# Preliminary Progress Review (PPR) Report of R25- 039


Project ID: R25 - 039

Project Title: Data-Privacy Focused Federated Learning Framework for Industrial IoT


| Student Details: | |
|---|---|
| Names: | Student IDs: |
| Nanayakkara Y.D.T. D | IT21826368 |
| Mendis H.R.M | IT21822612 |
| Weerasinghe K.M | IT21831904 |
| Dissanayaka K.D.A.R. A | IT21828348 |

Supervisor: Mr. Amila Seneratha

External Supervisor: Dr. Sanika Wijesekara

Co-Supervisor: Mr. Tharaniyawarma Kumaralingam


Date of Submission: 1st May 2025

# Table of Contents

# 1. Introduction

## 1.1. Purpose of the Document

*The purpose of this Preliminary Progress Review (PPR) Report for the 4th-year research project - R25-039 - is to provide an update on the work completed so far, key findings, any challenges faced, and adjustments to the research plan. This document aims to inform the project supervisor and other stakeholders of the current progress and to gather feedback for the next steps in the project.*

*Intended Audience:*

*Project Supervisors: To assess progress and provide guidance for the next phase of the research.*

*Team: To stay aligned and ensure everyone is on track with the project's goals.*

*External Reviewers (research paper): To evaluate progress and suggest improvements.*

*This report is an important part of ensuring the research stays on track and meets the objectives for the final year's project.*

## 1.2 Acronyms

| Acronym | Full Form |
|---|---|
| **ADRM** | Attack Detection and Resilience Module |
| **CNN** | Convolutional Neural Network |
| **DP** | Differential Privacy |
| **FL** | Federated Learning |
| **HE** | Homomorphic Encryption |
| **HMAC** | Hash-based Message Authentication Codes |
| **IIoT** | Industrial IoT (Industrial Internet of Things) |
| **IoT** | Internet of Things |
| **PPM** | Privacy Preservation Module |
| **SAM** | Secure Aggregation Module |
| **SCPEM** | Secure Communication and Protocol Enforcement Module |
| **SMC** | Secure Multiparty Computation |
| **SSS** | Shamir's Secret Sharing |
| **TC** | Threshold Cryptography |
| **TLS** | Transport Layer Security |

# 1. Overview

Initial Problem and Motivation

The project was necessary because industrial companies could not safely use the power of AI. Traditional machine learning required collecting all sensitive factory data into a single, central server, which posed major risks to data privacy and created a single point of failure. While FL offered a way to train AI models collaboratively without moving raw data, existing solutions were incomplete.

Our team recognized a critical gap: no single framework existed that successfully integrated and balanced the industrial requirements for robust **security, data privacy, efficient communication, and the resource limitations** of the IIoT devices.

Project Goal and Planned Development

The primary goal of this development was to create a **robust, integrated FL framework** that would allow industrial companies to safely build and use AI models in a highly secure manner. The system was intended to protect data across the entire training lifecycle and mitigate the catastrophic risks associated with security failures in industrial environments.

Key Planned Deliverables (What We Set Out to Build)

The development was structured to deliver a multi-layered security framework with the following core functionalities:

1. **Comprehensive Security Framework:** We planned to develop a foundational system using advanced cryptographic techniques, including **Secure Aggregation, Differential Privacy (DP), and Homomorphic Encryption (HE)**, to protect all data and model updates end-to-end.
2. **Robust IIoT Protocol:** A protocol was to be designed to enforce strict rules and use **mutual client/server validation** to actively detect and block common threats like **Byzantine and model poisoning attacks**, guaranteeing the integrity of the final AI model.
3. **Efficient System Design:** We intended to optimize the framework's architecture and apply techniques like **quantization** to ensure that communication and computation were highly efficient, allowing the system to operate effectively on resource-limited IIoT devices.
4. **Proven Performance:** The final step of the development was to be the rigorous testing and validation of the framework on real-world IIoT data to prove its effectiveness in security, privacy, and accuracy.

