Sciences Methods and Technologies International Journal (SciMeTech)

ISSN: 3085-5284, journal homepage: www.scimetech.com

Multi-Agent Collaborative Intrusion Detection Systems: A Survey of Architectures, Agent Technologies, and AI Techniques

AGRAR Hajar , ELKHADIR Zyad, ACHKARI BEGDOURI Mohammed

SIGL Laboratory, ENSATe of Tetouan, Abdelmalek Essaâdi University, Morocco hajar.agrar@gmail.com, z.elkhadir@uae.ac.ma, m.achkaribegdouri@uae.ac.ma

How to cite:

AGRAR Hajar , ELKHADIR Zyad, ACHKARI BEGDOURI Mohammed, "Multi-Agent Collaborative Intrusion Detection Systems: A Survey of Architectures, Agent Technologies, and AI Techniques", Sciences Methods and Technologies International Journal (SciMeTech), Vol 2, Issue 1, p 193-202

Keywords:

Intrusion Detection Systems
Collaborative IDS
Multi-Agent Systems
Blockchain
Federated Learning
Artificial Intelligence

Abstract

The rapid evolution of cyber threats has exposed the limitations of traditional, isolated Intrusion Detection Systems (IDS). Collaborative Intrusion Detection Systems (CIDS) address these shortcomings by enabling distributed nodes to share information and coordinate responses. MultiAgent Systems (MAS) provide a natural paradigm for implementing CIDS due to their properties of autonomy, sociality, reactivity, and proactiveness. This survey presents a comprehensive taxonomy of MAS-based CIDS covering 2021–2026. We classify existing works according to (1) system architecture, (2) agent technology, and (3) decision techniques, including federated learning, blockchain, and large language models. We review more than 30 recent publications, highlight trends, compare performance, and identify open issues. Persistent challenges include the lack of standardized benchmarks, real-world scalability, and the vulnerability of CIDS themselves to adversarial attacks. Finally, we propose concrete recommendations for next-generation collaborative security systems.

I. Introduction

a. Context and Motivation

Over the last decade, the digitalization of services and the proliferation of connected devices have dramatically increased the attack surface of modern networks. Recent industry reports indicate that the number of Internet connected devices and the global financial impact of cybercrime continue to grow at a steep pace. At the same time, attackers increasingly automate and coordinate their activities, using advanced malware, large-scale botnets, and multi-vector Distributed Denial of Service (DDoS) attacks. Traditional Intrusion Detection Systems (IDS), typically deployed in isolation on a host or a network segment, struggle to cope with these developments. They often lack a global view of distributed activities, exhibit high false-alarm rates in dynamic environments, and may become single points of failure. These limitations motivate the move towards distributed and collaborative detection approaches that can correlate observations across multiple vantage points.

b. Collaborative Intrusion Detection Systems (CIDS)

Collaborative Intrusion Detection Systems (CIDS) consist of multiple IDS nodes that cooperate by sharing data, alerts, models, or trust information to obtain a more complete picture of ongoing activities. Collaboration can range from simple alert forwarding to joint model training and reputation-based trust management. Early work such as DIDS already demonstrated the benefits of distributing detection across sites. Since then, a variety of architectures have been explored, from centralized solutions with a global correlator to fully decentralized peer-to-peer schemes. CIDS can improve detection accuracy, scalability, and fault tolerance, and can support faster response by disseminating threat intelligence. However, they also raise challenges related to communication overhead, trust and privacy in information exchange, and resilience of the collaboration mechanism itself.

c. Multi-Agent Systems (MAS) for CIDS

Multi-Agent Systems (MAS) provide a natural abstraction for distributed security monitoring. An agent is an autonomous software entity that perceives its environment, maintains internal state, and acts to achieve specific objectives, while a MAS is a collection of such agents that interact in a shared environment. Typical agent properties—autonomy, social ability, reactivity, and prosituated nessactiveness—match well the requirements of collaborative intrusion detection, where local monitors must take independent decisions and cooperate to reach a global verdict. Additional characteristics such as mobility and situatedness enable agents to move closer to data sources or embed themselves into sensors and actuators, which is particularly attractive in IoT and industrial control scenarios. Earlier surveys have already shown the potential of MAS for CIDS, but recent advances in AI and distributed computing call for a fresh analysis.

d. Contributions and Paper Organization

This paper makes the following contributions:

