electronics

MDPI

Article

FLIBD: A Federated Learning-Based IoT Big Data Management Approach for Privacy-Preserving over Apache Spark with FATE

Aristeidis Karras [1,*], Anastasios Giannaros [1], Leonidas Theodorakopoulos [2], George A. Krimpas [1], Gerasimos Kalogeratos [2], Christos Karras [1,*] and Spyros Sioutas [1]

[1] Computer Engineering and Informatics Department, University of Patras, 26504 Patras, Greece; giannaros@ceid.upatras.gr (A.G.); krimpas@upatras.gr (G.A.K.); sioutas@ceid.upatras.gr (S.S.)
[2] Department of Management Science and Technology, University of Patras, 26334 Patras, Greece; theodleo@upatras.gr (L.T.); gkalogeratos@upatras.gr (G.K.)
* Correspondence: akarras@ceid.upatras.gr (A.K.); c.karras@ceid.upatras.gr (C.K.)

Abstract: In this study, we introduce FLIBD, a novel strategy for managing Internet of Things (IoT) Big Data, intricately designed to ensure privacy preservation across extensive system networks. By utilising Federated Learning (FL), Apache Spark, and Federated AI Technology Enabler (FATE), we skilfully investigated the complicated area of IoT data management while simultaneously reinforcing privacy across broad network configurations. Our FLIBD architecture was thoughtfully designed to safeguard data and model privacy through a synergistic integration of distributed model training and secure model consolidation. Notably, we delved into an in-depth examination of adversarial activities within federated learning contexts. The Federated Adversarial Attack for Multi-Task Learning (FAAMT) was thoroughly assessed, unmasking its proficiency in showcasing and exploiting vulnerabilities across various federated learning approaches. Moreover, we offer an incisive evaluation of numerous federated learning defence mechanisms, including Romoa and RFA, in the scope of the FAAMT. Utilising well-defined evaluation metrics and analytical processes, our study demonstrated a resilient framework suitable for managing IoT Big Data across widespread deployments, while concurrently presenting a solid contribution to the progression and discussion surrounding defensive methodologies within the federated learning and IoT areas.

Keywords: federated learning; privacy preserving; data poisoning; Big Data systems; Apache Spark; FATE; IoT data management

Citation: Karras, A.; Giannaros, A.; Theodorakopoulos, L.; Krimpas, G.A.; Kalogeratos, G.; Karras, C.; Sioutas, S. FLIBD: A Federated Learning-Based IoT Big Data Management Approach for Privacy-Preserving over Apache Spark with FATE. *Electronics* **2023**, *12*, 4633. https://doi.org/10.3390/electronics12224633

Academic Editors: Chuan Zhang, Tong Wu and Weiting Zhang

Received: 18 October 2023
Revised: 8 November 2023
Accepted: 9 November 2023
Published: 13 November 2023

1. Introduction

Federated Learning (FL) is poised to play an essential role in extending the Internet of Things (IoT) and Big Data ecosystems by enabling entities to harness the computational power of private devices, thus safeguarding user data privacy [1]. Despite its benefits, FL is vulnerable to multiple types of assaults, including label-flipping and covert attacks [2–4]. The label-flipping attack specifically targets the central model by manipulating its decisions for a specific class, which can result in biased or incorrect results. To protect federated learning from data poisoning, researchers have devised techniques like differential privacy, which conceals user data details, and secure multi-party computation, which confidentially processes data across different sources. These methods strengthen FL's defences, preserving both data privacy and integrity [5–8]. To counter these threats, the research community has developed a plethora of privacy-preserving mechanisms and advanced techniques for improved model training, optimisation, and deployment while preserving the accuracy of the central model.

In the context of the Internet of Things and Big Data systems, Federated Learning (FL) has emerged as a vital paradigm for addressing the challenges associated with distributed data processing, privacy preservation, and resource utilisation. FL enables decentralised

machine learning on edge devices, facilitating efficient data processing without compromising privacy [9,10]. Federated Learning (FL) stands out as a pivotal approach to addressing the complex issues arising from the integration of the Internet of Things (IoT) with Big Data analytics. In the vast and dynamic landscape of the IoT, where Intelligent Transportation Systems (ITSs) stand as a cornerstone, FL is being adopted to navigate the complexities of privacy and data management. ITSs, which are a quintessential source of Big Data, benefit from FL's decentralised nature, enabling enhanced data privacy without compromising the utility of the information extracted from countless daily transactions [11]. In parallel, FL is advancing the field of healthcare, particularly in the management of Sexually Transmitted Infections (STIs), by providing a framework that respects patient confidentiality while harnessing large datasets for better disease control [12].

The potential of FL extends beyond healthcare into the automotive industry, particularly in the vehicular IoT. The system's ability to harness a multitude of end-user data for local model training makes it a promising alternative to traditional GPS navigation in urban environments. This is because it allows for the collection and processing of data points across diverse vehicular trajectories, thus enabling more-precise and context-aware navigation systems [13]. The privacy-centric design of FL ensures that sensitive user data remain on local devices, thereby reducing the risk of breaches and unauthorised access, a compelling advantage for massive user data applications [14].

Despite its significant advantages, FL is not without its vulnerabilities, with data poisoning attacks representing a salient threat to its security paradigm. These attacks compromise the integrity of learning models by injecting malicious data into the training process. Recent studies have shown that FL systems are susceptible to such threats, prompting an increase in research focused on fortifying these systems against potential breaches [15]. The research presented in [16,17] examined the security vulnerabilities of federated learning within IoT ecosystems and suggested measures to strengthen its security. Ultimately, as FL continues to be integrated across various sectors, ensuring the security of its systems against such attacks is imperative, warranting further investigation and development of sophisticated defence mechanisms, as illustrated by the extensive research in the studies [18–24,26].

In an FL scenario, the label-flipping attack is a significant concern because it can manipulate the central decisions of the model for a specific class, without reducing its overall accuracy. This manipulation can result in biased or incorrect results, which can have far-reaching consequences for businesses and individuals. To address this issue, the research community has developed various countermeasures to mitigate the risk of data poisoning attacks. The results of the experiments described in this paper provide valuable insights into the efficacy of these privacy-preserving mechanisms and their potential to mitigate the risk of data poisoning attacks in FL systems. The findings are particularly relevant for businesses that rely on FL for training machine learning models as they can help ensure the security and integrity of the models while promoting reliable results.

The contribution of this work is to initiate a comprehensive exploration of the security aspects of federated learning within the scope of IoT Big Data management using the perspective of the Federated Adversarial Attack for Multi-Task Learning (FAAMT) algorithm. Within our proposed model framework, we traversed the complex routes of managing large-scale data, ensuring that the inherent privacy features of federated learning are not jeopardised by potential adversarial attacks. Our research clearly outlines the potential vulnerabilities and susceptibilities of federated multi-task learning systems to data poisoning attacks, shedding light on essential insights to strengthen the robustness and security of the FLIBD approach. By integrating the robust, privacy-preserving Big Data management approach with a thorough analysis and mitigation of adversarial attacks via FAAMT and relevant defence mechanisms, this study creates a well-defined, stable foundation that ensures both the efficient management of IoT Big Data and the protection of collaborative, federated learning experiences in multi-task learning environments. Thus, our work presents a balanced combination of efficient data management and enhanced

security, driving the implementation of federated learning in real-world IoT and Big Data applications towards a secure, privacy-preserving future.

The remainder of this work is structured as follows. In Section 2, related work in the field of federated learning and data poisoning is surveyed and covered in detail along with state-of-the-art methods and data encryption applications. Section 3 outlines the methodology at both the theory and application levels, while Section 4 delves into the various types of attacks on federated learning. Section 5 highlights the experimental results and their findings. Finally, Section 7 concludes this work, and Section 7.1 provides a roadmap for future research directions.

2. Related Work

Federated Learning (FL) and data poisoning attacks are two crucial areas of research that have garnered significant attention in recent years. This section outlines the current progress and activities in these areas. FL is an innovative architecture designed to protect user privacy in machine learning environments in various areas [25–27]. It is commonly misunderstood, but this section provides examples to better understand its workings. For instance, when different companies aim to collaborate on a machine learning model training process, FL ensures that each company's local data remain internal by using encryption technology to transfer parameters between clients and the central server, thereby leading to the creation of a Global Model while preserving privacy.

Horizontal Federated Learning (HFL) is a subset of federated learning that splits datasets horizontally and removes data for training with the same user features, but different users [28,29]. This increases the number of user samples, but HFL may be vulnerable when the user attributes of two datasets overlap significantly, but the users themselves do not. To reduce private information exposure during the processing and transmission of components, HFL can employ homomorphic encryption systems [30–32], differential privacy mechanisms [33–35], and safe aggregation frameworks [36–38]. Other methods include blockchain-based FL [39,40] and multi-task FL [41,42].

Vertical Federated Learning (VFL) is used when the user attributes of two datasets partially overlap, but the users themselves overlap significantly [43–45]. This involves splitting the datasets vertically along a user/feature axis and removing data when users are identical, but user attributes vary. Approaches for VFL include SecureBoost, which suggests that all members input user attributes to train jointly, and a privacy-protecting logistic regression model based on VFL with parallelising objects for analysis and increased accuracy results [46–48].

Data poisoning attacks are an important topic of study in adversarial machine learning, which try to undermine machine learning algorithms [49,50]. These attacks have been studied on various machine learning models, including support vector machines [51], autoregressive models [52], collaborative filtering based on matrix factorisation [53], and neural networks for graph data [54–56]. In the context of multitask learning, Reference [57] offered a poisoning attack technique that differs from the situation in federated machine learning, where machine learning models are constructed based on datasets spread across various nodes/devices.

Federated machine learning is a rapidly growing field and offers opportunities for collaborative machine learning, but it also raises serious security and privacy concerns. To tackle these challenges, various defence strategies have been explored, including differential privacy, secure and robust aggregation, and outlier detection. Differential privacy has grown into an increasingly standard method for maintaining privacy in federated learning, but it can negatively impact the model's accuracy by introducing random noise to the data. Using secure and robust aggregation methods, such as Median-based aggregation, Trimmed Mean, Krum, and Bulyan, federated learning systems remain secure and robust. Outlier detection methods, such as rejection of adverse effects and Trim, identify and reject adversarial system interference proactively.

2.1. Data Security Encryption in Federated Learning: Relevant Literature and Approaches

Federated learning employs a distributed machine learning approach that facilitates training directly on devices, obviating the need to share sensitive data with a central server. Despite its advantages, concerns about data privacy and security persist in federated learning [58–61]. Various techniques, including homomorphic encryption, have been advanced to address these challenges. Homomorphic encryption is particularly noteworthy as it enables operations on encrypted data without necessitating their decryption, thereby maintaining data confidentiality.

In the study by [58], the researchers combined homomorphic encryption with cryptographic tools, such as masking and local model protection, to counter potential model inversion or reconstruction attacks, common threats in the domain of private financial or business data. Their findings affirmed that their proposed approach meets the data privacy standards. Another research work [59] conceptualised a blockchain-backed federated learning model for the Industrial Internet of Things (IIoT). This study introduced data protection schemes like distributed K-means clustering utilising differential privacy and homomorphic encryption and distributed AdaBoost with homomorphic encryption, emphasising multi-faceted data protection during the data and model sharing phases.

A different approach was presented in [60], where a system-level client selection method called Dubhe was introduced. This method allows clients to actively engage in training while ensuring their data privacy, using homomorphic encryption. The experiments revealed that Dubhe's performance, in terms of classification accuracy, is on par with the optimal greedy method, with minimal encryption and communication costs. Another mention is the study in [61], which offered an overview of challenges in federated learning and evaluated existing solutions, notably featuring homomorphic encryption.

Expanding on recent work in federated learning security, the study conducted by Fan et al. introduced a novel data-sharing scheme to enhance both security and efficiency [62]. Within this framework, three principal entities—the Cloud Service Provider (CSP), Data Provider (DP), and Data Requester (DR)—collaborate. Essentially, the DP uploads private data and potentially a re-encryption key to the CSP, allowing data re-encryption to specific user groups. Subsequently, the DR can request and decrypt this re-encrypted data using its private key. The scheme outlines eight critical algorithms, presenting an approach that holds significant promise for improving data-sharing protocols in FL ecosystems.

Following the exploration into federated learning security, another notable study detailed a sophisticated encryption algorithm tailored for plaintext images [63]. This process commences by segmenting the image into R, G, and B channels, with subsequent encryption operations applied uniformly across these channels. Key steps in this approach include the separation of channels, leveraging the 2D-LCLM complex chaotic system to generate pseudo-random sequences, and employing the Zhou Yi Eight Trigrams encryption rule to finalise the encryption. This method underscores the evolution of encryption techniques suitable for multimedia data in modern research landscapes.