# 2. Literature review

Securing communication and control channels in federated learning (FL) systems deployed across Industrial Internet of Things (IIoT) contexts has been a significant research topic. Modern studies repeatedly underline the need for identity-verified transport protocols and strong message validation to fight against eavesdropping, impersonation, tampering, and replay assaults on unreliable industrial networks. Implementations based on TLS 1.3 with mutual authentication are widely regarded as acceptable for restricted edge devices and servers, ensuring encrypted and authorized data exchange. To ensure integrity and freshness, many frameworks combine HMAC-based message authentication together with nonces or timestamps, prohibiting data tampering or replay under the high-latency and lossy conditions characteristic of industrial environments. Furthermore, role-based access control (RBAC) has been identified as crucial for restricting sensitive actions such as client enrollment, aggregation triggers, and model releases. Auditable decision logs increase accountability and integrate these systems with operational technology governance norms. Empirical work in this field generally analyzes handshake latency, CPU and memory overhead, and resilience to network-level threats, indicating that secure control techniques can be practical even under IIoT restrictions.

While federated learning inherently keeps raw data local, multiple studies have shown that gradients and model changes can nonetheless leak crucial operational information. To address this risk, Differential Privacy (DP) is widely recommended, using gradient clipping and calibrated noise injection to balance privacy with model utility. Homomorphic Encryption (HE) complements this by allowing encrypted updates to be aggregated without revealing individual contributions to the server. However, this strategy creates issues in ciphertext extension, computing expense, and bandwidth utilization. Comparative analyses of baseline FL, FL+DP, FL+HE, and hybrid FL+DP+HE configurations illustrate trade-offs among privacy, utility, and efficiency, offering design insights into parameter tuning such as ε-budgets, clipping norms, and encryption scheme choices based on device capability and data sensitivity. Collectively, this literature considers privacy methods not as independent add-ons but as part of an integrated operational plan that must sustain both compliance and industrial performance.

Another significant focus area is secure aggregation, which ensures that only the combined model update is revealed while safeguarding individual client contributions. Pairwise masking methods are widely employed to ensure this confidentiality, with Shamir's Secret Sharing (SSS) permitting threshold-based recovery when clients drop out. This technique permits mask reconstruction without degrading individual updates, assuring correctness even in the presence of unreliable connectivity, a prevalent concern in IIoT scenarios. Research prototypes maximize masking precision and messaging efficiency to accommodate heterogeneous devices and retain equivalent model accuracy to non-secure baselines. Experimental evaluations often examine resilience under high dropout rates, poisoning attacks, and backdoor scenarios, while monitoring communication latency and resource usage using realistic frameworks like gRPC and PyTorch.

To guard against malicious or false updates, anomaly detection and robust aggregation techniques are widely researched. These systems monitor incoming updates in real time, employing metrics such as gradient norms, updating similarity, and variation between rounds to detect anomalous behavior before aggregation. Once recognized, suspect updates may be down weighted, rejected, or quarantined, decreasing the risk of model poisoning or Byzantine defects. Many systems additionally incorporate recovery techniques like checkpoint rollback and adaptive retraining, ensuring model integrity and continuing operation even under adversarial conditions. Evaluations report measures such as detection accuracy, mitigation delay, and sustained model

performance under realistic IIoT scenarios. Importantly, this line of research underscores the necessity to co-tune detection sensitivity with privacy methods and transport security settings to ensure acceptable end-to-end latency and dependability within industrial service-level requirements.

- *Current state of the research problem.*

The Need for Federated Learning (FL)

- The growth of **Industrial Internet of Things (IIoT)** devices means that machine learning models are needed, but the traditional approach of centralizing all data poses significant risks, including privacy breaches and creating single points of failure.

