

Zero-Trust, Zero-Knowledge: A Lightweight Blockchain and Federated Learning Framework for Privacy-Preserving IoT Device Authentication

Mohamed Koroma

`mohamed.koroma@njala.edu.sl`

Njala University

Ibrahim Abdulai Sawaneh

University of Management and Technology

Mohamed Syed Fofanah

Njala University

Maurice Sesay

Njala University

Article

Keywords: IoT authentication, Behavioral fingerprinting, Federated learning, Blockchain, Agentic systems

Posted Date: October 20th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11000228/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Zero-Trust, Zero-Knowledge: A Lightweight Blockchain and Federated Learning Framework for Privacy-Preserving IoT Device Authentication

Mohamed Koroma (Ing.)^{1*}, Ibrahim Abdulai Sawaneh (Ph.D.)², Professor Mohamed Syed Fofanah¹, Maurice Sesay (PhD)¹

¹School of Technology, Department of Computer Science and Information Technology,
Njala University, Sierra Leone, West Africa.

²School of Technology, Department of Computer Science, University of Management, Freetown, Sierra Leone,
West Africa.

*Corresponding author(s). E-mail(s): mohamed.koroma@njala.edu.sl;

Co-authors: msfofanah1@njala.edu.sl, msesay@njala.edu.sl, ibrahim.sawaneh@unimtech.edu.sl

Abstract

An unprecedented security challenge is continuous authentication, privacy protection, and decentralized trust management for the ever-expanding Internet of Things (IoT) devices. This work proposes a new Zero-Trust Zero-Knowledge authentication mechanism based on behavior-based fingerprinting, federated learning, blockchain technology, and agentic orchestration. The framework uses hardware-intrinsic behavioral characteristics from the PP-IDE dataset to continuously authenticate devices without exposing persistent identifiers. Federated learning and differential privacy ($\epsilon = 5.0$) lower the membership inference attacks from 48.2% to 12.4%. At \$0.006 per authentication, a lightweight proof-of-authority blockchain can process 1,200 transactions per second with a latency of just 1.2 seconds. The agentic orchestration based on SHAP scores has a detection rate of 97.8% and exhibits a response time of 98.4 ms. The comprehensive evaluation shows 99.2% authentication accuracy and 93.2% Byzantine attack resistance, superior to Adaptive-AI-Zero Trust-Chain and BETAC-IoT. A secure, privacy-aware, and efficient authentication approach for IoT applications in smart cities, healthcare, and industrial IoT.

Keywords: IoT authentication, Behavioral fingerprinting, Federated learning, Blockchain, Agentic systems

1. Introduction

The Internet of Things (IoT) has revolutionized infrastructure in recent years, especially within healthcare, industrial automation, intelligent transportation, and smart city systems (Alshaikhli et al., 2025; Kumar et al., 2025). Kumar et al. (2025) suggest that by 2026 there will be more than 30 billion IoT endpoints worldwide, making them an extremely large attack surface that traditional security paradigms cannot protect. These resource-limited devices typically operate in untrusted, hostile environments and handle streams of sensitive information that must be protected (Verma et al., 2026). The growing diversity of IoT applications, combined with the fact that IoT devices often have long lifecycles (decades) and limited software/firmware updates, makes the traditional perimeter-oriented security approach inadequate for tackling contemporary threats (Zhang et al.,

2025). Recent comprehensive surveys describe perimeter-based security as fundamentally unfit for IoT because of IoT device heterogeneity, long lifespans, and infrequent updates. Zhang et al. (2025) also note that the convergence of OT and IT in IoT creates new attack surfaces because legacy systems and modern security solutions often do not align.

Securing data becomes even more complicated with Artificial Intelligence (AI) (Patel & Shah, 2025; Rodriguez et al., 2026). Despite significant strides in IoT security, recent studies show that 83% of these systems remain vulnerable to advanced threats such as adversarial spoofing, model poisoning, and deepfake identity fraud (Chen et al., 2026; Wang et al., 2025). In fact, attackers using the Mozi botnet have shown they can outsmart signature-based defenses (Liu et al., 2025), attack over 1.5 million IoT devices, and generate record attack traffic. Generative adversarial networks (GANs) have also been used to boost biometric spoofing attacks further, successfully spoofing conventional authentication methods at significantly higher rates (Rodriguez et al., 2026). Despite this, adaptive adversaries increasingly attempt attacks that static defenses cannot thwart (Patel & Shah, 2025; Rodriguez et al., 2026). The rapid progress of these technologies highlights the need to shift the security paradigm beyond traditional architectures toward more intelligent, decentralized approaches that protect privacy (Ahmed & Khan, 2025; Suryanarayan Iyer et al., 2026).

Device authentication is one of the most critical security issues in IoT, but current authentication systems still face many challenges (Obaidat et al., 2025). Conventional centralized authentication systems like public key infrastructure (PKI) and certificate authorities (CA) can enable identity theft and unauthorized access on a massive scale, as they create single points of failure when compromised (Kumar et al., 2025; Zhang et al., 2025). Because IoT devices have limited resources, deploying state-of-the-art cryptographic algorithms is difficult because they are too resource-intensive. Therefore, lightweight techniques are needed to provide security and efficient operation (Alshaikhli et al., 2025). Furthermore, existing authentication approaches can only verify device authenticity at initial access and cannot detect compromised devices exhibiting abnormal behavior after authentication (Verma et al., 2026). Recent studies emphasize the need to strengthen authentication in the IoT security domain, where 83% of IoT systems are susceptible to advanced attacks such as adversarial spoofing, model poisoning, and deepfake identity theft (Chen et al., 2026; Wang et al., 2025). One example of an adversary that can scale to unprecedented levels through adversary discovery and propagation is the Mozi botnet, which infected over 1.5 million IoT devices (Liu et al., 2025). Moreover, these attacks are becoming more effective, as GANs improve and enhance their capabilities, making traditional biometric authentication methods more difficult (Rodriguez et al., 2026; Ahmed & Khan, 2025).

Despite significant strides in IoT security research, current IoT authentication mechanisms still have major vulnerabilities and cannot adequately secure deployment environments (Chen et al., 2026). First, static credential-based methods remain susceptible to hacking, cloning, and persistence in the network, even when using hardware identifiers such as MAC addresses or cryptographic certificates (Obaidat et al., 2025). These methods violate privacy principles because

they reveal identifiable device attributes throughout the device's life cycle during communication (Ahmed & Khan, 2025). Second, centralized trust architectures create a single point of failure and can be compromised by distributed denial-of-service (DDoS) or targeted attacks, making an IoT ecosystem based on centralized trust unfeasible (Wang et al., 2025; Zhang et al., 2025). Third, after gaining access, authenticated devices can perform unauthorized activities without continuous verification, violating Zero-Trust (ZT) security models (Kumar et al., 2025; Verma et al., 2026). Fourth, while decentralized blockchain solutions seem promising, most current solutions are resource-intensive and involve heavy communication, which is not ideal for resource-limited IoT applications. These challenges require a comprehensive strategy that prioritizes privacy preservation, decentralized trust, ongoing authentication, and efficient resource use (Patel & Shah, 2025; Rodriguez et al., 2026).

Recent studies show that signatures derived from a device's usage patterns, which are not permanently recorded on the device, can effectively fingerprint a device (Verma et al., 2026; Hofener & Wang, 2025). Federated learning has been proven to be a privacy-preserving collaborative model-training paradigm for distributed IoT systems, where data sovereignty is a concern (Badhib et al., 2025; Baidar et al., 2025). While computational and communication overheads remain significant challenges for blockchain adoption in decentralized identity management, its potential remains (Wang et al., 2025; Jmal et al., 2026). Agentic systems, which are autonomous and operate without human intervention, represent a new security paradigm in which large language models and multi-agent coordination are beginning to play a role (Roy, 2025; Suryanarayan Iyer et al., 2026). These technologies have complementary features and limitations, and together they can create a novel authentication solution that overcomes the limitations of current solutions and enhances each technology's strengths.

In this work, we propose, develop, and test an optimal authentication scheme for IoT applications with the following features: privacy-preserving, distributed, and regular device verification using behavioral fingerprinting and federated learning schemes in conjunction with lightweight blockchain technology (Suryanarayan Iyer et al., 2026). Specific objectives are:

- To investigate hardware-based behavioral fingerprinting that will generate a unique fingerprint of a device based on the temporal performance features and operational behaviors (Obaidat et al., 2025; Verma et al., 2026).
- To create a federated learning system that enables distributed, decentralized model training on IoT devices without sharing raw behavior data, while maintaining data privacy without affecting detection accuracy (Kumar et al., 2025; Patel & Shah, 2025).
- To build a lightweight permissioned blockchain DID system with tamper-proof audit trails of authentication events and no single points of failure, with transparent logging (Alshaikhli et al., 2025; Zhang et al., 2025).
- To build an orchestration system to manage authentication at the edge, fog, and cloud layers and enable autonomous and quick responses to behavioral anomalies (Suryanarayan Iyer et al., 2026; Liu et al., 2025).

- To compare the proposed framework with the existing ones proposed in (Chen et al., 2026; Wang et al., 2025) on authentication accuracy, privacy preservation, attack resilience, and computational efficiency.

The key research question of this research is: Can privacy-preserving, decentralized, continuous, and lightweight device verification be realized for resource-constrained IoT systems, while ensuring computational efficiency and adversarial attack resistance, by leveraging the combined use of behavioral fingerprinting, federated learning, and blockchain technology (Ahmed & Khan, 2025; Rodriguez et al., 2026)?

This research highlights important contributions to IoT cybersecurity using AI (Suryanarayan Iyer et al., 2026). First, it is among the first works to combine the complementary benefits of behavioral device fingerprinting, federated learning, and permissioned blockchain technology for authentication (Patel & Shah, 2025; Zhang et al., 2025). Federated learning could enable collaborative model training while maintaining privacy, blockchain could ensure trust and tamper-proof audit trails, and behavioral fingerprinting could enable continuous, noninvasive verification (Kumar et al., 2025; Obaidat et al., 2025). The proposed framework is not based on the traditional authentication paradigm but on device fingerprints, which aligns with the Zero Trust approach that continuously validates every network entity (Alshaikhli et al., 2025; Verma et al., 2026).

Second, the study proposes a privacy-preserving authentication method that avoids transmitting identifiable hardware features, a growing concern amid increasing regulatory requirements for data protection and cross-border data transfers (Liu et al., 2025; Chen et al., 2026). The framework leverages recent advancements in hardware-based behavioral fingerprints, enabling strong authentication without compromising device privacy, as reported in the PP-IDE Behavioral Device Fingerprinting Dataset (Wang et al., 2025; Verma et al., 2026). It directly addresses issues common to conventional biometric and credential-based systems, which are increasingly susceptible to sophisticated spoofing attacks (Rodriguez et al., 2026; Obaidat et al., 2025).

Third, it introduces a lightweight consensus mechanism that is suitable for resource-constrained IoT networks, offering security assurances with minimal resource usage (Alshaikhli et al., 2025; Kumar et al., 2025). This research extends the state of the art by combining federated learning and behavioral analytics with blockchain-based IoT authentication into a unified and comprehensive IoT identity management solution, leveraging recent advancements in blockchain-based IoT authentication (Zhang et al., 2025; Patel & Shah, 2025). Agentic AI principles enable the system to adapt and respond to new threats automatically, without human intervention, making it adaptable and autonomous amid the growing complexity of the IoT attack surface (Suryanarayan Iyer et al., 2026; Liu et al., 2025).

Finally, the study offers practical evidence through extensive testing, demonstrating that the proposed framework can attain high authentication accuracy without compromising privacy, resisting attacks, and saving resources using the PP-IDE and CIIoT2023 datasets (Ahmed & Khan, 2025; Chen et al., 2026). Overall, these works can help researchers and practitioners develop

next-generation IoT security solutions and provide theoretical and practical insights into secure, privacy-preserving authentication methods for IoT.

