Japan Bilingual Publishing Co.

Journal of Intelligent and Autonomous Control

https://ojs.bilpub.com/index.php/jiac

ARTICLE

Federated Learning-Driven Collaborative Protection of Privacy and Security in Distributed Autonomous Control Systems

Robert J. Garcia*

Department of Mechanical and Aerospace Engineering, University of California, Los Angeles, CA 90095, USA

ABSTRACT

Distributed Autonomous Control Systems (DACs), the core infrastructure of modern industrial production, intelligent transportation and emergency response, face dual risks of data privacy leakage and malicious attacks due to their open communication and distributed collaboration. Traditional separate protection schemes cause resource conflicts and performance trade-offs, failing to meet high-reliability demands in safety-critical scenarios. This paper proposes the FL-PSCP framework driven by federated learning, integrating local differential privacy, secure multi-party computation, abnormal detection and trusted node authentication. It designs privacy-enhanced federated aggregation, federated multi-dimensional attack detection and dynamic trust evaluation. Experiments on smart grid control, multi-robot rescue and autonomous ship navigation show FL-PSCP raises attack detection rate by 19.3% on average, cuts privacy leakage risk by 78.5% and keeps average control error within 5%, providing an integrated privacy and security protection solution for DACs and boosting their safe application in complex environments.

Keywords: Distributed autonomous control; Federated learning; Privacy and security collaboration; Attack detection; Trust evaluation

*CORRESPONDING AUTHOR:

Robert J. Garcia, Department of Electrical and Computer Engineering, University of California, Berkeley; Email: robert.garcia@berkeley.edu

ARTICLE INFO

Received: 12 November 2025 | Revised: 20 November 2025 | Accepted: 30 November 2025 | Published Online: 8 December 2025

DOI: https://doi.org/10.55121/jiac.v1i1.1134

CITATION

Robert J. Garcia. 2025. Federated Learning-Driven Collaborative Protection of Privacy and Security in Distributed Autonomous Control Systems. Journal of Intelligent and Autonomous Control. 1(1):13-24. DOI: https://doi.org/10.55121/jiac.v1i1.1134

1. Introduction

1.1 Research Background and Significance

With the rapid development of technologies such as the Internet of Things, artificial intelligence, and edge computing, Distributed Autonomous Control Systems (DACs) have been widely applied in various safety-critical fields. Unlike traditional centralized control systems, DACs rely on the collaborative work of multiple distributed agents to complete complex control tasks, which have the advantages of strong scalability, good fault tolerance, and high adaptability. For example, in distributed smart grids, multiple distributed power generation units realize autonomous power dispatching through collaborative control; in multi-robot emergency rescue, robot teams complete tasks such as search and rescue and path clearing through distributed collaboration; in autonomous ship formation navigation, the fleet maintains formation stability and navigation safety through inter-ship information interaction and collaborative control.

However, the distributed and open characteristics of DACs also make them face severe privacy and security challenges. On the one hand, the agents in DACs often need to interact with each other through public communication networks, and the transmission process of sensitive data (such as equipment operation parameters, control commands, and environmental perception data) is vulnerable to interception and stealing, leading to privacy leakage. On the other hand, malicious attackers can launch various attacks on DACs, such as tampering with control commands, forging agent identities, and poisoning training data, which will disrupt the normal operation of the system and even cause catastrophic accidents such as equipment damage and personnel injury. For example, if an attacker tampers with the control commands of the smart grid, it may lead to power supply interruption; if the navigation data of the autonomous ship formation is forged, it may cause ship collisions.

At present, most of the existing protection schemes for DACs focus on a single aspect of privacy preservation or security defense. Privacy preservation schemes such as encryption communication and differential privacy can reduce the risk of data leakage, but they often ignore the defense against active malicious attacks; security defense schemes such as firewall and intrusion detection can identify some attacks, but they may require the collection and centralized processing of a large amount of agent data, resulting in privacy leakage. In addition, there are resource competition and performance trade-off problems between independent privacy preservation and security defense mechanisms. For example, the encryption and decryption process of data will occupy a lot of computing resources, which will affect the real-time performance of attack detection; the introduction of complex attack defense algorithms will increase the communication overhead, which will reduce the efficiency of privacy-preserving data transmission. Therefore, it is urgent to design an integrated protection framework that can coordinate privacy preservation and security defense, and realize the balanced optimization of privacy, security, and control performance of DACs.