The Gaps in Current FL Systems

While FL is the correct direction, current systems are insufficient for the IIoT environment because they lack a single, comprehensive solution. Specifically, no existing framework successfully integrates robust solutions for all these critical challenges in a unified system

**Security & Attack Resilience**

Existing studies explore defenses against poisoning and Byzantine attacks, but they typically **do not provide a holistic solution** that also accounts for privacy protection and communication efficiency.

**Data Privacy**

Techniques like **Differential Privacy (DP)** and **Homomorphic Encryption (HE)** have been integrated to provide formal privacy guarantees[c9. However, these solutions often come with **significant computational overhead**, making them challenging for resource-limited IIoT devices.

**Resource Efficiency**

Techniques like **model compression and hierarchical FL architectures** have been proposed to reduce overhead. However, these approaches **may introduce new security vulnerabilities** or trade off with model accuracy.

## 2.2. Identification and significance of the problem

The proposed research introduces a novel approach by unifying four critical modules secure command-and-control, privacy preservation, secure aggregation, and real-time attack defense into a single, co-tuned Federated Learning (FL) framework optimized for Industrial Internet of Things (IIoT) environments. While prior studies tend to focus on characteristics such as transport security, privacy, or model robustness, this work integrates all four into a cohesive and policy-driven architecture specifically built to manage heterogeneous devices, non-IID data, and unstable industrial networks.

The framework's originality resides in its holistic design and adaptability. It provides policy-controlled switching between FL, FL+DP, FL+HE, and hybrid privacy modes; resource-aware aggregation parameters that match device capabilities; role-based access control (RBAC) for sensitive operations; and customizable anomaly detection sensitivity for defense mechanisms. Each of these components is measured by end-to-end metrics including accuracy, convergence, bandwidth, latency, and resource usage to assure practicality in real-world IIoT implementations.

The scientific and technical challenge is enormous. Implementing such an integrated system demands secure certificate lifecycle management and perfect forward secrecy (PFS) at fleet scale, reconciling the computational and bandwidth overhead of homomorphic encryption (HE) with constrained industrial networks, and calibrating Differential Privacy (DP) to limit information leakage without degrading model accuracy. It also requires ensuring threshold-based recovery under frequent client dropouts and preserving resistance against poisoning and Byzantine assaults in dynamic, non-IID data distributions.

Beyond individual difficulties, these layers interact in nontrivial ways. For instance, DP noise may conceal anomaly detection signals, HE and secure aggregation bring latency that conflicts with authenticated transport and real-time defense budgets, and overly rigid RBAC can hinder speedy recovery if not designed with operational workflows in mind. Achieving stability, accuracy, and availability therefore demands careful cross-module co-design, coordination, and empirical validation.

Ultimately, our research intends to give defense-in-depth without losing performance or operability an outcome not yet established in literature. By combining coordinated mechanisms with systematic benchmarking and tuning guidance, the work addresses a clear gap: the absence of a deployment-grade, end-to-end FL solution that simultaneously satisfies confidentiality, integrity, privacy, robustness, and efficiency requirements in industrial-scale environments.

# 3. Research Question

How can an integrated Industrial IoT federated learning framework jointly provide secure and authenticated communication, tunable privacy of client updates, confidential and dropout, Robust aggregation, and real-time attack defense, while maintaining target accuracy, latency, bandwidth, and resource budgets across heterogeneous, unreliable industrial networks?

## Component Research question

**Secure Communication and Protocol Enforcement (SCPM)**

How can we design a low overhead, scalable communication and protocol based security layer for IIoT federated learning that uses TLS 1.3 mutual authentication, HMAC integrity, and rolebased access control to stop MitM, replay, and unauthorized commands without breaking latency and throughput on constrained devices?

**Privacy Preservation Module (PPM)**

How can Differential Privacy and Homomorphic Encryption individually and together be calibrated so that client update leakage is firmly restricted while model accuracy and communication/compute costs remain acceptable for real industrial workloads and networks?