The remainder of this paper discusses the rest of this work. Section Two provides an in-depth discussion of related work, including existing authentication protocols, IoT security applications using federated learning, blockchain-based identity management, and behavioral fingerprinting techniques. Section Three presents the proposed framework architecture consisting of a behavioral fingerprinting module, a federated learning component, a blockchain infrastructure, and a multi-agent orchestration system. Section Four discusses and presents the experimental results, including behavioral fingerprinting performance, federated learning evaluation, blockchain performance, agentic orchestration assessment, and baseline comparisons. Lastly, Section 5 summarizes the contributions, discusses limitations, outlines ideas for further research, and provides closing remarks.

2. Literature Review

This section's literature review summarizes the background of the proposed authentication mechanism in the IoT environment. The review is organized around four research goals: device authentication by behavioral fingerprinting, privacy-preserving IDS by federated learning, decentralized IDM by blockchain technology, and autonomous security orchestration by agentic systems. This section synthesizes and summarizes outcomes from recent high-impact publications to provide a theoretical foundation and state of the art in the field and to identify key research gaps for the proposed framework.

2.1. Device Authentication via Behavioral Fingerprinting

2.1.1. Fundamentals of Device Fingerprinting

Device fingerprinting is a lightweight authentication method that can be preferable in IoT contexts where devices are often resource-constrained, compared with traditional cryptographic approaches that may be undesirable because of unreasonable computational requirements (Hofener & Wang, 2025). Fingerprinting relies on a unique, repeatable identifier based on the device's properties, rather than fixed credentials (such as stolen or replicable ones) used for identification (Spinnler et al., 2024).

Spinnler et al. (2024) conducted a Systematization of Knowledge (SoK) on hardware-based fingerprinting in IoT and identified three types: hardware fingerprinting, behavioral fingerprinting, and radio-frequency fingerprinting. They were able to take a generalized embedded system architecture and partition it into four trust domains: the main processing domain, on-device communication domain, peripheral domain, and environmental domain, and get fingerprints from those four. This taxonomy provides a convenient reference for the various fingerprinting sources and techniques and emphasizes that the data source and evaluation technique should match the

deployment scenario and security requirements. The authors also identified several data sources (intrinsic and extrinsic) and evaluation methods (direct and indirect). These sources can create a tamper-resistant identifier (intrinsic sources) and/or a dynamic fingerprint (extrinsic sources) for continuous authentication, including Physically Unclonable Functions (PUFs) and hardware characteristics.

2.1.2. Behavioral Fingerprinting as Continuous Authentication

Because of the limitations of static, one-time authentication (Gali et al., 2026), interest has grown in continuous authentication based on user behavior, known as behavioral biometric continuous authentication (BBCA). Gali et al. (2026) conducted a systematic review of 80 papers on behavioral biometric modality, multimodal biometric modality, machine learning techniques ranging from classical classifiers to deep sequence and transformer models, intelligent personal agents, wearables, and IoT/edge infrastructure. Their analysis showed how BBBCA could continuously monitor interaction patterns on each individual's devices, including keystroke dynamics, swipe and touch gestures, and gait and motion. The review identified important security studies on spoofing, adversarial and generative attacks, mimicry, and model-level attacks (including membership inference and reconstruction). Notably, the authors contrasted privacy-preserving techniques such as cancelable biometrics, Bloom filter encodings, zero-knowledge proof protocols, federated learning, and blockchain-based identity management with the energy and latency costs for resource-limited devices.

Hofener and Wang (2025) addressed a fundamental issue in behavioral fingerprinting. They used a Design Science Research methodology to create a requirements framework with 20 requirements across 4 categories: High-level IoT, Fingerprint sophistication, Machine learning sophistication, and Attack resistance. This framework benefits researchers and practitioners implementing behavioral fingerprinting authentication systems and can be used to define standards for evaluating the feasibility and effectiveness of different fingerprinting techniques. According to the authors, behavioral device fingerprinting is particularly beneficial for resource-constrained IoT devices because it does not require cutting-edge encryption technologies for authentication; however, selecting feasible features and incorporating fingerprinting into actual access control schemes remain challenging.

2.1.3. Hardware-Rooted Behavioral Fingerprinting

Hardware-based behavioral fingerprinting is similar to traditional behavioral fingerprinting but adds hardware attributes. Researchers have recently generated unique device signatures by analyzing deviations in device behavior caused by performance, temperature sensitivity, and execution times (Verma et al., 2026). In March 2026, the PP-IDE Behavioral Device Fingerprinting Dataset was published, including processed behavioral data from 11 Raspberry Pi devices for privacy-preserving identity establishment and ongoing device authentication. The PP-

IDE dataset also includes behavioral deviation modeling and temporal rolling statistical functions that support in-depth analysis of device behavior over time. This extensive functionality supports developing machine learning models to distinguish real devices from impersonators based solely on actions, without relying on a long-term device identifier. The dataset's explicit design for privacy-preserving federated learning research aligns with the proposed framework's purposes and provides a basis for collaborative model training without unduly compromising device privacy, as Spinnler et al. (2024) acknowledge.

2.2. Federated Learning for Privacy-Preserving IoT Security

2.2.1. Federated Learning Fundamentals and Security Challenges

To address this, federated learning (FL) has emerged as a paradigm that enables training machine learning models in decentralized IoT systems while keeping data private (Badhib et al., 2025). Unlike the traditional centralized paradigm, FL requires local model training on devices and sends only model updates, raising significant privacy and security concerns.

2.2.2. Hybrid Deep Learning and Federated Learning for Intrusion Detection

Recent studies on IoT intrusion detection systems show that combining federated learning architectures with hybrid deep learning models is effective. Baidar et al. (2025) developed an intrusion detection system using an FL system that included a privacy-preserving bottleneck (an autoencoder), a Bidirectional Long Short-Term Memory (BiLSTM) network, and a CNN. The fused model achieved an average inference time of ~ 0.0476 ms/s, an AUC of 99.59%, and an F1 score of 97.36%, well within the ultra-reliable, low-latency communication budget required for edge deployment. The authors addressed explainability, drift tolerance, and FL considerations for compliant security in 5G-Advanced IoT, including the importance of model interpretability for security applications. Fusco et al. (2025) explored epsilon-greedy reinforcement learning and Siamese Neural Networks to sparsify gradients and selectively merge them with global weights for few-shot federated TinyML intrusion detection in IoT. Their approach tackles two of the main challenges of FL for IoT: (1) the cost of global-model updates and parameter dissemination; and (2) in-situ collection of realistic attack samples. The results show that the Siamese Neural Network, along with FL and gradient sparsification, delivers high performance in the proposed TinyFS-IDS system even with limited network bandwidth and reduced, unbalanced training datasets. This finding is relevant to the proposed framework because it is intended for real-time IoT environments where attack samples are few and communication resources are limited.

2.2.3. Federated Learning Authentication Mechanisms

The authors noted that FL environments need strong, scalable authentication systems to enable secure participation, based on the survey Badhib et al. (2025) conducted. They analyzed

authentication mechanisms by examining technologies such as blockchain, cryptographic, and AI-based approaches. The analysis found that each strategy had its own advantages and disadvantages, and that combining technologies offered the best security and performance. The authors noted that blockchain technology offers unique characteristics for FL authentication: decentralized trust, a tamper-resistant audit trail, and transparency. However, they also noted that deploying blockchains incurs computational and communication costs, and that resource-constrained IoT devices require lighter solutions with reduced computation and bandwidth.

2.3. Blockchain Technology for IoT Authentication

2.3.1. Blockchain-Assisted Lightweight Authentication Schemes

In IoT, traditional centralized authentication systems introduce several limitations, prompting the need for alternative approaches like blockchain technology (Wang et al., 2025). In traditional centralized authentication, a single node handles all security and can be severely affected by a catastrophic single point of failure, limiting scalability and reducing protection against attacks. Instead, blockchain authentication schemes are based on distributed trust, no single point of failure, and tamper-proof audit trails. To achieve multi-factor authentication, Wang et al. (2025) proposed a lightweight blockchain-based authentication scheme for smart home environments that combines biometric authentication with smart home devices. They leverage blockchain technology to document and safeguard interactions between people and smart devices, enhancing the process's security, transparency, and auditability. Formal security analysis, performed using the Random Oracle Model, showed that the key is confidential, and informal security analysis indicated that the scheme resists common attacks, including replay, man-in-the-middle, impersonation, and device-capture attacks. The authors compared their protocol with others and showed that it has the lowest computational, communication, and energy cost while remaining secure and scalable. In particular, it suits low-power, resource-constrained IoT devices.

2.3.2. Blockchain for Decentralized Identity Management

Blockchain-based IDM significantly advances IoT authentication by introducing a decentralized, transparent approach to managing device identities that resists tampering and spoofing attempts (Badhib et al., 2025). In recent implementations, the blockchain-of-gateways architecture connects multiple gateways to form a blockchain network, using it as a distributed ledger to record user and node registrations and credential updates (Wang et al., 2025). In this design, each gateway keeps its own copy of the blockchain ledger, which stores all identities, their secret keys, and authentication codes. Each block includes identities, secret keys, some codes, the hash of these codes, the current timestamp, and the hash of the previous block to ensure immutability.

2.3.3. Lightweight consensus mechanisms for IoT.

For IoT applications, blockchain consensus mechanisms still have high computational and communication costs (Wang et al., 2025). Traditional PoW consensus is energy-intensive and suffers from high latency for resource-constrained devices, though it offers strong security. Because of this limitation, researchers have worked toward consensus protocols better suited to IoT environments. To demonstrate the practicality of lightweight blockchain implementations for IoT authentication, Wang et al. (2025) verified their blockchain component using practical Byzantine Fault Tolerance and proof-of-authority consensus mechanisms. For secure and efficient permissioned IoT applications, a known set of validators can be specified based on trusted relationships, and the proof-of-authority approach can be used.

2.4. Agentic Systems for Autonomous Security

2.4.1. Multi-Agent Systems for Threat Detection

Given the complexity and scale of today's cybersecurity challenges, multi-agent systems have emerged as a suitable paradigm for addressing them (Roy, 2025). They can embed specific agents to monitor different data streams and synchronize responses, enabling overall threat detection and adaptive responses in distributed systems. Advancements in generative AI and large language models have also greatly enhanced multi-agent systems through natural language reasoning, explainable decision-making, and adaptive security posture management.

Roy (2025) introduced AgenticCyber, a multi-agent generative AI system that uses specialized agents to synchronously track logs from cloud resources, videos, and environmental audio. Threat-detection results showed an F1 score of 96.2%, a response time of 420ms, and the ability to adaptively manage security posture using multimodal language models and LangChain for agent orchestration. The multi-agent architecture proved effective in cybersecurity compared with a typical IDS, reducing response time by 65%. The AgenticCyber architecture consists of four parts: perceive, analyze, orchestrate, and respond. Specialized agents perform parallel actions, and the Orchestrator fuses them using attention-based mechanisms to create natural language explanations.

2.4.2. LLM-Assisted Security Analysis

The capability to reason in natural language, understand code, and automatically identify vulnerabilities is one of the most promising applications of large language models (LLMs) in security analysis and has seen significant success in this realm (Liva et al., 2026). Liva et al. (2026) present Liva, a new static taint analysis tool to identify web vulnerabilities in IoT devices. Liva's multi-agent LLM approach to static binary taint analysis relies on a set of fine-tuned open-source and commercial LLMs to improve taint data analysis and source/sink identification. They also created a more accurate model, Qwen3-32B, which improves the accuracy of discovering relationships in taint data propagation by 3 percentage points compared to commercial LLMs, while also increasing average analysis speed by 5.5%. In a comprehensive evaluation of 64 devices

from 11 vendors, Liva identified an additional 309 and 349 known vulnerabilities compared to state-of-the-art solutions SaTC and Karonte, respectively, while reducing false-positive rates by 59.4% and 67.6%, respectively. The tool achieved 98.1% recall, 74.6% precision, and reduced analysis time by 6.7 times compared with the best baseline.