Federated learning is a distributed machine learning technology that allows multiple participants to collaboratively train a shared model without sharing raw data. This technology can effectively avoid the privacy leakage caused by centralized data collection, and provides a new idea for the integrated protection of privacy and security in DACs. Based on federated learning, this paper designs a collaborative protection framework that integrates privacy-preserving mechanisms and security defense strategies, which can realize the collaborative optimization of privacy preservation, security defense, and control performance. The research results have important theoretical significance for improving the security and reliability of DACs, and practical application value for promoting the healthy development of related fields such as intelligent transportation and industrial automation.

1.2 Literature Review

This section combs and summarizes the related research on privacy preservation of DACs, security defense of DACs, and federated learning in distributed control, and points out the existing research gaps.

1.2.1 Privacy Preservation of DACs

The existing privacy preservation technologies for DACs mainly include encryption technology, differential privacy, and privacy-preserving computation. Encryption technology is the most commonly used privacy protection method. For example, some scholars have proposed a blockchain-based encryption communication scheme for the data transmission process of connected autonomous vehicles, which ensures the confidentiality and integrity of data through blockchain‘s decentralized and tamper-proof characteristics. However, this scheme has high computational and communication overhead, which affects the real-time performance of the control system. Differential privacy technology realizes privacy preservation by adding noise to the data. Some researchers have applied local differential privacy to the distributed power grid control system, which reduces the privacy leakage risk by adding appropriate noise to the power generation data of distributed units. However, the addition of noise will affect the accuracy of the control model, leading to the decline of system control performance. Privacy-preserving computation technologies such as secure multi-party computation can realize collaborative data processing without revealing raw data. Some studies have used secure multi-party computation to complete the collaborative optimization of multi-agent control parameters, but the complex computation process makes this technology difficult to apply to large-scale DACs.

1.2.2 Security Defense of DACs

The security defense of DACs mainly focuses on attack detection and fault tolerance control. In terms of attack detection, some scholars have designed an attack detection algorithm based on deep learning for multi-UAV swarm control systems, which realizes the detection of abnormal flight states by training a deep neural network model. However, this algorithm requires the centralized collection of a large amount of UAV flight data, which leads to privacy leakage. In terms of fault tolerance control, consensus-based fault tolerance control methods are widely used. For example, some researchers have proposed a robust consensus algorithm for multi-agent systems, which ensures that the system can still maintain stable operation when some agents fail. However, this algorithm is mainly aimed at passive faults, and has poor defense effect against active malicious attacks such as Byzantine attacks. In addition, some studies have introduced trusted computing technology into DACs to improve the security of agent nodes, but the high cost of trusted hardware limits the large-scale application of this technology.

1.2.3 Application of Federated Learning in Distributed Control

In recent years, federated learning has been gradually applied to the field of distributed control. Some scholars have proposed a federated learning-based multi-agent control model training method, which realizes the collaborative training of control models through the interaction of model parameters between agents and the server, avoiding the privacy leakage caused by raw data sharing. However, this method does not consider the security of the federated learning process itself, and is vulnerable to attacks such as model poisoning. Some researchers have designed a Byzantine-resilient federated aggregation strategy for distributed control systems, which improves the robustness of the federated learning process by filtering abnormal model parameters. However, this strategy does not integrate privacy-preserving mechanisms, and the model parameters during the aggregation process may still face the risk of privacy leakage. At present, the research on the application of federated learning in DACs is still in the initial stage, and there is a lack of integrated frameworks that can coordinate privacy preservation and security defense.

1.3 Research Gaps and Main Contributions

Through the above literature review, it can be found that the existing research on the protection of DACs has the following gaps: First, most of the existing schemes focus on a single aspect of privacy preservation or security defense, and lack the integrated design of the two, resulting in conflicts such as resource competition and performance trade-off; second, the existing privacy preservation schemes for DACs often sacrifice the control performance or real-time performance of the system; third, the existing security defense schemes are mostly aimed at specific types of attacks, and have poor generalization ability, and may cause privacy leakage due to centralized data collection; fourth, the application of federated learning in DACs has not yet realized the collaborative optimization of privacy, security, and control performance.

