Computing and Informatics, Vol. 44, 2025, 682–716, doi: 10.31577/cai_2025_3_682

OPTIMIZING SECURITY AND PERFORMANCE IN BLOCKCHAIN-ENHANCED FEDERATED LEARNING THROUGH PARTICIPANT SELECTION WITH ROLE DETERMINATION

Wafa Bouras

Computer Science, University of M'sila
PO Box 166 Ichebilia, M'sila, 28000, Algeria
&
LIM Laboratory, Souk Ahras University
Souk Ahras, Algeria
e-mail: wafa.bouras@univ-msila.dz

Kameleddine Heraguemi

Computer Science, University of M'sila
PO Box 166 Ichebilia, M'sila, 28000, Algeria
&
National School of Artificial Intelligence
Sidi Abd Allah, Algiers, Algeria
e-mail: kameleddine.heraguemi@ensia.edu.dz

Mohamed Benouis

Computer Science, University of M'sila
PO Box 166 Ichebilia, M'sila, 28000, Algeria
&
Human-Centered Artificial Intelligence
University of Augsburg, 286159 Augsburg, Germany
e-mail: mohamed.benouis@uni-a.de

Brahim Bouderah

Computer Science, University of M'sila
PO Box 166 Ichebilia, M'sila, 28000, Algeria
&
University Abdelhamid Ibn Badis-Mostaganem
Mostaganem, Algeria
e-mail: brahim.bouderah@univ-msila.dz

Samir Akrouf

Computer Science and Its Applications Laboratory of M'sila (LIAM)
University of M'sila
PO Box 166 Ichebilia, M'sila, 28000, Algeria
e-mail: samir.akrouf@univ-msila.dz

Abstract. Federated learning (FL) allows distributed devices to jointly train a global model while safeguarding the privacy of their local data. However, selecting and securing clients, especially in environments with potentially malicious participants, remains a critical challenge. This study proposes an innovative participant selection method to enhance both security and efficiency in centralized and decentralized FL frameworks. In the centralized framework, this method effectively excludes clients with weak privacy protections and optimization capabilities, thus increasing overall system security. For decentralized FL, a blockchain-supported approach is introduced, which further strengthens the robustness of the system. Using a dynamic role assignment algorithm, roles such as worker, validator, and miner are allocated based on security and performance metrics for each training round. The findings show that this method performs on a par with the scenarios free of malicious clients, demonstrating the value of blockchain technology in improving FL protocols. By addressing security vulnerabilities and improving training efficiency, this research contributes to the development of more secure and efficient FL systems, underscoring the importance of advanced participant selection and role assignment strategies.

Keywords: Federated learning, participant selection, blockchain, malicious attacks, distributed systems

Mathematics Subject Classification 2010: 68W99

1 INTRODUCTION

In the era of distributed machine learning, managing large volumes of training data across multiple devices poses significant challenges. Traditional approaches rely on centralized servers to handle data, raising concerns about privacy and security. Federated learning (FL) [1] emerged as a solution to these concerns, allowing devices to collaboratively train a global model without exposing their local data [2]. However, several vulnerabilities remain unaddressed in current FL implementations, particularly with participant selection and the security of data transmitted during training rounds.

Existing FL systems depend on participants periodically sharing model updates with a central server [3], yet they lack strong guarantees regarding the integrity and trustworthiness of these updates. Adverse conditions or malicious actors can corrupt the data, leading to compromised models. The literature lacks robust solutions to mitigate these risks and ensure both privacy and performance in FL.

To address this, we introduce a novel method that optimizes participant selection and role assignment in FL, leveraging blockchain technology to enhance security in decentralized frameworks. In centralized FL, our proposed method filters participants based on privacy and optimization capabilities, ensuring that only reliable clients contribute to the model. For decentralized FL, a blockchain-assisted framework ensures secure and dynamic role determination, assigning roles like worker, validator, and miner based on clients' security and performance levels. This approach guarantees robustness even in the presence of potential adversaries, providing performance levels comparable to environments without malicious clients.

Blockchain [4], with its tamper-resistant and traceable nature, forms the backbone of our decentralized FL framework, ensuring secure and transparent interactions between participants. By combining FL and blockchain, our system safeguards model updates during training, preserving both the integrity of the data and the privacy of participants.

The key contributions of this paper include:

- A participant selection algorithm that optimizes privacy and performance in centralized FL;
- A role determination algorithm for decentralized FL, enhancing security and performance using blockchain technology;
- A comprehensive evaluation of the proposed system in both centralized and decentralized settings, demonstrating its resilience against adversarial conditions.

The structure of the paper is as follows: Section 2 reviews the literature on FL and blockchain integration. Section 3 defines the challenges in participant selection and role determination in FL systems. Section 4 introduces the proposed participant selection algorithm for centralized FL, while Section 5 delves into the blockchain-assisted role determination method for decentralized FL. Finally, Section 6 presents

experimental results, and Section 7 concludes the paper with a summary of key findings and potential future research directions.

2 RELATED WORKS

In recent years, extensive research has been conducted to enhance the security, performance, and reliability of Federated Learning (FL) systems. This section reviews significant contributions in client selection methods, security and privacy mechanisms, and the integration of blockchain technology into FL, highlighting gaps that our proposed approach seeks to address.

2.1 Client Selection Methods

Client selection is a critical aspect of FL that influences model performance, convergence speed, and resource efficiency. Various strategies have been developed to optimize client selection, considering factors such as client heterogeneity, energy constraints, and security requirements.

2.1.1 Reputation-Based and Heterogeneity-Aware Methods

The PIRATE framework [5] marks a significant advancement in decentralized federated learning (FL) by leveraging consortium blockchain technology to implement a decentralized reputation system for client selection. This approach enhances the reliability of participating clients, ensuring that only trusted contributors influence the model training process.

In contrast, Oort [6] introduces an optimization mechanism designed for centralized FL, prioritizing participant selection based on processing time and accuracy. By employing the Oort executor, the framework streamlines FL coordination and enhances overall system performance. However, its centralized focus limits its applicability in addressing the challenges posed by client heterogeneity in decentralized settings.

Building upon Oort, the PISCES algorithm [7] introduces an asynchronous enhancement to the participant selection process, enabling straggling clients to contribute to model training. While this modification improves inclusivity, it does not adequately address privacy and security considerations, which are critical in many FL deployments.

The AFL framework [8] takes a probability-based approach to client selection in FL by evaluating the utility of each client's data. This method effectively reduces communication costs and improves the model efficiency while maintaining differential privacy to protect client data.

To address communication costs and data heterogeneity in mobile devices, the Hermes framework [9] provides a novel solution for federated learning. Hermes enables devices to learn personalized, structured sparse deep neural networks through structured pruning, significantly reducing communication overhead and improving

inference efficiency. Additionally, it incorporates a participant selection strategy by grouping devices based on their communication capabilities and data characteristics, optimizing participation for both efficiency and personalization.

The control mechanism proposed by [10] addresses client heterogeneity by dynamically adjusting the frequency of client selection, allowing lower-capacity clients to participate more frequently. Experiments on deep neural network (DNN) tasks using large-scale FEMNIST datasets demonstrate that including such clients enhances participation rates and improves model accuracy, particularly in centralized FL scenarios.

Finally, the approach described by [11] tackles the challenge of training federated deep learning models on mobile devices while preserving data privacy. This method introduces an optimal user selection strategy based on reputation scores, improving the efficiency and stability of federated learning, particularly in scenarios with non-IID and imbalanced data distributions.

2.1.2 Blockchain-Based Participant Selection

Blockchain technology has been increasingly integrated into FL to address trust and transparency challenges in participant selection. The decentralized nature of blockchain allows for secure and transparent client selection without relying on a centralized authority.

One example is the work by [12], which presents a blockchain-based approach to optimize edge node selection in FL settings. This method uses a consensus algorithm to evaluate the trustworthiness of clients based on their past performance and behavior, thereby ensuring that only reliable participants are selected. Using blockchain in this context enhances security and promotes fairness in the selection process.

Another significant contribution is by [13], which introduces a blockchain-based framework for secure and transparent participant selection in healthcare FL environments. Maintaining an immutable record of client contributions ensures that only legitimate participants are involved in the training process, thereby reducing the risk of adversarial attacks.

The Lotto framework [14] also leverages blockchain to implement secure random and informed selection algorithms. This method aligns client selection with honest behaviors while maintaining low computational overhead, making it resilient against malicious clients.

The article [15] examined the development of a federated learning (FL) framework that integrated blockchain technology to improve participant selection, auditability, and privacy. The framework addressed key challenges using blockchain and smart contracts to create a transparent and verifiable selection process. In this process, participants anonymously submitted their training losses, allowing for evaluation without compromising private information. The framework ensured that participants could not falsify their contributions or impersonate others by employing cryptographic methods such as commitment schemes and zero-knowledge proofs.

This approach not only balanced auditability and privacy but also enhanced training efficiency and model accuracy, making it a valuable contribution to the study of trust and privacy in decentralized machine learning.

2.1.3 Energy-Aware Methods

Energy efficiency is crucial, especially in mobile and edge FL environments. The REWAFL approach [16] optimizes participant selection by considering residual energy levels and wireless conditions. This method minimizes energy consumption while maintaining effective training, making it suitable for heterogeneous mobile networks. However, it primarily targets decentralized FL scenarios, with limited applicability to centralized systems.