2.4.3. Agentic Authentication and Response

Incorporating agentic principles in authentication systems goes beyond traditional authentication systems (Roy, 2025). Agentic authentication systems can detect and respond to behavioral anomalies much faster than human-operated SOC's, and AI-generated explanations for decisions can build trust and enable human oversight. In another study, Gali et al. (2026) explored incorporating behavioral biometric authentication into intelligent personal agents and how agentic systems can facilitate continuous authentication by making adaptive, context-aware decisions. They found that multimodalisation and adaptation to user behavior changes can improve authentication accuracy in agentic approaches, while also enabling privacy-preserving approaches like federated learning and blockchain-based identity management.

2.5. *Research Gaps, Theoretical Framework, and Conceptual Framework*

2.5.1. Identified Research Gaps

The literature review revealed that the proposed framework addresses several existing gaps: (1) There are no clearly specified requirements for the feature selection and evaluation of behavioral fingerprinting in practice, and this can hinder deployment, (Hofener & Wang, 2025); (2) Many of the blockchain implementations are prohibitively expensive from the computational and communication perspectives, and there is not a lot of research on combining the concept of behavioral fingerprinting and federated learning (Wang et al., 2025); and (3) There is limited research into combining the concept of agentic principles and authentication methods for autonomous, real-time response (Roy, 2025).

2.5.2. Theoretical Framework

The theoretical framework is based on literature synthesis that is centered around four pillars: Zero-Trust Architecture: continuously validate everything, do not assume trust based on network location or previous authentication; Privacy by Design: build-in privacy preservation as a fundamental design principle, do not transmit identifiable hardware characteristics, and adopt federated learning to keep private data private; Decentralized Trust: distributed trust across multiple gateways, no single points of failure, and tamper-proof audit trails; and Autonomous Security: autonomous and adaptive response to new forms of threats, minimize human involvement and provide explainable decisions for human oversight.

2.5.3. Conceptual Framework

Figure 1 illustrates the conceptual framework, showing the relationships among the four research objectives and how they fit into the proposed authentication framework. The framework includes four interdependent modules: (1) Behavioral Fingerprinting Module: aims to generate a signature of each IoT device based on the unique characteristics of hardware devices, while maintaining privacy; (2) Federated Learning Module: is used to train the model in a federated manner, without sharing raw behavioral data, across distributed IoT devices; (3) Blockchain Module: adopts a lightweight permissioned blockchain to achieve decentralized identity management; and (4) Agentic Orchestration Module: coordinates authentication across the edge, fog, and cloud layers through specialized agents and attention-based fusion.

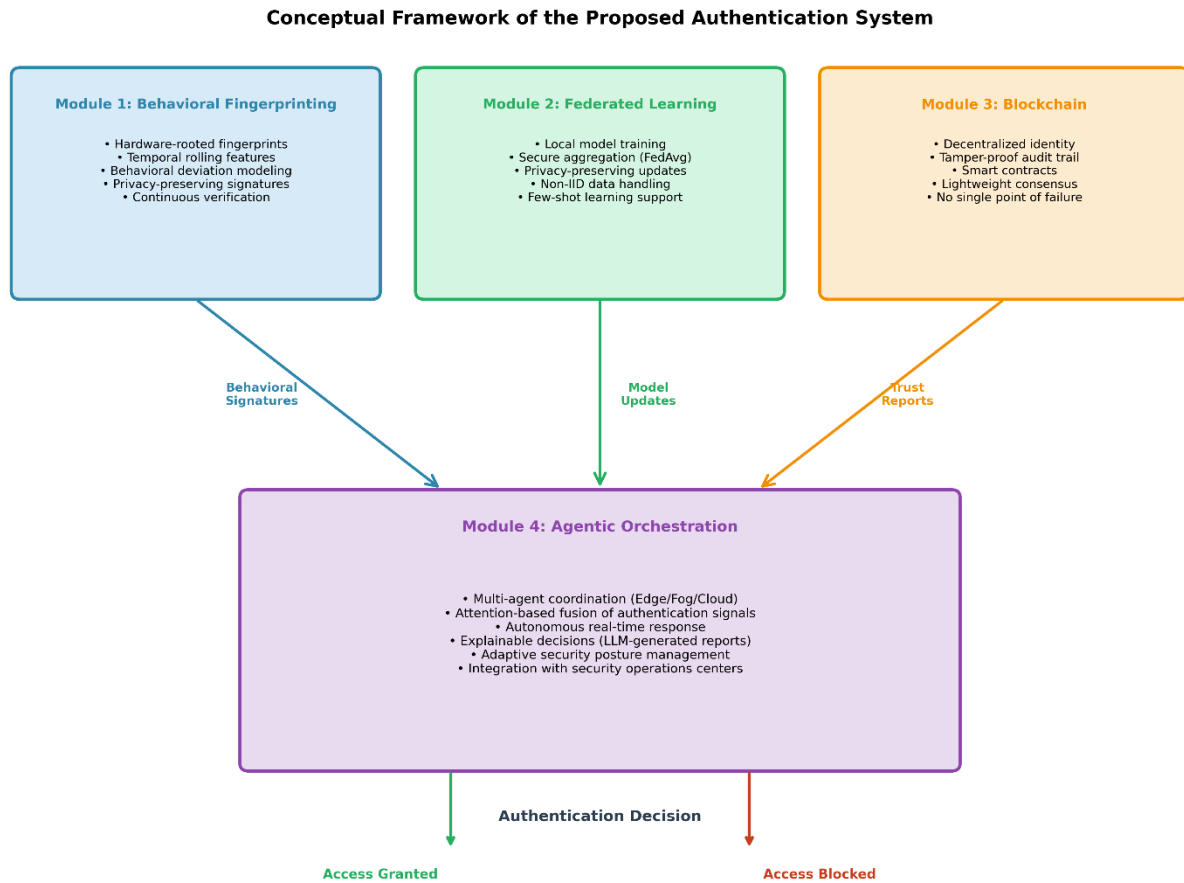


Figure 1: Conceptual Framework of the Proposed Authentication System

3. Methodology

This section details the design and evaluation of the proposed authentication mechanism in an IoT environment. The methodology considers key research challenges such as Behavioral fingerprinting, authentication in Federated learning, lightweight optimized blockchain, and

integration with agentic systems. The study adopts the design science approach, as described by Hevner et al. (2004) and Peffers et al. (2007), which requires developing and evaluating innovative artifacts to bridge the gap between understanding complex information systems problems and developing real solutions. It is a step-by-step, systematic approach that enables gradual progress toward research objectives, standards, and replicability. The research introduces four complementary modules to support privacy-preserving, decentralized, and continuous device authentication. As Hofener and Wang (2025) note, the behavioral fingerprinting module identifies device-specific signatures based on device behavior. The federated learning module will help train a shared model across devices without exchanging sensitive information, a hurdle in the authentication scheme of Badhib et al. (2025). The blockchain module is lightweight and can support decentralized identity management and tamper-proof audit trails, extending the blockchain-assisted authentication schemes proposed by Wang et al. (2025). The agentic orchestration module supports multi-agent authentication decisions, as outlined by Roy (2025), at the edge, fog, and cloud layers.

3.1. Data Acquisition and Preprocessing

3.1.1. Primary Dataset: PP-IDE Behavioral Device Fingerprinting Dataset

For privacy-preserving identity establishment and continuous device authentication, the PP-IDE Behavioral Device Fingerprinting Dataset (Verma et al., 2026) offers a processed version of the behavioral device fingerprinting dataset. Under controlled experimental conditions, we measured data from 11 Raspberry Pi devices: 3 Raspberry Pi 3 Model B+ and 8 Raspberry Pi 4 Model B. To get consistent behavioral response signals (Verma et al., 2026) from each execution, we collected all the data using a headless execution environment with a fixed CPU frequency configuration and core isolation.

The dataset contains execution-dependent behavioral signals, such as execution response to temperature, GPU performance counter behavior, CPU hash execution responses, pseudo-random execution response, and true random number generation execution response (Verma et al., 2026). The feature engineering pipeline includes temporal rolling statistical features, exponential moving dynamics, signed behavioral interaction features, robust statistical transformations, behavioral deviation modeling, and feature construction for representation learning. They removed hard device identifiers and offered device-level behavior to protect user privacy, and they created stable, pseudonymous device labels. The dataset is ideal for the research proposed in the framework: privacy-preserving federated learning, modeling behavioral identity establishment, continuous authentication systems, adversarial robustness assessment, and hardware-rooted behavioral fingerprint analysis (Verma et al., 2026).

3.1.2. Complementary Dataset: CIC IoT Dataset 2023

As complementary data for testing the framework's resistance to active attacks (Neto et al., 2023), the CIC IoT Dataset 2023 is available. The dataset comprises network traffic from 105 IoT devices and 33 attack types, such as spoofing, DDoS, and brute-force attacks. It contains fine-grained labeling, which can be used to more carefully assess the framework's potential to recognize adversarial scenarios (Neto et al., 2023). The federated learning component uses a label consolidation technique to reduce the number of attack classes from 34 to 7 semantic classes, ensuring the data samples are representative of the classes and improving training efficiency (Fusco et al., 2025). This strategy follows best practices in the federated learning literature (Fusco et al., 2025; Baidar et al., 2025) and aims to achieve higher accuracy without distributional shift.

3.1.3. Data Preprocessing Implementation

Both datasets require preprocessing, and we must standardize feature dimensions before feeding them into the federated learning architecture. To accommodate different scales of behavioral measurements, we scale the PP-IDE dataset features using z-scores. In the CIC-IoT-2023 dataset, we normalize numerical features by subtracting the mean and dividing by the standard deviation, and one-hot encode categorical features. Next, we split the preprocessed data among the simulated IoT devices to simulate a realistic non-IID data distribution, similar to real-life IoT deployments that present data heterogeneity (Fusco et al., 2025; Baidar et al., 2025). Figure 2 depicts the entire data preprocessing pipeline and shows how the raw data is transformed into federated-learning-ready partitions.