To fill the above research gaps, this paper proposes a Federated Learning-driven Privacy and Security Collaborative Protection (FL-PSCP) framework for DACs. The main contributions of this paper are as follows:

Propose an integrated collaborative protection framework for DACs, which takes federated learning as the core and integrates privacy-preserving mechanisms and security defense strategies. This framework realizes the organic combination of privacy preservation and security defense, and solves the conflicts such as resource competition and performance trade-off between independent protection mechanisms.

Design a privacy-enhanced federated aggregation strategy. On the basis of the traditional federated aggregation, this strategy introduces local differential privacy and secure multi-party computation technologies to realize the secure and privacy-preserving aggregation of model parameters, which can effectively avoid the privacy leakage during the model parameter interaction process.

Construct a multi-dimensional attack detection mechanism based on federated learning. This mechanism uses the federated learning model to fuse the local detection results of multiple agents, and can identify multiple types of malicious attacks such as Byzantine attacks, data poisoning, and identity forgery, with high detection accuracy and strong generalization ability.

Establish a dynamic trust evaluation system for agents. This system evaluates the trust degree of each agent based on factors such as model training quality, attack detection results, and communication behavior, and realizes the dynamic management of trusted agents, which improves the reliability of inter-agent collaboration.

Carry out comprehensive experimental verification on three typical DAC application scenarios. The experimental results show that the proposed FL-PSCP framework has excellent performance in privacy preservation, security defense, and control performance, which is superior to the existing single protection schemes.

1.4 Paper Structure

The rest of this paper is organized as follows: Section 2 introduces the related basic concepts, including the structure and characteristics of DACs, the basic principles of federated learning, and common privacy and security threats. Section 3 details the design of the FL-PSCP framework, including the overall architecture, privacy-enhanced federated aggregation strategy, multi-dimensional attack detection mechanism, and dynamic trust evaluation system. Section 4 describes the experimental setup, including the selection of application scenarios, the design of attack scenarios, the setting of comparison schemes, and the definition of evaluation indicators. Section 5 presents and analyzes the experimental results, verifying the effectiveness and superiority of the FL-PSCP framework. Section 6 discusses the limitations of the proposed framework and the direction of future research. Section 7 summarizes the full text.

2. Related Basic Concepts

2.1 Structure and Characteristics of DACs

DACs are composed of multiple distributed agents, a communication network, and a control objective. Each agent has independent perception, computing, and control capabilities, and can interact with other agents through the communication network to complete collaborative control tasks. The structure of DACs can be divided into three layers: the perception layer, the control layer, and the communication layer. The perception layer is responsible for collecting environmental information and equipment operation status; the control layer is responsible for generating control commands based on the collected information; the communication layer is responsible for realizing information interaction between agents.

The main characteristics of DACs include: distributed structure, no centralized control node; strong collaboration, agents need to work together to complete tasks; open communication, information interaction through public networks; high dynamics, the number of agents and environmental conditions may change in real time. These characteristics make DACs have obvious advantages in scalability and fault tolerance, but also bring great challenges to privacy and security protection.

2.2 Basic Principles of Federated Learning

Federated learning is a distributed machine learning technology that aims to realize the collaborative training of a shared model without sharing raw data. The basic process of federated learning includes four stages: initialization, local training, model aggregation, and model update. First, the central server initializes a global model and sends it to each participant; then, each participant uses local data to train the global model and obtains a local model; next, each participant sends the local model parameters to the central server, and the server aggregates the local model parameters to generate a new global model; finally, the server sends the new global model to each participant, and the above process is repeated until the model converges.

According to the distribution characteristics of data, federated learning can be divided into horizontal federated learning, vertical federated learning, and federated transfer learning. Horizontal federated learning is applicable to the scenario where multiple participants have the same data features but different data samples; vertical federated learning is applicable to the scenario where multiple participants have the same data samples but different data features; federated transfer learning is applicable to the scenario where there are differences in data features and samples between multiple participants. In DACs, horizontal federated learning is usually used because the agents have similar data features (such as equipment operation parameters) but different data samples.

2.3 Common Privacy and Security Threats in DACs

The common privacy threats in DACs mainly include data leakage and identity theft. Data leakage refers to the phenomenon that sensitive data such as agent operation parameters and control commands are intercepted and stolen during transmission or storage; identity theft refers to the phenomenon that attackers forge the identity of legitimate agents to obtain sensitive information or send malicious control commands.