2.1.4 Security-Enhanced Methods

Security in FL has been a growing concern, particularly in ensuring the integrity of model updates and protecting against adversarial attacks. VerifyNet [17] addresses this by providing a robust verification mechanism for server results, capable of handling user dropouts and operating under honest-but-curious settings. This centralized approach ensures data integrity but does not fully integrate privacy-preserving mechanisms.

FedRank [18], on the other hand, employs imitation learning to optimize client selection. By ranking clients based on their contributions through pairwise training, FedRank accelerates model convergence and enhances selection efficiency, prioritizing high-performing clients while reducing energy consumption.

2.1.5 Adaptive Resource Management

Managing resources efficiently in FL is critical for large-scale deployments. The Selective Aggregation of Models (SAM) approach [19] mitigates communication bottlenecks by allowing clients to selectively upload models, reducing overhead while maintaining accuracy. SAM is particularly effective in centralized FL, where communication costs are a significant constraint.

The FLOAT framework [20] further enhances resource management by dynamically optimizing FL parameters through multi-objective reinforcement learning. By reducing client dropouts and managing stragglers, FLOAT improves the efficiency of FL systems, particularly in environments with heterogeneous resources.

2.2 Blockchain-Enhanced FL Solutions

Blockchain technology is a decentralized and distributed ledger system that allows for secure and transparent record-keeping across a network of nodes. Originally introduced as the underlying technology for cryptocurrencies like Bitcoin, blockchain has since expanded its applications to various fields due to its unique properties.

At its core, a blockchain comprises blocks, each containing a set of transaction data. These blocks are linked together in a chronological chain using cryptographic hashes, which ensure that once data is added, it cannot be altered without impacting the entire chain. This feature guarantees the immutability and security of data, making blockchain highly resistant to tampering and fraud.

One of the key components of a blockchain network is its consensus mechanism. Since there is no central authority, consensus algorithms are used to ensure agreement among the distributed nodes. Common consensus models include Proof of Work (PoW), where participants solve complex mathematical problems to validate transactions, and Proof of Stake (PoS), where validators are chosen based on the number of tokens they hold. Both mechanisms play a critical role in ensuring the security and integrity of the network.

Blockchain's decentralized, transparent, and immutable nature has made it suitable for applications beyond digital currencies. It is widely used in industries such as finance, supply chain management, and healthcare. In particular, blockchain is essential for developing smart contracts and decentralized applications (dApps), which rely on these core features for secure and automated transactions.

In specialized fields like federated learning, blockchain offers additional benefits. For instance, it can provide a verifiable record of actions, such as unlearning requests, ensuring data compliance and privacy requirements are met in distributed learning frameworks. This enhances transparency and trust, two critical factors in managing sensitive data.

2.2.1 Reputation and Privacy Mechanisms

FedCure [21] introduces a personalized FL framework for Internet of Medical Things (IoMT) applications, leveraging blockchain to ensure data privacy while addressing device heterogeneity through edge computing. This approach reduces latency and improves model performance in healthcare scenarios, demonstrating blockchain's potential in privacy-preserving FL.

The PIRATE framework [5] also employs blockchain to create a decentralized, reputation-based client selection system, enhancing trust and transparency in FL processes. These solutions highlight the role of blockchain in promoting secure and reliable collaboration in FL environments.

2.2.2 Mitigating Adversarial Attacks and Dropouts

Lotto [14] addresses the vulnerability of FL to adversarial servers by using secure random and informed selection algorithms. This approach ensures that client selection aligns with honest behaviors, maintaining low computational overhead while safeguarding against manipulation. This method enhances the resilience of FL systems against malicious clients.

2.3 Security and Privacy Mechanisms in FL

Blockchain's decentralized, transparent, and immutable nature makes it well-suited for enhancing security and privacy in FL. By providing a verifiable record of actions, blockchain ensures data integrity and compliance, which is crucial in privacy-preserving FL frameworks. The work [22] introduces SLMFed as a mechanism for Incremental and FL (IFL) to update AI models in dynamic IoT environments, managing stage transitions and client selection by adopting a stage-based and layer-wise approach to periodic learning. The article [23] introduced FLIPS (FL using Intelligent Participant Selection), an innovative solution to improve the efficiency and accuracy of FL systems. Jia et al. [24] introduced a blockchain-enabled scheme for securely aggregating FL data in Industrial Internet of Things (IIoT) environments, incorporating differential privacy to protect data while optimizing model updates. This demonstrates blockchain's capability to enhance both security and privacy in FL.

The work [25] proposed a taxonomy for evaluating the trustworthiness of FL models, focusing on privacy, fairness, robustness, and accountability. This framework underscores the importance of secure and transparent mechanisms in FL, which blockchain can facilitate. The work [26] introduces the Privacy-aware Task Assignment (PA-TA) Problem, which aims to allocate tasks to workers while prioritizing privacy and utility maximization. Wang et al. [27] explored an unbiased client sampling strategy to accelerate the convergence speed of FL. The work [28] aimed to solve the problem of data heterogeneity in FL, suggesting a one-pass distribution sketch to fill this gap, maintaining ϵ-differential privacy.

Further, several studies [26, 5, 17] have shown that privacy preservation is a critical issue in FL. Techniques like homomorphic encryption (HE) and secure multiparty computation (SMC) have been proposed within FL frameworks to protect data privacy, as seen in the articles [29] and [24].

The article [12] introduced a blockchain-based approach to optimize edge node selection while preserving privacy in FL settings, while authors of [13] focused on ensuring security and reliability in sharing medical records, prioritizing patient privacy.

The works [24] and [30] extensively studied the integration of differential privacy into blockchain technology, highlighting its practical implications for enhancing privacy in various applications.

Building on these studies, our work seeks to enhance the security and privacy of FL systems by integrating synchronous FL protocols with proven blockchain methodologies. Our approach aims to provide a unified framework that balances performance, privacy, and security, addressing the limitations of existing methods.

2.4 Summary of Gaps and Contributions

The reviewed literature highlights significant advancements in federated learning (FL), particularly in client selection, energy efficiency, and security mechanisms.

However, most existing methods either prioritize performance optimization or privacy preservation, and fail to provide a comprehensive framework that balances both objectives in both centralized and decentralized FL environments. Notably, there is a lack of algorithms that simultaneously address both optimization and security in client selection. To fill this gap, we propose a novel algorithm that integrates optimization and security considerations, ensuring a robust and efficient participant selection process. Our approach also incorporates blockchain-enhanced role determination, enabling secure and dynamic participant selection while optimizing for privacy and system performance. Through extensive experiments using MNIST datasets, we demonstrate that our method not only improves model convergence but also enhances the overall security and privacy of FL systems, especially in heterogeneous environments.

3 SYSTEM MODEL AND PROBLEM FORMULATION

3.1 Problem Formulation

In centralized Federated Learning (FL) methods, selecting devices with malicious intent or poor performance can significantly delay the training process. These issues often arise from devices with inferior hardware, unreliable connections, or frequent dropouts. To mitigate these challenges, we propose an algorithm to identify and exclude malicious or underperforming devices. The primary objective is to ensure the selection of the most secure and effective devices, thereby enhancing both the efficiency and security of the FL framework.

In blockchain-enabled FL systems, nodes assume distinct roles that are crucial for maintaining system integrity. Validators ensure compliance with standards and authenticate updates through cryptographic methods, while miners are responsible for creating new blocks for FL transactions, a process that demands substantial computational resources. Workers train machine learning models on local datasets, preserving data privacy, and subsequently send their updates to a central aggregator or validator for integration into the global model. A key priority is to select secure and proficient nodes as workers, aligning role allocation with the system's requirements. Notably, in decentralized systems, this role determination method is executed following the consensus process within the blockchain, ensuring that role assignments are consistent with the system's validated state.

Building on recent research [24, 30, 32], which emphasizes the importance of privacy preservation in blockchain-based FL systems, we propose incorporating differential privacy into our algorithm. This enhancement aims to secure data transactions and foster the development of privacy-preserving methodologies in decentralized learning environments.

