans. Power Electron.* **2020**, *35*, 13714–13724. https://doi.org/10.1109/TPEL.2020.2995584.
77. Ali, Z.; Terriche, Y.; Su, C.L.; Tan, S.; Parise, G.; Rouhani, S.; Ahmad, S.; Elsisi, M. Distributed Resilient Control of Inverter-Based Resources in Cyber-Physical Shore-Power DC Shipboard Microgrids Against Cyberattacks. *IEEE Trans. Ind. Appl.* **2026**, *62*, 5795–5814. https://doi.org/10.1109/TIA.2025.3645020.
78. Nian, X.; Wei, Z.; Chen, Y.; Liu, J.; Niu, F.; Wen, Z. Resilient Nash Equilibrium Seeking for Cyber–Physical MASs Under Stochastic Trust Observation. *IEEE Internet Things J.* **2026**, *13*, 35257–35268. https://doi.org/10.1109/JIOT.2026.3697062.
79. Hassan, K.; Vettoretti, D.; Selvaratnam, D.; Sandberg, H.; Andrén, F. Resilient Power Inverter Network: A Hardware-in-the-Loop Simulation Case Study. In Proceedings of the IECON Proceedings (Industrial Electronics Conference), Chicago, IL, USA, 3–6 November 2024. https://doi.org/10.1109/IECON55916.2024.10905863.
80. Zhu, Q.; Bushnell, L.; Başar, T. Resilient Distributed Control of Multi-agent Cyber-Physical Systems. In *Control of Cyber-Physical Systems, Lecture Notes in Control and Information Sciences*; Springer: Berlin/Heidelberg, Germany, 2013; Volume 449. https://doi.org/10.1007/978-3-319-01159-2_16.
81. Sadabadi, M.; Cheng, X.; Cucuzzella, M.; Mayo-Maldonado, J.; Scherpen, J. Resilient Voltage Control for DC Microgrids under Integrity Cyberattacks on Voltage Sensors. *IEEE Trans. Control Syst. Technol.* **2026**, *34*, 2893–2900. https://doi.org/10.1109/TCST.2026.3702600.
82. Gao, B.; Bu, B.; Wang, X. A Comprehensive Resilient Control Strategy for CBTC Systems Through Train-to-Train Communications under Malicious Attacks. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 21015–21033. https://doi.org/10.1109/TITS.2022.3192510.
83. Zarandi, Z.; Sharifi, I. Detection and identification of cyber-attacks in cyber-physical systems based on machine learning methods. In Proceedings of the 2020 11th International Conference on Information and Knowledge Technology, IKT 2020, Tehran, Iran, 22–23 December 2020. https://doi.org/10.1109/IKT51791.2020.9345627.
84. Sadabadi, M. A Resilient-by-Design Distributed Control Framework for Cyber-Physical DC Microgrids. *IEEE Trans. Control Syst. Technol.* **2024**, *32*, 625–636. https://doi.org/10.1109/TCST.2023.3324530.
85. Zeng, R.; Li, Y.; Hu, S.; Liu, J.; Wu, H.; Chen, J.; Xu, Y.; Yang, X.; Cao, Y. Bi-Level Resilient Control Solution for Distributed Feeder Automation System Under Hybrid Attack. *IEEE Trans. Smart Grid* **2025**, *16*, 301–312. https://doi.org/10.1109/TSG.2024.3439751.
86. Zhu, M.; Martínez, S. On attack-resilient distributed formation control in operator-vehicle networks. *SIAM J. Control Optim.* **2014**, *52*, 3176–3202. https://doi.org/10.1137/110843332.

87. The, J.; Li, F.; Hashemi, M.; Fang, H. Model-free Resilient Controller Design based on Incentive Feedback Stackelberg Game and Q-learning. In Proceedings of the 2024 IEEE 7th International Conference on Industrial Cyber-Physical Systems, ICPS 2024, St. Louis, MO, USA, 12–15 May 2024. https://doi.org/10.1109/ICPS59941.2024.10640012.
