



Zero-Trust Privacy-Preserving Framework for Decentralized Cyber Threat Intelligence

**Research Methods & Professional Development
(ACFA7004)**

Submitted by

Md Eyamin Molla

Student ID: 2425028

MSc in Cybersecurity & Digital Forensics

University Of Wales Trinity Saint David

Submission Date: 07-05-2026

Abstract

The growing sophistication of cyber threats, along with more stringent data laws such as the General Data Protection Regulation (GDPR), presents a challenge to organizations in sharing information to identify threats while maintaining privacy duties. While current Federated Learning (FL) systems mitigate the privacy leakage of training data by not sharing it directly between nodes, many existing FL frameworks remain vulnerable to two specific threats: gradient inversion attacks and Byzantine model poisoning attacks.

This project aims to develop, deploy, and test a Zero-Trust Privacy-Preserving Federated Learning Framework for decentralized cyber threat intelligence. The framework will incorporate Adaptive Differential Privacy along with a tunable epsilon budget, Krum and Fed Median Byzantine-resilient aggregation methods, and an Identity Verification process, compliant with the NIST SP 800-207, for all connected nodes. The system is tested with the CIC-IDS2024 dataset in three scenarios: FL baseline, 2. FL with static Differential Privacy, 3. The complete FL-ADP-ZTA system.

The proposed framework will be evaluated by simulating gradient inversion attacks, MIA, and up to 20% of poisoned nodes (Byzantine nodes) for a thorough security assessment. The expected result is a framework that is capable of attaining an F1-score of 0.90 or higher for intrusion detection systems under a moderate privacy budget ($\epsilon \leq 1.0$), and will be able to decrease gradient reconstruction attack success by at least 70% compared to non-privacy-preserving federated learning. The proposal addresses an evident research gap by presenting a framework that can facilitate effective privacy and collaborative threat intelligence sharing in highly regulated sectors such as finance, health, and critical national infrastructures.

Contents

Abstract	2
1. Introduction	4
1.1 Context and Background	4
1.2 Research Problem	5
1.3 Research Aim and Objectives	5
1.4 Significance of the Research	6
2. Literature Review	6
2.1 Introduction to the Literature Review	6
2.2 Federated Learning	6
2.3 Differential Privacy	7
2.4 Zero Trust Architecture	8
2.5 Adversarial Attacks on Federated Learning	8
2.6 Literature Summary and Research Gap	9
3. Research Design and Methodology	10
3.1 Research Design	10
3.2 Research Methods	11
3.3 Research Materials	13
4. Expected Outcomes	13
5. Project Timeline	14
6. References	15

1. Introduction

1.1 Context and Background

Cybersecurity has become a major topic of concern in this digital era. Since cyber threats have been constantly developing over recent years, the threat landscape now includes many sophisticated cyber-attacks such as ransomware, botnets, zero-day attacks, and Advanced Persistent Threats (APTs). As such, the demand for a collaborative intelligence infrastructure in which threat data can be disseminated and shared between participating organizations has increased [1]. But this is where the conventional model of threat intelligence fails: the centralized model involves collecting all data, such as raw network logs, security events, etc., into a central storage system. Therefore, the centralized data repository is a high-value target for attackers, since the attackers may potentially obtain information from multiple entities from only one successful attack [2].

With the emergence of Artificial Intelligence (AI) and Machine Learning (ML) in the cybersecurity field, the detection of complex attacks has been enhanced. AI-based Intrusion Detection Systems (IDS) are capable of processing and monitoring millions of network events at the same time and can identify anomalies that a rule-based IDS cannot [3]. However, AI and ML models need a large amount of data for training to be effective. And these data are not easy to collect when much private data is restricted by various organizations because of different compliance requirements, such as the General Data Protection Regulation (GDPR) introduced in the EU, or the various industry-specific compliance standards in different industries, e.g., financial, healthcare sector [4]. This proposal strikes a balance between the need for collaborative AI-based threat detection and the legal responsibility to safeguard sensitive information.

Federated Learning (FL) is an AI framework that provides a potential solution to these problems [5]. FL was introduced in 2017 by McMahan et al. as a framework where stakeholders participate in collaborative AI model training. In this setup, the AI model is trained in-situ, at each machine that owns the data, and each of these machines then sends updates to the model to a central server. The server will update the global model according to the aggregation of these model updates. But researchers have shown that gradients shared in FL can be vulnerable to the reconstruction attack. Zhu et al. showed that the model gradients exchanged by FL models have potential risks of privacy breach, as a majority (90%) of the original training data can be recovered by adversaries based on gradient exchange information [6]. This suggests that using FL as it is currently implemented will not be enough to preserve privacy in the adversarial setting of cybersecurity threat intelligence.