Method	Selection Criteria	Key Features and Findings	Optimization	Privacy	Centralized/ Decentralized	Against Malicious Clients
PIRATE [5]	Reputation-based, Blockchain	Efficient reputation system, decentralized	✗	✓	Decentralized	✓
Oort [6]	Processing time, Accuracy	Optimizes participant selection, integrates with FL coordinator	✓	✗	Centralized	✗
Control Mechanism [10]	Selection frequency	Enhances participation of weaker clients, tested on FEMNIST datasets	✓	✗	Centralized	✗
VerifyNet [17]	Verification of server results	Robust under honest-but-curious security setting, supports dropout handling	✗	✓	Centralized	✓
SAM [19]	Selective Aggregation of Models	Efficient model aggregation, reduces communication overhead	✓	✗	Centralized	✗
REWAFL [16]	Residual energy, Wireless conditions	Optimizes energy consumption, suitable for mobile devices	✓	✓	Decentralized	✗
Ranking-based Client Selection [18]	Imitation learning, Efficiency	Improves efficiency through learning-based ranking	✓	✗	Centralized	✓
FLOAT [20]	Automated Tuning	Optimizes FL parameters automatically	✓	✗	Centralized	✗
Lotto [14]	Adversarial server resistance	Secure participant selection	✗	✓	Centralized	✓
Long-Term Client Selection [31]	Emulates full client participation	Long-term strategy for client selection	✓	✗	Centralized	✗
FedCure [21]	Heterogeneity-Aware, Blockchain	Personalized FL, intelligent healthcare applications	✗	✓	Decentralized	✓
Pisces [7]	Processing time, Accuracy	Optimizes participant selection, take advantage of struglers by adding asynchronous method	✓	✗	Centralized	✗
AFL [8]	Calculate probabilities depending on utilities	Probability-based client selection	✓	✗	Centralized	✗
Novel-reputation [11]	Calculate the reputation score for allowing participation	Trustscore for client selection	✓	✗	Centralized	✗
Hermes [9]	Communication overhead and improving inference efficiency	Group participant selection depending on communication	✓	✗	Centralized	✗
This Work	Data quality, Resource availability	Combines dynamic client selection with adaptive learning rates, proven on Mnist, privacy-preserving mechanisms	✓	✓	**Centralized/ Decentralized**	✓

Table 1. Comparison of existing client selection methods in federated learning (FL)

3.2 System Model

Our simulation models the dynamics of FL in both centralized and decentralized settings. Participants, called nodes, are tasked with training a neural network model on the MNIST dataset. The simulation dynamically assigns roles – workers, validators, miners, and potential adversaries – based on privacy and optimization levels conditions, with the role determination method in decentralized systems being contingent on the successful completion of blockchain consensus.

The iterative nature of FL is captured over multiple rounds. Workers contribute training updates, validators assess the quality of these updates, and miners aggregate them to improve the model. A blockchain securely records and stores these updates, ensuring both traceability and security.

Throughout the simulation, the system continuously monitors the model's accuracy, providing valuable insights into its performance over time. Bar charts visually depict the distribution of roles (workers, validators, miners) in each round, illustrating the collaborative efforts within the system.

The flexibility of our simulation framework allows researchers and practitioners to explore a variety of FL scenarios in both centralized and decentralized models. This adaptability is essential for experimenting with different aspects of FL and tailoring simulations to specific use cases and deployment scenarios.

4 PARTICIPANT SELECTION IN CENTRALIZED FL: BALANCING OPTIMIZATION AND PRIVACY

4.1 Overview of Participant Selection Process

In centralized Federated Learning (FL), the selection of participants is critical, balancing system performance and data privacy preservation. Potential participants are assessed by the FL server, acting as the coordinator, based on their technical capabilities and privacy safeguards. This process is summarized in Figure 1.

The selection metrics listed in Tables 2 and 3 were derived after an extensive review of related work in Section 2. Research consistently emphasizes that the technical and privacy criteria of devices significantly influence their performance in FL [6, 7, 11, 8]. The most commonly cited metrics in the literature include connectivity, computational resources, and security measures, which serve as the foundation for the metrics selected in this study.

4.2 Optimization Metrics

The optimization criteria focus on the technical aspects that influence a participant's ability to contribute effectively to the FL process. These include connectivity, battery life, storage capacity, and computational resources (RAM and CPU). The server evaluates these factors to ensure that selected participants can handle the resource-

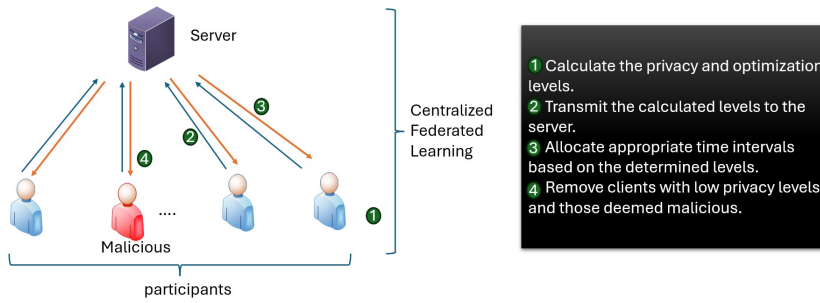

Figure 1. Graphical overview of the main steps of proposed participant selection method in centralized FL

intensive tasks involved in model training. Table 2 details the optimization metrics used in this evaluation.

Metric	**Description**
Connectivity Level	The type and strength of the device's network connection.
Battery Life	Battery level, with a threshold of 70 % as a key indicator.
Storage Capacity	Availability of sufficient storage space for the training process.
RAM	Adequacy of random-access memory for training tasks.
CPU	Suitability of the central processing unit for training requirements.
Priority	Preference given to devices based on proximity or other factors.

Table 2. Optimization metrics for participant selection

4.3 Privacy Metrics

Equally important are the privacy considerations, which ensure that participant data remains secure throughout the FL process. The server assesses various privacy metrics, including encryption strength, security protocols, firewall robustness, and attack vulnerability. These metrics help determine the privacy level of each participant, as outlined in Table 3.

Metric	Description
Encryption Algorithms	The effectiveness and strength of the encryption methods used.
Security Protocols	The security level provided by device and network protocols.
Firewall Robustness	The firewall's capability to control access and protect resources.
Vulnerability	The presence of known security weaknesses in the device.
Last Update Time	Recency of the last security update, indicating the currency of protective measures.

Table 3. Privacy metrics for participant selection

4.4 Refined Probabilistic Participant Selection Model

Beyond the basic participant selection algorithm, we propose a probabilistic model that adds complexity to the process, allowing for a more nuanced evaluation based on optimization and privacy levels. The probability $P(\text{selected})$ that a client will be selected is determined by the following formula:

$$P(\text{selected}) = \frac{1}{2}\left(\frac{1}{1+\exp(-(\text{privacyLevel}-8))} + \frac{1}{1+\exp(-(\text{optimizationLevel}-8))}\right). \tag{1}$$

This formula incorporates:

- **Sigmoid Function**: The sigmoid function is used to model the transition of selection probabilities based on 'privacyLevel' and 'optimizationLevel', providing a smooth, non-linear response to these metrics.
- **Thresholds**: Thresholds at $\theta_1 = 8$ and $\theta_2 = 8$ define significant cut-off points, marking where the probability of selection changes sharply.
- **Normalization**: A factor of $\frac{1}{2}$ ensures the combined probability remains within the range $[0, 1]$, balancing the impact of both privacy and optimization on selection.

4.5 Clarification of Threshold Basis ($\theta_1 = 8$ and $\theta_2 = 8$)

The thresholds $\theta_1 = 8$ and $\theta_2 = 8$ were chosen based on empirical evaluation and practical considerations. Here, we outline the reasoning behind this decision:

1. **Normalized Scale and Interpretation:** Devices are scored on a normalized scale of 0 to 10 across various metrics. A threshold of 8 reflects an 80 % performance level, which is a reasonable benchmark for identifying high-performing devices while accommodating slight imperfections.
2. **Empirical Validation:** During experimental trials, devices with scores ≥ 8 demonstrated consistent reliability in optimization and privacy:

   - **Optimization Level:** Ensures the device contributes effectively to model training.
   - **Privacy Level:** Meets necessary standards to safeguard data.

   Devices scoring below 8 showed deficiencies, leading to potential risks or inefficiencies.
3. **Balancing Inclusivity and Robustness:**

   - Lowering the threshold (e.g., $\theta < 8$) risks including devices with suboptimal capabilities.
   - Raising the threshold (e.g., $\theta > 8$) may exclude devices that are sufficiently capable, reducing overall participation.

   The choice of 8 balances these competing priorities.
4. **Practical Impact on Selection Probability:** The sigmoid function used in the formula emphasizes the transition around the threshold. At $x = 8$, the selection probability is 0.5, representing a moderate likelihood. Scores above 8 increase this probability, while those below 8 result in a rapid decline, reflecting their reduced suitability.
5. **Cumulative Metric Aggregation:** Each metric (e.g., battery life, encryption strength) contributes to the overall optimization and privacy levels:

$$\text{Optimization Level} = (\text{Battery} + \text{Storage} + \text{RAM} + \text{CPU} + \text{Priority})/5,$$

$$\text{Privacy Level} = (\text{Encryption} + \text{Security Protocols} + \text{Firewall} + \text{Vulnerability} + \text{Update Time})/5.$$

   Devices scoring below 8 in either metric are less likely to meet the system's reliability and security requirements.
6. **Example to Illustrate Threshold Application:** Consider a device with the following scores:

   - **Optimization Metrics:** Battery = 8.0, Storage = 9.0, RAM = 7.0, CPU = 6.5, Priority = 8.5.
   - **Privacy Metrics:** Encryption = 7.0, Security Protocols = 7.5, Firewall = 8.0, Vulnerability = 9.0, Update Time = 8.5.

The aggregated scores are:

$$\text{Optimization Level} = \frac{8.0 + 9.0 + 7.0 + 6.5 + 8.5}{5} = 7.8,$$

$$\text{Privacy Level} = \frac{7.0 + 7.5 + 8.0 + 9.0 + 8.5}{5} = 8.0.$$

Since the optimization level is below 8, the selection probability is lower:

$$P(\text{selected}) = \frac{1}{2}\left(\frac{1}{1+\exp(-(7.8-8))} + \frac{1}{1+\exp(-(8.0-8))}\right) \approx 0.4636.$$

This reflects the importance of meeting the thresholds to ensure both system reliability and security.