88. Zhao, Z.; Wang, T.; Peng, F.; Yu, J. Adaptive Fuzzy Resilient Decentralized Control for Nonlinear Large-Scale CPSs Under DoS Attacks. *IEEE Trans. Fuzzy Syst.* **2024**, *32*, 5899–5909. https://doi.org/10.1109/TFUZZ.2024.3434726.
89. Wang, Y.; Deng, C.; Liu, Y.; Wei, Z. A cyber-resilient control approach for islanded microgrids under hybrid attacks. *Int. J. Electr. Power Energy Syst.* **2023**, *147*, 108889. https://doi.org/10.1016/j.ijepes.2022.108889.
90. Petrillo, A.; Pescape, A.; Santini, S. A Secure Adaptive Control for Cooperative Driving of Autonomous Connected Vehicles in the Presence of Heterogeneous Communication Delays and Cyberattacks. *IEEE Trans. Cybern.* **2021**, *51*, 1134–1149. https://doi.org/10.1109/TCYB.2019.2962601.
91. Karimi, A.; Ahmadi, A.; Shahbazi, Z.; Shafiee, Q.; Bevrani, H. A Resilient Control Method against False Data Injection Attack in DC Microgrids. In Proceedings of the 2021 7th International Conference on Control, Instrumentation and Automation, ICCIA 2021, Tabriz, Iran, 23–24 February 2021. https://doi.org/10.1109/ICCIA52082.2021.9403594.
92. Xiahou, K.; Zou, K.; Wang, Y.; Lin, Z.; Liu, Y.; Liu, Z.; Wu, Q. Distributed Observer-Based Resilient Control of Cyber-Physical DC Microgrids Against Communication-Link and Local Attacks. *IEEE Trans. Smart Grid* **2025**, *16*, 5459–5471. https://doi.org/10.1109/TSG.2025.3593976.
93. Clarke, S.; Hwang, S.; Thapliyal, O.; Hwang, I. Distributed Denial-of-Service Resilient Control for Urban Air Mobility Applications. *J. Aerosp. Inf. Syst.* **2023**, *20*, 873–889. https://doi.org/10.2514/1.I011222.
94. Li, Z.; Li, Q.; Ding, D.W.; Xie, X. Event-based secure leader-following consensus for cyber-physical systems under denial-of-service attacks. *Int. J. Robust Nonlinear Control* **2022**, *32*, 6346–6360. https://doi.org/10.1002/rnc.6144.
95. Li, K.; Wang, Y.; Chen, H.; Gao, J. A Resilient Control Strategy in Connected Vehicles under Actuator Attacks. In Proceedings of the 2023 7th CAA International Conference on Vehicular Control and Intelligence, CVCI 2023, Changsha, China, 27–29 October 2023. https://doi.org/10.1109/CVCI59596.2023.10397258.
96. Usmani, G.; Rehan, M.; Wu, X.; Rehman, A. Resilient Consensus Control for Heterogeneous Multi-Agent Systems Subject to Two-Channel Stochastic Deception Attacks. *IEEE Trans. Consum. Electron.* **2026**, *72*, 2049–2060. https://doi.org/10.1109/TCE.2025.3645813.
97. Yang, Y.; Guo, L.; Li, X.; Li, J.; Liu, W.; He, H. A Data-Driven Detection strategy of False Data in Cooperative DC Microgrids. In Proceedings of the IECON Proceedings (Industrial Electronics Conference), Toronto, ON, Canada, 13–16 October 2021. https://doi.org/10.1109/IECON48115.2021.9589318.
98. Yu, M.; Wang, W.; Huang, J.; Wen, C.; Zhou, J. Decentralized Adaptive Secure Control of Uncertain Nonlinear Time-Varying Interconnected Systems Against Sensor and Actuator Attacks. *IEEE Trans. Cybern.* **2024**, *54*, 5012–5025. https://doi.org/10.1109/TCYB.2024.3373198.
99. Amullen, E.; Shetty, S.; Keel, L. Model-based resilient control for a multi-agent system against Denial of Service attacks. In Proceedings of the World Automation Congress Proceedings, Rio Grande, Puerto Rico, 31 July–4 August 2016. https://doi.org/10.1109/WAC.2016.7582963.