Differential Privacy (DP) [7] is an approach that aims to preserve the privacy of individual data records, proposed by Dwork et al. (2014), which adds noise to a mechanism's output, to prevent the leakage of individual information. Another cyber framework is Zero Trust Architecture (ZTA) [8], proposed by the U.S. National Institute of Standards and Technology (NIST), which assumes that all resources should not be trusted by default and are always under verification by both users and devices. The use of FL to protect AI model security has been proposed; DP for the protection of individual privacy has also been proposed, but there is not enough literature exploring the possibility of combining FL with DP, and combining the ZTA model

with the two of them, when used as an AI security framework for decentralized cyber threat intelligence [9].

1.2 Research Problem

The main question that this proposal addresses is that currently, there is no comprehensive and unified approach to provide collaborative AI-based threat detection while ensuring protection of organizational information from inference attacks and enforcing Zero Trust access control in a distributed network.

There are three underlying concerns that prompt this research question:

- First, the traditional IDS solutions usually rely on data aggregation in a centralized data store and hence have a single point of failure, and are incompatible with GDPR and other data protection regulations [4].
- Second, existing FL methods are vulnerable to gradient inversion attacks as demonstrated by Zhu et al. [6], as well as model poisoning attacks, in which attackers provide malicious updates to contaminate the global model [10].
- Third, while other studies like Pokhrel et al. [9] and Kawalkar & Bhoyar [11] have explored FL in conjunction with Zero Trust concepts, they treat data privacy and access control as orthogonal problems.

There is no work that has simultaneously considered Byzantine resilient aggregation, differentially-private federated learning, and Zero Trust validation in a framework specifically designed to handle cyber threat intelligence.

1.3 Research Aim and Objectives

Aim: Design, develop, and evaluate an integrated Zero-Trust Privacy-Preserving Federated Learning Framework for decentralized and collaborative cyber threat detection, which ensures mathematical security against gradient inference attacks and Byzantine model poisoning.

Objectives: In order to meet the above Aim, the following SMART objectives are formulated:

- By the end of Week 2, a systematic review of existing literature on Federated Learning, Differential Privacy, Zero Trust Architecture, and adversarial techniques in cybersecurity will be undertaken to identify existing research gaps.
- By the end of Week 4, a decentralized Federated Learning (FL) architecture for cyber threat analytics will be developed and deployed on the CIC-IDS2024 dataset [18], ensuring that data is stored only at simulated enterprise nodes.
- By the end of Week 6, an Adaptive Differential Privacy mechanism (with configurable epsilon values) will be integrated into the FL architecture, so that individual network traffic data is effectively anonymized.

- By the end of Week 7, the developed framework will be validated against Byzantine model poisoning attacks using Byzantine-robust aggregation algorithms (e.g., Krum [16], Fed Median) using up to 20% poisoned data (nodes).
- By the end of Week 8, the impact of the DP mechanism on the utility of the model will be validated using Membership Inference Attack (MIA) probes and standard classification accuracy metrics (e.g., accuracy, precision, recall, F1-score) to determine the optimal balance between utility and privacy.

1.4 Significance of the Research

This work is useful to various groups. For research, this study fills a gap in the existing literature (refer to [9,11]) for a holistic framework in the context of federated learning, adaptive detection, and zero-trust for cybersecurity.

From an application point of view, this work will help financial services companies, healthcare providers, and critical national infrastructure (CNI) to collaboratively detect and share threats, with the anonymity of client information guaranteed and without the GDPR and industry-specific policies acting as deterrents [4].

In terms of policy, this study can be a guide to support policymakers in realizing the 2022 National Cyber Strategy in the United Kingdom, which specifically calls for the importance of resilient, privacy-based, and trusted data-sharing [12].

2. Literature Review

2.1 Introduction to the Literature Review

This section provides a review of the existing body of scholarship applicable to this proposal, considering four interrelated areas: Federated Learning (FL) as a distributed AI approach, Differential Privacy (DP) as a formal privacy concept, Zero Trust Architecture (ZTA) as an access control paradigm, and the adversarial threat landscape within FL systems. The literature review will cover the literature published within a time frame of 2014-2026 (including the seminal literature and the latest empirical studies). This is to determine the state of the art, evaluate contributions and limitations of previous studies, and create the gap that this proposal fills.