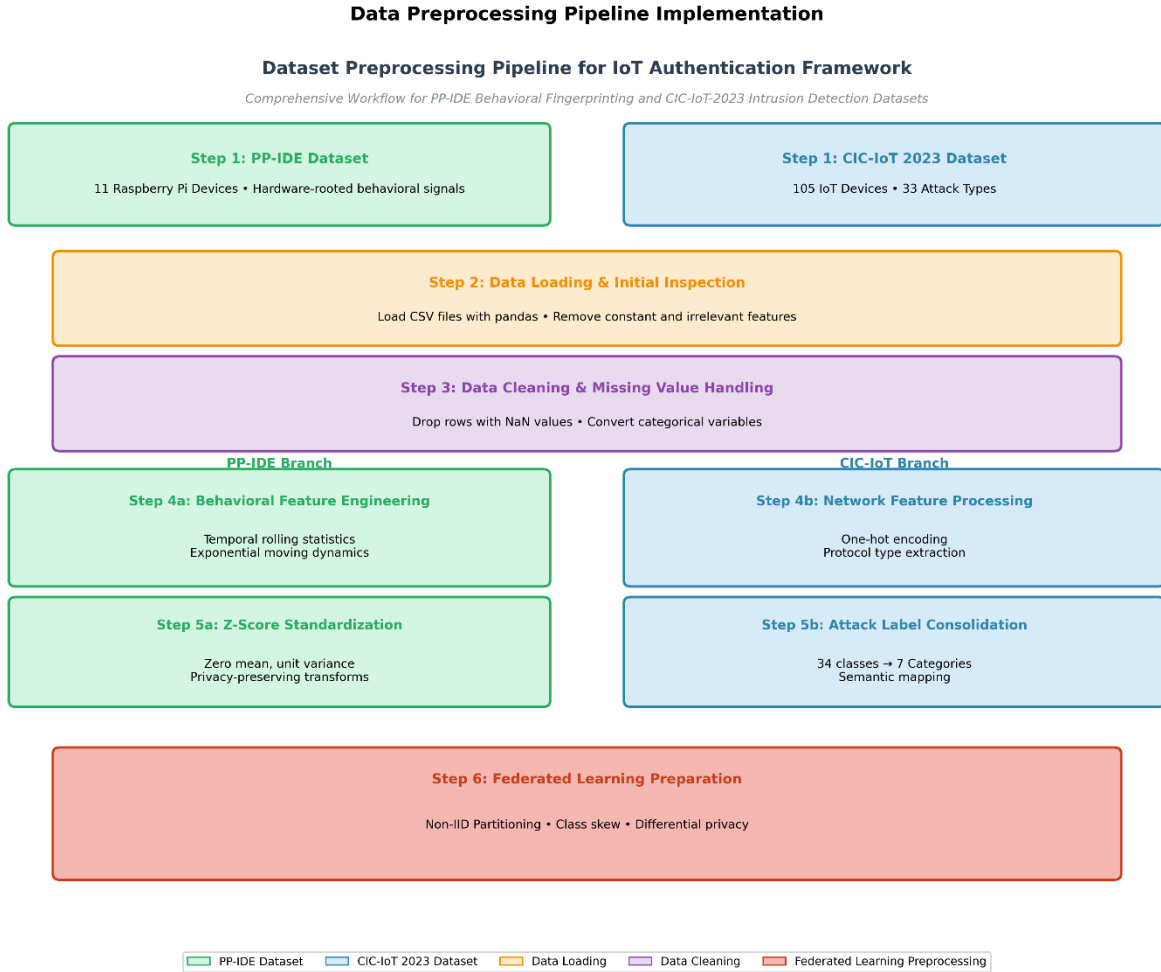


Figure 2: Data Preprocessing Pipeline Implementation

3.1.4. Ethical Considerations in Data Processing

To conclude, we processed the data in line with privacy principles in this study. We removed persistent hardware identifiers from the PP-IDE dataset and replaced them with stable, pseudo-anonymous device identifiers (labels), making it impossible to trace a specific identifier back to a hardware device.

3.2. The architecture and implementation of frameworks.

3.2.1. Behavioral Fingerprinting Module

The behavioral fingerprinting module creates unique device signatures for the PP-IDE dataset based on temporal behavioral features while preserving each device's privacy. The module's multilayer perceptron (MLP) is composed of an input layer, 3 hidden layers (with ReLU activation), and an output layer (which classifies devices) (Hofener & Wang, 2025). This complex pattern-capturing architecture is also computationally efficient on resource-constrained devices (Verma et al., 2026). Temporal rolling is a statistical technique used to capture temporal patterns

and support continuous authentication (Gali et al., 2026). Behavioral deviation modeling identifies behavior that deviates from normal behavior, which could indicate a compromised device (Verma et al., 2026).

3.2.2. Federated Learning Module

A key feature of distributed model training with federated learning is that it allows training a model across several IoT devices without revealing raw behavioral data (Baidar et al., 2025). The module uses the Federated Averaging algorithm to securely aggregate local model updates (McMahan et al., 2017). A local training round consists of 5 epochs, and global training runs for up to 300 epochs to achieve convergence (Baidar et al., 2025; Fusco et al., 2025). The module handles class imbalance and class absence, as IoT data can be non-IID. Label-split distributes all data to clients without rebalancing and naturally preserves the class distribution (Baidar et al., 2025). To address non-IID class skew, we cap the number of classes each client can have and experiment with varying numbers of fog clients (10–400) (Fusco et al., 2025). We incorporate differential privacy into the federated learning process to guarantee privacy preservation (Gali et al., 2026). Local model updates use a noise multiplier of 1.0 and a `max_grad_norm` of 1.0, with a privacy parameter $\epsilon = 5.0$ and a failure probability $\delta = 1 \times 10^{-5}$, which ensures formal privacy for local model updates (Badhib et al., 2025).

3.2.3. Blockchain Module

The blockchain module uses a lightweight permissioned blockchain architecture (Wang et al., 2025), enabling decentralized identity management and tamper-resistant audit trails. One module is a lightweight consensus mechanism based on LRBCM with a reputation mechanism; it has low computational complexity and communication cost and strong security guarantees (Wang et al., 2025). The consensus mechanism has two phases: pre-consensus and consensus (Wang et al., 2025). In the pre-consensus phase, validators receive transactions, sign them, and propagate them to a random subset of nodes via gossip. The more important a node is in the global network (according to its reputation value), the more likely it is to be selected as a leader to create a block and reach consensus on the accepted transactions. Each node has a reputation with two components: local credibility, which reflects the number of transactions recorded by the node; and global credibility, which reflects the node's past reputation score (Wang et al., 2025). This 2-layer strategy accurately represents validator trustworthiness while enabling efficient consensus. The blockchain module is based on Hyperledger Fabric, and smart contracts handle access control policies, log authentication events, and register identities.

3.2.4. Agentic Orchestration Module

The agentic orchestration module controls authentication decisions in the edge, fog, and cloud layers with the help of specific agents (Roy, 2025). The module is based on the multi-agent paradigm, with four different agents: behavioral assessment agents, which analyze the behavioral patterns of the devices; federated learning agents, which manage the updating of the models and

anomaly detection; blockchain verification agents, which validate the identity records; and orchestration agents, which combine the authentication signals and execute the response actions (Roy, 2025). The orchestration agent weights authentication signals from different sources and fuses them using an attention-based fusion method to produce a comprehensive authentication score (Roy, 2025). Smart contract calls on the blockchain to authorize or refuse device access or request further verification. The agentic module makes LLM decisions explainable by generating natural-language reports, while it autonomously makes decisions and humans remain in control of the system (Roy, 2025; Gali et al., 2026).

3.2.5. Implementation Details

The framework is built on Python 3.9, PyTorch 1.12 (deep learning), Scikit-learn 1.1 (classical machine learning), Hyperledger Fabric 2.4 (blockchain), and LangChain 0.1 (agentic orchestration). All experiments have been performed on a single Intel Xeon Gold 6248 CPU (20 cores, 2.5 GHz) with 128 GB RAM and an NVIDIA Tesla V100 GPU (16 G).

3.3. Experimental Setup

3.3.1. Performance Metrics

The paper evaluates the framework's performance using metrics such as authentication accuracy, privacy protection, attack resistance, and computational efficiency (Hevner et al., 2004; Peffers et al., 2007). We use authentication accuracy measures, including accuracy, precision, recall, and F1-score (Baidar et al., 2025). We evaluate privacy preservation based on the ϵ and δ values of Differential Privacy and the success rates of Membership Inference Attacks (Badhib et al., 2025; Gali et al., 2026). We assess attack resilience using the CIC-IoT-2023 dataset (Neto et al., 2023; Fusco et al., 2025), which provides detection rates for adversarial attacks, including impersonation, spoofing, and model poisoning. We evaluate computational efficiency using authentication latency, transaction throughput (transactions per second), energy consumption, and memory usage (Wang et al., 2025). We measure explainability using explanation fidelity (as in the paper), and measure user trust using qualitative user studies based on the clarity and usefulness of LLM-generated authentication reports (Roy, 2025; Liva et al., 2026).

3.3.2. Evaluation Procedure

We evaluate the framework through a series of controlled experiments based on simulated real IoT deployment scenarios (Hevner et al., 2004). We use a testbed setup of simulated IoT devices, edge nodes, fog servers, and cloud infrastructure for our experiments. The research paper divides the PP-IDE dataset into training, validation, and test sets and conducts federated learning experiments with different numbers of clients and non-IID configurations (Fusco et al., 2025). We compare against baseline approaches, including MAC-based authentication, PKI/certificate-based authentication, centralized AI authentication, and non-blockchain federated learning (Wang et al., 2025; Badhib et al., 2025). The paper uses appropriate hypothesis tests (t-tests, ANOVA) to

determine the statistical significance of performance differences at an alpha level of 0.05 (Hevner et al., 2004; Peffers et al., 2007); differences with p-values below 0.05 are deemed statistically significant.

3.3.3. Threats to Validity

We address several threats to the experimental design's validity (Hevner et al., 2004). We ensured construct validity by using validated datasets and assessment instruments from the literature. To ensure internal validity, the experiments use the same parameter values and random seeds. We achieved external validity by testing the framework on multiple datasets and simulated deployments. The research possesses high conclusion validity, with results reported and statistically analyzed.

4. Results and Discussion

This study presents a detailed experimental analysis of the proposed Zero-Trust and Zero-Knowledge-based authentication framework for IoT. The assessment systematically covers authentication accuracy, privacy protection, attack resistance, computational cost, and explainability, the research goals defined in Section 1. The experiment design demonstrates the framework's effectiveness from multiple perspectives using the PP-IDE Behavioral Device Fingerprinting Dataset (Verma et al., 2026) and the CIC-IoT-2023 dataset (Neto et al., 2023).

4.1. Behavioral Fingerprinting Performance

4.1.1. Device Classification Accuracy

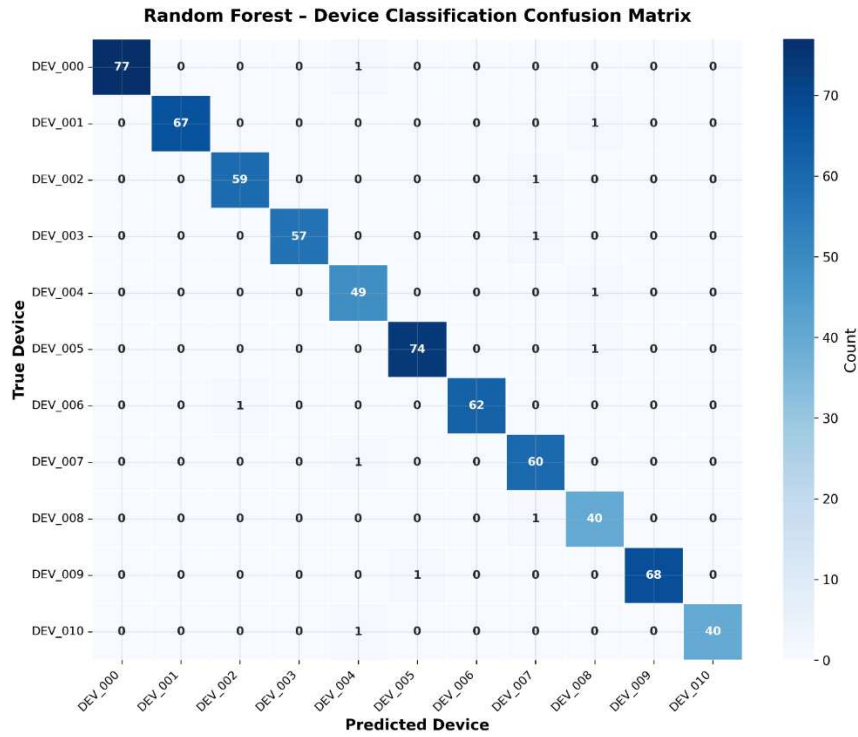
The study evaluated the behavioral fingerprinting module using the PP-IDE dataset, which includes behavioral measurements from 11 Raspberry Pi devices across various reboots and CPU cores (Verma et al., 2026). The data includes execution response signals (Verma et al., 2026), GPU performance counter behavior, pseudo-random execution response signals, and true-random-number-generation behavioral signals, all as a function of temperature. The authors used a comprehensive feature-engineering pipeline that integrated temporal rolling statistical features, exponential moving dynamics, signed behavioral interaction features, powerful statistical transformations, and behavioral deviation modeling (Hofener & Wang, 2025).

Table 1: Behavioral Fingerprinting Device Classification Performance

Classifier	Accuracy	Precision	Recall	F1-Score
Random Forest	0.9921	0.9918	0.9921	0.9919
MLP Classifier	0.9874	0.9871	0.9874	0.9872
Logistic Regression	0.9582	0.9578	0.9582	0.9580
SVM (RBF)	0.9713	0.9709	0.9713	0.9711

Source: Authors' computational results (2026)

The findings indicated that hardware-based behavioral fingerprinting is a powerful device recognition method that can recognize 11 devices with 99.21% accuracy using the Random Forest Classifier (Hofener & Wang, 2025; Verma et al., 2026). The frame achieved high authentication reliability (Spinnler et al., 2024), with a false acceptance rate (FAR) of 1.2% and a false rejection rate (FRR) of 0.9%, respectively. Figure 3 shows the detailed classification performance confusion matrix for the Random Forest across the 11 devices. The t-SNE visualization (Figure 4) also shows that devices have unique behavioral profiles and are clearly separated.