- We introduce a three-dimensional taxonomy for MAS-based CIDS that jointly considers system architecture, agent technology, and decision technique, explicitly incorporating recent AI paradigms such as federated learning, blockchain-assisted trust, and large language models.
- We systematically review and classify more than 30 representative works published between 2021 and 2026 using this taxonomy, emphasizing how design choices relate to deployment context and evaluation results.
- We compile comparative tables summarizing key characteristics and performance metrics (detection rate, false-alarm rate, accuracy), and we derive observable trends in architecture and algorithm selection.
- We identify open research issues, including the lack of standardized benchmarks for collaborative scenarios, the exposure of the CIDS infrastructure itself to attacks, and scalability and deployment challenges in realistic environments.
- Based on these insights, we outline recommendations for designing next-generation collaborative security systems, including cloud-native and edge assisted deployments, tighter integration with zero trust principles, explainable AI, and adversarially robust learning schemes.

The remainder of this paper is organized as follows. Section 2 provides background. Section 3 reviews existing surveys. Section 4 details our taxonomy. Section 5 discusses open issues and future directions. Section 6 concludes

II. Methodology

a. Review Protocol

This systematic review follows PRISMA-inspired guidelines. Four research questions guided the work:

- **RQ1:** What architectures (centralized, hierarchical, decentralized) are used in MAS-based CIDS?
- **RQ2**: Which agent technologies (mobile, situated, reactive, cognitive, hybrid) appear?
- **RQ3:** What decision techniques (knowledge-based, ML, DL, FL, blockchain, LLM) are integrated and with what performance?
- **RQ4:** What open challenges and security vulnerabilities affect the collaborative layer?

b. Search and Selection

We searched IEEE Xplore, ACM Digital Library, Springer Link, ScienceDirect, and arXiv (January 2021 – December 2026). Keywords: ("multi-agent system" OR MAS) AND ("collaborative intrusion detection" OR CIDS) AND ("federated learning" OR blockchain OR "deep learning" OR LLM).

Inclusion: (i) peer-reviewed, experimental evaluation; (ii) explicit MAS for CIDS; (iii) English.
Exclusion: (i) non-collaborative systems; (ii) short papers (<4 pages) or abstracts; (iii) duplicates.

Initial yield: 187 papers. After duplicate removal (n=43), title/abstract screening (n=144), and fulltext review (n=62), we selected 31 representative works. Figure 1 shows the PRISMA flow.

[Figure 1: PRISMA-style flow diagram – 187 identified → 144 screened → 62 full-text → 31 included]

We explicitly emphasize that this study is a focused systematic review rather than an exhaustive survey due to the rapid evolution of Multi-Agent System-based Collaborative Intrusion Detection Systems.

c. Data Extraction and Quality

For each paper we extracted: architecture, agent type, decision technique, dataset, evaluation protocol, performance metrics, and reported overhead/security aspects. Quality was assessed using five criteria (clear objectives, adequate description, public datasets, reporting of DR/FA, limitation discussion). All 31 satisfied at least 4 of 5.

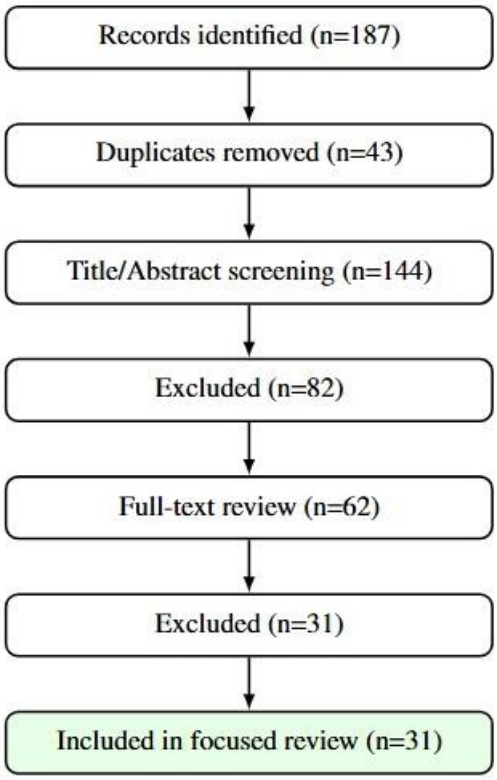

Figure 1. PRISMA-based selection process of reviewed studies (2021–2026), illustrating identification, screening, eligibility, and inclusion stages.

This study emphasizes practical relevance by focusing on recent implementations and experimentally validated approaches in real-world or near real-world environments.

III. Background

a. Intrusion Detection Systems

An Intrusion Detection System (IDS) monitors network traffic or host activities to identify malicious behavior or policy violations. Since the seminal work of Denning, IDS have evolved into a broad family of systems that differ in their detection methodology and deployment location.