2.2 Federated Learning

In their work in [5], McMahan et al. [5] proposed the concept of federated learning alongside Fed Avg, which trained a shared neural network over many thousands of heterogeneous mobile devices without moving the data [5]. They intended to share the updates from each client device and the central server as opposed to data samples, which is a major departure from the conventional paradigm of distributed machine learning, which relies on a centralized data repository [5].

The federated learning setting has recently been revisited in a structured cross-silo scenario [13] in which a few entities with large volumes of sensitive data are participating in a learning environment, and has been proposed as a viable solution for cybersecurity applications [13]. Some recent works have demonstrated that federated learning can be used for intrusion

detection systems [3], malicious malware classification [3], and network anomaly detection [3], which is appropriate because it could leverage a jointly learnt model while preserving data privacy [4].

While federated learning is a significant step towards the realization of distributed, privacy-aware learning, it is not without its weaknesses. For instance, Li et al. [10] demonstrated that an adversary in a federated learning system could subvert the model aggregation to cause the final global model to misclassify a particular pattern of malicious network traffic as well as reveal gradient information from other participating entities [10]. Fundamentally, Fed Avg [5] is predicated on all the participating devices in the federated system being trustworthy, which clearly is an unrealistic assumption in cybersecurity environments where the malicious actors may already have access to a compromised device or are even already within the organization's internal network. In fact, Kairous et al. [1] explicitly note the lack of robustness of federated learning, stating that it is perhaps the "most important open research challenge" in the area [1]. This provides a significant motivation to develop a federated intrusion detection system architecture that could be trained against both Byzantine and honest non-Byzantine attackers in a federated system.

2.3 Differential Privacy

Formalized by Dwork et al. [7], Differential Privacy provides a robust statistical guarantee that the output of any computation is unlikely to be sufficiently useful in revealing any given entry within the dataset. Differential Privacy adds calibrated Laplacian or Gaussian noise to model outputs or gradients according to a parameter called the privacy budget or epsilon. The lower the value of ϵ , the higher the amount of privacy, which leads to more noise and lower model accuracy [7]. However, the problem of employing Differential Privacy in a Federated Learning context is that it must find a balance between privacy and accuracy. Wang et al. [2] introduced an adaptive Differential Privacy mechanism that adapts the epsilon budget according to data sensitivity, achieving better detection accuracy than using a fixed Differential Privacy budget in a simulated cyber network. Their evaluation is based on a simulated cyber network with synthetic data. Furthermore, their approach does not consider Byzantine resilience, which we seek to address in this research. Another example of research on Differential Privacy applied in Federated Learning is the Membership Inference Attack defense surveyed by Hu et al. [14]. Hu et al. found that their proposed Differential Privacy mechanisms for Membership Inference Attack protection reduce privacy risk but with a high cost to classification accuracy; the privacy-accuracy trade-off is an ongoing problem in Federated Learning-based cybersecurity. Furthermore, neither of these Differential Privacy approaches in Federated Learning considers the proposed Zero Trust layer in this paper and would likely not work if a compromised node is trusted to train the Federated Learning model.

2.4 Zero Trust Architecture

Zero Trust Architecture (ZTA) is a cybersecurity strategy that assumes no one, no place, and nothing inside the corporate network boundary is trusted by default. NIST SP 800-207 states that Zero Trust is based on continuous verification, least-privilege access, strong identity validation, device posture assessment, and policy-based access control before granting access to resources [8]. It's a change from perimeter-based security, which means that internal users and systems are typically assumed to be trusted when they are allowed in.

Decentralized cyber threat intelligence is especially relevant in the case of Zero Trust, as Federated Learning requires nodes to report updates on the model to a central aggregator. An untrustworthy or malicious node can submit poisoned updates, try to manipulate the models, or try to manipulate the shared gradient. Thus, by verifying that each of the nodes in the Federated Learning network is authenticated, authorized, and continuously assessed before its contribution is incorporated into the global model, Zero Trust can enhance FL.

Recent studies have started to explore the relationship between Zero Trust and Federated Learning. A powerful Zero Trust framework utilizing blockchain-based Federated Learning and anomaly detection was proposed by Pokhrel et al. [9], and cloud security with the help of Federated Learning, AI-driven policies, blockchain, and Zero Trust principles was discussed by Kawalkar and Bhoyar [11]. However, the studies are not comprehensive enough to fully incorporate Zero Trust validation and adaptive DP and Byzantine robust aggregation in a single framework for Cyber threat Intelligence. This leaves a research gap. To fill this gap, the proposed study implements Zero Trust as a verification layer for controlling the participation of nodes; Differential Privacy for gradient information protection, and Byzantine-robust aggregation to mitigate the impact of poisoned model updates.