In the context of images, as previously noted, another study delved into the scope of blockchain security by highlighting a traceability model for the DAG blockchain system [64]. As images demand robust encryption methods, blockchains seek dependable verification systems. This model applies a witness mechanism, encompassing everything from unit addition to the blockchain to information retrieval for users. In this context, a "unit" serves as the primary data container, encapsulating vital information such as hash values and the public key of the uploader. A defining feature of this model is its steadfast commitment to data integrity. Once a unit is validated, modifying its content becomes challenging since all interconnected units would require alterations, ensuring that unauthorised changes are nearly impossible.

A study examined adaptable encryption and decryption frameworks for CKKS homomorphic encryption, enabling computation on encrypted data without decryption [65]. The CKKS scheme, reliant on the RLWE problem and approximate arithmetic, encodes real numbers into polynomials for encryption with a public key. After remote computa-

tion on encrypted data, the client decrypts the received results with its secret key. The paper highlights the importance of secure key generation and evaluation key creation for homomorphic operations. The proposed architectures aim to enhance the efficiency of CKKS encryption without delving into extensive data security details [65]. In conjunction with adaptable encryption for CKKS, the study expects future research to integrate homomorphic encryption with machine learning, potentially exploring HE-integrated federated learning for enhanced data privacy.

Ultimately, reviews of the literature on data security in federated learning emphasise strategies such as homomorphic encryption and blockchain to protect data in machine learning. Various methods are proposed to safeguard sensitive information, from cryptographic techniques to innovative data-sharing schemes, bolstering security while maintaining efficiency. The aim is to forge robust federated learning systems that guarantee data privacy, with emerging research exploring the integration of these security methods with homomorphic encryption for enhanced machine learning applications.

2.2. Privacy Threats in Federated Learning

Privacy challenges that are inherent within Federated Learning (FL) architectures, particularly pertaining to the extensive interaction across various participating entities, pose significant research concerns [66]. One such crucial threat is the Deep Gradients Leakage attack, which is strategically leveraged by an adversary who acts as an aggregator. Within this attack schema, the adversary capitalises on exploiting the gradients of the model with the underlying intent to extrapolate or infer the private data of the individual participants [67]. This form of attack underscores an intricate manipulation of data and, quite conspicuously, has a direct impact on the inherent privacy of data that are circulated within the FL architecture. Furthermore, the assimilation of Generative Adversarial Network (GAN) attacks within the adversarial framework often sees the attacker using GANs to meticulously generate data, which mirrors the private data of participants [67]. The data generation here is formulated such that it concurrently is metaphorically camouflaged, disguising it as legitimate data and, thus, introducing potential jeopardy to the integrity and confidentiality of the original data.

In conjunction with the previous attacks, the looming threat of poisoning attacks and inference attacks magnifies the privacy dilemma in FL. The former envisages a scenario wherein an adversarial participant malevolently injects data into the training process with the aim to subtly, yet systematically, manipulate the Global Model, consequently propagating erroneous inferences [68]. This perturbation of the learning process is not just detrimental to the model accuracy, but also corrodes the authenticity of predictions, potentially cascading to flawed decision-making processes. On the other hand, inference attacks extrapolate this issue, where an adversary infers the private data of other participants through a strategic exploration of the Global Model [68]. Moreover, range inference attacks refine this adversarial strategy by attempting to ascertain the range of the private data of participants, providing a discreet, yet robust mechanism to violate privacy norms [69]. Thus, these types of attacks lead to a wide and deep invasion of privacy, calling for new and effective strategies to mitigate them.

Addressing privacy threats in Federated Learning (FL) invokes a multilayered approach where technologies like blockchain and Trusted Execution Environments (TEEs) have been significant, yet not entirely impervious to certain attack vectors. The incorporation of blockchain technology serves as a decentralised ledger, which aids in securing transparent record-keeping and transactions, thereby mitigating single-point failures and ensuring a level of accountability within the FL paradigm [67]. Concurrently, TEEs ensure secure computational zones, safeguarding data and models during the computation and offering a protective shield against a spectrum of threats.

However, it is essential to acknowledge that current FL protocols exhibit certain deficiencies in rendering an all-encompassing security framework, thus spotlighting a crucial necessity for more-detailed and in-depth research in this arena [70]. The pressing

requirement pivots around designing a security architecture that seamlessly blends robustness and comprehensiveness while also manifesting adaptability to the ever-evolving threat landscape. This ensures sustained, privacy-preserving FL operations among the intricate and dynamic cyber–physical interactions in large-scale IoT systems [71]. This establishes a prolific domain for continuing and prospective research, with a dedicated focus on embedding a sophisticated and well-articulated balance of security and privacy within the FL paradigm. Ensuring that innovative solutions are not merely confined to theoretical frameworks, but extend to practical viability in real-world deployments also stands as a pivotal aspect. Such a pursuit not only enriches the academic discourse around privacy-preserving mechanisms in FL, but also contributes substantively to the operational robustness of FL in large-scale systems where data privacy and security are paramount.

2.3. Privacy and Security in Federated Learning Systems: State-of-the-Art Defensive Mechanisms

Federated Learning (FL) emerges as a pivotal methodology, enabling Machine Learning (ML) application directly on devices while safeguarding sensitive and private information from unwarranted disclosure and tracking. Despite its innovative approach, FL's security framework invites further scrutiny, especially within sectors managing exceptionally sensitive data, such as the healthcare industry [72–74]. Vulnerabilities in FL, including susceptibility to model poisoning, data heterogeneity, and model inversion attacks, possess the potential to undermine the efficacy of the Global Model [72–76].

Various defensive tactics have been introduced to counter these threats, such as the implementation of robust aggregation algorithms, deploying Multi-Party-Computation (MPC)-based secure aggregation, and the utilisation of trained autoencoder-based anomaly-detection models [73–75]. Notably, several state-of-the-art defences against model poisoning attacks, including FedCC, Krum, and Trimmed Mean, have been articulated in the existing literature [75,77].

Nevertheless, these strategies often provide solutions that are parallel and non-intersecting with respect to individual attacks or concerns. Moreover, the meticulous orchestration of collusion among malicious participants can subtly reduce the bias triggered in the poisoned local model—minimising disparities from the poison-free model. This subtlety becomes critical in facilitating stealthy backdoor attacks and eluding a myriad of top-tier defence strategies currently available in FL [76]. Thus, a void exists, signalling an exigent need for additional research aimed at devising potent and encompassing defensive mechanisms to bolster the security infrastructure of FL systems.

Romoa stands out in the arena of Federated Learning (FL) as it applies a logarithm-based normalisation strategy, steering clear of the pitfalls associated with scaled gradients that originate from nefarious entities. This strategic model aggregation method acts as a bulwark against model poisoning attacks, a pertinent concern in the FL framework, where several decentralised nodes collaborate in model training. The stabilisation and integrity of the model during its training phase are crucial to ensure the derived insights and applications remain valid and reliable. Hence, Romoa not only addresses the immediate concerns related to malicious activities in FL, but also underscores the necessity of innovatively confronting challenges to uphold the robustness of decentralised learning models.

Concurrently, Robust Federated Aggregation (RFA) demonstrates a divergent, yet equally significant methodology, emphasising the utilisation of a weighted average of gradients to fortify FL systems against Byzantine attacks. The pertinence of resisting such attacks is elevated in sensitive domains, such as healthcare, where the accuracy and reliability of models can directly impact decision-making and outcomes. RFA, through its adept handling of gradients and ensuring the integrity of the aggregation process, helps to sustain the credibility of FL in environments where malicious actors might seek to destabilise the model. Thus, RFA emerges not merely as a defensive mechanism, but as a vital cog ensuring the seamless operation of FL systems, especially where the veracity of aggregated models is critical.

While Romoa and RFA significantly advance the security mechanisms within FL systems, the journey towards a thoroughly secure, decentralised learning environment remains ongoing and necessitates continual research and development. This becomes particularly poignant in the realms of the IoT and Big Data, where large volumes of data are processed and analysed across various nodes. The challenge extends beyond merely defending against known threats to preemptively identifying and mitigating potential future vulnerabilities within the FL paradigm. The continual evolvement of defence mechanisms, in tandem with the evolution of threats, underscores the dynamic and complex nature of securing FL systems. Therefore, it is imperative for the research community to remain engaged in a persistent exploration of innovative strategies and mechanisms to safeguard FL against a spectrum of threats, ensuring its viability and trustworthiness in diverse applications across varied domains.

Navigating through the landscape of defensive mechanisms in federated learning, a variety of strategies have been spotlighted, each exemplifying unique approaches and methodologies towards mitigating adversarial endeavours. The summary presented in Table 1 encapsulates a selection of these mechanisms, illustrating the diversity and specificity with which each strategy is forged and employed. Notably, while strategies like FedCC and Krum emphasise robust aggregation algorithms and server-side defences, respectively, others like FL-WBC introduce client-based strategies to shield the federated learning model from adversarial attacks. This table not only serves as a confluence of varied defensive strategies, but also underscores the multifaceted nature of the challenges that federated learning systems encounter in maintaining model integrity and privacy preservation. Thus, the mechanisms detailed in Table 1 establish a foundation for a more-detailed and -subtle examination of the architectural and functional aspects of these defences, thereby facilitating subsequent research and progression in the spectrum of secure and robust federated learning. This overview of different mechanisms seeks to build a basic understanding, which will help direct future research and development paths in the field.

Table 1. Summary of defensive mechanisms in federated learning systems.

Mechanism	Description	Reference
FedCC	Employs a robust aggregation algorithm, mitigating both targeted and untargeted model poisoning or backdoor attacks, even in non-IID data scenarios.	[75]
Krum	Acts as a server-side defence via an aggregation mechanism, but may be susceptible to Covert Model Poisoning (CMP).	[77]
Trimmed Mean	Server-side aggregation similar to Krum, yet also potentially prone to CMP, aiming to resist model poisoning attacks.	[77]
MPC-Based Aggregation	Mitigates model inversion attacks by employing a trained autoencoder-based anomaly-detection model during aggregation.	[74]
FL-WBC	A client-based strategy that minimises attack impacts on the Global Model by perturbing the parameter space during local training.	[78]
Romoa	Utilises a logarithm-based normalisation to manage scaled gradients from malicious vehicles, resisting model poisoning attacks.	[79]
RFA	Utilises a weighted average of gradients to resist Byzantine attacks, aiming to establish a robust aggregation method.	[80]

2.3.1. Limitations of Current Defensive Mechanisms in Federated Learning

The development and implementation of defensive mechanisms in Federated Learning (FL) have been imperative for ensuring secure and robust model training in decentralised learning environments. Nonetheless, prevailing methods manifest substantial limitations, hampering their optimal functionality and efficacy in practical scenarios.

Computational and Communication Overhead

A significant limitation is the computational burden imposed on Edge Devices (EDs), which often possess restricted computational resources. This limitation arises from the heavy computation overhead, which is further exacerbated when these devices are tasked with conducting complex learning or defensive processes. Concurrently, the communication overhead is another crucial aspect, predominantly related to the uploading of converged local models' parameters to a centralised server, where aggregation is performed. This not only demands substantial communication resources, but also exposes the system to potential communication-related vulnerabilities [81].

Knowledge Preservation and Incremental Learning

Moreover, a crucial challenge is associated with the guarantee of preserved knowledge, especially in the context of incremental learning over new local datasets. The current defensive mechanisms may jeopardise the acquired knowledge due to the absence of a solid framework that ensures the stability and plasticity of the learned models during subsequent learning phases. This issue is particularly prominent when FL systems encounter novel data, and adaptive learning becomes crucial to preserving and updating the Global Model appropriately [82].

Security and Privacy Concerns

Security concerns, including susceptibility to various attacks (such as model poisoning, privacy inference, and Byzantine attacks), underscore another essential limitation. While privacy preservation is a cornerstone of FL, ensuring robust defence against intricate attack strategies, particularly those exploiting the decentralised nature of FL, remains a pressing concern. The vulnerabilities related to privacy inference attacks, which aim to infer sensitive information from the shared model updates, and Byzantine attacks, where malicious actors disseminate falsified updates, are notably challenging to mitigate comprehensively [82]. This notwithstanding, novel approaches, namely Romoa and RFA, have been proposed to address some of these challenges by introducing advanced aggregation methods designed to resist various attacks while ensuring robust model training [81].

2.4. Problem Statement

Navigating the management of voluminous data derived from the Internet of Things (IoT) environment, coupled with ensuring privacy within expansive systems through Federated Learning (FL), highlights a complex and multi-dimensional challenge. Specifically, our discourse pivots on the following predominant axes of difficulties:

- **Massive and rapid IoT data:** IoT environments are characterised by the generation of immense volumes of data (D_i) at an astounding velocity, making effective and efficient data management imperative to prevent systemic bottlenecks and to ensure sustained operational performance.
- **Preserving privacy with FL:** Ensuring the data (D_i and D_t) remain localised and uncompromised during Global Model training in FL demands robust methodologies to prevent leakage or unintended disclosure of sensitive information through model updates (**W**).
- **Label-flipping attacks:** These attacks, wherein labels of data instances are maliciously altered ($\tilde{D}i$ and $\tilde{D}s$), present a pronounced threat to model integrity and efficacy in

FL. Here, the design and implementation of defensive mechanisms that can detect, mitigate, or recover from such attacks is of paramount importance.

$$\mathcal{H}(\tilde{D}i, \tilde{D}s; \mathbf{W}) \rightarrow \min_{\mathbf{W}} L(D_i, D_t, \mathbf{W}) \quad (1)$$

where $\mathcal{H}$ is the high-level function that seeks to minimise the loss function L given the perturbed data and weight matrix, ensuring the learned model $\mathbf{W}$ is resilient against the attack.