Algorithmic Integration: This model mirrors the selection process by assigning higher probabilities to clients with 'privacyLevel' and 'optimizationLevel' above 8. Clients with intermediate levels (4 to 8) have moderate probabilities, while those below the thresholds are unlikely to be selected. This probabilistic approach adds flexibility to the selection process, improving its robustness.

Optimization Level	**Privacy Level > 8**	**4 < Privacy Level < 8**	**Privacy Level < 4**
Level > 8	Selected (time = 0)	Selected (time = 0)	Eliminated
4 < Level < 8	Selected (time = +1)	Selected (time = +1)	Eliminated
Level < 4	Selected (time = +2)	Selected (time = +2)	Eliminated

Table 4. Participant selection based on optimization and privacy levels

As illustrated by Algorithm 1, the server selects clients based on optimization and privacy metrics derived from the probabilistic model outlined previously. This algorithm evaluates each client's suitability by considering their optimization and privacy levels, as specified by the problem formulation. Clients are then categorized into different selection tiers based on these metrics. The assignment of these tiers is intended to enhance the overall performance of the Federated Learning system by strategically allocating more time to clients with higher metrics, thereby improving system efficiency and effectiveness.

4.6 Experimental Scenarios and Results

The experiment delineates four distinct scenarios, each involving a cohort of 100 clients over 20 rounds. The first scenario embodies a centralized FL approach, where clients collaboratively train a global model under a unified and coordinated framework. In contrast, the second scenario introduces a more adversarial environment by

Algorithm 1 Participant Selection in FL

```
function SelectParticipants(clients)
    selectedClients ← []
    for client in clients do
        privacyLevel ← GetPrivacyLevel(client)
        optimizationLevel ← GetOptimizationLevel(client)
        if privacyLevel > 8 then
            selectedClients.append((client, t1))      ▷ Worker with requested time
        else if 4 < privacyLevel ≤ 8 then
            selectedClients.append((client, t2))  ▷ Worker with requested time +
1
        else                                                          ▷ Eliminated
        end if
        if optimizationLevel > 8 then
            selectedClients.append((client, t1))      ▷ Worker with requested time
        else if 4 < optimizationLevel ≤ 8 then
            selectedClients.append((client, t2)) ▷ Worker with requested time + 1
        else                                                          ▷ Eliminated
        end if
    end for
    return selectedClients
end function
```

incorporating malicious clients into the centralized FL paradigm. These malicious clients aim to disrupt the training process or manipulate the model parameters, thereby challenging the integrity and efficacy of the FL setup.

Expanding upon this adversarial scenario, the third experimental setup incorporates differential privacy techniques. Differential privacy serves as a robust privacy-preserving mechanism by introducing controlled noise or perturbations to the model updates, thereby safeguarding the privacy of individual client data while still allowing for meaningful model training and inference. By integrating differential privacy into the centralized FL process, the study aims to evaluate its effectiveness in mitigating the privacy risks associated with data aggregation and model updates.

Furthermore, the fourth scenario introduces an innovative participant selection method alongside the utilization of differential privacy. This approach involves dynamically selecting participants for model training based on various criteria, such as trustworthiness, performance history, or adherence to privacy-preserving protocols. By incorporating participant selection mechanisms into the FL framework, the experiment seeks to optimize the participant pool's composition while simultaneously preserving individual client data's privacy through integrating differential privacy techniques.

These experimental scenarios comprehensively explore diverse strategies for enhancing FL systems' robustness, privacy, and security. Through rigorous experimen-

tation and analysis, the study aims to elucidate effective approaches for mitigating the impact of malicious adversaries, preserving data privacy, and optimizing participant selection in centralized FL settings.

4.6.1 Experimental Setup and Evaluation Metrics

To provide a clear understanding of the experimental process, we describe the hardware/software environments, key parameters, and evaluation metrics used. The experiments were conducted on a system equipped with an Intel Core i7 processor, 16 GB RAM, and an NVIDIA RTX 3070 GPU, running Windows 11. The framework was implemented using Python 3.9, with PyTorch for model training. A custom, self-implemented private blockchain was designed specifically for this study, ensuring full control over the blockchain's consensus mechanism, structure, and integration with the federated learning framework. Differential privacy mechanisms were incorporated using the Opacus library.

Key experimental parameters include a learning rate of 0.01, batch size of 64, and 10 epochs for each client in every round. Evaluation metrics include model accuracy, robustness against adversarial attacks, and privacy preservation efficiency across all scenarios. While these details are secondary to the study's main findings, they provide important context for the experimental process and results interpretation.

4.7 Results Analysis: A Formal Examination of Findings

As shown in Figures 2 and 3, the method used to choose clients in FL plays a critical role in achieving optimal performance while protecting data privacy. By employing sophisticated optimization algorithms and privacy-preserving mechanisms, FL systems can choose clients whose data helps the model perform better, all while minimizing the risk of privacy breaches.

When comparing test accuracy across FL iterations, both the "FL with Client Selection" and "Standard FL" scenarios show similar performance, but the "FL with Client Selection" scenario shows steady improvement in accuracy over time. This highlights the effectiveness of the client selection process, which uses optimization strategies to choose clients whose data improves model accuracy without compromising privacy.

Overall, these findings highlight the importance of optimizing and prioritizing privacy in the client selection process in FL. FL systems can maintain client data confidentiality while identifying clients and contributing to model refinement through optimization and privacy-preserving mechanisms. By integrating various elements, we achieve optimal performance and foster trust and collaboration among participants, ultimately enhancing FL's effectiveness and dependability in real-world scenarios.

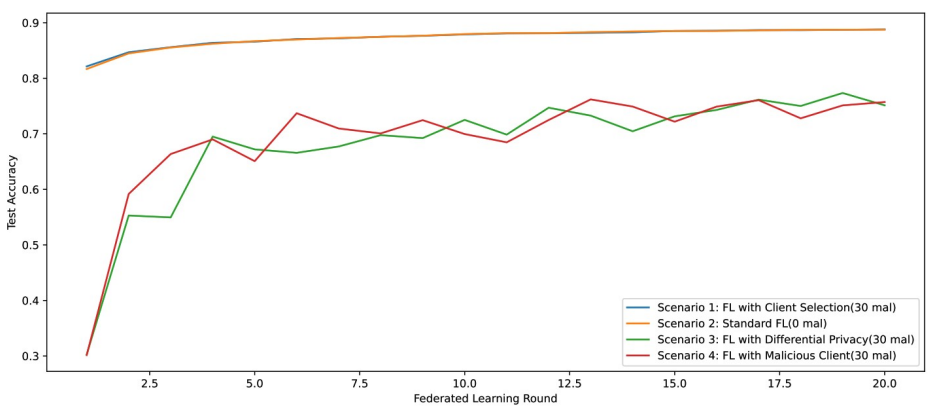

Figure 2. Results of centralized participant selection under IID-data

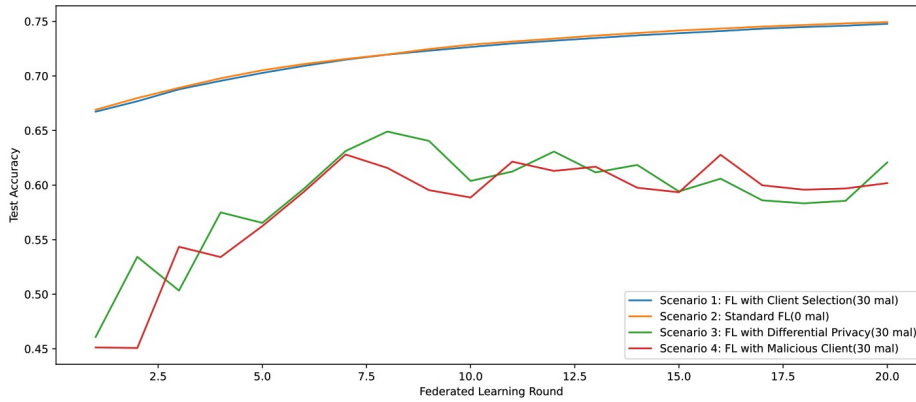

Figure 3. Results of centralized participant selection under non-IID-data

4.8 Comparison with Previous Work

The proposed method integrates both privacy and optimization objectives, marking a significant departure from previous approaches that focused on either one or the other. The experiments were conducted using the MNIST dataset, a widely used benchmark for image classification. A simple neural network (SimpleNN) with two hidden layers was employed to evaluate the performance of the participant selection methods. The scenario involved malicious clients intentionally providing incorrect updates, simulating a real-world federated learning challenge.

Table 5 presents a comparison between our centralized method and previous works. The algorithms compared, including Random Selection, Oort, Pisces, and Hermes, were chosen because they also aim to optimize federated learning by collecting and utilizing device information, such as computational efficiency and data quality, for participant selection. However, these methods focus primarily on optimization without considering privacy or security.

By contrast, our proposed method balances security and optimization, demonstrating its effectiveness in scenarios with malicious clients. The results show that our approach achieves substantial improvements, with an accuracy of 90.39 %, significantly outperforming state-of-the-art optimization-focused methods.