2.5 Adversarial Attacks on Federated Learning

Two specific classes of adversarial attacks are relevant to our investigation. First, gradient inversion attacks [6] have shown (Zhu et al.) that it is feasible for an attacker to reconstruct private data from shared gradients using their mathematical properties. They demonstrated that in an FL system that is only protected by standard data obfuscation, it is possible to recover over 90% of the original training data [6], invalidating the privacy claims of the naive FL approach. The second type of attack is Byzantine attacks, also referred to as model poisoning [15], in which the attacker injects poisoned updates to the global model to cause the global model to become less accurate (Fang et al.). Blanchard et al. [16] introduced the Krum algorithm as the first formal algorithmic defense against this problem, showing that a single step of robust aggregation is enough to eliminate the effects of the Byzantine attacks. However, Krum was evaluated on image classification tasks under the assumption of an IID data distribution, which is an unrealistic assumption on the network traffic data that is commonly seen in practical settings. In addition, Li et al. [10] extended the application of Byzantine-robust aggregation to a class-imbalanced setting, while also not investigating DP and Zero Trust verification alongside aggregation, a particularly detrimental omission. While the model poisoning attack would be thwarted by Byzantine-robust aggregation, verified but malicious nodes are able to execute gradient inversion attacks even before being removed by aggregation

mechanisms. Therefore, a co-designed multi-layered defense is a promising solution to mitigate these attacks.

2.6 Literature Summary and Research Gap

Table 1 below summarizes the key contributions and limitations of the literature reviewed.

Table 1: Summary of Key Literature

Author(s) & Year	Focus Area	Key Contribution	Limitation / Gap
McMahan et al. (2017) [5]	Federated Learning	Introduced the FedAvg algorithm for distributed ML training without sharing raw data	No privacy guarantees against gradient inference attacks
Dwork et al. (2014) [7]	Differential Privacy	Formal mathematical definition of DP; introduced the epsilon-delta privacy budget framework	Focused on static datasets; not adapted for dynamic cybersecurity environments
Zhu et al. (2019) [6]	FL Privacy Attacks	Demonstrated that gradient inversion attacks can reconstruct up to 90% of the original training data from model updates	Attack tested in vision tasks; limited evaluation in network traffic/cybersecurity contexts
Pokhrel et al. (2024) [9]	Decentralized FL & Zero Trust	Proposed an architecture combining FL security, differential privacy, and Zero Trust principles	Frameworks are treated independently; lacks integrated adversarial validation
Li et al. (2024) [10]	Byzantine-robust FL aggregation	Proposed contribution-wise Byzantine-robust aggregation for class-balanced federated learning	Does not integrate adaptive DP or Zero Trust verification for cyber threat intelligence
Kawalkar & Bhoyar (2024) [11]	Cloud Security & FL	Demonstrated that Zero Trust and FL can improve cloud security and privacy	Both mechanisms were studied separately; the integrated real-time security framework remains underexplored
Wang et al. (2024) [2]	Adaptive DP in Cyber Analytics	Proposed adaptive differential privacy mechanisms that tune the epsilon budget based on data sensitivity	Evaluation limited to simulated environments; Byzantine-resilience not addressed
Fang et al. (2020) [15]	Byzantine Model Poisoning	First systematic study of local model poisoning attacks against Byzantine-robust FL methods	Evaluated on general ML tasks; not applied to cyber threat intelligence or combined with DP/ZTA
Bai et al. (2024) [14]	Membership Inference Attacks	Developed defenses against Membership Inference Attacks (MIA) in federated settings	Defenses reduce model accuracy significantly; the trade-off balance needs further optimization

However, a review of the literature reveals a significant and enduring gap: while Fed Avg, DP, and ZTA have been investigated independently and in some instances in part combined, there is no existing literature that has explored an integrated solution for gradient inference attacks, Byzantine model poisoning, and Zero Trust node verification that is specific to cybersecurity and evaluated on real traffic [18]. As the above literature review highlights, an important area of research remains: whereas Federated Learning (FL), Differential Privacy (DP), and Zero

Trust Architecture (ZTA) have been widely investigated separately and in some limited combinations [9][11], an integrated and holistic assessment of these three techniques has never been performed. Existing research either considers such solutions to be modular "bolt-on" components [9], only evaluates their performance in a non-cybersecurity context [2], or only considers them under conditions when no adversary is present [10][14]. To address this, we will implement an integrated approach for gradient inference attacks, Byzantine model poisoning, and Zero Trust node verification and evaluate its performance using the CIC-IDS2024 benchmark data [18].