- **Ensuring scalability:** Addressing the scale, by ensuring the developed FL model not only counters privacy and security threats, but also scales efficiently to manage and process expansive and diverse IoT Big Data.
- **Technological integration and novelty:** While FATE offers a promising federated learning framework and Apache Spark offers fast in-memory data processing, exploring and integrating these technologies in an innovative manner that cohesively addresses IoT Big Data management challenges within an FL paradigm becomes crucial.

$$\mathcal{H}(D_i, D_t; \mathbf{W}, \mathbf{\Omega}) \rightarrow \min_{\mathbf{W},\mathbf{\Omega}} L(D_i, D_t, \mathbf{W}, \mathbf{\Omega}) \quad (2)$$

where $\mathcal{H}$ is aimed at minimising the loss function L concerning the data, the weight matrix $\mathbf{W}$, and the model relationship matrix $\mathbf{\Omega}$, ensuring a harmonised functionality between the integrated technologies and, also, enabling scalable and efficient data processing and model training across federated nodes.

The core objective of FLIBD is to diligently construct a framework that mitigates the identified challenges. This venture not only seeks to resolve current issues, but also aspires to craft a model that leads, moulds, and enhances forthcoming technological and methodological progress within the sphere of privacy-preserving data management in expansive IoT deployments, leveraging federated learning.

2.5. Proposed Architecture

FLIBD formulates an insightful architecture, intending to skilfully manage voluminous IoT-originated data whilst concurrently ensuring meticulous privacy preservation across large-scale systems, achieved by integratively employing Federated Learning (FL), Apache Spark, and FATE. The fundamental layers and their concomitant functionalities are delineated below and represented in Figure 1:

1. **IoT data harvesting and initial processing layer:**
   - Dynamic data acquisition: implements a strategic approach to dynamically harvest, categorise, and preliminarily process extensive IoT data, utilising Apache Spark's proficient in-memory computation to adeptly handle both streaming and batch data.
2. **In situ model training and data privacy layer:**
   - Intrinsic local model training: employs FATE to facilitate localised model training, reinforcing data privacy by ensuring data are processed in situ.
   - Data and model security mechanism: integrate cryptographic and obfuscation techniques to safeguard data and model during communication, thus fortifying privacy and integrity.
3. **Federated learning and secure model consolidation layer:**
   - Privacy-aware federated learning: engages FATE to promote decentralised learning, which encourages local model training and securely amalgamates model updates without necessitating direct data exchange.
   - Model aggregation and resilience: establishes a secure aggregation node that amalgamates model updates and validates them against potential adversarial actions and possible model poisoning.

4. **Global Model enhancement and feedback integration layer:**
   - Deploy, Enhance, and Evaluate: apply the Global Model to enhance local models, instigating a comprehensive evaluation and feedback mechanism that informs subsequent training cycles.
5. **Adaptive scalability and dynamic management layer:**
   - Dynamic scalability management: utilises Apache Spark to ensure adaptive scalability, which accommodates the continuous data and computational demands intrinsic to vast IoT setups.
   - Proactive system management: implements AI-driven predictive management and maintenance mechanisms, aiming to anticipate potential system needs and iteratively optimise for both performance and reliability.

In essence, the FLIBD architecture aspires to function as a robust foundation for upcoming advancements in privacy-preserving data management within the continuously evolving IoT environment. It makes an effort to navigate present challenges with viable solutions, whilst concurrently establishing a robust framework beneficial to encouraging future research and development in privacy-preserving methodologies for managing IoT Big Data across large-scale scenarios.

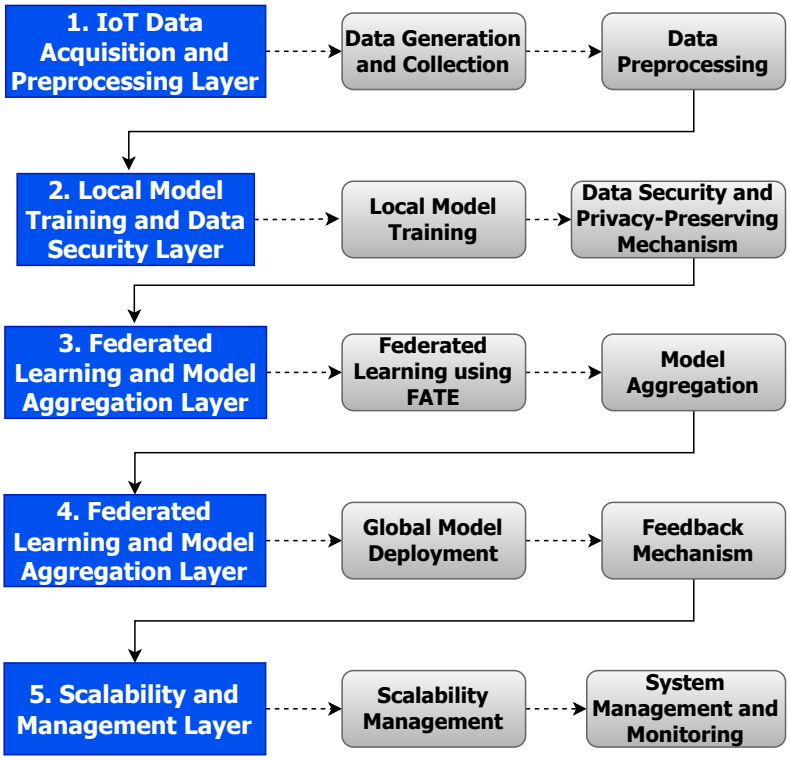


Figure 1. The basic architecture.

3. Methodology

In this section, we provide a comprehensive overview of the steps taken in the data poisoning attack strategies that we propose for federated machine learning. Initially, the concept of federated multi-task learning functions as a general framework for multi-task learning in a federated environment is given. We then present the mathematical formulation of our poisoning attack and describe how to further optimise the model. The abbreviations for the variables used in our work are given in Table 2. To make it simpler for the reader to follow our methodology and comprehend the underlying principles, we introduce and elucidate these notations in depth throughout the paper. Moreover, we provide a representation of our methodology, which includes plots, to help the reader understand the steps involved in our proposed attack.

Table 2. Notation for utilised variables.

Symbol	Interpretation
D_i	Uninfected data for node i
D_t	Uninfected data for target node t
$\tilde{D}i$	Infected data for node i
$\tilde{D}s$	Infected data for attacking node s
$\mathcal{T}$	Set of target nodes
$\mathcal{S}$	Set of attacking nodes
$\mathcal{H}$	Upper level function
$\mathbf{W}$	Weight matrix
$\mathbf{\Omega}$	Model relationship matrix

Objectives of Federated Machine Learning

In this section, we address the issue of data poisoning attacks in federated multi-task learning. We introduce three types of attacks based on real-world scenarios and propose a bilevel formulation to determine the optimal attacks. The goal of the attacker is to degrade the performance of a set of target nodes $\mathcal{T}$ by injecting corrupted or poisoned data into a set of source attacking nodes $\mathcal{S}$. The objective of federated machine learning is to develop a model from data generated by n distinct devices denoted as $\mathcal{D}_i$ with $i \in [n]$, which may have different data distributions. To address this, separate models $\mathbf{w}_1, \ldots, \mathbf{w}_n$ are trained on individual local datasets. The focus of this paper is on a horizontal (sample-based) federated learning model known as federated multi-task learning [83], which is a general multi-task learning framework applied in a federated setting. The federated multi-task learning model can be represented as:

$$\sum_{i=1}^{N} \sum_{j=1}^{M_i} \mathcal{L}i\Big(\mathbf{w}i^\top \mathbf{x}i, j, \mathbf{y}i, j\Big), \tag{3}$$

regularised by the trace of the product $\mathbf{W\Omega W}^\top$ scaled by λ_1 and the Frobenius norm of $\mathbf{W}$ squared scaled by λ_2:

$$\lambda_1 \mathrm{Tr}\Big(\mathbf{W\Omega W}^\top\Big) + \lambda_2 |\mathbf{W}|_F^2. \tag{4}$$

The symbol Tr denotes the trace, and $|\mathbf{W}|_F$ denotes the Frobenius norm of $\mathbf{W}$.

Here, $(\mathbf{x}_i^j, \mathbf{y}_i^j)$ is the j-th sample of the i-th device and m_i represents the number of clean samples in the i-th device. The matrix $W \in \mathcal{R}^{d \times n}$, with the i-th column being the weight vector for the i-th device, comprises the weight vectors $\mathbf{w}_1, \ldots, \mathbf{w}_n$. The relationships among the devices are modelled by $\Omega \in \mathcal{R}^{n \times n}$, and the regularisation terms are controlled by the parameters $\lambda_1 > 0$ and $\lambda_2 > 0$.

In the context of federated machine learning, finding the $\mathbf{\Omega}$ matrix can be computed separately from the data, as it is independent of them. A key contribution of [83] was the creation of an effective method for distributed optimisation to update the $\mathbf{W}$ matrix. This was accomplished by incorporating the ideas of distributed primal–dual optimisation, which were previously outlined in [84].

Let $m = \sum_{i=1}^{n} m_i$ and $\mathcal{X} = \mathrm{Diag}(\mathcal{X}1, \ldots, \mathcal{X}n) \in \mathbb{R}^{nd \times m}$. Holding the $\mathbf{\Omega}$ matrix constant, the optimisation problem can be re-expressed in its dual form with the dual variables $\alpha_i \in \mathbb{R}^{m_i}$ as follows:

$$\min \boldsymbol{\alpha}, \mathbf{W}, \mathbf{\Omega} \sum i = 1^n \sum j = 1^{m_i} \mathcal{L}_i\Big(-\boldsymbol{\alpha}_i^\top \mathbf{x}_i^j\Big) + \lambda_1 \mathrm{Tr}\Big(\mathbf{W\Omega W}^\top\Big). \tag{5}$$

where Tr denotes the trace function and λ_1 is a hyperparameter. The optimisation process of finding the $\mathbf{\Omega}$ matrix is separable from the data, allowing it to be calculated at a central location. The study by [83] contributed a robust method for distributed optimisation to refine the $\mathbf{W}$ matrix, building upon the concepts of distributed primal–dual optimisation described in [84]. The optimisation involves the conjugate dual function $\mathcal{L}i^*$ of $\mathcal{L}i$ and the term λ_1 that

regularises $\mathrm{Tr}(\mathbf{W}\mathbf{\Omega}\mathbf{W}^\top)$. The variable α_i^j in $\oslash \in \mathcal{R}^n$ represents the j-th sample $(\mathbf{x}_i^j, \mathbf{y}_i^j)$ from the i-th device. In this work, we designate $D_i = (\mathbf{x}_i^j, \boldsymbol{\alpha}_i, \mathbf{y}_i^j) \mid \mathbf{x}_i^j \in \mathbb{R}^d, \boldsymbol{\alpha}_i \in \mathbb{R}^{m_i}, \mathbf{y}_i^j \in \mathbb{R}^{m_i}$ as the uncorrupted data in device i. The malicious data injected into node i are represented as:

$$\tilde{D}_i = (\tilde{X}_i, \tilde{\alpha}_i, \tilde{y}_i) \mid \tilde{X}_i \in \mathbb{R}^{d \times \tilde{n}_i}, \tilde{\alpha}_i \in \mathbb{R}^{\tilde{n}_i}, \tilde{y}_i \in \mathbb{R}^{\tilde{n}_i} \tag{6}$$

where $\tilde{n}_i$ is the number of injected samples for node i. If $i \notin \mathcal{S}$, then $\tilde{D}_i = \emptyset$, meaning that $\tilde{n}_i = 0$. The three types of attacks we focused on for FL systems are:

- Direct attack: $\mathcal{T} = \mathcal{S}$. The attacker can directly inject data into all the target nodes, exploiting a vulnerability during data collection. For instance, in the case of mobile phone activity recognition, an attacker can provide counterfeit sensor data to directly attack the target mobile phones (nodes).
- Indirect attack: When no direct data injection is possible into the targeted nodes, symbolised by $\mathcal{T} \cap \mathcal{S} = \emptyset$, the attacker can exert influence on these nodes indirectly. This is achieved by introducing tainted data into adjacent nodes, exploiting the communication protocol to subsequently impact the target nodes. Such an attack leverages the interconnected nature of the network, allowing for the propagation of the attack effects through the established data-sharing pathways.
- Hybrid attack: This form of attack represents a combination of both direct and indirect methods, wherein the aggressor has the capability to contaminate the data pools of both the target and source nodes at once. By employing this dual approach, the attacker enhances the potential to disrupt the network, manipulating data flows and learning processes by simultaneously injecting poisoned data into multiple nodes, thereby magnifying the scope and impact of the attack across the network.