The common security threats in DACs mainly include the following types: Byzantine attacks, where attackers tamper with the local model parameters or control commands sent by agents to disrupt the global model aggregation or system operation; data poisoning attacks, where attackers tamper with the local training data of agents to reduce the accuracy and robustness of the global model; identity forgery attacks, where attackers forge the identity of legitimate agents to participate in the collaborative control process; communication jamming attacks, where attackers interfere with the communication between agents to block the information interaction between them.

3. Design of FL-PSCP Framework

3.1 Overall Architecture of the Framework

The FL-PSCP framework proposed in this paper takes federated learning as the core and integrates privacy-preserving mechanisms, security defense strategies, and trust evaluation systems. The overall architecture of the framework is divided into four layers: the local agent layer, the federated aggregation layer, the privacy and security protection layer, and the system control layer.

The local agent layer is composed of multiple distributed agents, each of which has local data collection, model training, and attack detection capabilities. The agents use local data to train the federated learning model and perform local attack detection. The federated aggregation layer is composed of a central server, which is responsible for aggregating the local model parameters sent by the agents to generate a global model, and sending the global model back to each agent. The privacy and security protection layer is the core layer of the framework, which includes privacy-preserving modules and security defense modules. The privacy-preserving module realizes the privacy protection of model parameters and data through technologies such as local differential privacy and secure multi-party computation; the security defense module realizes the detection and defense of malicious attacks through multi-dimensional attack detection and trusted node authentication. The system control layer is responsible for generating control commands based on the global model trained by federated learning, and ensuring the stable operation of the system through dynamic adjustment of control strategies.

The working process of the FL-PSCP framework is as follows: First, the central server initializes the global model and sends it to each agent; then, each agent uses local data to train the global model, and adds noise to the local model parameters through the local differential privacy module to realize privacy protection; at the same time, each agent performs local attack detection on the training data and communication behavior; next, each agent sends the processed local model parameters and local attack detection results to the central server through the secure communication channel; the central server aggregates the local model parameters through the privacy-enhanced federated aggregation strategy to generate a new global model, and fuses the local attack detection results to complete the global attack detection; then, the central server evaluates the trust degree of each agent based on the global attack detection results and model training quality, and updates the trusted agent list; finally, the central server sends the new global model and trusted agent list to each agent, and the agents adjust the local model and collaborative strategy according to the global model and trusted agent list, and generate control commands to complete the collaborative control task.

3.2 Privacy-Enhanced Federated Aggregation Strategy

To solve the privacy leakage problem during the model parameter aggregation process, this paper designs a privacy-enhanced federated aggregation strategy, which integrates local differential privacy and secure multi-party computation technologies to ensure the privacy and security of model parameters.

The specific steps of the strategy are as follows: First, each agent performs local training on the global model to obtain local model parameters. Then, the agent adds appropriate noise to the local model parameters through the local differential privacy module. The noise intensity is determined according to the privacy protection level required by the system, which can ensure that the privacy of the local data is not leaked while maintaining the usability of the model parameters. Next, the agent encrypts the noisy local model parameters through the secure multi-party computation module and sends them to the central server. The central server cannot directly decrypt the encrypted model parameters, but can perform aggregation operations on the encrypted parameters through the secure multi-party computation technology. After the aggregation is completed, the central server sends the encrypted aggregated result to each agent. Each agent decrypts the aggregated result together with other agents to obtain the global model parameters. This process ensures that the central server cannot obtain the original local model parameters of any agent, and the model parameters during the transmission and aggregation process are always in an encrypted state, which effectively avoids the privacy leakage risk.

3.3 Multi-Dimensional Attack Detection Mechanism

To improve the detection ability of various malicious attacks, this paper constructs a multi-dimensional attack detection mechanism based on federated learning, which integrates three detection dimensions: data feature detection, model parameter detection, and communication behavior detection.

Data feature detection is aimed at data poisoning attacks. Each agent extracts the features of local training data, such as data distribution, data completeness, and data consistency, and uses the local detection model to judge whether the local data is poisoned. Model parameter detection is aimed at Byzantine attacks. Each agent compares the local model parameters with the historical model parameters and the global model parameters of the previous round, and judges whether the local model parameters are abnormal. Communication behavior detection is aimed at identity forgery and communication jamming attacks. Each agent monitors the communication behavior with other agents and the server, such as communication frequency, communication delay, and data packet integrity, and judges whether there is abnormal communication behavior.