3. Research Design and Methodology

3.1 Research Design

We employ the Design Science Research (DSR) methodology [17] for this research because the goal of the proposed research is to develop and evaluate a novel artifact (the integrated Zero-Trust FL Framework) and not, as would be the case with observation or descriptive studies, to examine or explain a specific phenomenon. The DSR cycle includes the phases of problem identification, design of an artifact, implementation, and evaluation according to pre-defined standards [17]. Notably, this research approach is a standard for cybersecurity systems design research, where research contributions often come in the form of building and assessing prototyped mechanisms [3]. The research methodology will follow a quantitative and experimental approach. The proposed system will be implemented using real network traffic datasets and simulated network environments. The system's performance will be tested against established criteria, which include the accuracy of detection, robustness of privacy, and resiliency to attacks.

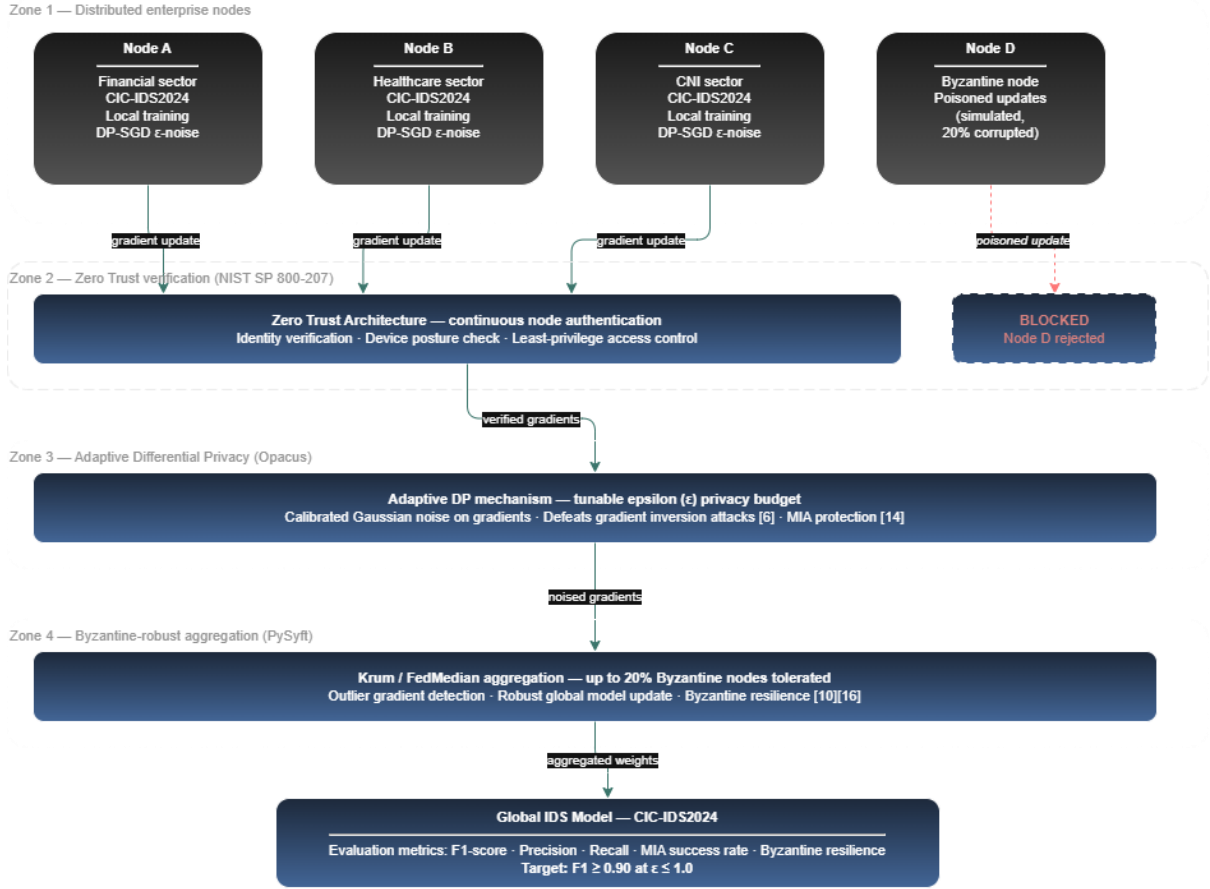


Figure 1: Zero-Trust Privacy-Preserving FL Framework (FL-ADP-ZTA)

The architecture shows that in enterprise nodes, local IDS models are trained from distributed CIC-IDS2024 data with no raw data shared (Zone 1). Then, noise gradient passes to Zero Trust node authentication, where Byzantine nodes will be filtered out and no longer be submitted to participate (Zone 2). After that, the verified gradient will pass through the adaptive differential privacy mechanism with a tunable epsilon budget (Zone 3), and the Byzantine-robust aggregation method Krum/Fed Median is used as the last mechanism before generating the global IDS model (Zone 4 and Zone 5).