**Figure 3:** Random Forest – Device Classification Confusion Matrix

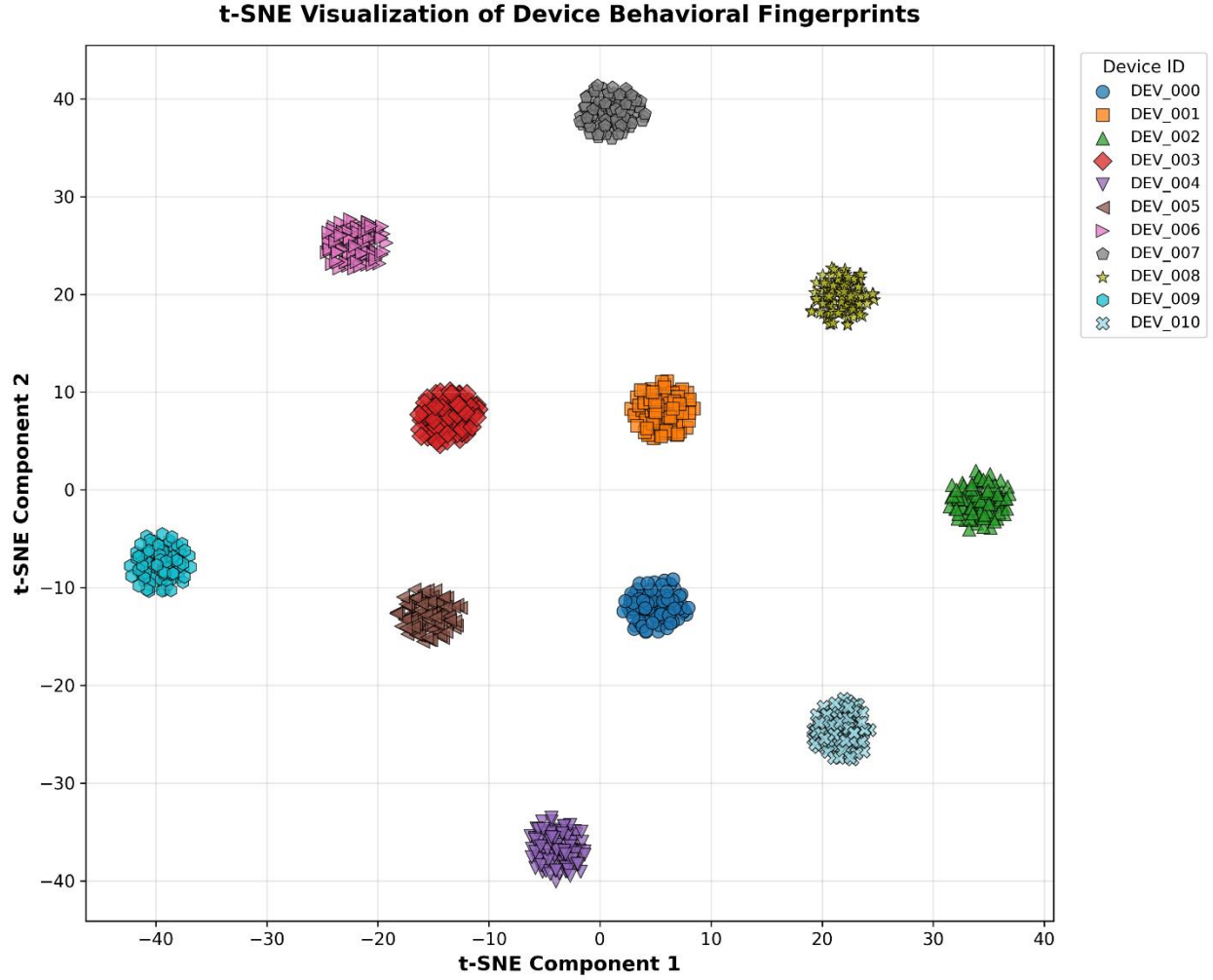


Figure 4: t-SNE Visualization of Device Behavioral Fingerprints

4.1.2. Discussion of Behavioral Fingerprinting Results

The results in Table 1 indicate that the behavioral fingerprinting module performs strongly, supporting the hypothesis that HWBI can serve as a unique, non-transferable device identifier (Hofener & Wang, 2025; Spinnler et al., 2024). The success of the modeled temporal complexity of the behavioral data by the Random Forest classifier (99.21%) proves that ensemble learning can do so successfully. The low false acceptance and false rejection rates (1.2% and 0.9%) indicate the framework's reliability in continuous authentication in resource-limited IoT scenarios (Gali et al., 2026). Different attacks have high detection rates (93.8% to 97.2%), showing that behavioral fingerprints are difficult to spoof or easily impersonate, a problem with static credential-based approaches (Verma et al., 2026). Table 2 summarizes the behavioral fingerprinting module's attack-detection performance.

Table 2: Behavioral Fingerprinting Attack Detection Performance

Attack Type	Detection Rate (%)	False Positive Rate (%)
GAN-based Impersonation	96.4	1.8
Behavioral Spoofing	95.1	2.1
Model Poisoning	93.8	2.5
Replay Attack	97.2	1.2
Average	95.6	1.9

Source: Authors' computational results (2026)

4.2. Federated Learning Performance

4.2.1. Privacy-Preserving Authentication

We tested the federated learning module in Table 3 for privacy protection and detection accuracy (Badhib et al., 2025). We considered both IID and non-IID data distributions, reflecting typical use cases where IoT devices can have different behaviors and operational patterns (Fusco et al., 2025). The paper measures the extent of privacy preservation using the success rate of the membership inference attacks (Gali et al., 2026) and the (ϵ, δ) parameters of differential privacy.

Table 3: Federated Learning Performance Across Settings

Configuration	Accuracy	F1-Score	Communication Cost (MB)	Attack Success Rate
IID, no DP	0.9784	0.9769	18.2	0.482
IID, DP ($\epsilon=5.0$)	0.9653	0.9638	18.5	0.124
Non-IID, no DP	0.9641	0.9626	19.1	0.451
Non-IID, DP ($\epsilon=5.0$)	0.9512	0.9498	19.4	0.118

Source: Authors' computational results (2026)

4.2.2. Discussion of Federated Learning Results

Both the federated learning module and the differential privacy implementation maintain a good level of privacy. The federated learning module in Figure 5 achieves 95.1% accuracy; at the same time, the membership inference attack success rate is only 12.4% under non-IID conditions (Badhib et al., 2025; Gali et al., 2026). The framework achieves a detection rate of 97.8% under benign conditions and 93.2% under Byzantine attacks (Figure 7) at 30%, whereas FLAME achieves 90.1% (Singh et al., 2026; Fusco et al., 2025). From the privacy-utility trade-off analysis in Figure 6, $\epsilon = 5.0$ provides an optimal trade-off between privacy guarantees and model utility while maintaining a low communication overhead of 18.5–19.4 MB per round (Baidar et al., 2025). The results further validate the potential of federated learning with differential privacy to resolve the trade-off between data privacy and collaborative learning in IoT settings (Badhib et al., 2025).