**Secure Aggregation Module (SAM)**

How can a layered secure aggregation algorithm that uses pairwise masking with Shamir's Secret Sharing ensure the server learns just the global sum and still tolerate client dropouts and diverse device limits without harming convergence or final accuracy?

**Attack Defense and Resilience Module (ADRM)**

How can a lightweight, realtime anomaly detection and reaction pipeline be integrated into federated training to reliably detect, isolate, and recover from poisoning and Byzantine updates at IIoT scale while keeping latency low and model quality high?

# 4. Research Objectives

The main objective of this research is to devise, execute, and assess a comprehensive Federated Learning (FL) framework for Industrial Internet of Things (IIoT) settings that ensures concurrent safeguarding across all key security and privacy aspects. The framework seeks to safeguard command-and-control traffic, maintain client update confidentiality, guarantee dropout-resilient and privacy-preserving aggregation, and facilitate real-time attack defense, all while ensuring acceptable accuracy, latency, bandwidth, and resource utilization across diverse devices and unstable industrial networks. A supplementary objective is to create deployment-ready tuning guidelines that co-optimize security, privacy budgets, aggregation thresholds, and detection sensitivity, delivering explicit empirical reports on accuracy, convergence, and operational overheads under realistic industrial workloads.

The initial component, Secure Communication and Protocol Enforcement, emphasizes the development of a low-overhead, authenticated transport layer utilizing TLS 1.3 with mutual authentication, HMAC-based message integrity, and freshness validation to safeguard all control and model-update exchanges. It incorporates role-based access control (RBAC) to govern sensitive operations between clients and servers. This objective encompasses the validation of certificate lifecycle processes namely provisioning, rotation, and revocation and the quantification of their impacts on latency, throughput, and availability in adversarial scenarios, including man-in-the-middle, replay, and unauthorized-command attacks.

The second component, Privacy Preservation, aims to mitigate information leakage from client updates via Differential Privacy (DP) with adjustable privacy budgets and gradient clipping, while additionally providing an optional Homomorphic Encryption (HE) pathway for encrypted aggregation when applicable on IIoT networks and edge devices. This module's evaluation entails a comparative analysis of baseline FL, FL+DP, FL+HE, and hybrid FL+DP+HE configurations on representative industrial datasets, determining parameter sets that optimally reconcile privacy, accuracy, and computational or communication overhead.

The third component, Secure Aggregation, aims to ensure confidentiality and resilience against dropout by employing a two-tier aggregation protocol that obscures individual updates via pairwise masking and utilizes Shamir’s Secret Sharing (SSS) to reconstruct aggregate contributions from absent clients. This objective highlights resource-aware masking accuracy to accommodate device diversity and seeks to exhibit model equivalence with non-secure benchmarks while methodically assessing scalability, latency, and bandwidth performance.

The final component, Attack Defense and Resilience, emphasizes the development of a lightweight, real-time anomaly detection and mitigation pipeline that detects and isolates poisoning and Byzantine attacks during training while minimizing latency. The module incorporates strong aggregation, checkpoint rollback, and adaptive retraining mechanisms to reinstate secure operational states after identifying anomalies. Evaluation metrics encompass detection precision and recall, time-to-mitigation, and sustained model accuracy in non-IID data distributions and IIoT-scale implementations.

The objectives collectively seek to establish a complete defense-in-depth framework that integrates secure communication, privacy protection, secure aggregation, and real-time resilience into a practical and empirically validated solution for industrial settings.

# 5. Research methodology

- 1. The Research Method: System Development and Empirical Evaluation

The core method selected is the **development and implementation of a robust, modular system** , followed by an **empirical evaluation** to validate its effectiveness. The system is designed as a **Federated Learning Framework** that integrates various components to ensure secure, private, and efficient collaborative model training.