The objective to maximally impair the functioning of the target nodes is structured as a two-tiered optimisation challenge, in line with the model proposed by [55]:

$$\max_{\tilde{D}s|s\in\mathcal{S}} \sum t|t \in \mathcal{T}\mathcal{L}t(D_t, w_t), \text{s.t.,} \quad \min \alpha, \mathbf{W}, \mathbf{\Omega} \sum_{\ell=1}^{m} \frac{1}{n_\ell + \tilde{n}\ell} \mathcal{L}\ell(D_\ell \cap \tilde{D}_\ell) + \lambda_1 \mathcal{R}(X\alpha). \tag{7}$$

Here, $\tilde{D}_s|s \in \mathcal{S}$ denotes the set of injected data for the source attacking nodes. The upper-level problem is defined as maximising the performance degradation of the target nodes and is denoted as $\mathcal{H}$. The secondary optimisation issue, addressed by the second condition, concerns a federated multi-task learning scenario in which the training dataset comprises a combination of unaltered and compromised data points.

$$\begin{aligned} &\max_{\tilde{D}_s \subset \mathcal{S}} \sum_{t\in\mathcal{T}} \mathcal{L}_t(D_t, w_t), \\ &\text{s.t.,} \min_{\substack{\alpha, \mathbf{W}, \\ \mathbf{\Omega}}} \left(\sum_{\ell=1}^{m} \frac{1}{n_\ell + \tilde{n}_\ell} \mathcal{L}_\ell(D_\ell \cup \tilde{D}_\ell, \mathbf{W}) + \lambda_1 \mathcal{R}(X\alpha) + \lambda_2 \|\mathbf{W}\|_F^2 \right), \\ &\quad \|\alpha_i\|_2 \leq C, \quad \forall i \in \{1, \ldots, n\}. \end{aligned} \tag{8}$$

The adjusted optimisation challenge, as specified in Equation (8), is detailed as follows:

- The aim, symbolised by $\max_{\tilde{D}_s|s\in\mathcal{S}} \sum_{t|t\in\mathcal{T}} \mathcal{L}_t(D_t, w_t)$, seeks to escalate the collective loss for all target nodes within the group $\mathcal{T}$, with $\tilde{D}_s$ representing the data manipulated by the attackers.
- The initial limitation strives to diminish the total loss for all learning tasks under the federation, depicted by $\min_{\alpha,\mathbf{W},\mathbf{\Omega}} \left(\sum_{\ell=1}^{m} \frac{1}{n_\ell + \tilde{n}_\ell} \mathcal{L}_\ell(D_\ell \cup \tilde{D}_\ell, \mathbf{W}) + \lambda_1 \mathcal{R}(X\alpha) + \lambda_2 \|\mathbf{W}\|_F^2 \right)$. This includes both the original and tampered data, with $\mathcal{L}_\ell$ denoting the loss for each task ℓ.
- To deter overfitting, which is fitting the model too closely to a limited set of data points, the model includes penalty terms $\lambda_1 \mathcal{R}(X\alpha)$ and $\lambda_2 \|\mathbf{W}\|_F^2$. These terms penalise the model's complexity, balanced by the parameters λ_1 and λ_2.

- A constraint $\|\alpha_i\|_2 \leq C$ ensures that the model's parameters α do not exceed a certain threshold C, maintaining the model's general performance and stability.

In essence, this framework streamlines the complex aspects of federated learning when tasked with handling both unmodified and tampered data, a situation often arising in extensive data settings such as the IoT.

4. Attacks on Federated Learning

This section presents the newly developed FAAMT algorithm, designed to identify the most-strategic approaches for launching attacks within federated learning environments. The algorithm builds upon the commonly used approach in data poisoning attack research, based on related works, and utilises a projected stochastic gradient ascent method to effectively raise the empirical loss of the target nodes, thereby reducing their performance in classification or regression tasks.

In the health IoT domain, a notable innovation is the DPFL-HIoT model, which applies Gradient Boosting Trees (GBTs) to detect healthcare fraud [85]. This method effectively safeguards patient data privacy while still enabling the machine learning model training process. Considering the sensitivity of health information and the difficulties in merging such data, FL emerges as a practical AI-powered tool. It serves the vast amounts of medical data stored in healthcare institutions, crucial for enabling smart health services like telehealth, diagnostics, and ongoing patient monitoring. FL, thus, offers a critical solution that allows for the use of these data while prioritising privacy.

A recent study described a method for executing tensor decomposition while safeguarding privacy within federated cloud systems [86]. This method is detailed extensively, alongside a comprehensive analysis of its complexity. It employs the Paillier encryption scheme and secure multi-party computation to ensure the confidentiality of the decomposition process. Additionally, the paper evaluates the precision of its predictive ratings alongside the computational and communication overhead incurred by users. The analysis of complexity includes the unfolding of tensors, the application of the Lanczos algorithm to these unfolded matrices, the determination of truncated orthogonal bases for tridiagonal matrices, and the construction of the core tensor using encrypted data. The findings within this paper validate the method's efficacy in maintaining privacy while facilitating tensor decomposition in a federated cloud context.

An innovative strategy has been developed to address the complexity of IoT device data and privacy concerns in Recurrent Neural Network (RNN) models [87]. This approach introduces a Differentially Private Tensor-based RNN (DPTRNN) framework dedicated to maintaining the confidentiality of sensitive user information within IoT applications. The proposed DPTRNN model is designed to tackle the challenges of dealing with heterogeneous IoT data and privacy issues in RNN models. Employing a tensor-based algorithm for back-propagation through time, enhanced with perturbation techniques to ensure privacy, the efficacy of the DPTRNN model is evaluated against video datasets. The results indicated that the model surpasses previous methods in terms of accuracy while providing a higher level of privacy protection.

4.1. Strategies for Attacking via Alternating Minimisation

The task of refining the attack strategy within bilevel problems presents a significant hurdle, mainly attributed to the intricate and non-convex characteristics of the subordinate problem. In our bilevel construct, the dominant problem is a straightforward primal issue, whereas the subordinate issue reveals a complex, non-linear, and non-convex nature. To navigate these complexities, we adopted an alternating minimisation technique. This method involves a cyclic update of the injected data to amplify the impact on the function $\mathcal{H}$, which quantifies the performance deterioration of the targeted nodes. This technique is instrumental in addressing the challenging non-convexity of the subordinate problem, paving the way for an enhanced attack stratagem tailored to federated learning contexts.

By implementing this alternating minimisation, we capitalise on its inherent flexibility to fine-tune our approach iteratively, aligning our injected data with the optimal points of disruption. Each iteration brings us closer to a strategy that can subtly, yet substantially weaken the model's performance from within. Moreover, this method's iterative nature allows us to monitor and adjust the attack in real-time, reacting to changes and defences put in place by the network, ensuring that our attack remains effective throughout the learning process. Such adaptability makes it an indispensable tool in the arsenal for compromising federated learning systems.

To tackle the non-convex nature of the lower-level problem in the bilevel optimisation of the attack problem, we adopted an alternating minimisation approach, where we optimise the features of the injected data, denoted by $\left(\tilde{D}_{s,i}, \alpha s, i\right)$, by fixing the labels of the injected data. The updates to the injected data $\tilde{D}_{s,i}$ are obtained via the projected stochastic gradient ascent method as shown below:

$$\left(\tilde{D}_{s,i}\right)^{k} \leftarrow \text{Proj}\,\mathbf{X}\left(\left(\tilde{D}_{s,i}\right)^{k-1} + \eta_1 \nabla_{\left(\tilde{D}_{s,i}\right)^{k-1}} \mathcal{H}\right), \tag{9}$$

Here, η_1 denotes the learning rate, k indexes the current iteration, and $\mathbf{X}$ defines the set of all permissible injected data, as outlined by the initial restriction in the primary problem $\mathcal{H}$. The projection operator $\text{Proj}\,\mathbf{X}(\cdot)$ ensures that the injected data lie within the ℓ_2-norm ball of radius r. The corresponding dual variable $\alpha s, i$ is updated accordingly as $\tilde{D}_{s,i}$ evolves:

$$\alpha_{s,i}^{k} \leftarrow \alpha_{s,i}^{k-1} + \Delta\alpha_{s,i} \tag{10}$$

where $\Delta(\alpha_{s,i})$ is the step in the dual space that maximises the upper-level problem $\mathcal{H}$. The equation for computing the gradient of the t-th target node in Equation (9) is derived using the chain rule.

$$\nabla_{\hat{x}_a^t} \mathcal{L}t(D_t, w_t) = \nabla w_t \mathcal{L}t(D_t, w_t) \cdot \nabla \hat{x}_a^{\pm} w_t. \tag{11}$$

The gradient of the objective function at the upper level with respect to both $\hat{x}_a^t$ and w_t is given by $\nabla_{\hat{x}_a^t} \mathcal{L}_t(D_t, w_t) = \nabla_{w_t} \mathcal{L}_t(D_t, w_t) \times \nabla_{\hat{x}_a^t} w_t$. The first term on the right-hand side, $\nabla_{w_t} \mathcal{L}_t(D_t, w_t)$, is readily computable as it solely relies on the loss function $\mathcal{L}_t(\cdot)$. Conversely, the second term, $\nabla_{\hat{x}_a^t} w_t$, hinges on the optimal conditions of the lower-level problem as set out in Equation (7).

In order to ascertain $\nabla_{\hat{x}_a^t} w_t$, we commence by fixing the parameter Ω in the lower-level issue to sidestep its constraints, leading us to the dual problem stated as:

$$\min_{\alpha, \mathbf{W}} \sum_{\ell=1}^{m} \frac{1}{n_\ell + \tilde{n}_\ell} \mathcal{L}_\ell(D_\ell \cup \tilde{D}_\ell) + \lambda_1 \mathcal{R}(X\alpha) \tag{12}$$

Considering the optimality conditions of the lower-level problem as constraints within the upper-level problem, we treat Ω as a constant matrix during the gradient determination. Furthermore, given that $\mathcal{R}(X\alpha)$ is continuously differentiable, we correlate $\mathbf{W}$ and α through the relationship $w_\ell(\alpha) = \nabla\mathcal{R}(X\alpha)$. Lastly, we proceed with updating the dual variable α and calculating the ensuing gradients as shown in Equation (9).

To update the dual variable α, we seek to maximise the dual objective with a least-squares loss or hinge loss function by updating each element of α individually. The optimisation problem in Equation (11) can be reformulated into a constrained optimisation problem for node ℓ as:

$$\min_{\alpha} \sum_{\ell=1}^{m} \frac{1}{n_\ell + \hat{n}\ell} \left(\mathcal{L}\ell^\star\left(-\alpha_\ell^i\right) + \mathcal{L}\ell^\star\left(-\hat{\alpha}\ell^{i'}\right)\right) + \lambda_1 \mathcal{R}^\star(X[\alpha_\ell; \hat{\alpha}_\ell]) \tag{13}$$

To enable distributed computation across multiple nodes, the optimisation problem in Equation (10) can be separated into smaller, data-local subproblems. This is achieved by making a quadratic approximation of the dual problem. At each step k, two samples are chosen randomly, one from the original clean data (i.e., $i \in 1, \ldots, n_\ell$) and one from the injected data (i.e., $i' \in 1, \ldots, \hat{n}\ell$). The updates for both α_ℓ^i and $\hat{\alpha}_\ell^{i'}$ in node ℓ can be computed as:

$$\left(\alpha_\ell^i\right)^k = \left(\alpha_\ell^i\right)^{k-1} + \Delta\alpha_\ell^i, \left(\hat{\alpha}\ell^{i'}\right)^k = \left(\hat{\alpha}\ell^{i'}\right)^{k-1} + \Delta\hat{\alpha}_\ell^{i'}, \tag{14}$$

where $\Delta\alpha_\ell^i$ and $\Delta\hat{\alpha}_\ell^{i'}$ are the stepsizes that maximise the dual objective in Equation (15) when all other variables are fixed. To achieve maximum dual ascent, we optimise:

$$\Delta\alpha_\ell^i = \arg\min_{a\in\mathbb{R}} \mathcal{L}^*\ell\left(-\left(\alpha\ell^i + a\right)\right) + a\left\langle w_\ell(\alpha_\ell), x_\ell^i\right\rangle + \frac{\lambda}{2}\left|x_\ell^i a\right| M_\ell^2 \tag{15}$$

$$\Delta\hat{\alpha}\ell^{i'} = \arg\min \hat{a} \in \mathbb{R}\mathcal{L}^*\ell\left(-\left(\hat{\alpha}\ell^{i'} + \hat{a}\right)\right) + \hat{a}\left\langle w_\ell(\hat{\alpha}\ell), x\ell^{i'}\right\rangle + \frac{\lambda}{2}\left|x_\ell^{i'}\hat{a}\right| M_\ell^2 \tag{16}$$