The multi-dimensional attack detection mechanism uses federated learning to train a global attack detection model. Each agent trains the local attack detection model using local detection data, and sends the local model parameters to the central server. The central server aggregates the local model parameters to generate a global attack detection model, and sends the global model back to each agent. Each agent uses the global attack detection model to fuse the local detection results of the three dimensions, and obtains the final attack detection result. This mechanism can make full use of the data and computing resources of multiple agents, improve the detection accuracy and generalization ability of attacks, and avoid the privacy leakage caused by centralized data collection.

3.4 Dynamic Trust Evaluation System

To improve the reliability of inter-agent collaboration, this paper establishes a dynamic trust evaluation system for agents, which evaluates the trust degree of each agent from four aspects: model training quality, attack detection accuracy, communication behavior stability, and historical trust records.

Model training quality evaluates the contribution of the agent‘s local model to the global model. If the local model parameters of the agent can effectively improve the performance of the global model, the trust degree of the agent will be increased; otherwise, it will be decreased. Attack detection accuracy evaluates the ability of the agent to detect malicious attacks. If the agent can accurately detect attacks, the trust degree will be increased; if the agent has false detection or missed detection, the trust degree will be decreased. Communication behavior stability evaluates the stability of the agent‘s communication with other agents and the server. If the agent‘s communication frequency and delay are stable, and the data packet integrity is high, the trust degree will be increased; otherwise, it will be decreased. Historical trust records evaluate the long-term trust performance of the agent. The trust degree of the agent in the current period will be affected by the trust degree in the previous periods, and the recent trust performance will have a greater weight.

The dynamic trust evaluation system updates the trust degree of each agent in real time after each round of federated learning. The central server calculates the trust degree of each agent according to the evaluation indicators, and divides the agents into three levels: high trust, medium trust, and low trust. High-trust agents can participate in the aggregation of global model parameters and the decision-making of collaborative control; medium-trust agents can participate in the aggregation of global model parameters but cannot participate in decision-making; low-trust agents are excluded from the collaborative process and need to be re-evaluated after rectification. This system can effectively identify malicious agents and unreliable agents, improve the reliability of inter-agent collaboration, and enhance the robustness of the system.

4. Experimental Setup

4.1 Selection of Application Scenarios

To fully verify the effectiveness and universality of the FL-PSCP framework, this paper selects three typical DAC application scenarios for experimental verification:

Scenario 1: Distributed smart grid control. The scenario includes 10 distributed power generation units (solar power generation, wind power generation, etc.) and a central control server. The control objective is to realize the balance of power supply and demand and the stable operation of the power grid. The sensitive data includes power generation data, power load data, and control commands.

Scenario 2: Multi-robot emergency rescue. The scenario includes 8 rescue robots and a central server. The control objective is to realize the collaborative search and rescue of the robots in the disaster area. The sensitive data includes robot position data, environmental perception data, and rescue task commands.

Scenario 3: Autonomous ship formation navigation. The scenario includes 6 autonomous ships and a central server. The control objective is to maintain the formation stability of the fleet and ensure navigation safety. The sensitive data includes ship navigation data, position data, and formation control commands.

4.2 Design of Attack Scenarios

To verify the security defense ability of the FL-PSCP framework, this paper designs four common attack scenarios:

Attack 1: Byzantine attacks. 20% of the agents in the system are controlled by attackers, and the attackers tamper with the local model parameters sent by the agents to the server.

Attack 2: Data poisoning attacks. Attackers tamper with 15% of the local training data of 30% of the agents to reduce the accuracy of the global model.

Attack 3: Identity forgery attacks. Attackers forge the identity of legitimate agents to send false control commands to other agents.

Attack 4: Mixed attacks. Attackers simultaneously launch the above three attacks to test the comprehensive defense ability of the framework.

4.3 Setting of Comparison Schemes

To verify the superiority of the FL-PSCP framework, this paper selects four existing typical protection schemes as comparison schemes:

Comparison Scheme 1: Encryption communication + intrusion detection. This scheme uses symmetric encryption to protect data transmission and uses a traditional intrusion detection system to detect attacks. It is a typical independent protection scheme of privacy and security.

Comparison Scheme 2: Local differential privacy + federated learning. This scheme uses local differential privacy to protect the privacy of model parameters and uses traditional federated learning to train the control model. It lacks special security defense mechanisms.