Experiments conducted under a scenario with malicious clients		
Algorithm	**Objective**	**Accuracy**
Random selection [1]	Optimization	68.9 %
Oort [6]	Optimization	16.7 %
Pisces [7]	Optimization	28.2 %
Hermes [9]	Optimization	23.8 %
Our proposed work	**Security + Optimization**	**90.39 %**

Table 5. Comparison between our centralized method and previous works

4.9 Discussion and Future Work

The experimental results confirm that our probabilistic participant selection model, which balances optimization and privacy metrics, significantly improves the performance and security of centralized FL systems. While the current model shows promising results, it opens avenues for future research. These include extending the model to decentralized FL environments, exploring the impact of dynamic threshold adjustments, and integrating additional security layers such as homomorphic encryption or secure multi-party computation. Further, real-world applications could be explored, particularly in domains where both performance and privacy are critical, such as healthcare and finance. As FL continues to evolve, the ability to dynamically adapt participant selection criteria based on real-time data and evolving threats will become increasingly important.

This section presents an enhanced participant selection method for centralized FL that simultaneously optimizes system performance and maintains high privacy standards. By introducing a probabilistic model, we have demonstrated how careful consideration of both optimization and privacy metrics can significantly improve the security and effectiveness of FL systems. The approach shows potential for broader applications and lays the groundwork for future enhancements in participant selection methods.

Section 5 elaborates on this proposed framework, detailing how blockchain technology can be leveraged to facilitate decentralized decision-making in client selection.

By decentralizing the process, the framework aims to improve transparency, security, and efficiency in node selection, ultimately leading to improved model performance and data privacy preservation.

5 ROLE DETERMINATION IN BLOCKCHAIN FL: ENHANCING PARTICIPANT SELECTION WITHIN THE BLOCKCHAIN FRAMEWORK

5.1 Introduction to the Proposed Work

Building upon the Blockchain-Assisted Federated Learning (FL) methodology outlined in the previous section, this section introduces a dynamic role determination method. This method categorizes nodes based on their privacy and optimization levels, enhancing the efficiency and security of the FL process (see Figure 4). The mechanism is implemented on a private permissioned blockchain, which ensures enhanced security and privacy throughout the FL process. The choice of a private blockchain is critical as it provides controlled access and governance, which is essential for maintaining system integrity. Notably, experiments have shown that the consensus algorithm choice does not significantly impact the effectiveness of the role determination mechanism. The main steps are illustrated in Figure 1.

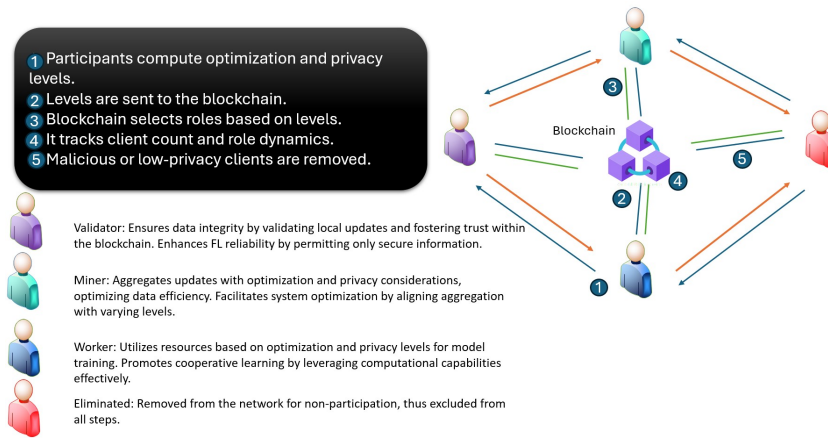

Figure 4. Graphical overview of the main steps of the proposed Role Determination method in Blockchain FL

5.2 Role Assignment Based on Optimization and Privacy Levels

Nodes are categorized according to their optimization and privacy values. Optimization factors such as connectivity, battery life, storage, RAM, and CPU are assessed

alongside privacy considerations, including encryption, security measures, authentication protocols, firewall integrity, and vulnerability resistance. These criteria form the basis for stratifying nodes into distinct roles within the FL framework.

5.3 Blockchain-Aided Role Assignment

Nodes are assigned the roles based on their optimization and privacy levels, as shown in Table 6. This role assignment is executed within the framework of a private permissioned blockchain. Nodes with higher optimization and privacy values are assigned critical roles such as model training (worker role), while nodes with lower privacy values are excluded from certain tasks. Nodes with intermediate privacy values assume roles with extended responsibilities. A private blockchain ensures controlled and secure role assignments, safeguarding data privacy and system integrity.

Levels	Privacy Level > 8	$4 <$ Privacy Level < 8	Privacy Level < 3
optimization level > 8	worker	worker	worker
$4 <$ optimization level < 8	miner	worker	worker
optimization level < 3	validator	worker	eliminated

Table 6. Role assignment based on optimization and privacy levels in blockchain-assisted decentralized FL

5.4 Probabilistic Role Assignment Formula

To assign roles probabilistically based on 'privacyLevel' and 'optimizationLevel', we define the following formula. This formula incorporates the conditions from the algorithm and integrates probabilistic elements for a more nuanced role assignment.

Let:

- $\theta_{\text{privacy}} = 8$ (threshold for privacy level),
- $\theta_{\text{optimization}} = 8$ (threshold for optimization level),
- $\alpha = 4$ (lower threshold for privacy and optimization levels).

The probability of each role R is defined as follows.

5.4.1 Worker

$$P(\text{worker}) = \frac{1}{2}\left(\frac{1}{1+\exp(-(\text{privacyLevel} - \theta_{\text{privacy}}))} + \frac{1}{1+\exp(-(\text{optimizationLevel} - \theta_{\text{optimization}}))}\right). \tag{2}$$

5.4.2 Miner

$$P(\text{miner}) = \frac{1}{2}\left(\frac{1}{1+\exp(-(\text{privacyLevel}-\theta_{\text{privacy}}))}\right.$$
$$\left.\cdot\frac{1}{1+\exp(-(\text{optimizationLevel}-\alpha))}\right). \tag{3}$$

5.4.3 Validator

$$P(\text{validator}) = \frac{1}{2}\left(\frac{1}{1+\exp(-(\text{privacyLevel}-\theta_{\text{privacy}}))}\right.$$
$$\left.\cdot\frac{1}{1+\exp(-(\text{optimizationLevel}-\alpha))}\right). \tag{4}$$

5.4.4 Eliminated

$$P(\text{eliminated}) = 1 - \left(P(\text{worker}) + P(\text{miner}) + P(\text{validator})\right). \tag{5}$$

5.5 Algorithm for Dynamic Role Assignment

Algorithm 2 formalizes the role assignment process within the Blockchain-Assisted FL (BAFL) context. This algorithm dynamically determines participant roles based on optimization and privacy levels, contributing to the orchestration of roles within the FL paradigm.

5.6 Detailed Overview of Roles and System Integration

5.6.1 Validator Role and Privacy Integrity

The validator plays a critical role in maintaining system privacy integrity by validating local updates and establishing a trust layer within the blockchain framework. This process ensures that only authenticated and secure information permeates the FL ecosystem, enhancing the overall reliability of the model training process.

5.6.2 Miner Role and Data Aggregation Efficiency

Miners aggregate validated data, tailoring the aggregation process to the optimization and privacy levels of participating devices. By effectively managing computational resources, miners contribute to overall system optimization.

Algorithm 2 Assign Role Algorithm for Blockchain-Assisted FL

```
function assignRole(optimizationLevel, privacyLevel)
    if optimizationLevel > 8 then
        if privacyLevel > 8 then
            return "worker"
        else if 4 < privacyLevel < 8 then
            return "worker"
        else
            return "worker"
        end if
    else if 4 < optimizationLevel < 8 then
        if privacyLevel > 8 then
            return "miner"
        else if 4 < privacyLevel < 8 then
            return "worker"
        else
            return "worker"
        end if
    else
        if privacyLevel > 8 then
            return "validator"
        else if 4 < privacyLevel < 8 then
            return "worker"
        else
            return "eliminated"
        end if
    end if
end function
```

5.6.3 Worker Role and Temporal Resource Allocation

Workers responsible for model training engage in tasks commensurate with their optimization and privacy levels. This allocation ensures meaningful contributions to the FL process, leveraging the collective computational prowess of participating devices.

5.6.4 Computational Interplay and System Synergy with Election Process

The computational interplay between validators and miners is facilitated through an election process, aligning collaboration with specific device characteristics. This election fosters a synergistic system architecture, ensuring optimized computational efforts and privacy considerations.

5.6.5 Adaptive Role Adjustment and System Resilience

The dynamic approach adjusts to changing circumstances, altering roles in line with evolving techniques and security concerns. This adaptability ensures that the FL process remains responsive and maintains continued optimization and compliance with privacy.

5.6.6 Blockchain-Aided Trust Establishment and System Governance

Blockchain integration within the FL architecture fosters a collaborative governance model, ensuring a balanced distribution of responsibilities between validators, miners, and workers. This approach improves system robustness and resilience against vulnerabilities.

5.6.7 Privacy-Centric Data Transmission and Continuous Monitoring

The system emphasizes privacy-centric data transmission through secure communication channels and strict authentication protocols. Continuous monitoring and evaluation ensure alignment with the system objectives, identifying potential bottlenecks and optimization opportunities.

5.7 Experiments

The experimental design consists of four scenarios, each involving 100 clients participating in 20 rounds. The first situation showcases a typical decentralized FL approach, where clients work together to train a global model without any malicious intervention. In contrast, the second scenario presents an environment with malicious clients in the decentralized FL framework. These malicious clients seek to disrupt training and manipulate model parameters, which undermines the integrity and effectiveness of the FL paradigm.