Here, M is a symmetric positive definite matrix and $M_\ell \in \mathbb{R}^{d\times d}$ is the ℓ-th diagonal block of M. The inverse of M is expressed as

$$M^{-1} = \Omega + \frac{\lambda_2}{\lambda_1} I_{md\times md}. \tag{17}$$

where $\Omega := \Omega \otimes I_{d\times d} \in \mathbb{R}^{md\times md}$ and λ_1 and λ_2 are positive constants. For the least-squares loss function, the closed-form solution for $\Delta\alpha_\ell^i$ is:

$$\Delta\alpha_\ell^i = \frac{y_\ell^i - \left(x_\ell^i\right)^\top x_\ell^i \alpha_\ell^i - 0.5\left(\alpha_\ell^i\right)^{k-1}}{0.5 + \lambda_1\left|x_\ell^i\right|\mathrm{M}\ell^2}. \tag{18}$$

For the hinge loss, $\Delta\hat{\alpha}_\ell^{i^r}$ can be computed as follows:

$$\Delta\hat{\alpha}\ell^{i^r} = y\ell^{i^r} \max\left(0, \min\left(1, \frac{1 - \left(x_\ell^{i^r}\right)^\top x_\ell^{i^r}\left(\alpha_\ell^{k-1}\right) y_\ell^{i^r}}{\lambda_1\left|x_\ell^{i^r}\right| {M_\ell}^2} + \left(\alpha_\ell^{k-1}\right) y_\ell^{i^r}\right)\right) - \left(\alpha_\ell^{k-1}\right) \tag{19}$$

The gradient associated with each of the injected data $\hat{x}_s^i$ for the corresponding target node t is expressed as:

$$\nabla_{(\hat{x}_s^s)}\mathcal{L}_t\left((w_t)^\top x_t^j, y_t^j\right) = 2\left((w_t)^\top x_t^j - y_t^j\right) x_t^j \cdot \Delta\hat{\alpha}_s^i \Omega(t,s) \tag{20}$$

For the hinge loss, the gradient is:

$$\nabla_{(\hat{x}_s^u)}\mathcal{L}_t\left((w_t)^\top x_t^j, y_t^j\right) = y_t^j x_t^j \cdot \Delta\alpha_s^i \Omega(t,s) \tag{21}$$

The algorithm for the attack on federated learning is given in Algorithm 1.

The Federated Adversarial Attack for Multi-Task Learning (FAAMT) Algorithm 1 is an essential mechanism designed to test the robustness of federated learning systems against malicious attacks. It systematically integrates adversarial samples into a network's nodes, specifically within a multi-task learning setting. Through iterative updates governed by a central learning rate and convergence criteria, FAAMT evaluates the vulnerability of federated learning structures. This assessment is essential for strengthening the defence mechanisms of such systems against advanced adversarial strategies.

Algorithm 1 Federated Adversarial Attack for Multi-Task Learning (FAAMT).

procedure FAAMT($T, S, \hat{n}_s$)
  Randomly initialise the current state $f(\hat{X}_0, \hat{v}_0, \hat{y}_0) \in S$
  Initialise the set of injected data as $\hat{D} = \hat{D}_0$, and set the iteration counter $k = 0$
  Set the global learning rate α and tolerance ϵ for convergence
  while $k <$ maximum number of iterations **and** not converged **do**
    # of parallel computations for each node
    for all nodes $i = 1$ to m in parallel **do**
      Calculate the solution δv_i for node i using Equation (15) or Equation (16)
      Update the local variables $v_i \leftarrow v_i + \alpha \cdot \delta v_i$
      if $i \in S$ **then**
        Calculate the solution $\delta \hat{v}_i$ for node i using Equation (15) or Equation (16)
        Update the adversarial variables $\hat{v}_i \leftarrow \hat{v}_i + \alpha \cdot \delta \hat{v}_i$
      end if
    end for
    Aggregate local updates to update Global Model parameters W_k
    Check for convergence with a predefined criterion, e.g., $||W_k - W_{k-1}|| < \epsilon$
    Increase counter $k \leftarrow k + 1$
    # of injected data point updates for source nodes
    for all source nodes $s = 1$ to S in parallel **do**
      Update the injected data point $\hat{x}_s$ using Equation (7)
    end for
    Update the set of injected data as $\hat{D} \leftarrow \hat{D} \cup \{\hat{x}_s\}_{s=1}^{S}$
  end while
  if converged **then**
    return Successfully converged with adversarial examples $\hat{D}$
  else
    return Algorithm reached maximum iterations without convergence
  end if
end procedure

Further Analysis

To attack the federated learning model, we focused on the bilevel optimisation problem. The upper-level problem aims to find the optimal attack strategy, while the lower-level problem is to update the model weights. The bilevel problem can be expressed as:

$$\max_{\tilde{D}_{s,i}, \alpha s, i} \quad \mathcal{H}(D, W; \tilde{D}_{s,i}, \alpha s, i) \quad \text{subject to} \quad |\tilde{D}_{s,i}|2 \leq r, \quad W = \arg\min W' \mathcal{L}(D \cup \tilde{D}_{s,i}, W'; \alpha s, i) \tag{22}$$

For the non-convex nature of the lower-level problem, we used the alternating minimisation approach to optimise the bilevel problem. We initiate by fixing the labels of the injected data and optimising the features of the injected data, denoted by $(\tilde{D}_{s,i}, \alpha s, i)$, as follows:

$$\left(\tilde{D}_{s,i}\right)^k \leftarrow \text{Proj}\,\mathbf{X}\left(\left(\tilde{D}_{s,i}\right)^{k-1} + \eta_1 \nabla_{\left(\tilde{D}_{s,i}\right)^{k-1}} \mathcal{H}\right), \tag{23}$$

To compute the gradient, we first derive the gradient of the target node loss function $\mathcal{L}_t(D_t, W)$ with respect to the model weights W:

$$\nabla_W \mathcal{L}t(D_t, W) = \sum j = 1^{n_t} \nabla_w \ell(w_t^\top x_t^j, y_t^j), \tag{24}$$

where $\ell(\cdot)$ is the loss function, x_t^j and y_t^j are the j-th data point and its label in the target node dataset, and n_t is the number of data points in the target node dataset. Next, we need to compute the gradient of the upper-level objective with respect to the injected data $\tilde{D}_{s,i}$:

$$\nabla_{\tilde{D}_{s,i}} \mathcal{H} = \sum_{t=1}^{T} \nabla_{\tilde{D}_{s,i}} \mathcal{L}_t(D_t \cup \tilde{D}_{s,i}, W). \tag{25}$$

To calculate this gradient, we utilise the chain rule, obtaining:

$$\nabla_{\tilde{D}_{s,i}} \mathcal{H} = \sum_{t=1}^{T} \nabla_W \mathcal{L}t(D_t, W) \cdot \nabla \tilde{D}_{s,i} W. \tag{26}$$

For the lower-level problem, we update the model weights W using gradient descent with learning rate η_2:

$$W^k \leftarrow W^{k-1} - \eta_2 \nabla_W \mathcal{L}(D \cup \tilde{D}_{s,i}, W^{k-1}; \alpha s, i). \tag{27}$$

After updating the features of the injected data, we fix the features and optimise the labels of the injected data:

$$\alpha s, i^k \leftarrow \text{Proj}\, \mathcal{Y}\left(\alpha s, i^{k-1} + \eta_3 \nabla_{\alpha s, i^{k-1}} \mathcal{H}\right), \tag{28}$$

where Proj $\mathcal{Y}$ projects the labels into the feasible set $\mathcal{Y}$.

To compute the gradient of the upper level objective with respect to the labels of the injected data, we have:

$$\nabla_{\alpha s,i} \mathcal{H} = \sum_{t=1}^{T} \nabla_{\alpha s,i} \mathcal{L}_t(D_t \cup \tilde{D}_{s,i}, W). \tag{29}$$

We apply the chain rule to obtain the following expression:

$$\nabla_{\alpha s,i} \mathcal{H} = \sum_{t=1}^{T} \nabla_W \mathcal{L}t(D_t, W) \cdot \nabla \alpha s, iW. \tag{30}$$

Finally, the alternating minimisation algorithm consists of iterating through these updates for the features of the injected data, the labels of the injected data, and the model weights until convergence. By focusing on the gradients of the loss function, the attacker can effectively manipulate the model's behaviour, leading to a successful attack.

5. Experimental Results

5.1. System Specifications of Experiments

The ensuing experimental design, crucial for substantiating the efficacy of FLIBD, was scrupulously devised to accurately emulate a pragmatic IoT Big Data management context. This, in turn, allows for a discerning evaluation of the approach's potency in safeguarding privacy within expansive systems. In particular, the experiments were enacted within a high-performance computing environment, marked by its robust computational and storage capabilities, thereby being impeccably aligned for navigating through the complexities of resource-demanding federated learning tasks among handling copious IoT Big Data. A comprehensive exposition of the essential system specifications leveraged for the experiments is proffered in Table 3.

The above system was orchestrated to emulate large-scale system scenarios for federated learning with a lens toward IoT Big Data management. The CPU, equipped with an extensive number of cores, alongside the robust RAM, facilitates efficient parallel processing capabilities, ensuring expedited computational procedures. The GPU, boasting a substantive video memory, becomes quintessential in efficiently managing and processing massive datasets, especially under operations requiring enhanced parallelisation, such as

model training in federated learning frameworks. The ultra-fast NVMe storage plays a critical role in minimising data retrieval and storage times, thus mitigating potential bottlenecks arising from disk I/O operations. Furthermore, the experiment leveraged Apache Spark with Federated AI Technology Enabler (FATE) to enable secure computations and model training, allowing data privacy preservation through federated learning mechanisms.

Table 3. System specifications for experimental evaluation.

Component	Specification
CPU	AMD 5950X
RAM	64GB DDR4
Storage	2TB NVMe
GPU	3090Ti (24GB VRAM)
Operating System	Ubuntu 18.04 LTS
Network Interface	1GBPS
Software Framework	Apache Spark with FATE

5.2. Experimental Evaluation of Attacking Mechanism

In the ever-expanding domain of Federated Learning (FL), the imperative to guard against adversarial attacks whilst ensuring accurate model predictions has become paramount. The ensuing analysis offers a rigorous examination of the Federated Adversarial Attack for Multi-Task Learning (FAAMT) algorithm (Algorithm 1). This adversarial approach, developed with meticulous attention to the intricacies of multi-task learning, seeks to diminish the predictive accuracy of various federated learning methods, even those that are conventionally recognised for their robustness against adversarial machinations.

To unravel the efficacy of FAAMT, it was assessed against several widely acknowledged federated learning methods and defences, such as the Global Model, Median Model, Trimmed Mean Model, Krum, Bulyan, RFA, and Romoa. Each method typically exhibits substantial resistance to conventional adversarial tactics, establishing them as formidable entities in federated learning defences. However, through the vision of FAAMT, we sought to decipher whether their robustness holds steadfast or gradually attenuates under the sophisticated attack mechanisms unleashed by the algorithm.

The outcomes derived from applying the FAAMT method (Figure 2) show the performance variations among the federated learning mechanisms previously discussed, unmasking their potential susceptibilities. This analytical exploration, thus, enriches our comprehension of the resilience landscape among adversarial challenges within federated learning frameworks, underscoring the specific vulnerabilities and strengths inherent in each approach.

Intriguingly, the derived results emanating from the visual representation expose a compelling narrative pertaining to the potency of the FAAMT mechanism. A discerning glance at the plot elucidates a conspicuous decrement in the predictive accuracies of the defended federated learning methods under attack, reinforcing the assertion that FAAMT has managed to permeate their defensive veils to a certain degree.

Certain defence mechanisms, which previously heralded unwavering resilience in the face of adversarial onslaughts, now showcase perceptible vulnerabilities when entwined with the FAAMT. This not only lays bare latent susceptibilities within these methods, but also accentuates the detail and sophistication embedded within the FAAMT algorithm. Notably, certain defences exhibited a more-pronounced susceptibility, suggesting a hierarchical differentiation in robustness among the examined methods and emphasising the necessity for an incessant evolution in defence strategies.

As the realm of federated learning continues to evolve, so too will the complexity and cunning of adversarial attacks. Thus, the illumination of FAAMT's ability to infiltrate these robust defence mechanisms propels the discourse on the perennial battle between adversarial attacks and defence methods forward. It beckons the scientific community to

explore further, excavate deeper, and innovate with a renewed zeal to safeguard federated learning models against progressively shrewd adversarial endeavours. It is within this iterative process of attack and defence that the future robustness and reliability of federated learning models will be forged, ensuring the integrity and durability of decentralised learning mechanisms in real-world applications.