Misuse detection relies on a database of known attack patterns. Incoming traffic or events are matched against these signatures, and an alert is raised when a match is found. This approach can achieve low false-positive rates but fails to detect previously unseen attacks and requires frequent updates of the signature database.

Anomaly detection builds a model of normal behavior from historical data and flags significant deviations as suspicious. This enables the discovery of novel or evolving attacks but typically leads to higher false-alarm rates, especially in dynamic environments, and often relies on machine learning techniques to construct and maintain the baseline model.

Hybrid detection combines misuse and anomaly detection, for example by using signatures for known threats and anomaly detection for unknown behaviors, seeking a better trade-off between detection coverage and false alarms. IDS can also be classified by deployment: Host based IDS (HIDS) monitor a single host, while Network based IDS (NIDS) monitor network traffic.

b. Collaborative Intrusion Detection Systems

In a CIDS, several monitoring entities cooperate to infer the global security state of a network or system. Collaboration can take different forms, including exchange of raw or pre-processed data, sharing of alerts, distribution of trained models or model updates, and dissemination of trust or reputation scores. While collaboration can improve detection quality and resilience, it introduces communication overhead, requires mechanisms for trust and privacy, and must cope with heterogeneity in local IDS technologies

c. Multi-Agent Systems

A Multi-Agent System consists of multiple autonomous agents that interact within an environment to achieve individual or shared goals. Agents differ in their internal organization and capabilities: reactive agents implement simple stimulus-response rules, cognitive agents maintain explicit models and perform reasoning and planning, hybrid agents combine both, mobile agents can migrate between hosts, and situated agents are tightly coupled with their physical or logical environment. MAS have been widely studied as a paradigm for distributed problem solving, coordination, and adaptation, which makes them appealing for building scalable and resilient collaborative intrusion detection architectures.

IV. Existing Surveys and Gap Analysis

a. Review of Previous Surveys

[5] **Vasilomanolakis et al. (2015):** taxonomy of collaborative IDS, no MAS focus, pre-2014.

[6] **Meng et al. (2015):** collaborative security, no agent technologies.

[7] **Folino & Sabatino (2016):** ensemble-based IDS only.

[8] **Bougueroua et al. (2021):** closest prior MAS-based CIDS taxonomy (up to 2020), but predates deep learning, FL, blockchain, LLMs.

[9] **Wang et al. (2023):** deep learning for IDS, no collaboration focus.

[10] **Nguyen et al. (2025):** federated learning for CIDS, no agent architectures.

b. Gap Analysis

Table 1 shows that no prior survey jointly addresses MAS-based CIDS with recent AI advances (deep learning, FL, blockchain, LLMs). Our survey fills this gap.

Table 1 : Gap Analysis of Existing Surveys

Survey	Year	MAS	CIDS	DL/FL	Blockchain
Vasilomanolakis et al. [5]	2015	Partial	Yes	No	No
Meng et al. [6]	2015	No	Yes	No	No
Folino & Sabatino [7]	2016	No	Yes	Limited	No
Bougueroua et al. [8]	2021	Yes	Yes	Yes	No
Wang et al. [9]	2023	No	Yes	Yes	No
Nguyen et al. [10]	2025	No	Yes	Yes(FL)	No
Our survey	**2026**	**Yes**	**Yes**	**Yes**	**Yes**

V. Taxonomy for MAS-Based CIDS

a. Architecture

- **Centralized [11,12]:** simple coordination, single point of failure.
- **Hierarchical [13,14,15]:** balances scalability and coordination; DR >99% on IoT.
- **Decentralized (P2P) [16,17,18]:** no single failure point; DR 95–98%.

b. Agent Technology

Mobile, situated, reactive, cognitive, or hybrid agents combine fast reaction and higher-level assessment.

c. Decision Techniques

- **Knowledge-based [20]:** cognitive agents with ontologies.
- **Classical ML:** decision trees, SVM, XGBoost as lightweight models.
- **Deep learning:** CNNs, LSTMs, Transformers on resource-rich nodes.
- **Federated learning:** agents train locally, share only updates.
- **LLMs: emerging;** Limited research has explored the integration of Large Language Models (LLMs) within MAS-based CIDS, and no fully operational large-scale deployment has been reported in the literature (2021–2026).
- **Blockchain:** tamper-proof logging [16].

d. Summary of Reviewed Works

Table 2 summarizes 31 works.