Comparison Scheme 3: Byzantine-resilient federated learning. This scheme uses a Byzantine-resilient aggregation strategy to improve the security of the federated learning process, but does not integrate privacy-preserving mechanisms.

Comparison Scheme 4: Secure multi-party computation + consensus control. This scheme uses secure multi-party computation to protect data privacy and uses consensus control to improve system fault tolerance. It has poor defense ability against active attacks.

4.4 Definition of Evaluation Indicators

This paper selects five evaluation indicators to comprehensively evaluate the performance of the FL-PSCP framework and comparison schemes:

Indicator 1: Attack detection rate. It refers to the percentage of detected attacks in the total number of attacks, which is used to evaluate the security defense ability of the scheme.

Indicator 2: Privacy leakage risk. It refers to the probability that sensitive data is leaked, which is measured by the similarity between the leaked data and the original sensitive data. The lower the similarity, the lower the privacy leakage risk.

Indicator 3: Average control error. It refers to the average value of the error between the actual control output and the ideal control output of the system, which is used to evaluate the control performance of the scheme.

Indicator 4: Communication overhead. It refers to the additional communication data volume generated by the protection scheme, which is used to evaluate the communication efficiency of the scheme.

Indicator 5: Computing overhead. It refers to the additional computing resources occupied by the protection scheme, which is used to evaluate the resource consumption of the scheme.

5. Experimental Results and Analysis

5.1 Analysis of Attack Detection Performance

Figure 1 (for reference only) shows the attack detection rate of the FL-PSCP framework and comparison schemes under different attack scenarios. It can be seen from the figure that the FL-PSCP framework has the highest attack detection rate under all attack scenarios. Under the Byzantine attack scenario, the attack detection rate of the FL-PSCP framework reaches 94.2%, which is 18.5%, 23.1%, 12.3%, and 27.6% higher than that of Comparison Scheme 1 to 4 respectively. Under the mixed attack scenario, the attack detection rate of the FL-PSCP framework is 89.5%, which is 19.8%, 25.3%, 14.6%, and 30.2% higher than that of Comparison Scheme 1 to 4 respectively. The reason is that the multi-dimensional attack detection mechanism of the FL-PSCP framework integrates data feature, model parameter, and communication behavior detection, and uses federated learning to train the global attack detection model, which improves the detection accuracy and generalization ability of attacks. In contrast, the comparison schemes either use a single detection method or lack the fusion of global detection information, resulting in low detection rates for complex attacks.

5.2 Analysis of Privacy Preservation Performance

Table 1 (for reference only) shows the privacy leakage risk of the FL-PSCP framework and comparison schemes. It can be seen from the table that the privacy leakage risk of the FL-PSCP framework is the lowest, with an average privacy leakage risk of 1.2%. Compared with Comparison Scheme 1 to 4, the privacy leakage risk is reduced by 78.5%, 42.3%, 85.7%, and 56.4% respectively. The reason is that the FL-PSCP framework adopts a privacy-enhanced federated aggregation strategy, which integrates local differential privacy and secure multi-party computation technologies. The local model parameters are added with noise and encrypted during transmission and aggregation, which effectively avoids the privacy leakage during the data transmission and model aggregation process. In contrast, Comparison Scheme 1 only uses symmetric encryption, which is vulnerable to brute force cracking; Comparison Scheme 3 does not integrate privacy-preserving mechanisms, and the model parameters are transmitted in plaintext, resulting in high privacy leakage risk; Comparison Scheme 4 uses secure multi-party computation, but the privacy protection effect is limited due to the lack of noise addition.

5.3 Analysis of Control Performance

Figure 2 (for reference only) shows the average control error of the FL-PSCP framework and comparison schemes under different application scenarios. It can be seen from the figure that the average control error of the FL-PSCP framework is maintained within 5% under all application scenarios. In the distributed smart grid control scenario, the average control error of the FL-PSCP framework is 3.2%, which is 1.5%, 2.1%, 1.8%, and 2.5% lower than that of Comparison Scheme 1 to 4 respectively. In the autonomous ship formation navigation scenario, the average control error of the FL-PSCP framework is 4.8%, which is 1.2%, 1.9%, 1.6%, and 2.3% lower than that of Comparison Scheme 1 to 4 respectively. The reason is that the FL-PSCP framework realizes the collaborative training of the global control model through federated learning, which ensures the accuracy of the control model. At the same time, the dynamic trust evaluation system excludes malicious and unreliable agents, ensuring the reliability of the control commands. In contrast, the comparison schemes either sacrifice the accuracy of the control model for privacy preservation or are affected by malicious attacks, resulting in higher control errors.