100. Saeidipour, M.; Fani, B.; Keyvani-Boroujeni, B.; Sadeghkhani, I. Cyber-Resilient Distributed Secondary Control for Grid-Forming Inverters in AC Microgrids Using Short-Time Correlation Transform. *IET Renew. Power Gener.* **2026**, *20*, e70229. https://doi.org/10.1049/rpg2.70229.
101. Xu, K.; Niu, Y.; Zhang, Z. Resilient Control Strategy for Smart Grid Under Delayed Sporadic Measurements: Attack Estimation via Hybrid Observer. *IEEE Trans. Autom. Sci. Eng.* **2025**, *22*, 19175–19186. https://doi.org/10.1109/TASE.2025.3593480.
102. Mishra, M.; Mishra, S.; Mahish, P. Resilient Voltage Restoration Scheme for AC Microgrids under Cyber Attack. In Proceedings of the 2025 IEEE Energy Conversion Congress and Exposition Asia: Shaping a Greener Future with Power Electronics, ECCE-Asia 2025, Bengaluru, India, 11–14 May 2025. https://doi.org/10.1109/ECCE-Asia63110.2025.11112463.
103. Vu, T.; Mukherjee, S.; Adetola, V. Resilient Communication Scheme for Distributed Decision of Interconnecting Networks of Microgrids. In Proceedings of the 2023 IEEE Power and Energy Society Innovative Smart Grid Technologies Conference, ISGT 2023, Washington, DC, USA, 16–19 January 2023. https://doi.org/10.1109/ISGT51731.2023.10066341.
104. Yang, H.; Dai, L.; Ma, Y.; Qiang, Z.; Xia, Y.; Liu, G.P. Resilience Distributed MPC for Dynamically Coupled Multiple Cyber–Physical Systems Subject to Severe Attacks. *IEEE Trans. Cybern.* **2025**, *55*, 3193–3205. https://doi.org/10.1109/TCYB.2025.3569377.
105. Wu, S.; Ye, D.; Shao, X. Event-based distributed resilient control strategy for microgrids subject to disturbances and hybrid attacks. *Appl. Math. Comput.* **2023**, *459*, 128273. https://doi.org/10.1016/j.amc.2023.128273.
106. 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*, 5250–5261. https://doi.org/10.1109/TASE.2026.3663295.
107. Deng, C.; Wen, C. MAS-Based Distributed Resilient Control for a Class of Cyber-Physical Systems with Communication Delays under DoS Attacks. *IEEE Trans. Cybern.* **2021**, *51*, 2347–235. https://doi.org/10.1109/TCYB.2020.2972686.

108. Zhang, K.; Dong, X.; Zhang, H.; Su, R. Zero-Sum Optimal Control for a Cyber–Physical System in Unreliable Communications via Secure Decentralized Policy Iteration. *IEEE Internet Things J.* **2025**, *12*, 27339–27349. https://doi.org/10.1109/JIOT.2025.3562385.
109. Xiao, X.; Zhou, Q.; Wang, F.; Huang, W. Three-stage Defensive Framework for Distributed Microgrid Control Against Cyberattacks. *J. Mod. Power Syst. Clean Energy* **2022**, *10*, 1669–1678. https://doi.org/10.35833/MPCE.2021.000333.
110. Hassan, K.; Selvaratnam, D.; Sandberg, H. On Resilience Guarantees by Finite-Time Robust Control Barrier Functions With Application to Power Inverter Networks. *IEEE Open J. Control Syst.* **2024**, *3*, 497–513. https://doi.org/10.1109/OJCSYS.2024.3487408.
111. Yemini, M.; Nedić, A.; Goldsmith, A.J.; Gil, S. Resilient Distributed Optimization for Multiagent Cyberphysical Systems. *IEEE Trans. Autom. Control* **2025**, *70*, 3952–3967. https://doi.org/10.1109/TAC.2025.3532791.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.