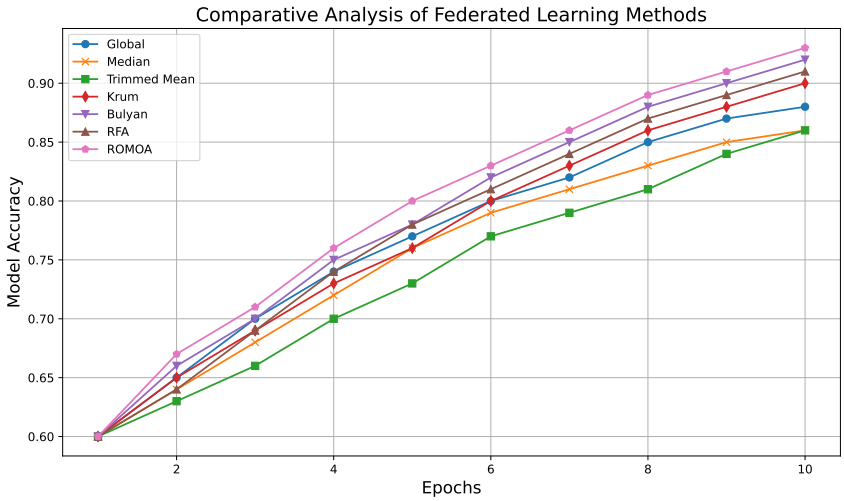
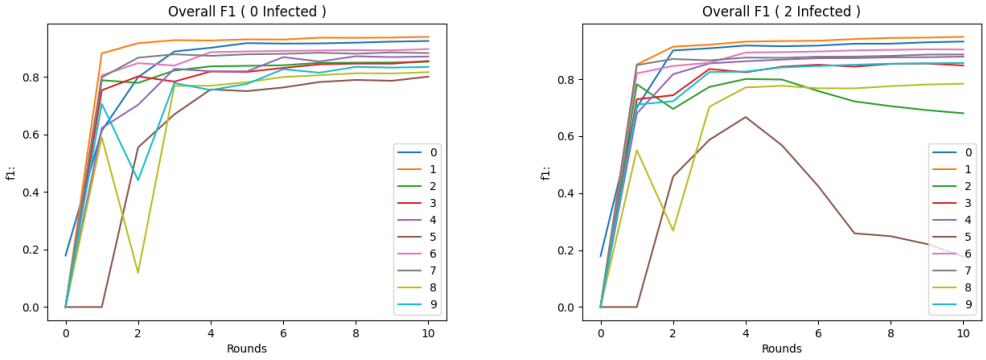


Figure 2. Efficacy assessment of FAAMT against various federated learning defence mechanisms.

5.3. Experimental Evaluation of Defences

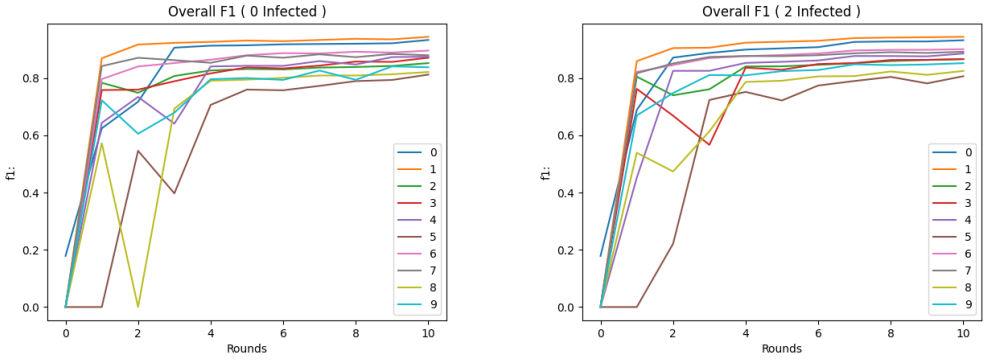
This section will analyse the plots produced during the experimental evaluation of the implemented defences based on the proposed attack. In each case, the maximum number of clients was 10 and the number of attackers in the system varied from none to 5 clients, i.e., 50% of the system. A higher percentage than this was not considered as there is no point in implementing a defence on a system that malicious users mostly own. The section will start with the results of the attack on a network using the standard FedAvg aggregation method and then split by the percentage of exposure. For the experiments, we utilised the widely used CIFAR-10 dataset, which contains 32×32 images, which are classified into various object categories like birds, cats, and aeroplanes. The training dataset contains 5000 images for each of the 10 classes. The purpose of this analysis was to determine the effectiveness of the implemented defences against the proposed attack and to identify any vulnerabilities or limitations. The experimental evaluation aimed to provide insights into the defences' strengths and weaknesses under different conditions. The experimental results of this analysis may guide the development of future defences and enhance the security of federated learning systems. We evaluated the methods of the Median, Trimmed Mean, Krum, and Bulyan before the attack and afterwards. Before the label-flipping attack, we observed that a very simple logistic regression model in this task had significant accuracy, approaching 98%. In other words, it is a relatively tractable and simple way to implement this task. This is illustrated in Figure 3a.

From above, we can see the same results as previously, but with the difference that, now, 20% of the users in our system were malicious. We can see from the F1-score that there was a significant decrease in accuracy, which can be catastrophic. In the end, we also provide the F1-score for the case where 50% of the clients were malicious. In Figure 3b, we can see that the Digit 5 class had a significant decrease even with 20% of the clients being malicious. An important by-product of this attack is that the class to which we converted Digit 5 (i.e., Class 2) also had a significant reduction in its F1-score. In Figure 4a, we can observe that the convergence in the F1-score was not as good as in the simple FedAvg method, which started from the third round, while the Median started from the fourth or even fifth round. In Figure 4b, we can see that, even with a method as simple as the Median, we can identify an increase in the F1-score that was immediate. Also, the convergence was similar.



(**a**) F1-score for each digit before attacking. (**b**) F1-score for each digit after attacking.

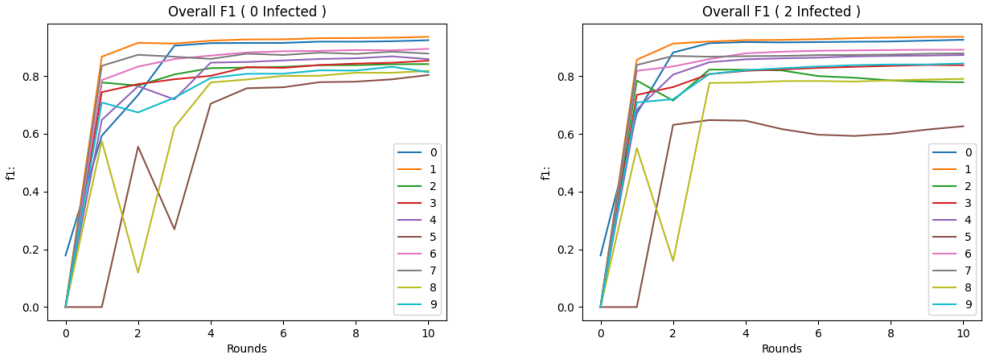
Figure 3. Performance of the Global Model after each round of FL.



(**a**) F1-score for each digit before attacking. (**b**) F1-score for each digit after attacking.

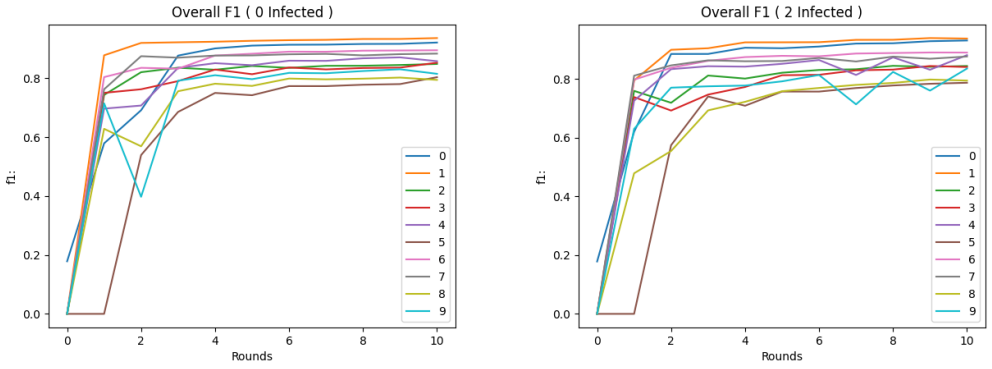
Figure 4. Performance of the Median Model after each round of FL.

For the Trimmed Mean method, the results are shown in Figure 5a. Similar to the previous method, convergence occurred later. The Trimmed Mean method (Figure 5b) was observed to not perform as well, but this may be attributed to the fact that the customer sample after trimming the edges was small. For the Krum method, the results are shown in Figure 6a. Compared to the previous method, it appeared to have higher convergence. However, after the attack (Figure 6b), it appeared that it lowered while maintaining higher results than the Trimmed Mean method. For the Bulyan method, the results are shown in Figure 7a. Here, the convergence was similar to the Trimmed Mean method, but lower than the Krum model. However, the attack on the model (Figure 7b) appeared to have a more-robust performance.



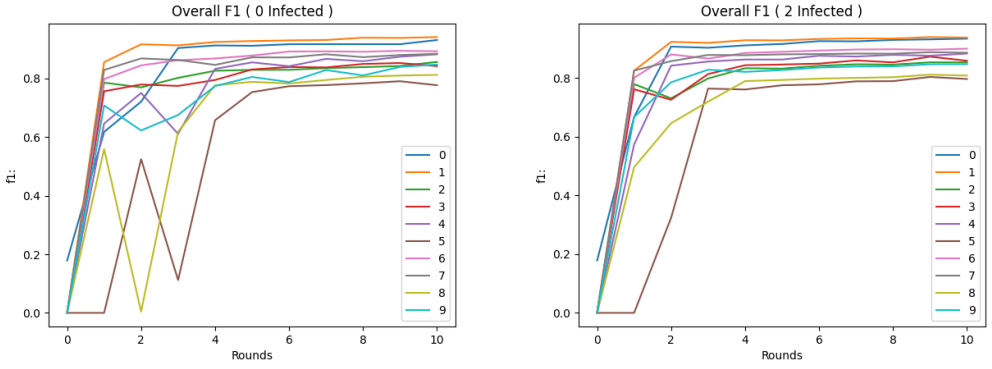
(**a**) F1-score for each digit before attacking. (**b**) F1-score for each digit after attacking.

Figure 5. Performance of the Trimmed Mean Model after each round of FL.



(**a**) F1-score for each digit before attacking. (**b**) F1-score for each digit after attacking.

Figure 6. Performance of the Krum model after each round of FL.



(**a**) F1-score for each digit before attacking. (**b**) F1-score for each digit after attacking.

Figure 7. Performance of the Bulyan model after each round of FL.

Evaluation Metrics:

To rigorously assess the effectiveness of the Federated Adversarial Attack for Multi-Task Learning (FAAMT) and the robustness of various federated learning defensive mechanisms, we employed two pivotal metrics: *accuracy* and the *F1-score*. The accuracy metric gauges the overall performance of the model by evaluating the ratio of correctly predicted instances to the total instances and is defined as

$$\text{Accuracy} = \frac{\text{True Positives (TPs)} + \text{True Negatives (TNs)}}{\text{Total Instances (P+N)}} \tag{31}$$

where TPs and TNs denote the number of true positive and true negative predictions, respectively, and P and N represent the total actual positive and negative instances, respectively. The F1-score, the harmonic mean of precision and recall, offers a balanced perspective, especially pivotal when dealing with imbalanced datasets, and is expressed as:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{32}$$

where

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{False Positives (FP)}} \tag{33}$$

and

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{False Negatives (FN)}} \tag{34}$$

Utilising these metrics, the ensuing sections deliberate the comprehensive evaluation of the FAAMT and the subjected defence mechanisms, shedding light on their respective strengths and vulnerabilities amidst adversarial incursions.

The results for the proposed attack and defence mechanisms are shown in Table 4, where we assess the accuracy and F1-score for the attack (FAAMT method), as well as the defence mechanism behaviour. Lastly, we evaluated the performance of both the attack and defence on 2 clients infected out of 10.

Table 4. Metrics assessing the performance of attack and defence mechanisms.

Method	Attack (A)		Defence (D)		Affected Clients (D)
	Accuracy	F1-Score	Accuracy	F1 Score	out of 10
Global Model	0.88	0.84	0.85	0.81	2
Median Model	0.86	0.82	0.87	0.83	2
Trimmed Mean Model	0.86	0.81	0.86	0.82	2
Krum	0.90	0.84	0.88	0.83	2
Bulyan	0.93	0.80	0.89	0.83	2

6. Discussion

The experimental analysis presented delved into the complex interactions between various federated learning defence mechanisms and a sophisticated adversarial attack, the FAAMT. The results from the high-performance computational environment underscored the imperative of robust defences in the realm of IoT Big Data management. The experiments revealed that even well-established defences, such as Krum and Bulyan, exhibit vulnerabilities when confronted with the FAAMT, hinting at the subtleties in the defence mechanisms that must be addressed. This observation is crucial, for it demonstrates that the current landscape of federated learning is not impervious to novel attack strategies and must continuously evolve.