- Conceptual Framework and Identified Variables

The conceptual framework is based on a **multi-layered architecture** where security and privacy are guaranteed at every stage of the FL process. The system is composed of four interconnected modules that work in synergy:

| Module | Core Functionality and Techniques Used |
|---|---|
| **Secure Communication and Protocol Enforcement Module (SCPEM)** | Acts as the "rule-book" for data exchange, ensuring integrity and confidentiality. It uses **TLS 1.3 with Mutual Authentication** to prevent unauthorized access and **HMAC** for data integrity. |
| **Attack Detection and Resilience Module (ADRM)** | A critical component that actively monitors for a wide range of cyber threats, including **poisoning attacks and Byzantine attacks**. |
| **Privacy Preservation Module (PPM)** | Provides end-to-end privacy guarantees by implementing **Differential Privacy (DP)** (adding noise) and **Homomorphic Encryption (HE)** (computing on encrypted data). |
| **Secure Aggregation Module (SAM)** | Ensures the central server cannot inspect individual raw updates[cite: 112]. It utilizes a **dual-protocol approach** combining **Shamir's Secret Sharing (SSS)** and the **Bonawitz Protocol**. |

*Identified Matrices/Variables*

The framework's performance is validated based on four key metrics:

**Security:** Measured by resilience against attacks (e.g., **Backdoor Attack Success Rate** ) and the number of **Blocked Attacks**.

**Privacy:** Measured via **Privacy Audits** and the effect of DP/HE on the aggregated model distribution.

**Accuracy:** Measured by **Global Model Test Accuracy (%)** across training rounds.

**Efficiency/Resilience:** Measured by **Aggregation Success Rate vs. Participant Dropout Rate** to demonstrate fault tolerance.

7. **Sources for test data and analysis,**

For the simulation setup, the **CIFAR-10 dataset** was used. A **Convolutional Neural Network (CNN)** model was employed to represent an industrial Internet of Things space.

**Must change to an Industrial IIOT dataset**

- *Data Collection Procedures*

The research simulates the iterative **Federated Learning** process where clients perform local computations and send only model updates to a central server. Specific procedures for secure data handling include:

**Client Authentication:** Clients and the server must present and validate a digital certificate using **mutual authentication** (TLS 1.3) to prevent unauthorized devices from joining the process.
**Data Integrity:** Every message includes a **Hash-based Message Authentication Code (HMAC)**, which the server recalculates upon receipt. If the calculated HMAC does not match the received one, the message is immediately discarded as corrupted or tampered with.
**Private Submission:** Clients' model updates are first privatized using **Differential Privacy** and then encrypted using **Homomorphic Encryption** before being sent.
**Secure Aggregation:** Each client's update is broken into multiple shares using **Shamir's Secret Sharing** to prevent a single point of failure and sent to the server for computation on encrypted shares.

- *Data analysis methods to be used*

The collected experimental data is analyzed using the following methods:

- **Anomaly Detection:** The ADRM uses a two-pronged approach to identify suspicious client behavior before model aggregation:
  - **Statistical Analysis:** Calculating a "fingerprint" of updates based on the average and spread of weights.
  - **Deviation Scoring:** Assigning a "suspiciousness score" to each update and flagging it if the score exceeds a predetermined threshold.
- **Comparative Accuracy Plotting:** Global model accuracy is plotted against the number of training rounds, comparing the framework (with ADRM/SCPEM) against baseline methods (FedAvg) to demonstrate resilience against attacks.
- **Privacy Audits:** Conducted to formally measure the trade-off between the achieved privacy level and the resulting model utility, ensuring privacy mechanisms function as intended.
- **Robustness Evaluation:** The system's robustness is empirically evaluated by testing resilience against simulated client dropouts and malicious behavior, with the results compared against theoretical failure points.

# 6. Anticipated benefits

## Benefits to users

- Better Data Privacy:
  By combining tunable Differential Privacy and optional Homomorphic Encryption, model updates can protect sensitive operational information. This helps organizations meet regulatory and compliance requirements while keeping model accuracy and communication speed at acceptable levels.