3.2 Research Methods

The technical core approach is a Federated Learning + Adaptive Differential Privacy (FL-ADP) system. The rationale behind this combination is simple, based on the gap in the research: FL provides a data residency and GDPR compliance [5], and Adaptive DP is a tunable epsilon budget that hides gradient information and defeats gradient inversion attacks [2]. The implementation of zero trust is based on a non-stop node authentication and verification layer, which is based on NIST 800-207 [8], to prevent compromised nodes from reaching the aggregator.

The research method is experimental. Three conditions in which the experiments will be tested are: (1) a control FL system in the absence of privacy; (2) FL with static differential privacy; and (3) the complete FL-ADP-ZTA integrated system. This controlled comparison will enable

isolating the contribution of each component, which is a requirement of rigorous scientific assessment [17].

An approach using Red Teaming will be used to conduct adversarial validation. For Gradient inversion attacks, we will apply the approach of Zhu et al. [6]. For Membership Inference Attacks (MIA), we will apply the protocol by Li et al. [14]. For Byzantine model poisoning, up to 20% of the simulated FL nodes will be configured to become malicious and will submit corrupted model updates. The effects on global model accuracy will then be assessed both before and after applying Byzantine-robust aggregation based on the experimental protocol used by Li et al. [10].

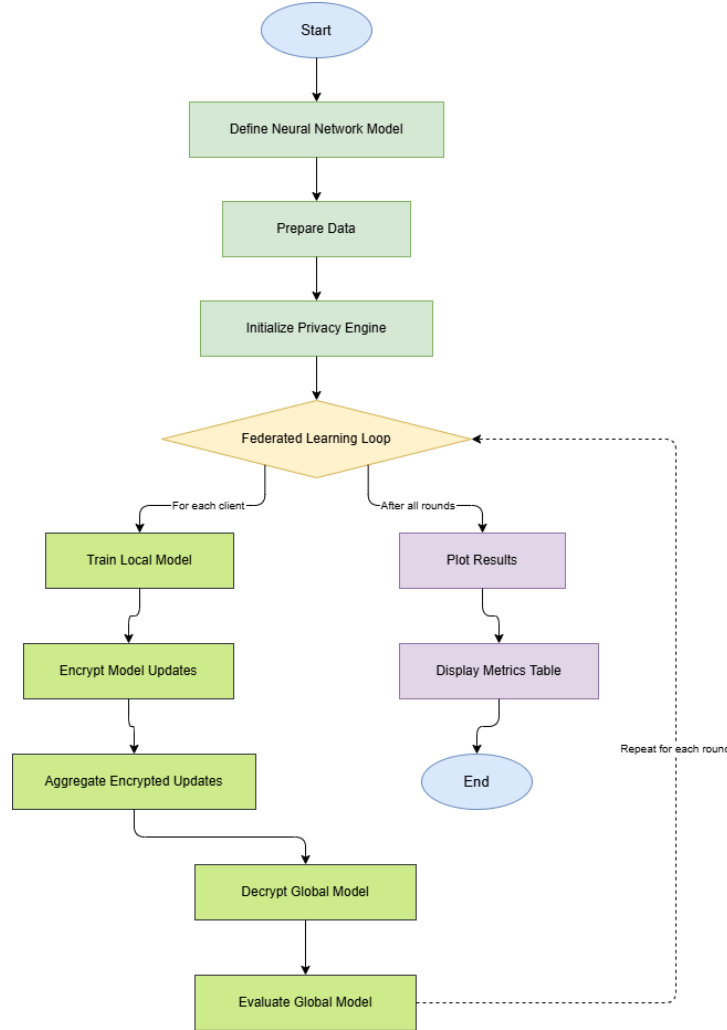


Fig. 2. Experimental workflow of the proposed privacy-preserving federated learning system for cyber threat intelligence.

The evaluation metrics that will be used are:

- Model Utility: Accuracy of classification, precision, recall, and F1-score on the IDS task with different levels of epsilon budget.
- Privacy Strength: Success rate of Membership Inference Attacks (low scores are better) and gradient reconstruction accuracy (low scores are better).