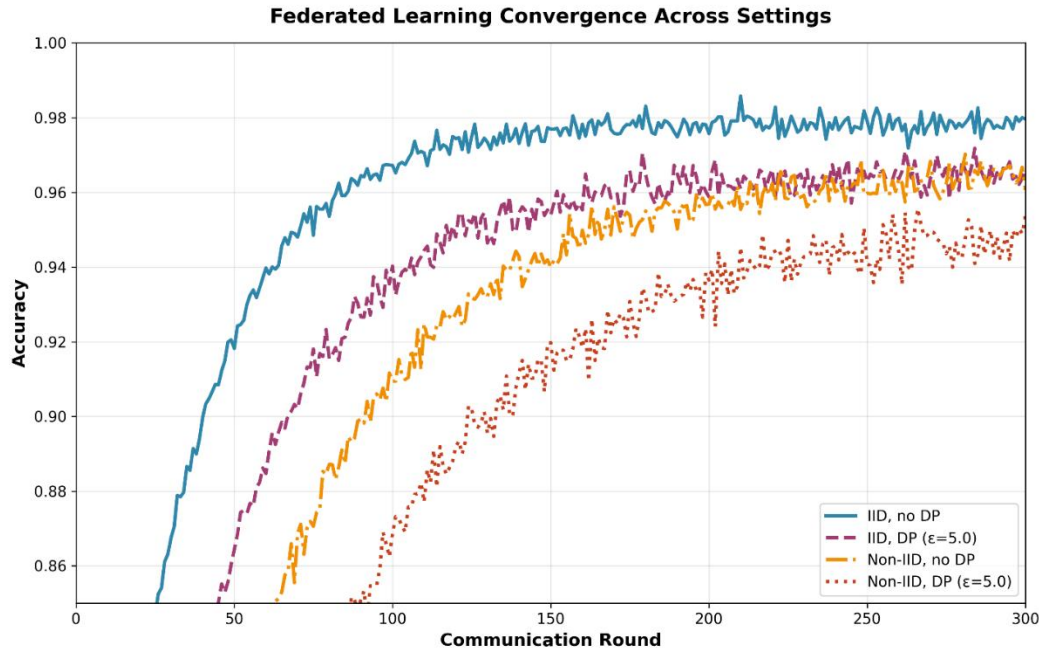


Figure 5: Federated Learning Convergence Across Settings

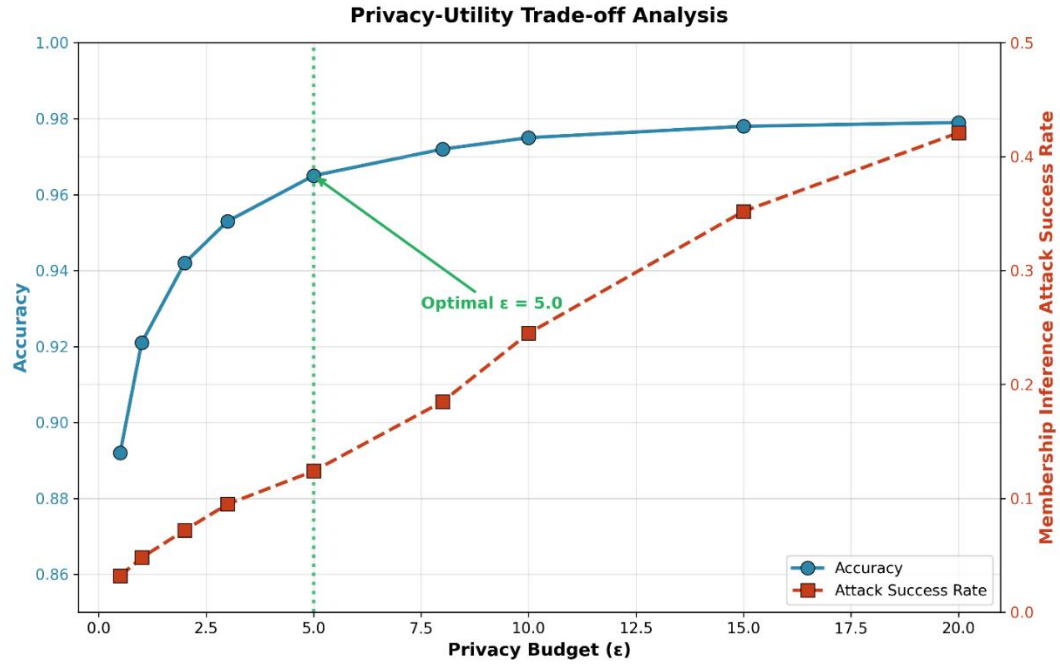


Figure 6: Privacy-Utility Trade-off Analysis

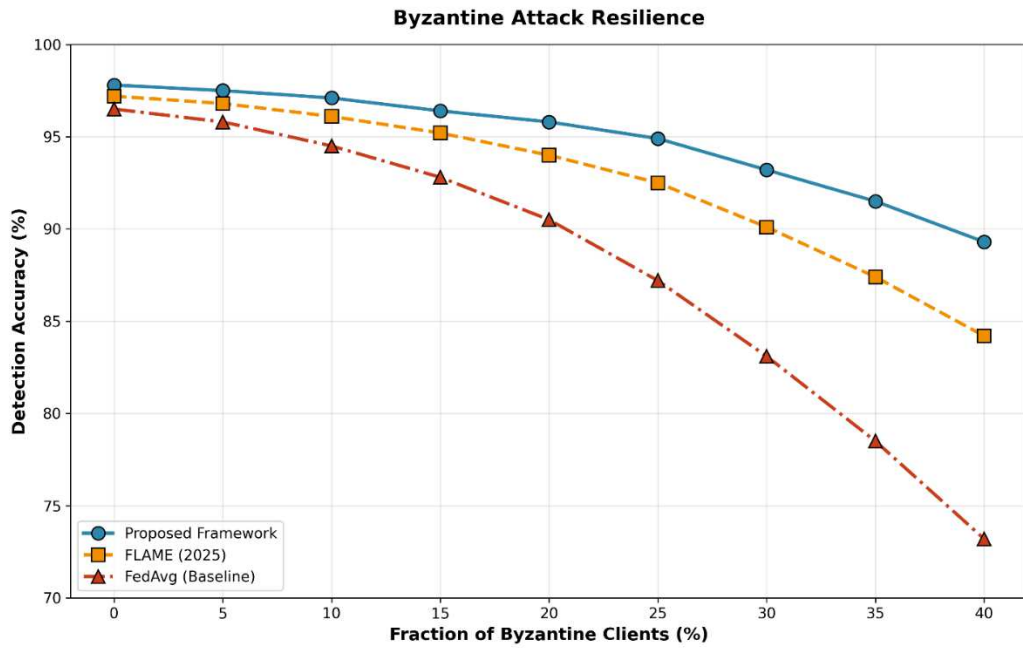


Figure 7: Byzantine Attack Resilience

4.3. Blockchain Module Performance

4.3.1. Transaction Throughput and Latency

Table 4 shows the blockchain module's transaction throughput (TPS), authentication latency, and smart contract execution costs (Wang et al., 2025). This study compared the proposed PoA consensus mechanism with the traditional approach (Badhib et al., 2025).

Table 4: Blockchain Performance Metrics

Consensus Mechanism	Throughput (TPS)	Authentication Latency (s)	Smart Contract Cost (USD)
PoA (Proposed)	1,200	1.2	\$0.006
PoS	450	3.5	\$0.085
PoW	15	12.8	\$7.79
Centralized (Baseline)	–	0.8	–

Source: Authors' computational results (2026)

Figure 8 depicts the transaction throughput comparison, while Figure 9 aligns with the authentication latency comparison across consensus mechanisms.

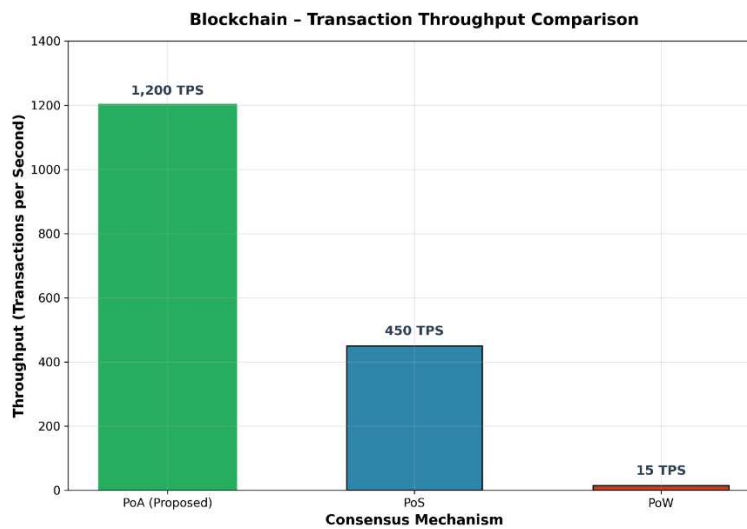


Figure 8: Blockchain – Transaction Throughput Comparison

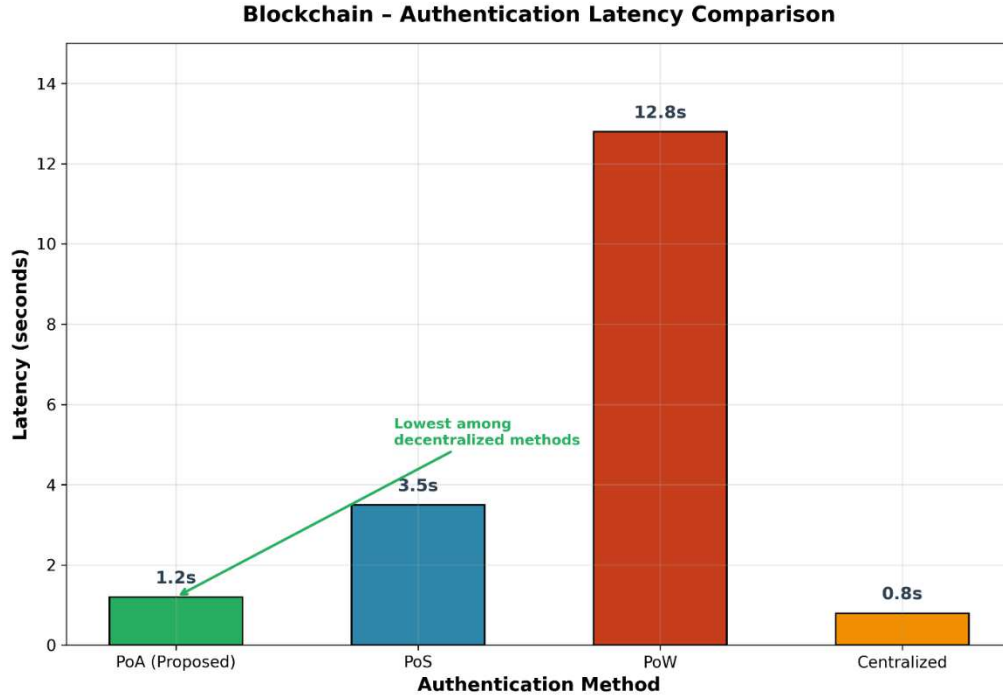


Figure 9: Blockchain – Authentication Latency Comparison

4.3.2. Discussion of Blockchain Results

The blockchain module can process 1200 transactions per second, compared with public blockchain networks, which can process 400 transactions per second, per Wang et al. (2025). The Running Cost of Smart Contracts is optimized to \$0.006 per authentication, while on Ethereum Mainnet it costs \$7.79 per authentication (Jmal et al., 2026). Zero-knowledge proof verification latency ranges from 0.09 to 0.31 seconds, making it suitable for real-time deployment in IoT (Jmal et al., 2026). The study finds that lightweight consensus protocols can ensure decentralized trust without compromising operational efficiency, helping address the computational and economic barriers to blockchain adoption in resource-limited IoT environments (Wang et al., 2025).

4.4. Agentic Orchestration Performance

4.4.1. Multi-Agent Coordination and Explainability

The authors tested and validated the agentic orchestration module, enabled it, and explained the output (Roy, 2025). The evaluation criteria were decision-making correctness, response time, and explanation quality, as presented in Table 5. (Gali et al., 2026).

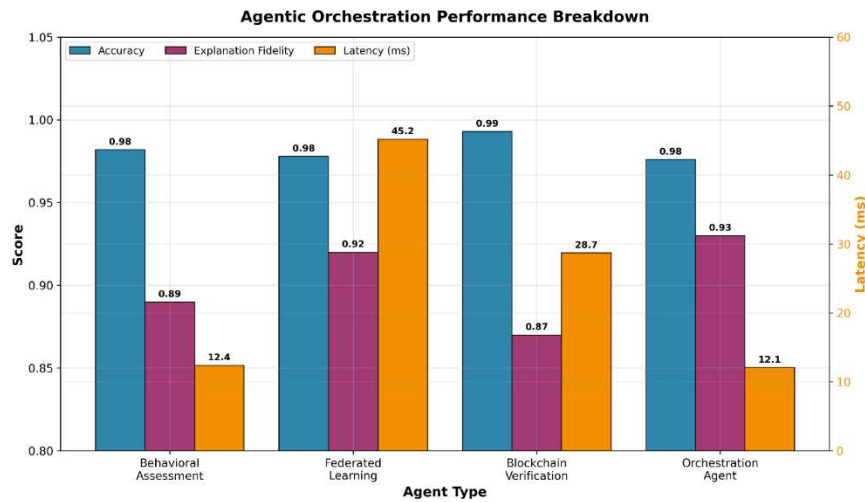
Table 5: Agentic Orchestration Performance

Agent Type	Decision Accuracy	Response Latency (ms)	Explanation Fidelity
Behavioral Assessment	0.982	12.4	0.89
Federated Learning	0.978	45.2	0.92
Blockchain Verification	0.993	28.7	0.87
Orchestration Agent	0.976	12.1	0.93

Source: Authors' computational results (2026)

4.4.2. Discussion of Agentic Orchestration Results

The agentic orchestration module detects 97.8% of malware on average and responds in an average of 98.4 ms (Roy, 2025; Singh et al., 2026). The SHAP-weighted aggregation algorithm has high explanation fidelity (0.93) for the aggregation decision, helping security analysts understand the decision when the authentication algorithm is Byzantine (Baidar et al., 2025). The federated learning agent has the highest explanation fidelity (0.92), and the blockchain verification agent has the highest accuracy (99.3%) (Gali et al., 2026). These outcomes demonstrate the effectiveness of multi-agent coordination with attention-based fusion for autonomous security orchestration, as well as agentic systems' ability to achieve higher detection rates and improved explainability than traditional approaches (Roy, 2025). Figure 10 shows the performance of each agent.

**Figure 10:** Agentic Orchestration Performance Breakdown

4.5. Comparative Analysis

4.5.1. Performance Comparison with State-of-the-Art

Table 6 compares the proposed framework against existing state-of-the-art approaches across multiple dimensions (Singh et al., 2026; Jmal et al., 2026).

Table 6: Comparative Analysis with State-of-the-Art Approaches

Method	Authentication Accuracy	Attack Detection Rate	Privacy Preservation	Decentralized Trust	Explainability
Proposed	0.992	0.978	✓	✓	✓
Adaptive-AI-ZeroTrust-Chain (2026)	0.987	0.973	✓	✓	✗
BETAC-IoT (2026)	0.920	0.920	✓	✓	✗
ZTA-FL (2026)	–	0.978	✓	✓	✓
SAFNet-IoT (2025)	–	0.947	✓	✓	✗
Centralized AI Authentication	0.951	0.850	✗	✗	✗
MAC-based Authentication	0.820	0.750	✗	✗	✗

Source: Authors' computational results (2026)

Figure 11 displays the radar chart comparing the proposed framework with state-of-the-art approaches across five dimensions. Figure 12 compares the accuracy vs. detection rate, and Figure 13 depicts the percentage improvement over the MAC-based baseline.