- Secure and robust Aggregation:
  Pairwise masking makes sure that the server only learns about model updates as a whole, and Shamir's Secret Sharing (SSS) lets clients who drop out during training recover their threshold. This maintains accuracy parity with non-secure baselines even under device heterogeneity and unstable connectivity.

- Real-Time Integrity and Resilience:
  An easy-to-use anomaly detection and strong aggregation pipeline finds and stops poisoning and Byzantine attacks in real time. It uses automated isolation, rollback, and adaptive retraining to keep the model available with little extra latency.

- Better visibility and control over operations:
  End-to-end monitoring and secure workflows covering certificates, policies, and client enrollment—provide administrators with actionable metrics (accuracy, convergence, latency, bandwidth, CPU/memory) for dependable, maintainable deployments.

- Optimized Bandwidth and Faster Insights: On-device training and compact model updates reduce bandwidth consumption and accelerate insight generation, making the framework well-suited for intermittent and low-rate industrial communication links.

## Contribution to the body of knowledge

- Unified End-to-End Architecture:
  A co-designed framework is introduced that integrates secure transport, tunable privacy, confidential and dropout-robust aggregation, and real-time attack defense. The work empirically explores cross-module trade-offs such as Differential Privacy noise versus anomaly-detection accuracy and security overheads versus latency budgets under IIoT constraints.

- Comprehensive Comparative Study:
  Evaluation of FL, FL+DP, FL+HE, and FL+DP+HE configurations on industrial datasets maps the privacy–utility–efficiency frontier and provides parameter guidelines for ε budgets, clipping norms, and encryption settings tailored to heterogeneous devices and networks.

- Novel Two-Layer Secure Aggregation Protocol:
  A practical aggregation method combining pairwise masking for confidentiality and Shamir's Secret Sharing for dropout recovery is implemented and validated, including resource-aware precision control to support constrained hardware without degrading correctness or accuracy.

- Integrated Real-Time Defense Pipeline:
  A lightweight detection-and-mitigation system is built and benchmarked using precision, recall, time-to-mitigation, and accuracy retention metrics. It orchestrates automated responses—down-weighting, rejecting, or quarantining malicious updates—alongside checkpoint rollback and adaptive retraining.

- Specialized Secure Communication Blueprint:
  A tailored communication layer for IIoT federated learning is engineered with TLS 1.3 mutual authentication, HMAC integrity, RBAC enforcement, and certificate lifecycle management, evaluated for latency, throughput, and availability under realistic adversarial scenarios.

- Reproducible Evaluation Framework:
  A standardized testing and reporting environment is developed to measure accuracy, convergence, confidentiality, integrity, robustness to dropouts and attacks, and resource/communication overhead. This contributes a repeatable benchmarking methodology for industrial FL research and deployment.

# 7. Scope and specified deliverables / expected research outcome

- 1. Scope (What the Software Product(s) Will Do)

The core deliverable is a comprehensive, functional, and integrated **Federated Learning (FL) framework** specifically designed for IIoT environments. This software product's primary function is to enable secure and private collaborative training of AI models across distributed industrial devices.

| **Component/Functionality** | **Explanation of Deliverable** |
|---|---|
| **Secure Aggregation (SAM)** | The software will securely aggregate model updates from multiple clients using **Shamir's Secret Sharing (SSS)**. It will prevent the central server from ever viewing any single client's raw update, ensuring privacy. |
| **End-to-End Privacy (PPM)** | The product integrates **Differential Privacy (DP)** and **Homomorphic Encryption (HE)** into the training workflow. This guarantees that model updates are protected both with noise addition (DP) and computation on encrypted data (HE). |
| **Attack Resilience (ADRM)** | The system will actively monitor for and **block malicious model updates** sent by compromised clients. It will successfully defend against known threats like **Model Poisoning** and **Byzantine Attacks** by using statistical anomaly detection. |
| **Secure Communication (SCPEM)** | The framework will enforce a strict protocol requiring **mutual authentication** (TLS 1.3) between the server and all clients. It ensures that only authorized, verified IIoT devices can participate in the collaborative training process. |
| **Resource Efficiency** | The final code base will incorporate optimization techniques (e.g., quantization) to demonstrate **reduced communication overhead** and be suitable for deployment on resource-constrained IIoT devices. |