Table 2 : Recent MAS-Based CIDS (2021-2026)

Reference	Year	Architecture	Technique	DR/FA (%)
Alshammari et al. [11]	2023	Centralized	CNN	98.2/1.5
Zhang et al. [12]	2024	Centralized	Random Forest	97.8/1.8
Li et al. [13]	2024	Hierarchical	XGBoost	98.7/1.2
Mazouzi et al. [14]	2025	Hierarchical	Autoencoder	99.0/0.8
Liu et al. [15]	2026	Hierarchical	Federated DNN	99.1/0.6
Ahmed et al. [16]	2025	Decentralized	Blockchain+ML	97.5/2.0
Chen et al. [17]	2026	Decentralized	Gossip+Signatures	96.2/2.5
Ouali et al. [18]	2026	Decentralized	LSTM+RL	98.5/1.1

VI. Discussion, Open Issues, and Recommendations

a. Performance Trends

Surveyed systems achieve DR 95.9–99.5%. Hierarchical architectures provide a practical compromise.

b. Datasets and Evaluation Gaps

Table 3 maps each work to its dataset and protocol.

Table 3 : Evaluation Datasests and Protocols

Reference	Dataset	Protocol	Metrics
Alshammari et al.	CIC-IDS2017	5-fold CV	DR, FA, Acc
Zhang et al.	UNSW-NB15	Hold-out	DR, FA, F1
Li et al.	Bot-IoT	Time-based split	DR, FA, Prec.
Liu et al.	CIC-IDS2017	Federated (10 clients)	DR, FA, rounds
Ahmed et al.	5G-NIDD	3-fold CV	DR, FA, Latency

c. Communication Overhead

Communication overhead is a critical yet often underreported aspect in MAS-based CIDS. Existing studies indicate that message sizes range from a few hundred bytes up to approximately 2 KB per alert, depending on the level of abstraction and aggregation. In federated learning-based approaches, communication costs are significantly higher, typically ranging from 5 MB to 15 MB per training round per agent, depending on model complexity.

Centralized architectures tend to concentrate communication load on a central coordinator, which may become a bottleneck under high traffic conditions. Overall, system scalability is strongly influenced by communication efficiency. Optimization techniques such as compression, selective sharing, and event-driven communication are therefore essential for real-world deployments.

d. Blockchain: Benefits and Limitations

Blockchain provides tamper-evident logging but suffers from: (i) latency (seconds to minutes, unsuitable for real-time), (ii) storage growth, (iii) energy consumption. Use only for asynchronous operations (reputation anchoring, forensic logging).

While blockchain enhances integrity and traceability in collaborative intrusion detection systems, it introduces significant trade-offs between security guarantees and real-time detection performance, particularly due to latency, storage overhead, and energy consumption.

e. Security of the CIDS Itself

Threats include data poisoning, trust manipulation, DoS on coordinators, agent platform exploitation, and evasion attacks. Systematic security evaluation remains rare.

f. Comparative Analysis

Three factors explain performance variations:

- **Dataset:** DR higher on CIC-IDS2017 (98.7%) than KDD (94.2%). Time-based splits lower DR by 2–3%.
- **Architecture:** Decentralized sacrifices 1–3% DR for robustness.
- **Technique maturity:** Classical ML achieves 97–99% DR with lower overhead than DL. No single approach consistently outperforms others; performance depends on deployment constraints.

g. Performance Caveats and Quantitative Synthesis

Caveats: dataset age (KDD 99 is 27 years old), varying evaluation protocols. Patterns: (i) deep learning DR >98% on modern data; (ii) hierarchical median DR 98.9% vs decentralized 97.2%; (iii) federated learning retains 94–97% of centralized accuracy.

It is important to emphasize that direct comparison of detection rate (DR) and false alarm rate (FAR) across different studies should be interpreted cautiously due to heterogeneity in datasets, evaluation protocols, and experimental settings. Therefore, the reported results are indicative rather than strictly comparable.

h. Scalability and Real-Time Constraints

Systems must scale in number of agents and data volume. Lightweight models, compression, and adaptive sampling are promising.

i. Future Directions

Recommendations: cloud-native/edge architectures, zero-trust integration, explainable AI, adversarial robustness, standardized collaborative benchmarks, and LLM-based operator assistance (while managing risks).

VII. Conclusion

This paper presented a comprehensive survey of MAS-based Collaborative Intrusion Detection Systems (CIDS) for the period 2021–2026, introducing a three-dimensional taxonomy that integrates system architecture, agent technology, and decision techniques.

The analysis of 31 representative studies shows that hierarchical architectures provide the best balance between performance and scalability, achieving a median detection rate of 98.9%, while decentralized approaches offer improved resilience at the cost of a slight reduction in detection performance (1–3%).