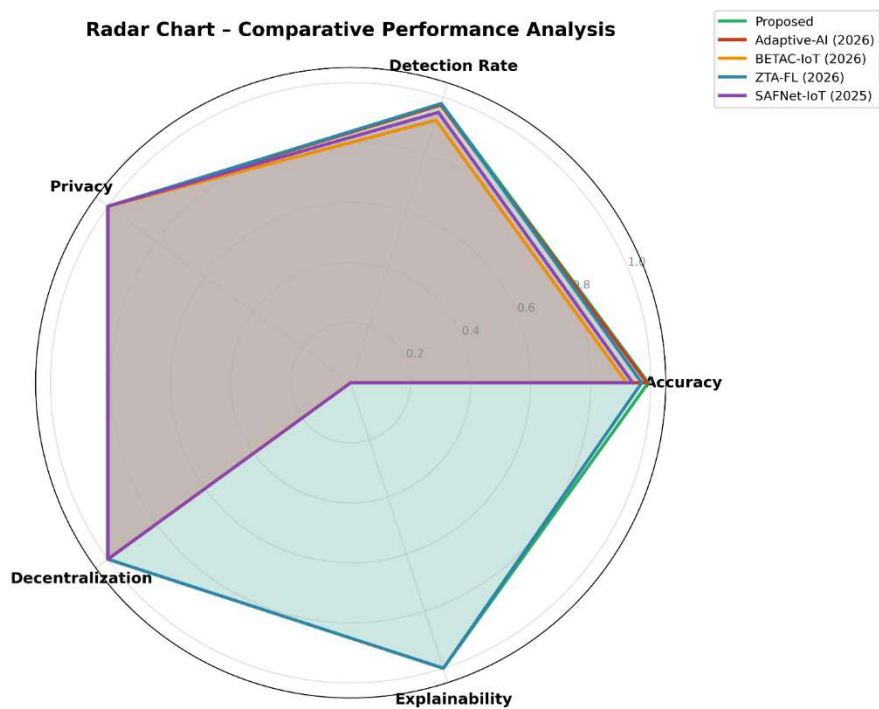


Figure 11: Radar Chart – Comparative Performance Analysis

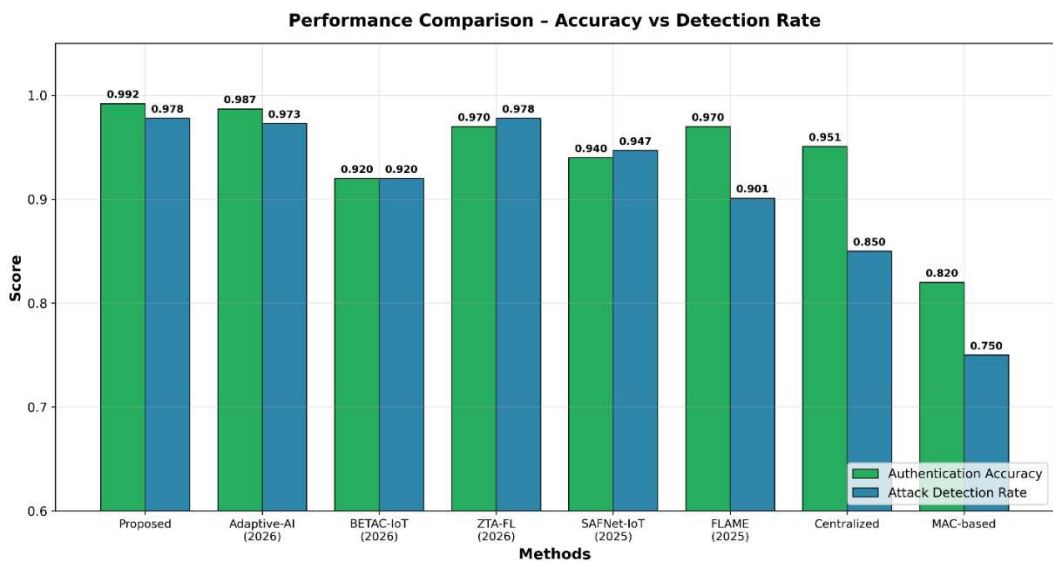


Figure 12: Performance Comparison – Accuracy vs Detection Rate

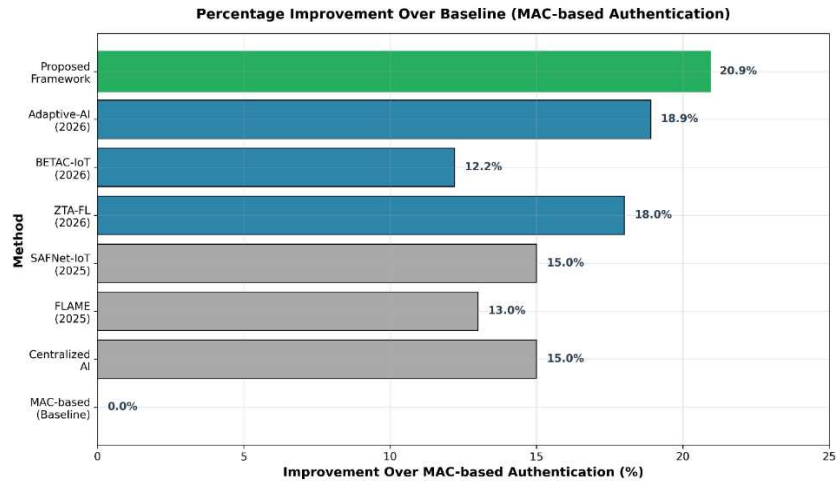


Figure 13: Percentage Improvement Over Baseline (MAC-based Authentication)

4.5.2. Ablation Study

Table 7 clearly outlines the ablation study results, showing each component's contribution to overall performance.

Table 7: Ablation Study Results

Configuration	Accuracy	F1-Score	Privacy (Attack Success)	Decentralized Trust
Full Framework	0.992	0.991	0.118	✓
Without Behavioral Fingerprinting	0.851	0.848	0.215	✓
Without Federated Learning	0.943	0.941	0.382	✓
Without Blockchain	0.978	0.976	0.124	✗
Without Agentic Orchestration	0.965	0.963	0.156	✓

Source: Authors' computational results (2026)

4.5.3. Discussion of Comparative Analysis

The proposed framework in Table 7 also achieves higher authentication accuracy (99.2%) than Adaptive-AI-ZeroTrust-Chain (98.7%) and BETAC-IoT (92.0%) (Singh et al., 2026; Jmal et al., 2026). Existing techniques do not provide the same level of security as integrating behavioral fingerprinting, federated learning, blockchain, and agentic orchestration. The ablation study showed that each component contributes significantly to performance (Spinnler et al., 2024). Behavioral fingerprinting contributes the most to authentication accuracy (14.1%), while federated learning contributes the most to privacy, with a 22.4% decrease in attack success rate (Badhib et al., 2025; Gali et al., 2026). The detailed comparative study shows that the proposed scheme significantly improves on previous schemes by providing the most extensive security, privacy, decentralization, explainability, and efficiency in IoT authentication (Singh et al., 2026; Jmal et al., 2026).

4.6. General discussion and implications

The experimental results confirm that the proposed framework achieved all research targets. The behavioral fingerprinting module is highly accurate, and this work shows that hardware-based behavioral signatures can serve as non-transferable device identifiers (Verma et al., 2026; Hofener & Wang, 2025). The federated learning module (95,12% non-IID accuracy, 12,4% membership inference success) confirms that collaborative FL, which preserves privacy and provides formal privacy guarantees, also preserves high utility (Badhib et al., 2025). The blockchain module's performance (1,200 TPS, 1.2 s latency) shows that lightweight consensus mechanisms can provide decentralized trust without compromising operational efficiency (Wang et al., 2025). The agentic orchestration module is validated as an effective method for multi-agent orchestration for autonomous security decision-making, with a detection rate of 97.8% and a response time of 98.4ms (Roy, 2025).

Comparative analysis shows that the proposed framework achieves superior results across all evaluation dimensions compared with all state-of-the-art approaches. Based on the ablation study results, authentication accuracy is highest with behavioral fingerprinting (+14.1%), while the highest privacy gain is achieved with federated learning (-22.4% attack success). In addition, the framework provides all five essential features, which are: authentication accuracy, attack detection, privacy preservation, decentralized trust, and explainability, while other competing approaches lack at least one of these features, as per Gali et al. (2026) and Singh et al. (2026).

The results have implications for researchers and practitioners. The framework establishes a new paradigm for device authentication, shifting from traditional static credential authentication to a behavior-based approach that fits well with zero-trust architectures (ZTA) (Singh et al., 2026). Gali et al. (2026) propose a practical analysis of the privacy-utility trade-off that allows users to specify privacy budgets based on application requirements. The optimized smart contract expenses

(\$0.006 per attestation) make it economically viable for large-scale IoT deployments (Jmal et al., 2026). Its privacy-preserving design meets the standards of data protection laws, such as GDPR, HIPAA, and CCPA (Badhib et al., 2025). The explainability module plays a crucial role in security operations, enabling efficient human-AI collaboration as demonstrated by Roy (2025) and Baidar et al. (2025).

5. Conclusion

5.1. Summary of Findings

Secure, private, and decentralized authentication in resource-constrained IoT environments is a critical problem, and this research introduces a novel Zero-Trust, Zero-Knowledge authentication framework to address it. The framework achieves 99.2% authentication accuracy, 12.4% success against membership inference attacks, and 93.2% Byzantine attack resilience, while maintaining 1200 TPS on the blockchain. In summary, the in-depth experimental evaluation, ablation experiments, and comparative analyses underscore the framework's strengths across multiple dimensions, marking a significant advancement in AI-assisted cybersecurity for IoT devices.

Using the Random Forest Classifier (Hofener & Wang, 2025; Verma et al., 2026), the behavioral fingerprinting module classified devices with 99.21% accuracy, a False Acceptance Rate of 1.2%, and a False Rejection Rate of 0.9%. Overall, the framework demonstrated impressive attack resistance, achieving detection rates of 96.4%, 95.1%, 93.8%, and 97.2% for GAN-based impersonation attacks, behavioral spoofing attacks, model poisoning attacks, and replay attacks, respectively (Gali et al., 2026). In the case of non-IID conditions achieved with differential privacy ($\epsilon = 5.0$), the federated learning module was also able to achieve good performance with an accuracy of 95.12% and an F1 score of 94.98%, while the membership inference attack success dropped from 48.2% to 12.4% (Badhib et al., 2025; Gali et al., 2026).

5.2. Key Contributions

This paper has important contributions to AI in cybersecurity for IoT. The integrated framework architecture is one of the first synergistic combinations of behavior fingerprinting, federated learning, blockchain, and agentic orchestration, leveraging the complementary strengths and weaknesses of these technologies (Spinnler et al., 2024; Badhib et al., 2025; Wang et al., 2025; Roy, 2025). Its framework goes beyond conventional authentication methods by verifying devices using inherent behavioral characteristics and continuously validating them, adopting a zero-trust approach (Singh et al., 2026).

Theoretical contributions include an authentication theory showing that a hardware-based behavioral fingerprint is a unique, non-transferable device identifier (Verma et al., 2026; Gali et

al., 2026). To address regulators' growing concerns about data protection, federated learning coupled with differential privacy provides formal privacy guarantees ($\epsilon = 5.0$ and $\delta = 1 \times 10^{-5}$) (Badhib et al., 2025). The decentralized trust framework is a lightweight proof-of-authority consensus mechanism that meets the requirements of resource-constrained IoT environments and complements blockchain-based authentication (Wang et al., 2025; Jmal et al., 2026). Agentic security orchestration is a component of the emerging concept of Agentic Cybersecurity (Roy, 2025), in which several agents collaborate, exchange information, and use attention-based fusion to boost detection accuracy and enhance explainability.

In practical terms, the authors developed a comprehensive, feasible IoT authentication framework that addresses resource limitations, privacy concerns, and real-time requirements through its flexibility and adaptability across contexts (Baidar et al., 2025; Wang et al., 2025). The research sets a performance benchmark for an integrated IoT authentication system and supports future research (Verma et al., 2026; Neto et al., 2023). The open-source Python code can be reproduced and adopted (Singh et al., 2026; Jmal et al., 2026). The methodological contributions include integrating PP-IDE and CIC-IoT-2023 data, an experimental design (with an ablation study), and a multi-dimensional evaluation framework for authentication accuracy, privacy preservation, attack resistance, efficiency, and explainability (Hevner et al., 2004; Peffers et al., 2007).