- 2. Limitations (What the Software Product(s) Will NOT Do)

- The scope is intentionally constrained to a research prototype and does not encompass full commercial deployment or certain advanced functionalities:
- **Not a Commercial Product:** The delivered software is a **proof-of-concept/research prototype**. It will not include a fully polished user interface, robust error handling required for commercial use, or compatibility with a wide variety of commercial IIoT platforms (beyond the simulated environment).
- **Limited Cryptographic Optimization:** While the framework implements HE, it will **not fully optimize the computational overhead** of the cryptographic protocols to a commercial-ready level. Latency caused by HE on extremely low-power devices is a recognized area for future work.
- **Limited Heterogeneous Support:** The framework, as a research output, is tested primarily on a specific model architecture (e.g., CNN for image/time-series data). It **will not** automatically support or be validated for a vast array of different heterogeneous models or industrial tasks without further modification.

- **No Defense Against Zero-Day Attacks:** The **ADRM** is designed to counter known and simulated attack patterns (poisoning, Byzantine). It is **not guaranteed to defend** against novel, **zero-day cyber-attacks** that use completely unknown methodologies.

# 8. Research constraints

These conditions placed boundaries on the development and evaluation of the Data-privacy based Federated Learning Framework for Industrial IoT:

## 1. Computational and Cryptographic Overhead

- The most significant constraint was the **heavy computational load** imposed by the chosen privacy mechanisms:

- **Homomorphic Encryption (HE):** HE guarantees absolute privacy but is computationally expensive. This limited the complexity and size of the AI models that could be used for testing, as the researchers had to ensure the encrypted operations could complete within a reasonable simulation time.
- **IIoT Device Limitations:** The need for the framework to run on simulated **resource-constrained IIoT devices** (limited CPU, memory, and battery) directly constrained the size of the neural network model, the frequency of communication, and the choice of cryptographic parameters.

## 2. Time and Scope Limitations

- **Limited Development Time:** The fixed duration of the research period constrained the team to develp a **prototype framework**. It was impossible to develop a fully optimized, commercial-grade product.
- **Focus on Core Functionality:** The team had to focus exclusively on validating the security, privacy, and efficiency mechanisms. This meant the project **did not include** the development of a user-friendly deployment tool or robust error-handling mechanisms required for real-world enterprise software.

## 3. Data and Generalization Constraints

- **Homogeneous Testing Environment:** The evaluation was conducted using a single, well-known dataset (CIFAR-10) and a single type of model (CNN). This limits the proven **generalizability** of the framework. The researchers could not definitively prove its efficiency across a vast, heterogeneous network of IIoT devices running different models and tasks (e.g., NLP, complex sensor fusion).

- **Simulated Environment:** The project was tested in a **simulated environment** rather than on live, operational industrial systems. This constraint means the empirical results do not account for all the unpredictable latency, dropped packets, and real-world noise inherent in a physical industrial network.

## 4. Security Scope Limitations

- **Known Attack Vectors:** The attack detection module (ADRM) was designed and validated against **known and simulated attack types** (Model Poisoning, Byzantine Attacks). The research was constrained by the inability to predict or defend against truly novel **zero-day cyber-attacks**.