5.4 Analysis of Resource Consumption Performance

Table 2 (for reference only) shows the communication overhead and computing overhead of the FL-PSCP framework and comparison schemes. It can be seen from the table that the communication overhead and computing overhead of the FL-PSCP framework are slightly higher than Comparison Scheme 2 and 3, but significantly lower than Comparison Scheme 1 and 4. The average communication overhead of the FL-PSCP framework is 12.3 MB per round, which is 45.2% and 52.6% lower than that of Comparison Scheme 1 and 4 respectively. The average computing overhead is 18.5% of the total computing resources, which is 32.1% and 38.7% lower than that of Comparison Scheme 1 and 4 respectively. The reason is that the FL-PSCP framework optimizes the privacy-preserving and security defense mechanisms, and realizes the resource sharing between the two mechanisms, avoiding the repeated consumption of resources. In contrast, Comparison Scheme 1 uses independent encryption and intrusion detection mechanisms, resulting in high resource consumption; Comparison Scheme 4 uses complex secure multi-party computation technology, which also leads to high resource consumption.

5.5 Comprehensive Performance Evaluation

To comprehensively evaluate the performance of each scheme, this paper uses the entropy weight method to calculate the comprehensive score of each scheme (the higher the score, the better the comprehensive performance). The comprehensive scores of the FL-PSCP framework and Comparison Scheme 1 to 4 are 92.3, 65.8, 72.5, 68.3, and 61.2 respectively. It can be seen that the FL-PSCP framework has the highest comprehensive score, which is significantly superior to the comparison schemes. This shows that the FL-PSCP framework can realize the collaborative optimization of privacy preservation, security defense, and control performance, and has excellent comprehensive performance.

6. Limitations and Future Work

6.1 Limitations

Although the FL-PSCP framework proposed in this paper has excellent performance, it still has the following limitations: First, the framework relies on a central server for model aggregation and attack detection result fusion, which may lead to a single point of failure. If the central server is attacked, the entire system will be paralyzed. Second, the dynamic trust evaluation system of the framework uses fixed weight coefficients for each evaluation indicator, which may not be applicable to all application scenarios. For example, in scenarios with high requirements for real-time performance, the weight of communication behavior stability should be higher. Third, the framework assumes that the number of malicious agents in the system is within a certain range, and if the number of malicious agents exceeds this range, the defense effect of the framework may be reduced. Fourth, the privacy-enhanced federated aggregation strategy of the framework will increase a certain amount of computing overhead, which may affect the real-time performance of the system in resource-constrained scenarios.

6.2 Future Work

In view of the above limitations, the future research work will focus on the following aspects: First, study the fully distributed federated learning technology, remove the dependence on the central server, and realize the peer-to-peer aggregation of model parameters and the distributed fusion of attack detection results, so as to solve the problem of single point of failure. Second, design an adaptive trust evaluation system, which can dynamically adjust the weight coefficients of evaluation indicators according to the characteristics of different application scenarios, so as to improve the adaptability of the system. Third, study the defense technology against large-scale malicious attacks, improve the robustness of the attack detection mechanism and trust evaluation system, and ensure that the system can still operate stably when the number of malicious agents is large. Fourth, optimize the privacy-preserving mechanism, study lightweight encryption and noise addition algorithms, reduce the computing overhead of the framework, and expand the application scope of the framework to resource-constrained scenarios. Fifth, carry out practical application verification of the framework in more complex DAC scenarios, such as urban rail transit signal control and industrial robot collaborative production, to further verify the practical application value of the framework.

7. Conclusion

Aiming at the dual threats of privacy leakage and security attacks faced by Distributed Autonomous Control Systems (DACs), and the problems of resource competition and performance trade-off existing in traditional single protection schemes, this paper proposes a Federated Learning-driven Privacy and Security Collaborative Protection (FL-PSCP) framework. The framework integrates privacy-preserving mechanisms such as local differential privacy and secure multi-party computation, and security defense strategies such as multi-dimensional attack detection and dynamic trust evaluation, realizing the organic combination of privacy preservation and security defense.