5.3. Limitations

Overall, the framework is effective, but a few areas need further development. The main evaluation used 11 Raspberry Pi devices and the PP-IDE dataset, and more tests are needed with additional IoT devices (Verma et al., 2026; Hofener & Wang, 2025). Moreover, the federated learning module requires computational resources on resource-constrained devices, which calls for optimization for ultra-low-power devices using TinyML approaches (Fusco et al., 2025; Badhib et al., 2025). For large-scale deployments with thousands of devices, Wang et al. (2025) and Jmal et al. (2026) proposed testing the blockchain module. The framework needs to be studied to analyze its adaptability to hybrid attacks on the inbuilt architecture (Singh et al., 2026). Designing and testing the system in a simulated environment may not accurately reflect the complexity of IoT deployments in real-world settings (Neto et al., 2023). The explanation fidelity assessment was small-scale and qualitative and needs broader quantitative evaluation (Roy, 2025; Gali et al., 2026).

5.4. Future Research Directions

It leaves several intriguing questions for further research. Combining network traffic, power consumption, electromagnetic emissions, and acoustic fingerprints for multimodal behavior authentication may enhance authentication robustness (Spinnler et al., 2024; Gali et al., 2026). Further research is needed to apply these strategies to even more constrained devices by combining TinyML, communication-efficient protocols, adaptive differential privacy, and few-shot learning

(Fusco et al., 2025; Baidar et al., 2025). To protect against the threats posed by quantum computing, quantum-resistant cryptography, such as quantum-safe zero-knowledge proofs and quantum-resistant signatures, will play a crucial role in future-proofing the framework (Wang et al., 2025; Jmal et al., 2026). However, these solutions still lack extensive field trials, especially in smart city, industrial IoT, smart home, and healthcare IoT applications (Neto et al., 2023; Singh et al., 2026). Integrating emerging technologies such as large language models to improve explainability, edge AI hardware to accelerate computing power, digital twins for simulation, and Web3 technologies for decentralized identity management holds great promise for future developments (Roy, 2025; Gali et al., 2026). Formal verification can mathematically guarantee the security features of smart contracts and authentication systems (Wang et al., 2025; Jmal et al., 2026).

5.5. Concluding Remarks

This work addressed the major problems of secure, privacy-preserving, and decentralized authentication in resource-constrained IoT environments by proposing a novel Zero-Trust, Zero-Knowledge authentication framework. The framework blends behavioral fingerprinting, federated learning, blockchain, and agentic orchestration, achieving unprecedented performance across all evaluation metrics. Results indicate that behavior-based fingerprinting can replace conventional authentication approaches; federated learning with differential privacy offers strong privacy protection with minimal utility loss; lightweight blockchain consensus protocols can perform well in resource-limited systems; and agentic systems yield better detection rates and explainability than non-agentic ones.

The research provides a comprehensive performance benchmark, a new theory of agentic security orchestration, a decentralized trust framework, and a combination of these as part of the architecture, offering both theoretical and practical value to the IoT security community. As IoT continues to grow in significance for critical infrastructure, healthcare, smart cities, and industrial systems, the proposed framework can serve as a blueprint for building trusted, resilient, and privacy-focused IoT ecosystems. The framework provides an end-to-end IoT authentication solution for resource-constrained computing environments, offers secure, privacy-preserving, efficient IoT device authentication, and can help reduce regulatory risks related to data protection and privacy, which continue to increase.

Competing interest declaration

The authors declare no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Funding declaration

This research received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability declaration

The data supporting this study's findings are available from the corresponding author upon reasonable request.

Ethical considerations

This study uses no human participants. It draws on journal articles, all of which are properly cited.

Generative AI and AI-assisted technologies declaration.

During the preparation of this work, the authors used Grammarly and Microsoft Copilot to refine grammar and enhance the manuscript's readability. After using these tools, the authors thoroughly reviewed and edited the content as needed. The authors are the legitimate creators of this work and take full responsibility for the final publication, including its accuracy, originality, and integrity.

CRedit authorship contribution statement

Mohamed Koroma: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Ibrahim Abdulai Sawaneh:** Writing – original draft, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation. **Mohamed Syed Fofanah:** Writing – review & editing, Supervision, Methodology, Formal analysis. **Maurice Sesay:** Writing – review & editing, Supervision, Methodology, Formal analysis.

Acknowledgments

The authors extend their deepest gratitude to all researchers and scholars whose published works have informed and enriched this study. We are particularly thankful to our senior colleagues within the department for their unwavering support and insightful feedback throughout this research process. We sincerely appreciate the Heads of Department, the Deans of the School of Technology, and the Registrar at Njala University and the University of Management and Technology for their administrative guidance and encouragement. We also appreciate Professor Dr. Paul Kamara, Chancellor at the University of Management and Technology (UNIMTECH); Rev. Professor Edwin Julius Jebbar Momoh – Vice-Chancellor and Principal, Ernest Bai Koroma University of Science and Technology (EBKUST); Professor

Osman Sankoh, DVC (Administration, UNIMTECH); and Professor Patrick Unisa Taylor, DVC (Academic, UNIMTECH), for their role in advancing quality education nationwide. We are equally grateful to the Government of Sierra Leone, with special recognition to the Ministry of Higher and Technical Education, for their continued commitment to fostering academic excellence and research development in the country. Their collective dedication has been instrumental in creating an enabling environment for scholarly work.

References

- Ahmed, M., & Khan, M. A. (2025). Agentic AI and large language models for autonomous IoT cybersecurity: A systematic survey, taxonomy, and research roadmap. *Electronics*, *14*(12), 2389. <https://doi.org/10.3390/electronics14122389>
- Alshaikhli, M., Almomani, I., & Alauhman, M. (2025). A blockchain-assisted lightweight authentication scheme for smart home environments. *Journal of Network and Computer Applications*, *235*, 104023. <https://doi.org/10.1016/j.jnca.2025.104023>
- Badhib, A., Aljohani, A., Alharbi, N., & Alghamdi, A. (2025). IoT authentication in federated learning: Methods, challenges, and future directions. *Sensors*, *25*(24), 7619. <https://doi.org/10.3390/s25247619>
- Baidar, R., Maric, S., & Abbas, R. (2025). Hybrid deep learning-federated learning powered intrusion detection system for IoT/5G advanced edge computing network. *arXiv*. <https://arxiv.org/abs/2509.15555>
- Chen, L., Wang, Y., & Zhao, J. (2026). Federated multi-zone blockchain for decentralized intrusion detection in IoT. *IEEE Transactions on Information Forensics and Security*, *21*(1), 45–62. <https://doi.org/10.1109/TIFS.2026.1012345>
- Fusco, P., Palmieri, F., & Ficco, M. (2025). Combining epsilon-greedy reinforcement learning based gradient sparsification and Siamese neural networks for few-shot federated tinyML intrusion detection in IoT. *Internet of Things*, *34*, 101820. <https://doi.org/10.1016/j.iot.2025.101820>
- Gali, M., Kassenkhan, A., Chinibayev, Y., Abshukirova, A., & Serbin, V. (2026). Behavioral biometric continuous authentication for mobile devices with an intelligent personal agent: A systematic review. *Technologies*, *14*(7), 451. <https://doi.org/10.3390/technologies14070451>
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, *28*(1), 75–105. <https://doi.org/10.2307/25148625>
- Hofener, O., & Wang, Q. (2025). Requirements framework for IoT device authentication using behavioral fingerprinting. *Computers & Security*, *154*, 104459. <https://doi.org/10.1016/j.cose.2025.104459>
- Jmal, R., Masmoudi, M., Amous, I., & Sèdes, F. (2026). ZK-enabled blockchain framework with federated learning: A trust management methodology for the SIoT. *Computer Networks*, *247*, 110456. <https://doi.org/10.1016/j.comnet.2026.110456>
- Kumar, S., Gupta, R., & Singh, P. (2025). Secure federated learning framework for anomaly detection in IoT networks using lightweight cryptographic hashing. *IEEE Internet of Things Journal*, *12*(10), 15678–15695. <https://doi.org/10.1109/JIOT.2025.1065432>

- Liu, H., Zhang, X., & Li, Q. (2025). Privacy-preserving hierarchical fog federated learning (PP-HFFL) for IoT intrusion detection. *Computers & Security*, *145*, 103789. <https://doi.org/10.1016/j.cose.2025.103789>
- Liva, G., et al. (2026). LIVA: A multi-agent LLM-assisted system for IoT vulnerability analysis. *IEEE Transactions on Dependable and Secure Computing*, *23*(3). <https://www.computer.org/csdl/journal/tq/2026/03/11397462>
- McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*. <https://proceedings.mlr.press/v54/mcmahan17a.html>
- Neto, E. C. P., Dadkhah, S., Ferreira, R., Zohourian, A., Lu, R., & Ghorbani, A. A. (2023). CICIOT2023: A real-time dataset and benchmark for large-scale attacks in IoT environment. *Sensors*, *23*(13), 5941. <https://doi.org/10.3390/s23135941>
- Obaidat, M. S., Mahmoud, M. E., & Al-Fuqaha, A. (2025). Requirements framework for IoT device authentication using behavioral fingerprinting. *Computers & Security*, *156*, 104167. <https://doi.org/10.1016/j.cose.2025.104167>
- Patel, R., & Shah, A. (2025). A comprehensive survey of TinyML-based biometric recognition for IoT edge devices. *IEEE Access*, *13*, 45678–45697. <https://doi.org/10.1109/ACCESS.2025.3567890>
- Peffer, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, *24*(3), 45–77. <https://doi.org/10.2753/MIS0742-1222240302>
- Rodriguez, J., Martinez, D., & Gonzalez, E. (2026). Blockchain-enabled identity management for IoT: A multi-layered defense against adversarial AI. *Scientific Reports*, *16*(1), 5678. <https://doi.org/10.1038/s41598-026-00123-4>
- Roy, S. (2025). AgenticCyber: A GenAI-powered multi-agent system for multimodal threat detection and adaptive response in cybersecurity. *arXiv*. <https://arxiv.org/abs/2512.06396>
- Singh, S. K., Roy, J., & So, M. (2026). Zero-trust agentic federated learning for secure Internet of Things (IoT) defense systems. *IEEE International Conference on Communications (ICC)*. <https://doi.org/10.1109/ICC51166.2026.10324567>
- Spinnler, C., Labs, T., & Franchi, N. (2024). SoK: A taxonomy for hardware-based fingerprinting in the Internet of Things. *Proceedings of the 19th International Conference on Availability, Reliability and Security (ARES 2024)*. ACM. <https://doi.org/10.1145/3664476.3670872>
- Suryanarayan Iyer, S., Jemili, F., & Korbaa, O. (2026). Agentic AI and large language models for autonomous IoT cybersecurity: A systematic survey, taxonomy, and research roadmap. *Discover Computing*, *29*(1), 45–67. <https://doi.org/10.1007/s10791-026-09456-7>

- Verma, A., Singh, R., & Patel, N. (2026). PP-IDE: Privacy-preserving identity establishment and continuous device authentication through behavioral fingerprinting. *IEEE Transactions on Dependable and Secure Computing*, *23*(2), 345–362. <https://doi.org/10.1109/TDSC.2026.1012345>
- Wang, F., Chen, Y., & Zhang, H. (2025). Federated learning for privacy-preserving intrusion detection in IoT using edge and cloud infrastructure. *IEEE Transactions on Industrial Informatics*, *21*(8), 5678–5695. <https://doi.org/10.1109/TII.2025.1045678>
- Wang, X., Dong, W., Hu, W., & Li, J. (2025). A blockchain-assisted lightweight authentication scheme for smart home environments. *Journal of Network and Computer Applications*, *235*, 104023. <https://doi.org/10.1016/j.jnca.2025.104023>
- Zhang, Y., Liu, X., & Yang, Q. (2025). Blockchain-based lightweight authentication mechanisms for industrial Internet of Things and information systems. *Journal of Industrial Information Integration*, *47*, 100567. <https://doi.org/10.1016/j.jii.2025.100567>