# 9. Project plan

| Phase | Tasks/Activities | Key Deliverables | Suggested Timeframe |
|---|---|---|---|
| **I. Preparation & Review** | Initial review of existing system. Define clear objectives for "optimization." | **Project Scope Document** (defining what "optimized" means). Checklist of existing errors. | **1 Week** |
| **II. User Testing & Feedback** | Conduct **User Testings** (surveys, interviews, usability sessions). Collect and analyze all feedback. | **User Testing Report** (with categorized findings and prioritized issues). | **1-2 Weeks** |
| **III. Iteration & Refinement** | **Fix the Errors and adjust** the system based on the User Testing Report. Implement and rigorously test all fixes. | Fully tested, **optimized system version**. Updated code/system documentation. | **2-3 Weeks** |
| **IV. Internal Review & Practice** | Prepare materials for an internal **Presentation 2** (focusing on optimization changes). Conduct internal review of the adjusted system. | **Presentation 2** slides and script. Review sign-off/approval. | **1 Week** |
| **V. Demonstration Readiness** | Final checks, testing, and polish of the adjusted system. Create scripts and scenarios for the **Demonstration**. | **Demostration ready** system/build. Demo script and run-of-show. | **1 Week** |
| **VI. Documentation** | Write and finalize the **Research Paper** documenting the problem, methodology, optimization process, and results. | Complete, peer-reviewed **Research Paper** draft. | **2-4 Weeks** |
| **VII. Finalization & Delivery** | Prepare slides and talking points for the **Final Presentation**. Rehearse the presentation and demonstration. | Final, rehearsed **Final Presentation** materials. | **1 Week** |

# 10. References

[1] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. Agüera y Arcas, "Communication-efficient learning of deep networks from decentralized data," in Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), 2017[1].

[2] J. Konečný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Khan, and A. T. Suresh, "Federated learning: Strategies for communication cost reduction," arXiv preprint arXiv: 1610.05492, 2016[2].

[3] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," IEEE Signal Processing Magazine, vol. 37, no. 4, pp. 115-125, 2020[3].

[4] K. Bonawitz, V. Ivanov, B. Kreuter, A. Marcedone, H. B. McMahan, S. Patel, et al., "Practical secure aggregation for federated learning on user-held data," arXiv preprint arXiv: 1712.07119, 2017[4].

[5] Y. Aono, T. Hayashi, T. Ohara, and S. Sasaki, "Privacy-preserving deep learning via homomorphic encryption," in Proceedings of the 2017 Asia Conference on Computer and Communications Security (ASIACCS), 2017[5].

[6] J. Bell and R. Lai, "Private federated learning using homomorphic encryption," in Proceedings of the 2019 IEEE Global Communications Conference (GLOBECOM), 2019[6].

[7] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang. "Deep learning with differential privacy." in Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS), 2016[7].

[8] Y. Wei, W. Chen, J. Zhang, and S. Guo, "DP-FedAvg: Differential Privacy Enhanced Federated Averaging for Industrial IoT," Sensors, vol. 21, no. 3, p. 856, 2021[8].

[9] A. N. Bhagoji, S. Chakraborty, P. Mittal, and S. Calo, "Analyzing the robustness of federated learning against poisoning attacks," in Proceedings of the 36th International Conference on Machine Learning (ICML), 2019[9].

[10] C. Xie, S. Wang, T. Long, and X. Wang, "FL-GA: A Robust and Secure Federated Learning Framework against Poisoning Attacks," in Proceedings of the 2018 IEEE International Conference on Big Data, 2018[10].

[11] Y. Cao, R. Gu, Y. Wang, and G. Chen, "Byzantine-Resilient Federated Learning for Industrial IoT with Robust Aggregation," IEEE Transactions on Industrial Informatics, 2022[11].

[12] V. A. Nguyen and N. T. Binh, "A Hierarchical Federated Learning Framework for Industrial IoT in Edge Computing," Journal of Network and Computer Applications, vol. 182, p. 103038, 2021[12].

[13] S. Zeng, Y. Li, and D. Gao, "A Survey on Federated Learning in Industrial IoT: Challenges, Applications, and Future Directions," IEEE Internet of Things Journal, 2022[13].

[14] M. Wang, J. Wang, J. Liu, and Z. Li, "Resource-Efficient Federated Learning for Industrial IoT with Data Reduction and Model Compression," IEEE Transactions on Industrial Informatics, 2020[14].