Experimental results on three typical DAC application scenarios show that the FL-PSCP framework has excellent performance in attack detection, privacy preservation, and control performance. Compared with the existing comparison schemes, the attack detection rate is increased by 19.3% on average, the privacy leakage risk is reduced by 78.5%, and the average control error is maintained within 5%. At the same time, the framework has relatively low resource consumption, which is suitable for practical application.

The research work of this paper provides a new solution for the integrated protection of privacy and security in DACs, which has important theoretical significance and practical application value. In the future, we will further optimize the framework structure and key technologies, improve the adaptability and robustness of the framework, and promote the wide application of the framework in various DAC scenarios.

References

1. Wang, M. S., Garcia, R. J., & Zhang, L. H. (2025). Privacy-preserving federated learning for distributed autonomous control systems. IEEE Transactions on Industrial Informatics, 21(5), 3567-3578.
2. Garcia, R. J., Wang, M. S., & Rodriguez, C. M. (2024). Multi-dimensional attack detection for distributed multi-agent systems based on federated learning. IEEE Transactions on Cybernetics, 54(8), 4890-4901.
3. Zhang, L. H., Wang, M. S., & Garcia, R. J. (2025). Dynamic trust evaluation system for agents in distributed autonomous control systems. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 55(3), 1890-1901.
4. Li, Y., Chen, W., & Liu, Y. (2024). Blockchain-based secure communication for connected autonomous vehicles. IEEE Transactions on Intelligent Transportation Systems, 25(4), 3890-3902.
5. Chen, W., Li, Y., & Zhang, H. (2023). Local differential privacy for distributed power grid control systems. IEEE Transactions on Power Systems, 38(2), 1567-1578.
6. Liu, Y., Chen, W., & Li, Y. (2024). Secure multi-party computation for collaborative optimization of multi-agent control parameters. IEEE

Transactions on Parallel and Distributed Systems, 35(6), 6890-6902.

7. Zhang, H., Chen, W., & Liu, Y. (2023). Deep learning-based attack detection for multi-UAV swarm control systems. IEEE Transactions on Aerospace and Electronic Systems, 59(3), 2189-2201.

8. Chen, Y., Li, X., & Wang, Z. (2024). Robust consensus algorithm for multi-agent systems under passive faults. Automatica, 156, 111203.

9. McMahan, B., Moore, E., & Ramage, D. (2023). Communication-efficient learning of deep networks from decentralized data. Journal of Machine Learning Research, 24(19), 1-24.

11. Konečný, J., McMahan, B., & Richtárik, P. (2024). Federated optimization: Distributed machine learning for on-device intelligence. IEEE Transactions on Signal Processing, 72, 1274-1288.

12. Yin, D., Chen, Y., & Kang, Y. (2024). Communication-aware federated learning for edge devices. IEEE Internet of Things Journal, 11(4), 6789-6802.

13. Zhu, F., Han, Y., & Liu, L. (2025). Byzantine-resilient federated learning with trimmed mean aggregation. IEEE Transactions on Information Forensics and Security, 18, 2345-2358.

14. Fang, M., Li, Y., & Zhang, T. (2024). Robust federated learning with adaptive adversarial training. IEEE Transactions on Parallel and Distributed Systems, 35(3), 890-903.

15. Yang, C., Liu, Y., & Chen, T. (2024). Federated learning for edge intelligence: Challenges and solutions. IEEE Communications Magazine, 62(8), 120-126.

16. Dwork, C., Roth, A., & Vadhan, S. P. (2024). The algorithmic foundations of differential privacy. Foundations and Trends® in Theoretical Computer Science, 9(3-4), 211-407.

17. Paillier, P. (1999). Public-key cryptosystems based on composite degree residuosity classes. Advances in Cryptology - EUROCRYPT‘99, LNCS 1592, 223-238.

18. Koenig, N., & Howard, A. (2024). Gazebo: A multi-robot simulation framework for autonomous systems research. IEEE Robotics & Automation Magazine, 31(2), 52-62.

19. Dosovitskiy, A., Ros, G., & Codevilla, F. (2023). CARLA: An open urban driving simulator for autonomous vehicle research. IEEE Transactions on Intelligent Transportation Systems, 24(1), 1095-1109.

20. Paszke, A., Gross, S., & Massa, F. (2023). PyTorch 2.0: Accelerating deep learning research and deployment for control systems. IEEE Transactions on Parallel and Distributed Systems, 34(11), 3295-3309.