Notably, the FAAMT elucidated a significant reduction in the F1-scores of the federated learning models, which was particularly pronounced for certain classes within the CIFAR-10 dataset. This degradation in performance suggests that, while current defences offer a degree of resilience, they are not entirely foolproof against such calculated adversarial intrusions. This is a stark reminder that the security of federated learning models remains a moving target, requiring ongoing research and refinement to mitigate the risk of adversarial exploitation.

The experimental evaluation also offered insights into the performance of various defence strategies post-attack. Methods like the Median and Trimmed Mean, traditionally considered less sophisticated than Krum or Bulyan, displayed a delayed convergence, indicating a slower recovery from adversarial disturbances. This behaviour highlights the need for a balance between the complexity of defence algorithms and their adaptability in the face of adversarial manoeuvres. The nuanced decrease in the performance of the Krum and Trimmed Mean Models post-attack further accentuates the need for dynamic defence strategies capable of coping with the evolving nature of adversarial threats.

In contrast, the Bulyan method showcased a relatively more-robust performance under adversarial conditions, suggesting that certain combinations of defence methodologies may offer enhanced resilience. The experimental results suggested a promising direction for future investigation: examining combined or multi-level defence strategies that could offer stronger protection against adversarial attacks. The field of federated learning, especially within the context of the IoT and Big Data, is, thus, poised at a crucial juncture where the iterative process of developing and testing new defence strategies is not just beneficial, but necessary to safeguard the integrity of decentralised learning models against increasingly sophisticated adversarial tactics.

7. Conclusions

In our investigation, we presented the Federated Adversarial Attack for Multi-Task Learning (FAAMT) algorithm, which significantly exposed the susceptibilities of federated multi-task learning frameworks to data poisoning. Our results provided clear evidence that FAAMT can critically undermine the training process, leading to a noticeable deterioration in model accuracy and performance. This highlights an urgent necessity for the establishment of robust defence strategies within such systems.

As we look to the horizon of federated multi-task learning, it is imperative to address the challenges presented by the FAAMT algorithm. The development and in-depth examination of defence strategies such as secure aggregation, robust outlier detection, and the integration of differential privacy mechanisms are pivotal areas requiring immediate attention. The future landscape should include extensive evaluations of these defences' effectiveness against sophisticated data poisoning tactics with respect to multi-task learning environments.

The fine-tuning of FAAMT can lead to many potential enhancements. Its adaptability could be tested against various data poisoning strategies to assess impacts on user data confidentiality. Furthermore, its extension to other federated learning paradigms, such as federated transfer learning, could provide significant insights. Investigating the resilience of optimisation techniques like gradient clipping related to adversarial interventions in multi-task scenarios could yield beneficial strategies for maintaining system integrity. Additionally, adversarial tactics that specifically target individual tasks or users, while minimising disruption to other system components, could reveal specific system vulnerabilities and build room for the enhancement of privacy-preserving measures.

In summary, this study highlighted a critical and immediate need to reinforce the security and privacy protocols of federated multi-task learning systems. The safe and effective application of these systems in practical settings demands the creation and implementation of robust defence mechanisms. Addressing these open challenges is essential to advancing the adoption of federated multi-task learning, ensuring that it can be confidently deployed to facilitate collaborative learning processes across heterogeneous devices, all while maintaining a steady defence of user privacy in the interconnected ecosystems of the IoT and Big Data.

7.1. Future Work

The prospect of combining Federated Learning (FL) and Automated Machine Learning (AutoML) opens new opportunities for enhancing data privacy and optimising Big Data management. The integration of FL with AutoML, as envisioned in the studies [88,89], provides a robust foundation for the development of the Federated Adversarial Attack for Multi-Task Learning (FAAMT) algorithm. This algorithm aims to address the complexities of multi-task learning within a federated framework, where the goal is to enable the collaborative training of models across multiple tasks while ensuring data privacy and robustness against adversarial attacks. The FAAMT approach could leverage the privacy-preserving nature of FL to protect data across distributed networks and the efficiency of AutoML to optimise learning tasks, thereby ensuring that the integrity of multi-task learning is maintained even in the presence of potential adversarial threats.

The integration of edge intelligence and edge caching mechanisms outlined in the first study with the advanced hyperparameter optimisation techniques from the second study provides the foundational elements necessary for the further development of FAAMT. The algorithm could harness the computational power available at the network edge, alongside sophisticated AutoML methods, to effectively manage the multi-faceted challenges of Big Data in a federated context. This synergy promises to deliver not only more-personalised and immediate data processing at the edge, but also a robust framework for multi-task models that are inherently more resistant to adversarial attacks. In such a way, the FAAMT algorithm has the potential to become a quintessential part of smart, decentralised networks,

enabling them to remain resilient in the face of evolving cyber threats while making the most of the vast datasets generated across IoT environments.

The development of the FAAMT algorithm points us in a new direction for research that incorporates the protective features of federated learning with the dynamic fine-tuning abilities of AutoML. By utilising both, future work can focus on improving this algorithm to better manage the complex tasks of learning multiple things at once in environments where there may be hostile attempts to disrupt learning. This will include creating better methods to spot and stop such attacks in networks where data are shared across different computers, making algorithms that can learn different tasks more effectively and efficiently, and ensuring that the system can handle the unevenly distributed data that are often found in everyday situations. Further studies can also investigate how to best spread out the work of computing across the edges of the networks, making the most of local computing for AutoML tasks and further protecting the privacy of data within the federated learning setup. As the FAAMT algorithm improves even more, it could be used in many areas, turning multi-task learning in networks where data are shared across many computers into not just a possibility, but a strong method for keeping data private and safe in a world that is more and more connected.

Author Contributions: A.K., A.G., L.T., G.A.K., G.K., C.K. and S.S. conceived of the idea, designed and performed the experiments, analysed the results, drafted the initial manuscript and revised the final manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kairouz, P.; McMahan, H.B.; Avent, B.; Bellet, A.; Bennis, M.; Bhagoji, A.N.; Bonawitz, K.; Charles, Z.; Cormode, G.; Cummings, R.; et al. Advances and open problems in federated learning. *Found. Trends Mach. Learn.* **2021**, *14*, 1–210. [CrossRef]
2. Jebreel, N.M.; Domingo-Ferrer, J.; Sánchez, D.; Blanco-Justicia, A. Defending against the label-flipping attack in federated learning. *arXiv* **2022**, arXiv:2207.01982.
3. Jiang, Y.; Zhang, W.; Chen, Y. Data Quality Detection Mechanism Against Label Flipping Attacks in Federated Learning. *IEEE Trans. Inf. Forensics Secur.* **2023**, *18*, 1625–1637. [CrossRef]
4. Li, D.; Wong, W.E.; Wang, W.; Yao, Y.; Chau, M. Detection and mitigation of label-flipping attacks in federated learning systems with KPCA and K-means. In Proceedings of the 2021 8th International Conference on Dependable Systems and Their Applications (DSA), Yinchuan, China, 5–6 August 2021; pp. 551–559.
5. Cheng, Y.; Liu, Y.; Chen, T.; Yang, Q. Federated learning for privacy-preserving AI. *Commun. ACM* **2020**, *63*, 33–36. [CrossRef]
6. Truex, S.; Baracaldo, N.; Anwar, A.; Steinke, T.; Ludwig, H.; Zhang, R.; Zhou, Y. A hybrid approach to privacy-preserving federated learning. In Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security, 2019; pp. 1–11. Available online: https://arxiv.org/abs/1812.03224 (accessed on 18 October 2023).
7. Wei, K.; Li, J.; Ding, M.; Ma, C.; Yang, H.H.; Farokhi, F.; Jin, S.; Quek, T.Q.; Poor, H.V. Federated learning with differential privacy: Algorithms and performance analysis. *IEEE Trans. Inf. Forensics Secur.* **2020**, *15*, 3454–3469. [CrossRef]
8. Pan, K.; Feng, K. Differential Privacy-Enabled Multi-Party Learning with Dynamic Privacy Budget Allocating Strategy. *Electronics* **2023**, *12*, 658. [CrossRef]
9. Karras, A.; Karras, C.; Giotopoulos, K.C.; Tsolis, D.; Oikonomou, K.; Sioutas, S. Peer to Peer Federated Learning: Towards Decentralized Machine Learning on Edge Devices. In Proceedings of the 2022 7th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM), Ioannina, Greece, 23–25 September 2022; pp. 1–9. [CrossRef]
10. Abreha, H.G.; Hayajneh, M.; Serhani, M.A. Federated Learning in Edge Computing: A Systematic Survey. *Sensors* **2022**, *22*, 450. [CrossRef]
11. Kaleem, S.; Sohail, A.; Tariq, M.U.; Asim, M. An Improved Big Data Analytics Architecture Using Federated Learning for IoT-Enabled Urban Intelligent Transportation Systems. *Sustainability* **2023**, *15*, 15333. [CrossRef]
12. Javed, A.R.; Hassan, M.A.; Shahzad, F.; Ahmed, W.; Singh, S.; Baker, T.; Gadekallu, T.R. Integration of blockchain technology and federated learning in vehicular (iot) networks: A comprehensive survey. *Sensors* **2022**, *22*, 4394. [CrossRef]