- **Adversarial Resilience:** It is a classification performance that is reduced when Byzantine nodes participate in more than 20 percent in classification, compared between baseline FL and the proposed framework.
- **Computational Overhead:** Round communication cost and training time, to measure feasible deploy ability.

3.3 Research Materials

Dataset: CIC-IDS2024 is the main dataset, created by the Canadian Institute for Cybersecurity [18]. This dataset was chosen due to the fact that it represents the latest standard regarding the latest types of cyber-attacks, such as DDoS, brute-force, API injection, and network infiltration patterns. **Reproducibility and comparability:** The publicly available, established benchmark is used to ensure reproducibility and comparability with the existing literature. The data set will be divided into simulated local nodes to recreate the non-IID (non-independent and identically distributed) nature of data distributions in real cross-organizational federated environments.

Software: Python 3.10 will be the major programming language. PySyft [19] will be used to implement the FL framework and offers a privacy-centric FL library with the ability to support differential privacy and integration of ZTA. Adaptive Differential Privacy mechanisms will be implemented based on the Opacus library [20], which offers GPU-optimized DP-SGD (Differentially Private Stochastic Gradient Descent). The parts of the neural network will be based on PyTorch. GitHub version control of code and fixed random seeds will make experimental reproducibility possible.

Hardware: Computationally intensive rounds of FL training will be performed on Google Collab Pro (NVIDIA A100 GPU) with a development and debugging workstation (RTX 3080, 32GB RAM) as a supplement. Computational overhead of DP-SGD [20] requires the use of GPU resources.

Ethical Considerations: CIC-IDS2024 is a completely synthetic dataset consisting of simulated network traffic that has no personally identifiable information (PII) [18]. There are no human subjects in this study. A form of ethics has been provided separately as required by the university. The data will be kept on encrypted storage approved by the university and will be deleted at the end of the project.

4. Expected Outcomes

It is expected that the combined FL-ADP-ZTA framework will excel over baseline FL across all three evaluation criteria. In particular, we expect the following results:

- The framework will attain an F1-score of at least 0.90 on DDoS and brute-force detection at a moderate privacy budget ($\epsilon \leq 1.0$), showing that cyber threat detection performance is possible even under stringent privacy requirements.
- The success rate of gradient inversion attacks will be at least 70% lower than in unsecured FL, confirming the effectiveness of the adaptive DP approach in preserving the privacy of gradients.

- IDS performance will be within 5% of the non-adversarial baseline with 20% Byzantine nodes, showing Byzantine-resilience.
- Overhead costs will be measured and put into context, offering empirical insights for the deployment of the framework.

The research implication would be a validated integrated model that builds on the previous studies, which have limited combinations [9][11]. Practically, it will be a template for deployment for organizations in regulated environments that want to participate in threat intelligence sharing with no privacy concerns.

Known limitations are the use of simulated rather than operational network environments, which might not entirely reflect diversity in real organizational data distributions. Further, the privacy-utility trade-off at low epsilon ($\epsilon < 0.1$) may prevent timely network threat detection, which will be empirically determined and evaluated.

5. Project Timeline

Table 2 presents a Gantt chart showing the planned phases and milestones for this research project over ten weeks.

Table 2: Project Gantt Chart (10-Week Plan)

Phase / Task	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10
Phase 1: Literature Review & Setup										
Phase 2: Dataset Prep (CIC-IDS2024)										
Phase 3: FL Architecture Design										
Phase 4: DP Integration & Testing										
Phase 5: Adversarial Evaluation (MIA/Byzantine)										
Phase 6: Results Analysis										
Phase 7: Write-up & Submission										

The key deliverables are:

- (1) a complete list of research papers and a research environment set up by end of Week 2;
- (2) a working Federated Learning prototype using the CIC-IDS2024 dataset by end of Week 4;
- (3) an integrated Federated Learning - Adaptive Differential Privacy - Zero Trust Access control system with successful unit tests by end of Week 6;
- (4) a set of results including adversarial evaluation by end of Week 8;
- (5) a final report on Week 10.