Building upon this adversarial scenario, the third experimental setup extends the analysis to incorporate differential privacy techniques. Differential privacy serves as a robust privacy-preserving mechanism, introducing noise or perturbations to the model updates to prevent the inference of sensitive information about individual clients. By integrating such privacy-enhancing measures into the FL process, the study aims to assess their efficacy in mitigating the influence of malicious clients and preserving data privacy in a decentralized setting.

Furthermore, the fourth scenario introduces a novel approach by integrating a private blockchain into the FL framework. This scenario addresses the challenges posed by malicious clients and leverages the blockchain's inherent properties, such as transparency, immutability, and decentralized consensus mechanisms, to facilitate role determination within the FL process. By employing private blockchain-based role determination alongside differential privacy measures, the experiment aims to explore synergistic effects in enhancing the security and privacy aspects of FL, particularly in the presence of malicious adversaries.

These experimental scenarios comprehensively explore various strategies for ensuring the robustness, integrity, and privacy of FL systems in adversarial environments. Through rigorous experimentation and analysis, the study endeavors to shed light on effective approaches for securing decentralized learning frameworks against malicious threats while safeguarding the privacy of participant data.

5.7.1 Experimental Setup and Evaluation Metrics

To provide a clear understanding of the experimental process, we describe the hardware/software environments, key parameters, and evaluation metrics used. The experiments were conducted on a system equipped with an Intel Core i7 processor, 16 GB RAM, and an NVIDIA RTX 3070 GPU, running Windows 11. The framework was implemented using Python 3.9, with PyTorch for model training. A custom, self-implemented private blockchain was designed specifically for this study, ensuring full control over the blockchain's consensus mechanism, structure, and integration with the federated learning framework. Differential privacy mechanisms were incorporated using the Opacus library. Key experimental parameters include a learning rate of 0.01, batch size of 64, and 10 epochs for each client in every round. Evaluation metrics include model accuracy, robustness against adversarial attacks, and privacy preservation efficiency across all scenarios. While these details are secondary to the study's main findings, they provide important context for the experimental process and results interpretation.

5.8 Results Analysis: A Formal Examination of Findings

The main emphasis of the analysis is on studying accuracy dynamics across multiple rounds in the context of FL, as shown in Figure 5. Four distinct scenarios are delineated, each characterized by varying degrees of decentralization and the presence of malicious clients at a rate of approximately 20 %. These scenarios are outlined as:

1. Decentralized FL with No Malicious Client,
2. Decentralized FL with Malicious Client (20 %),
3. Blockchain-Assisted FL with Malicious Client (20 %),
4. Blockchain FL Role Determination with Malicious Client (20 %).

Empirical observations reveal a hierarchical distribution of accuracy performances among these scenarios. Notably, the decentralized FL scenario with No Malicious Client exhibits remarkable accuracy, peaking at around 97 %. Conversely, scenarios involving malicious clients, namely decentralized FL and blockchain-assisted FL, demonstrate more modest accuracy ranges, fluctuating between 10 % and 17 % over consecutive rounds.

It is important to note that the experiments were conducted under identical conditions, using the MNIST dataset and a simple neural network (SimpleNN) model. This ensures consistency and allows for a fair comparison across different scenarios.

The clear difference in accuracy across scenarios can be attributed to the power of sophisticated role-determination methods. These methods strategically assign devices with better capabilities as "workers". This deters malicious clients and offers a mitigation strategy if they appear. The observed trends highlight the critical role of advanced role determination in optimizing FL. It leads to a significant improvement in accuracy, especially for decentralized and blockchain-based systems.

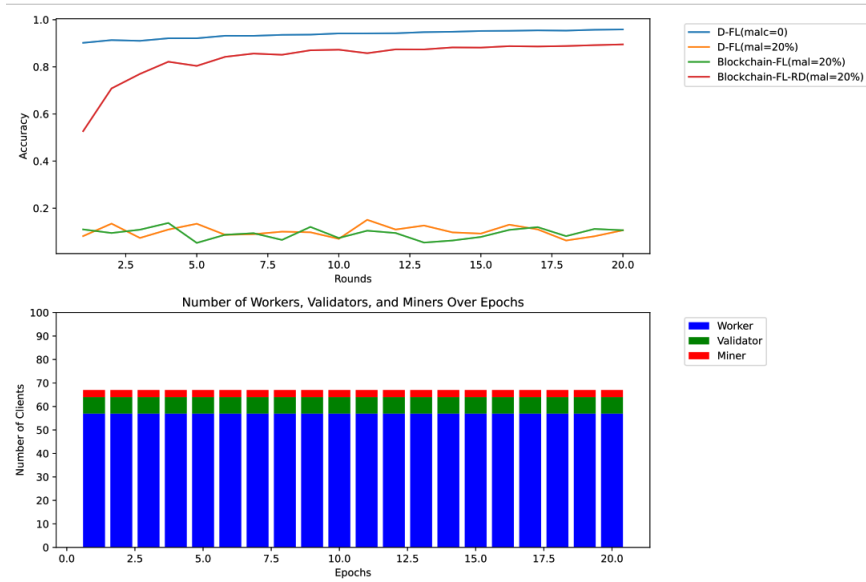

Figure 5. Results of blockchain role determination compared to other scenarios for IID

Figure 6 showcases the enhanced role determination method's performance under non-IID (non-independent and identically distributed) data conditions. This method is a diagnostic tool that assesses data compatibility for collaborative training with diverse models. It plays a crucial role in achieving superior performance metrics, even in non-IID settings, by effectively identifying and mitigating outliers or anomalies within the dataset.

The refined role determination method is a powerful filter for federated learning (FL). It examines data compatibility and eliminates outliers, boosting the overall robustness and resilience of the FL process to variations in data distributions. This method dynamically assigns roles based on optimization and privacy considerations, fostering a more cohesive FL environment, especially when dealing with non-IID data challenges.

The analysis revealed that the proposed method significantly improves both accuracy and security compared to traditional decentralized FL methods, particularly in environments with malicious clients. As depicted in Table 7, the proposed method

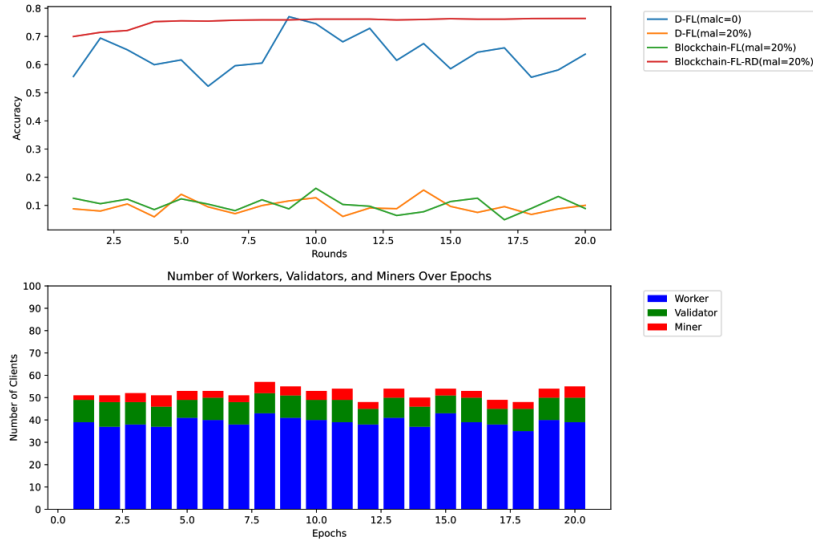

Figure 6. Results of blockchain role determination compared to other scenarios for non-IID

outperforms existing methods by achieving higher accuracy while maintaining security.

Algorithm	**Objective**	**Accuracy**
Random Selection [1]	Optimization	18.9 %
Oort [6]	Optimization	16.4 %
Active Federated Learning [8]	Optimization	18.4 %
novel-reputation [11]	Security	67.7 %
Proposed Method	**Security + Optimization**	**87.39 %**

Table 7. Comparative results of previous works and the proposed method

5.9 Conclusion and Future Directions

The role determination method, supported by blockchain technology, presents a secure and scalable solution for participant selection in FL systems. Future research will explore expanding this approach to more diverse and larger-scale FL environments and refine probabilistic role assignment formulas to further enhance adaptability and efficiency.

6 EXPERIENCES OF ROLE DETERMINATION ON DIFFERENT MODELS

Tests on dynamic role determination in a blockchain framework demonstrate its remarkable robustness. The system can optimize performance even with fixed node characteristics and environmental conditions. The results presented in Table 8 provide evidence that dynamic role determination is highly effective in enhancing performance and reducing negative consequences.

6.1 Adaptive Nature of Dynamic Role Determination

The dynamic role-determination mechanism proves its worth even in challenging circumstances. Despite fixed node characteristics and a stable environment, the system remains adaptable through continuous rule adjustments, enabling real-time performance optimization. Even without external changes, the system identifies and eliminates inefficiencies, ensuring optimal functionality.

6.2 Optimization of Performance in Challenging Scenarios

Table 8 showcases the system's resilience, even under worst-case conditions. Dynamic role determination maintains performance on par with other scenarios, demonstrating the effectiveness of adaptive rule adjustments. By dynamically allocating resources and optimizing performance, these adjustments significantly mitigate the impact of adverse conditions on overall outcomes.