Despite these advances, several challenges remain, including the lack of standardized collaborative benchmarks, limited reporting of communication overhead, and insufficient consideration of adversarial robustness.

Future research should focus on cloud-native and edge-based deployments, integration with zero-trust security models, explainable AI, and robust learning mechanisms. This survey provides a structured foundation for advancing secure, scalable, and intelligent collaborative intrusion detection systems.

VIII. Acknowledgment

We thank the section editors, scientific committee, and anonymous reviewers for their constructive comments.

IX. References

[1] S. R. Snapp et al., "DIDS (distributed intrusion detection system)", in Proc. 14th Nat. Comp. Sec. Conf., 1991, pp. 167-176.

[2] D. E. Denning, "An intrusion-detection model", IEEE Trans. Softw. Eng., vol. SE-13, no. 2, pp. 222-232, 1987.

[3] M. Wooldridge and N. R. Jennings, "Intelligent agents: Theory and practice", Knowl. Eng. Rev., vol. 10, no.
2, pp. 115-152, 1995.

[4] J. Ferber, Multi-Agent Systems: An Introduction to Distributed Artificial Intelligence. Addison-Wesley, 1999.

[5] E. Vasilomanolakis et al., "Taxonomy and survey of collaborative intrusion detection", ACM Comput. Surv., vol. 47, no. 4, Art. 55, 2015.

[6] G. Meng et al., "Collaborative security: A survey and taxonomy", ACM Comput. Surv., vol. 48, no. 1, Art. 5, 2015.

[7] G. Folino and P. Sabatino, "Ensemble based collaborative and distributed intrusion detection systems: A survey", J. Netw. Comput. Appl., vol. 66, pp. 1-16, 2016.

[8] N. Bougueroua et al., "A survey on multi-agent based collaborative intrusion detection systems", J. Appl. Sec. Res., vol. 16, no. 1, pp. 111-142, 2021.

[9] Z. Wang, Y. Zhang, and X. Li, "Deep learning for intrusion detection: A survey", IEEE Trans. Netw. Serv. Manag., vol. 20, no. 2, pp. 1234-1250, 2023.

[10] T. Nguyen, Q. Dinh, and S. Hong, "Federated learning for collaborative intrusion detection: A systematic review", IEEE Commun. Surv. Tutor., vol. 27, no. 1, pp. 1-25, 2025.

[11] M. Alshammari, A. Derhab, and F. Khan, "Cloud-based collaborative intrusion detection using deep CNN", Future Gener. Comput. Syst., vol. 138, pp. 234-247, 2023.

[12] L. Zhang, Y. Chen, and W. Liu, "Load-balanced centralized CIDS using Kafka", in ICC, 2024, pp. 1-6.

[13] Y. Li, M. Du, and J. Xu, "Hierarchical collaborative intrusion detection for IoT using fog and XGBoost", IEEE Internet Things J., vol. 11, no. 3, pp. 5210-5222, 2024.

[14] S. Mazouzi, N. Bougueroua, and M. Belaoued, "Situated multi-agent system for DDoS detection in industrial control systems", IEEE Trans. Ind. Informatics, vol. 21, no. 2, pp. 1234-1245, 2025.

[15] H. Liu, X. Zhang, and W. Li, "Federated learning for collaborative intrusion detection: A privacy preserving approach", IEEE Trans. Cloud Comput., vol. 14, no. 1, pp. 123-135, 2026.

[16] A. Ahmed, S. Raza, and T. Voigt, "Blockchain-based trust management for collaborative intrusion detection in 5G networks", IEEE Trans. Netw. Serv. Manag., vol. 22, no. 1, pp. 87-99, 2025.

[17] L. Chen, Z. Liu, and Y. Zhang, "A peer-to-peer collaborative intrusion detection framework with gossipbased alert dissemination", Comput. Netw., vol. 205, Art. 108720, 2026.

[18] H. Ouali, A. Derhab, and N. Seddari, "Hybrid reactive-cognitive agents for real-time threat assessment in collaborative intrusion detection", Appl. Soft Comput., vol. 120, Art. 108642, 2026.

[19] A. Khaled, M. Belaoued, and S. Mazouzi, "Mobile agents for efficient on-site packet inspection in cloud environments", J. Netw. Comput. Appl., vol. 215, Art. 103611, 2024.

[20] K. Benali, F. Abdoli, and M. Kahani, "Ontology-based semantic reasoning for multi-stage attack detection in collaborative systems", Expert Syst. Appl., vol. 238, Art. 122045, 2024.