6. References

- [1] P. Kairouz and H. B. McMahan *et al.*, "Advances and Open Problems in Federated Learning," *Foundations and Trends® in Machine Learning*, vol. 14, no. 1–2, pp. 1–210, 2021, doi: <https://doi.org/10.1561/22000000083>.
- [2] Z. Wang, X. Yu, Q. Huang, and Y. Gong, "An Adaptive Differential Privacy Method Based on Federated Learning," *arXiv preprint arXiv:2408.08909*, 2024. [Online]. Available: <https://arxiv.org/abs/2408.08909>
- [3] A. A. Ferrag *et al.*, "Intrusion Detection Based on Federated Learning: A Systematic Review," *ACM Computing Surveys*, 2025, doi: <https://doi.org/10.1145/3731596>.
- [4] European Parliament and Council, "Regulation (EU) 2016/679 of the European Parliament and of the Council (General Data Protection Regulation)," *Official Journal of the European Union*, L 119, pp. 1–88, May 2016. [Online]. Available: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679>
- [5] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," in *Proc. 20th Int. Conf. Artificial Intelligence and Statistics (AISTATS)*, PMLR, vol. 54, pp. 1273–1282, 2017. [Online]. Available: <https://proceedings.mlr.press/v54/mcmahan17a.html>
- [6] L. Zhu, Z. Liu, and S. Han, "Deep Leakage from Gradients," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 32, 2019. [Online]. Available: <https://arxiv.org/abs/1906.08935>
- [7] C. Dwork and A. Roth, "The Algorithmic Foundations of Differential Privacy," *Foundations and Trends® in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, 2014, doi: <https://doi.org/10.1561/04000000042>.
- [8] S. Rose, O. Borchert, S. Mitchell, and S. Connelly, "Zero Trust Architecture," National Institute of Standards and Technology, Gaithersburg, MD, NIST SP 800-207, Aug. 2020, doi: <https://doi.org/10.6028/NIST.SP.800-207>.
- [9] S. R. Pokhrel, L. Yang, S. Rajasegarar, and G. Li, "Robust Zero Trust Architecture: Joint Blockchain based Federated Learning and Anomaly Detection based Framework," *arXiv preprint arXiv:2406.17172*, 2024. [Online]. Available: <https://arxiv.org/abs/2406.17172>
- [10] Y. Li, W. Ding, H. Chen, W. Bao, and D. Yuan, "Contribution-wise Byzantine-robust aggregation for Class-Balanced Federated Learning," *Information Sciences*, vol. 667, p. 120475, May 2024, doi: <https://doi.org/10.1016/j.ins.2024.120475>.
- [11] S. A. Kawalkar and D. B. Bhoyar, "Design of an Efficient Cloud Security Model through Federated Learning, Blockchain, AI-Driven Policies, and Zero Trust Frameworks,"

International Journal of Intelligent Systems and Applications in Engineering, vol. 12, no. 10s, pp. 378–388, 2024. [Online]. Available: <https://ijisae.org/index.php/IJISAE/article/view/4387>

[12] HM Government, "National Cyber Strategy 2022," Cabinet Office, London, Dec. 2021. [Online]. Available: <https://www.gov.uk/government/publications/national-cyber-strategy-2022>

[13] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated Learning: Challenges, Methods, and Future Directions," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 50–60, May 2020, doi: <https://doi.org/10.1109/MSP.2020.2975749>.

[14] L. Bai, H. Hu, Q. Ye, H. Li, L. Wang, and J. Xu, "Membership Inference Attacks and Defenses in Federated Learning: A Survey," *ACM Computing Surveys*, vol. 57, no. 4, Article 89, 2025, doi: 10.1145/3704633.

[15] M. Fang, X. Cao, J. Jia, and N. Z. Gong, "Local Model Poisoning Attacks to Byzantine-Robust Federated Learning," in *Proc. 29th USENIX Security Symposium*, pp. 1605–1622, 2020. [Online]. Available: <https://arxiv.org/abs/1911.11815>

[16] P. Blanchard, E. M. El Mhamdi, R. Guerraoui, and J. Stainer, "Machine Learning with Adversaries: Byzantine Tolerant Gradient Descent," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017. [Online]. Available: <https://proceedings.neurips.cc/paper/2017/hash/f4b9ec30ad9f68f89b29639786cb62ef-Abstract.html>

[17] A. Hevner, S. March, J. Park, and S. Ram, "Design Science in Information Systems Research," *MIS Quarterly*, vol. 28, no. 1, pp. 75–105, 2004, doi: <https://doi.org/10.2307/25148625>.

[18] Canadian Institute for Cybersecurity, "Intrusion Detection Datasets," University of New Brunswick, 2024. [Online]. Available: <https://www.unb.ca/cic/datasets/index.html>

[19] OpenMined, "PySyft: A Library for Easy Federated Learning," *GitHub*, 2019. [Online]. Available: <https://github.com/OpenMined/PySyft>

[20] A. Yousefpour *et al.*, "Opacus: User-Friendly Differential Privacy Library in PyTorch," *arXiv preprint arXiv:2109.12298*, 2021. [Online]. Available: <https://arxiv.org/abs/2109.12298>