6.3 Resilience to Adverse Conditions

Even in the worst-case scenario, where frequent node elimination due to low privacy or optimization levels occurs, dynamic role determination is crucial for system resilience. The system adapts to these changing conditions by dynamically reallocating roles and resources, maintaining functionality, and ensuring stability and reliability under pressure.

6.4 Continuous Improvement and Iterative Adaptation

The iterative process of determining dynamic roles promotes continuous improvement within the system. Through continuous evaluation and real-time feedback-driven adjustments to the rules, the system identifies areas for improvement and implements corresponding changes. This iterative adaptation ensures that the system remains attuned to evolving requirements and maintains its adaptive capabilities over time.

6.5 Experimental Setup and Evaluation Metrics

To ensure reproducibility and provide a clear understanding of the experimental process, we briefly describe the hardware/software environments, experimental parameters, and evaluation metrics used.

The experiments were conducted on a system equipped with an Intel Core i7 processor, 16 GB RAM, and an NVIDIA RTX 3070 GPU, running Windows 11. The blockchain framework was implemented using Python 3.9, leveraging PyTorch for machine learning tasks. Key experimental parameters include the learning rate, batch size, and epoch count, which were adjusted to fit the specific requirements of each dataset and model.

The evaluation metrics focused on model accuracy across different scenarios, with specific attention to performance under "Best Case", "Average Case", and "Worst Case" conditions. While these details may not be central to the study's conclusions, they provide additional context to validate the experimental results.

Datasets and Models	SimpleNN	MLP	CNN	Net0
Best Case				
MNIST	83.12 %	85.9 %	82.29 %	87.39 %
FASHION MNIST	72.55 %	72.38 %	72.54 %	77.02 %
CIFAR-10	29.26 %	29.14 %	30.24 %	37.61 %
Average Case				
MNIST	78.28 %	79.59 %	77.6 %	86.77 %
FASHION MNIST	71.43 %	70.88 %	71.11 %	76.6 %
CIFAR-10	29.61 %	29.05 %	25.83 %	37.19 %
Worst Case				
MNIST	52.61 %	60.96 %	57.28 %	79.92 %
FASHION MNIST	52.04 %	55.34 %	57.49 %	73.07 %
CIFAR-10	28.65 %	21.99 %	15.63 %	37.17 %

Table 8. Results of experiments on different case scenarios

7 CONCLUSION AND FUTURE WORK

Our comprehensive investigation into federated learning (FL) scenarios, encompassing both independent and identically distributed (IID) and non-independent and identically distributed (Non-IID) settings, has unveiled valuable insights into participant selection, optimization strategies, privacy concerns, and the unique role determination methodologies employed within a blockchain-assisted FL framework.

The impressive performance of FL with strategic client selection underscores the crucial role of well-defined participant selection strategies in boosting overall accuracy. Conversely, the consistent improvement achieved by standard FL reinforces its dependability within the FL framework. However, the declining accuracy

observed in FL with differential privacy and FL with a malicious client emphasizes the intricate challenges of balancing privacy preservation and security within FL environments.

Integrating formalized metrics for optimization and privacy levels into the participant selection algorithm exemplifies a sophisticated decision-making process that skillfully balances technological capabilities and privacy safeguards. The resulting role assignments, outlined within the algorithmic framework and further detailed in the assigned roles table, highlight the intricate hierarchy established by the FL server based on a combination of optimization and privacy levels.

Significantly, introducing blockchain-assisted FL adds another layer of complexity to role determination. The interplay between participants, shaped by their optimization preferences and privacy concerns, illuminates the intricate nature of role allocation, collaborative learning optimization, and privacy maintenance within a blockchain-facilitated setting.

The role determination method proposed in this study leverages optimization and privacy considerations to allocate tasks effectively within the FL framework. By defining clear roles for nodes and integrating blockchain technology, this approach enhances the efficiency, security, and reliability of FL systems. Validators authenticate local updates to maintain data integrity, miners aggregate validated data, and workers focus on model training within a secure and optimized setting. Integrating blockchain technology fortifies the FL process with enhanced transparency and security measures.

In summary, our research demonstrates that strategic participant selection and robust optimization and privacy metrics are pivotal in achieving high performance in federated learning environments. The dynamic role determination method within a blockchain-assisted FL framework showcases a sophisticated approach to managing the intricate balance of efficiency, security, and privacy. This holistic approach ensures the FL ecosystem's robustness, reliability, and adaptability, ultimately advancing secure and efficient collaborative learning paradigms.

Future research should focus on advanced privacy-preserving mechanisms like homomorphic encryption, enhancing scalability, developing strategies to detect and mitigate sophisticated attacks, improving the adaptability of role determination algorithms, applying the framework to diverse real-world applications, creating standards for interoperability, and developing incentive mechanisms for honest participation. Addressing these areas will push the boundaries of federated learning, making it more secure, efficient, and widely applicable.

8 DECLARATIONS

Funding: No funding was received to carry out this study.

Conflict of interest/Competing interests: The authors declare no potential conflict of interest.

Ethics approval and consent to participate: Not applicable.

Consent for publication: All authors of this study have agreed to be published in your journal.

Data availability: The datasets analyzed during the current study are publicly available in the MNIST dataset repository (`http://yann.lecun.com/exdb/mnist/`).

Author contribution: All authors contributed equally to this work.

REFERENCES

[1] McMahan, H. B.—Moore, E.—Ramage, D.—Agera y Arcas, B.: Communication-Efficient Learning of Deep Networks from Decentralized Data. CoRR, 2016, doi: 10.48550/arXiv.1602.05629.

[2] Bonawitz, K.—Eichner, H.—Grieskamp, W.—Huba, D.—Ingerman, A.—Ivanov, V.—Kiddon, C.—Konečný, J.—Mazzocchi, S.—McMahan, B.—Van Overveldt, T.—Petrou, D.—Ramage, D.—Roselander, J.: Towards Federated Learning at Scale: System Design. In: Talwalkar, A., Smith, V., Zaharia, M. (Eds.): Proceedings of Machine Learning and Systems 1 (MLSys 2019). 2019, pp. 374–388, `https://proceedings.mlsys.org/paper_files/paper/2019/file/7b770da633baf74895be22a8807f1a8f-Paper.pdf`.

[3] Nishio, T.—Yonetani, R.: Client Selection for Federated Learning with Heterogeneous Resources in Mobile Edge. ICC 2019 – IEEE International Conference on Communications, 2019, pp. 1–7, doi: 10.1109/ICC.2019.8761315.

[4] Mendis, G. J.—Sabounchi, M.—Wei, J.—Roche, R.: Blockchain as a Service: An Autonomous, Privacy Preserving, Decentralized Architecture for Deep Learning. CoRR, 2018, doi: 10.48550/arXiv.1807.02515.

[5] Zhou, S.—Huang, H.—Chen, W.—Zhou, P.—Zheng, Z.—Guo, S.: PIRATE: A Blockchain-Based Secure Framework of Distributed Machine Learning in 5G Networks. IEEE Network, Vol. 34, 2020, No. 6, pp. 84–91, doi: 10.1109/MNET.001.1900658.

[6] Lai, F.—Zhu, X.—Madhyastha, H. V.—Chowdhury, M.: Oort: Efficient Federated Learning via Guided Participant Selection. 15th USENIX Symposium on Operating Systems Design and Implementation (OSDI 2021), 2021, pp. 19–35, `https://www.usenix.org/system/files/osdi21-lai.pdf`.

[7] Jiang, Z.—Wang, W.—Li, B.—Li, B.: Pisces: Efficient Federated Learning via Guided Asynchronous Training. Proceedings of the 13th Symposium on Cloud Computing (SoCC '22), ACM, 2022, pp. 370–385, doi: 10.1145/3542929.3563463.

[8] Goetz, J.—Malik, K.—Bui, D.—Moon, S.—Liu, H.—Kumar, A.: Active Federated Learning. CoRR, 2019, doi: 10.48550/arXiv.1909.12641.

[9] Li, A.—Sun, J.—Li, P.—Pu, Y.—Li, H.—Chen, Y.: Hermes: An Efficient Federated Learning Framework for Heterogeneous Mobile Clients. Proceedings of the 27th Annual International Conference on Mobile Computing and Networking (MobiCom '21), ACM, 2021, pp. 420–437, doi: 10.1145/3447993.3483278.

[10] Zhao, J.—Chang, X.—Feng, Y.—Liu, C. H.—Liu, N.: Participant Selection for Federated Learning with Heterogeneous Data in Intelligent Transport System. IEEE Transactions on Intelligent Transportation Systems, Vol. 24, 2023, No. 1, pp. 1106–1115, doi: 10.1109/TITS.2022.3149753.

[11] Wang, Y.—Kantarci, B.: A Novel Reputation-Aware Client Selection Scheme for Federated Learning Within Mobile Environments. 2020 IEEE 25th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD), 2020, pp. 1–6, doi: 10.1109/CAMAD50429.2020.9209263.

[12] Qammar, A.—Naouri, A.—Ding, J.—Ning, H.: Blockchain-Based Optimized Edge Node Selection and Privacy-Preserved Framework for Federated Learning. Cluster Computing, Vol. 27, 2024, No. 3, pp. 3203–3218, doi: 10.1007/s10586-023-04145-0.

[13] Javed, L.—Anjum, A.—Yakubu, B. M.—Iqbal, M.—Moqurrab, S. A.—Srivastava, G.: ShareChain: Blockchain-Enabled Model for Sharing Patient Data Using Federated Learning and Differential Privacy. Expert Systems, Vol. 40, 2023, No. 5, Art. No. e13131, doi: 10.1111/exsy.13131.

[14] Jiang, Z.—Ye, P.—He, S.—Wang, W.—Chen, R.—Li, B.: Lotto: Secure Participant Selection Against Adversarial Servers in Federated Learning. Proceedings of the 33rd USENIX Security Symposium (USENIX Security '24), 2024, pp. 343–360, https://www.usenix.org/system/files/usenixsecurity24-jiang-zhifeng.pdf.

[15] Zeng, H.—Zhang, M.—Liu, T.—Yang, A.: A Federated Learning Framework with Blockchain-Based Auditable Participant Selection. Computers, Materials & Continua, Vol. 79, 2024, No. 3, doi: 10.32604/cmc.2024.052846.

[16] Li, Y.—Qin, X.—Geng, J.—Chen, R.—Hou, Y.—Gong, Y.—Pan, M.—Zhang, P.: REWAFL: Residual Energy and Wireless Aware Participant Selection for Efficient Federated Learning over Mobile Devices. IEEE Transactions on Mobile Computing, Vol. 23, 2024, No. 10, pp. 9487–9501, doi: 10.1109/TMC.2024.3365477.

[17] Xu, G.—Li, H.—Liu, S.—Yang, K.—Lin, X.: VerifyNet: Secure and Verifiable Federated Learning. IEEE Transactions on Information Forensics and Security, Vol. 15, 2020, pp. 911–926, doi: 10.1109/TIFS.2019.2929409.

[18] Tian, C.—Shi, Z.—Qin, X.—Li, L.—Xu, C.: Ranking-Based Client Selection with Imitation Learning for Efficient Federated Learning. Proceedings of the 41st International Conference on Machine Learning (ICML), 2024, doi: 10.48550/arXiv.2405.04122.

[19] Shi, Y.—Fan, P.—Zhu, Z.—Peng, C.—Wang, F.—Letaief, K. B.: SAM: An Efficient Approach with Selective Aggregation of Models in Federated Learning. IEEE Internet of Things Journal, Vol. 11, 2024, No. 11, pp. 20769–20783, doi: 10.1109/JIOT.2024.3373822.

[20] Khan, A. F.—Khan, A. A.—Abdelmoniem, A. M.—Fountain, S.—Butt, A. R.—Anwar, A.: FLOAT: Federated Learning Optimizations with Automated Tuning. Proceedings of the Nineteenth European Conference on Computer Systems (EuroSys '24), ACM, 2024, pp. 200–218, doi: 10.1145/3627703.3650081.

[21] Sachin, D. N.—Basava, A.—Hegde, S.—Abhijit, C. S.—Ambesange, S.: FedCure: A Heterogeneity-Aware Personalized Federated Learning Framework for Intel-

ligent Healthcare Applications in IoMT Environments. IEEE Access, Vol. 12, 2024, pp. 15867–15883, doi: 10.1109/ACCESS.2024.3357514.

[22] You, L.—Guo, Z.—Zuo, B.—Chang, Y.—Yuen, C.: SLMFed: A Stage-Based and Layer-Wise Mechanism for Incremental Federated Learning to Assist Dynamic and Ubiquitous IoT. IEEE Internet of Things Journal, Vol. 11, 2024, No. 9, pp. 16364–16381, doi: 10.1109/JIOT.2024.3353793.

[23] Bhope, R. A.—Jayaram, K. R.—Venkatasubramanian, N.—Verma, A.—Thomas, G.: FLIPS: Federated Learning Using Intelligent Participant Selection. Proceedings of the 24th International Middleware Conference (Middleware '23), ACM, 2023, pp. 301–315, doi: 10.1145/3590140.3629123.

[24] Jia, B.—Zhang, X.—Liu, J.—Zhang, Y.—Huang, K.—Liang, Y.: Blockchain-Enabled Federated Learning Data Protection Aggregation Scheme with Differential Privacy and Homomorphic Encryption in IIoT. IEEE Transactions on Industrial Informatics, Vol. 18, 2022, No. 6, pp. 4049–4058, doi: 10.1109/TII.2021.3085960.

[25] Sánchez, P. M. S.—Celdrán, A. H.—Xie, N.—Bovet, G.—Pérez, G. M.—Stiller, B.: FederatedTrust: A Solution for Trustworthy Federated Learning. Future Generation Computer Systems, Vol. 152, 2024, pp. 83–98, doi: 10.1016/j.future.2023.10.013.

[26] Du, L.—Cheng, P.—Zheng, L.—Xi, W.—Lin, X.—Zhang, W.—Fang, J.: Dynamic Private Task Assignment under Differential Privacy. 2023 IEEE 39th International Conference on Data Engineering (ICDE), 2023, pp. 2740–2752, doi: 10.1109/ICDE55515.2023.00210.

[27] Wang, L.—Guo, Y.—Lin, T.—Tang, X.: DELTA: Diverse Client Sampling for Fast Federated Learning. In: Oh, A., Naumann, T., Globerson, A., Saenko, K., Hardt, M., Levine, S. (Eds.): Advances in Neural Information Processing Systems 36 (NeurIPS 2023). Curran Associates, Inc., 2023, pp. 47626–47668, https://proceedings.neurips.cc/paper_files/paper/2023/file/949c57d30f8791e3ae42646081b3c102-Paper-Conference.pdf.

[28] Liu, Z.—Xu, Z.—Coleman, B.—Shrivastava, A.: One-Pass Distribution Sketch for Measuring Data Heterogeneity in Federated Learning. In: Oh, A., Naumann, T., Globerson, A., Saenko, K., Hardt, M., Levine, S. (Eds.): Advances in Neural Information Processing Systems 36 (NeurIPS 2023). Curran Associates, Inc., 2023, pp. 15660–15679, https://proceedings.neurips.cc/paper_files/paper/2023/file/32c2f3e0a44d55820da7fbcee0a1d95c-Paper-Conference.pdf.

[29] Zhang, L.—Xu, J.—Vijayakumar, P.—Sharma, P. K.—Ghosh, U.: Homomorphic Encryption-Based Privacy-Preserving Federated Learning in IoT-Enabled Healthcare System. IEEE Transactions on Network Science and Engineering, Vol. 10, 2023, No. 5, pp. 2864–2880, doi: 10.1109/TNSE.2022.3185327.

[30] Hassan, M. U.—Rehmani, M. H.—Chen, J.: Differential Privacy in Blockchain Technology: A Futuristic Approach. Journal of Parallel and Distributed Computing, Vol. 145, 2020, pp. 50–74, doi: 10.1016/j.jpdc.2020.06.003.

[31] Li, Q.—Miao, J.—Zhao, P.—Zhou, L.—Ji, S.—Zhou, B.—Liu, F.: Emulating Full Client Participation: A Long-Term Client Selection Strategy for Federated Learning. CoRR, 2024, doi: 10.48550/arXiv.2405.13584.

[32] Pauu, K. T.—Wu, J.—Fan, Y. et al.: Differential Privacy and Blockchain-Empowered Decentralized Graph Federated Learning Enabled UAVs for Disaster Response. IEEE Internet of Things Journal, 2023, pp. 20930–20947, doi: 10.1109/JIOT.2023.3332216.

Wafa Bouras is Ph.D. student at M'sila University, where she also worked as a temporary teacher. She holds her Master's degree in networks and technologies of information and communication. Her research focuses on blockchain technology and cybersecurity, particularly in enhancing the security of distributed systems. She is passionate about exploring innovative solutions to real-world cybersecurity challenges and actively participates in academic discussions and conferences.

Kameleddine Heraguemi is the Director of Networks and Digital Development at the Ministry of Higher Education and Scientific Research and an Assistant Professor at the National School of Artificial Intelligence, Sidi Abdallah, Algeria. He received his Ph.D. in computer science from the University Ferhat Abbas Setif1 (UFAS1) and his Master's in computer science from the Mohamed-Cherif Messaadia University, Souk-Ahras, Algeria, in 2017 and 2012, respectively. His research interests include data mining, artificial intelligence, and evolutionary computing.

Mohamed Benouis is a computer researcher specializing in machine learning and data mining. He holds his Bachelor's degree in computer science from the University of Science and Technology Houari Boumediene, Algiers, Algeria. Currently pursuing his Ph.D., he is dedicated to advancing knowledge in artificial intelligence and developing innovative algorithms to solve real-world problems. With a strong focus on research, his contributions to the field are highly regarded, and he actively mentors aspiring researchers in computer science.

Brahim Bouderah is an accomplished academic and administrator. With his Ph.D. from the University of U.F.A. Setif, he has made significant research contributions, particularly in computing and mathematics. Currently serving as the Rector of the University Abdelhamid Ibn Badis Mostaganem, he has demonstrated exceptional leadership and expertise in academic and administrative roles.

Samir Akrouf is a renowned researcher in the field of computer science, with a focus on software engineering and distributed systems. He obtained his Ph.D. in computer science from the University of Sciences and Technology Houari Boumediene, Algiers, Algeria. He has contributed significantly to developing efficient algorithms and methodologies for distributed computing and software development.