13. Kong, Q.; Yin, F.; Xiao, Y.; Li, B.; Yang, X.; Cui, S. Achieving Blockchain-based Privacy-Preserving Location Proofs under Federated Learning. In Proceedings of the ICC 2021—IEEE International Conference on Communications, Montreal, QC, Canada, 14–23 June 2021; pp. 1–6. [CrossRef]
14. Math, S.; Tam, P.; Kim, S. Reliable Federated Learning Systems Based on Intelligent Resource Sharing Scheme for Big Data Internet of Things. *IEEE Access* **2021**, *9*, 108091–108100. [CrossRef]
15. Nuding, F.; Mayer, R. Data poisoning in sequential and parallel federated learning. In Proceedings of the 2022 ACM on International Workshop on Security and Privacy Analytics, 2022; pp. 24–34. Available online: https://dl.acm.org/doi/abs/10.1145/3510548.3519372 (accessed on 18 October 2023).
16. Tolpegin, V.; Truex, S.; Gursoy, M.E.; Liu, L. Data poisoning attacks against federated learning systems. In Proceedings of the Computer Security–ESORICS 2020: 25th European Symposium on Research in Computer Security, ESORICS 2020, Guildford, UK, 14–18 September 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 480–501.
17. Sun, G.; Cong, Y.; Dong, J.; Wang, Q.; Lyu, L.; Liu, J. Data poisoning attacks on federated machine learning. *IEEE Internet Things J.* **2021**, *9*, 11365–11375. [CrossRef]
18. Li, J.; Guo, W.; Han, X.; Cai, J.; Liu, X. Federated Learning based on Defending Against Data Poisoning Attacks in IoT. *arXiv* **2022**, arXiv:2209.06397.
19. Xu, W.; Fang, W.; Ding, Y.; Zou, M.; Xiong, N. Accelerating Federated Learning for IoT in Big Data Analytics with Pruning, Quantization and Selective Updating. *IEEE Access* **2021**, *9*, 38457–38466. [CrossRef]
20. Fu, A.; Zhang, X.; Xiong, N.; Gao, Y.; Wang, H.; Zhang, J. VFL: A verifiable federated learning with privacy-preserving for big data in industrial IoT. *IEEE Trans. Ind. Inform.* **2020**, *18*, 3316–3326. [CrossRef]
21. Can, Y.S.; Ersoy, C. Privacy-preserving federated deep learning for wearable IoT-based biomedical monitoring. *ACM Trans. Internet Technol. (TOIT)* **2021**, *21*, 1–17. [CrossRef]
22. Zhang, Y.; Zhang, Y.; Zhang, Z.; Bai, H.; Zhong, T.; Song, M. Evaluation of data poisoning attacks on federated learning-based network intrusion detection system. In Proceedings of the 2022 IEEE 24th Int Conf on High Performance Computing & Communications; 8th Int Conf on Data Science & Systems; 20th Int Conf on Smart City; 8th Int Conf on Dependability in Sensor, Cloud & Big Data Systems & Application (HPCC/DSS/SmartCity/DependSys), Hainan, China, 18–20 December 2022; pp. 2235–2242. [CrossRef]
23. Singh, S.; Rathore, S.; Alfarraj, O.; Tolba, A.; Yoon, B. A framework for privacy preservation of IoT healthcare data using Federated Learning and blockchain technology. *Future Gener. Comput. Syst.* **2022**, *129*, 380–388. [CrossRef]
24. Albelaihi, R.; Alasandagutti, A.; Yu, L.; Yao, J.; Sun, X. Deep-Reinforcement-Learning-Assisted Client Selection in Nonorthogonal-Multiple-Access-Based Federated Learning. *IEEE Internet Things J.* **2023**, *10*, 15515–15525. [CrossRef]
25. Yang, Q.; Liu, Y.; Chen, T.; Tong, Y. Federated machine learning: Concept and applications. *ACM Trans. Intell. Syst. Technol. (TIST)* **2019**, *10*, 1–19. [CrossRef]
26. Alam, T.; Gupta, R. Federated Learning and Its Role in the Privacy Preservation of IoT Devices. *Future Internet* **2022**, *14*, 246. [CrossRef]
27. Dhiman, G.; Juneja, S.; Mohafez, H.; El-Bayoumy, I.; Sharma, L.K.; Hadizadeh, M.; Islam, M.A.; Viriyasitavat, W.; Khandaker, M.U. Federated Learning Approach to Protect Healthcare Data over Big Data Scenario. *Sustainability* **2022**, *14*, 2500. [CrossRef]
28. Aono, Y.; Hayashi, T.; Wang, L.; Moriai, S. Privacy-preserving deep learning via additively homomorphic encryption. *IEEE Trans. Inf. Forensics Secur.* **2017**, *13*, 1333–1345.
29. Chen, Y.R.; Rezapour, A.; Tzeng, W.G. Privacy-preserving ridge regression on distributed data. *Inf. Sci.* **2018**, *451*, 34–49. [CrossRef]
30. Fang, H.; Qian, Q. Privacy Preserving Machine Learning with Homomorphic Encryption and Federated Learning. *Future Internet* **2021**, *13*, 94. [CrossRef]
31. Park, J.; Lim, H. Privacy-Preserving Federated Learning Using Homomorphic Encryption. *Appl. Sci.* **2022**, *12*, 734. [CrossRef]
32. Angulo, E.; Márquez, J.; Villanueva-Polanco, R. Training of Classification Models via Federated Learning and Homomorphic Encryption. *Sensors* **2023**, *23*, 1966. [CrossRef]
33. Shen, X.; Jiang, H.; Chen, Y.; Wang, B.; Gao, L. PLDP-FL: Federated Learning with Personalized Local Differential Privacy. *Entropy* **2023**, *25*, 485. [CrossRef]
34. Wang, X.; Wang, J.; Ma, X.; Wen, C. A Differential Privacy Strategy Based on Local Features of Non-Gaussian Noise in Federated Learning. *Sensors* **2022**, *22*, 2424. [CrossRef]
35. Zhao, J.; Yang, M.; Zhang, R.; Song, W.; Zheng, J.; Feng, J.; Matwin, S. Privacy-Enhanced Federated Learning: A Restrictively Self-Sampled and Data-Perturbed Local Differential Privacy Method. *Electronics* **2022**, *11*, 4007. [CrossRef]
36. McMahan, H.B.; Ramage, D.; Talwar, K.; Zhang, L. Learning differentially private recurrent language models. *arXiv* **2017**, arXiv:1710.06963.
37. So, J.; Güler, B.; Avestimehr, A.S. Turbo-Aggregate: Breaking the Quadratic Aggregation Barrier in Secure Federated Learning. *IEEE J. Sel. Areas Inf. Theory* **2021**, *2*, 479–489. [CrossRef]
38. Xu, G.; Li, H.; Liu, S.; Yang, K.; Lin, X. VerifyNet: Secure and Verifiable Federated Learning. *IEEE Trans. Inf. Forensics Secur.* **2020**, *15*, 911–926. [CrossRef]
39. Kim, H.; Park, J.; Bennis, M.; Kim, S.L. Blockchained On-Device Federated Learning. *IEEE Commun. Lett.* **2020**, *24*, 1279–1283. [CrossRef]

40. Mahmood, Z.; Jusas, V. Blockchain-Enabled: Multi-Layered Security Federated Learning Platform for Preserving Data Privacy. *Electronics* **2022**, *11*, 1624. [CrossRef]
41. Liu, H.; Zhou, H.; Chen, H.; Yan, Y.; Huang, J.; Xiong, A.; Yang, S.; Chen, J.; Guo, S. A Federated Learning Multi-Task Scheduling Mechanism Based on Trusted Computing Sandbox. *Sensors* **2023**, *23*, 2093. [CrossRef] [PubMed]
42. Mortaheb, M.; Vahapoglu, C.; Ulukus, S. Personalized Federated Multi-Task Learning over Wireless Fading Channels. *Algorithms* **2022**, *15*, 421. [CrossRef]
43. Du, W.; Atallah, M.J. Privacy-preserving cooperative statistical analysis. In Proceedings of the Seventeenth Annual Computer Security Applications Conference, New Orleans, LA, USA, 10–14 December 2001; pp. 102–110.
44. Du, W.; Han, Y.S.; Chen, S. Privacy-preserving multivariate statistical analysis: Linear regression and classification. In Proceedings of the 2004 SIAM International Conference on Data Mining, Lake Buena Vista, FL, USA, 22–24 April 2004; pp. 222–233.
45. Khan, A.; ten Thij, M.; Wilbik, A. Communication-Efficient Vertical Federated Learning. *Algorithms* **2022**, *15*, 273. [CrossRef]
46. Hardy, S.; Henecka, W.; Ivey-Law, H.; Nock, R.; Patrini, G.; Smith, G.; Thorne, B. Private federated learning on vertically partitioned data via entity resolution and additively homomorphic encryption. *arXiv* **2017**, arXiv:1711.10677.
47. Schoenmakers, B.; Tuyls, P. Efficient binary conversion for Paillier encrypted values. In Proceedings of the Advances in Cryptology-EUROCRYPT 2006: 24th Annual International Conference on the Theory and Applications of Cryptographic Techniques, St. Petersburg, Russia, 28 May–1 June 2006; Springer: Berlin/Heidelberg, Germany, 2006; pp. 522–537.
48. Zhong, Z.; Zhou, Y.; Wu, D.; Chen, X.; Chen, M.; Li, C.; Sheng, Q.Z. P-FedAvg: Parallelizing Federated Learning with Theoretical Guarantees. In Proceedings of the IEEE INFOCOM 2021—IEEE Conference on Computer Communications, Vancouver, BC, Canada, 10–13 May 2021; pp. 1–10. [CrossRef]
49. Barreno, M.; Nelson, B.; Sears, R.; Joseph, A.D.; Tygar, J.D. Can machine learning be secure? In Proceedings of the 2006 ACM Symposium on Information, Computer and Communications Security, Taipei, Taiwan, 21–24 March 2006; pp. 16–25.
50. Huang, L.; Joseph, A.D.; Nelson, B.; Rubinstein, B.I.; Tygar, J.D. Adversarial machine learning. In Proceedings of the 4th ACM Workshop on Security and Artificial Intelligence, Chicago, IL, USA, 21 October 2011; pp. 43–58.
51. Chu, W.L.; Lin, C.J.; Chang, K.N. Detection and Classification of Advanced Persistent Threats and Attacks Using the Support Vector Machine. *Appl. Sci.* **2019**, *9*, 4579. [CrossRef]
52. Chen, Y.; Hayawi, K.; Zhao, Q.; Mou, J.; Yang, L.; Tang, J.; Li, Q.; Wen, H. Vector Auto-Regression-Based False Data Injection Attack Detection Method in Edge Computing Environment. *Sensors* **2022**, *22*, 6789. [CrossRef]
53. Jiang, Y.; Zhou, Y.; Wu, D.; Li, C.; Wang, Y. On the Detection of Shilling Attacks in Federated Collaborative Filtering. In Proceedings of the 2020 International Symposium on Reliable Distributed Systems (SRDS), Shanghai, China, 21–24 September 2020; pp. 185–194. [CrossRef]
54. Alfeld, S.; Zhu, X.; Barford, P. Data poisoning attacks against autoregressive models. In Proceedings of the AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016; Volume 30.
55. Biggio, B.; Nelson, B.; Laskov, P. Poisoning attacks against support vector machines. *arXiv* **2012**, arXiv:1206.6389.
56. Zügner, D.; Akbarnejad, A.; Günnemann, S. Adversarial attacks on neural networks for graph data. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018; pp. 2847–2856. Available online: https://arxiv.org/abs/1805.07984 (accessed on 18 October 2023).
57. Zhao, M.; An, B.; Yu, Y.; Liu, S.; Pan, S. Data poisoning attacks on multi-task relationship learning. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; Volume 32.
58. Dhiman, S.; Nayak, S.; Mahato, G.K.; Ram, A.; Chakraborty, S.K. Homomorphic Encryption based Federated Learning for Financial Data Security. In Proceedings of the 2023 4th International Conference on Computing and Communication Systems (I3CS), Shillong, India, 16–18 March 2023; pp. 1–6.
59. Jia, B.; Zhang, X.; Liu, J.; Zhang, Y.; Huang, K.; Liang, Y. Blockchain-Enabled Federated Learning Data Protection Aggregation Scheme With Differential Privacy and Homomorphic Encryption in IIoT. *IEEE Trans. Ind. Inform.* **2022**, *18*, 4049–4058. [CrossRef]
60. Zhang, S.; Li, Z.; Chen, Q.; Zheng, W.; Leng, J.; Guo, M. Dubhe: Towards data unbiasedness with homomorphic encryption in federated learning client selection. In Proceedings of the 50th International Conference on Parallel Processing, Lemont, IL, USA, 9–12 August 2021; pp. 1–10.
61. Guo, X. Federated Learning for Data Security and Privacy Protection. In Proceedings of the 2021 12th International Symposium on Parallel Architectures, Algorithms and Programming (PAAP), Xi'an, China, 10–12 December 2021; pp. 194–197. [CrossRef]
62. Fan, C.I.; Hsu, Y.W.; Shie, C.H.; Tseng, Y.F. ID-Based Multireceiver Homomorphic Proxy Re-Encryption in Federated Learning. *ACM Trans. Sens. Netw.* **2022**, *18*, 1–25. [CrossRef]
63. Kou, L.; Wu, J.; Zhang, F.; Ji, P.; Ke, W.; Wan, J.; Liu, H.; Li, Y.; Yuan, Q. Image encryption for Offshore wind power based on 2D-LCLM and Zhou Yi Eight Trigrams. *Int. J. Bio-Inspired Comput.* **2023**, *22*, 53–64. [CrossRef]
64. Li, L.; Li, T. Traceability model based on improved witness mechanism. *CAAI Trans. Intell. Technol.* **2022**, *7*, 331–339. [CrossRef]
65. Lee, J.; Duong, P.N.; Lee, H. Configurable Encryption and Decryption Architectures for CKKS-Based Homomorphic Encryption. *Sensors* **2023**, *23*, 7389. [CrossRef]
66. Ge, L.; Li, H.; Wang, X.; Wang, Z. A review of secure federated learning: Privacy leakage threats, protection technologies, challenges and future directions. *Neurocomputing* **2023**, *561*, 126897. [CrossRef]
67. Zhang, J.; Zhu, H.; Wang, F.; Zhao, J.; Xu, Q.; Li, H. Security and privacy threats to federated learning: Issues, methods, and challenges. *Secur. Commun. Netw.* **2022**, *2022*, 2886795. [CrossRef]

68. Zhang, J.; Li, M.; Zeng, S.; Xie, B.; Zhao, D. A survey on security and privacy threats to federated learning. In Proceedings of the 2021 International Conference on Networking and Network Applications (NaNA), Lijiang City, China, 29 October–1 November 2021; pp. 319–326. [CrossRef]
69. Li, Y.; Bao, Y.; Xiang, L.; Liu, J.; Chen, C.; Wang, L.; Wang, X. Privacy threats analysis to secure federated learning. *arXiv* **2021**, arXiv:2106.13076.
70. Asad, M.; Moustafa, A.; Yu, C. A critical evaluation of privacy and security threats in federated learning. *Sensors* **2020**, *20*, 7182. [CrossRef]
71. Manzoor, S.I.; Jain, S.; Singh, Y.; Singh, H. Federated Learning Based Privacy Ensured Sensor Communication in IoT Networks: A Taxonomy, Threats and Attacks. *IEEE Access* **2023**, *11*, 42248–42275. [CrossRef]
72. Benmalek, M.; Benrekia, M.A.; Challal, Y. Security of federated learning: Attacks, defensive mechanisms, and challenges. *Rev. Sci. Technol. L'Inform. Série RIA Rev. D'Intell. Artif.* **2022**, *36*, 49–59. [CrossRef]
73. Arbaoui, M.; Rahmoun, A. Towards secure and reliable aggregation for Federated Learning protocols in healthcare applications. In Proceedings of the 2022 Ninth International Conference on Software Defined Systems (SDS), Paris, France, 12–15 December 2022; pp. 1–3.
74. Abdelli, K.; Cho, J.Y.; Pachnicke, S. Secure Collaborative Learning for Predictive Maintenance in Optical Networks. In Proceedings of the Secure IT Systems: 26th Nordic Conference, NordSec 2021, Virtual Event, 29–30 November 2021; Proceedings 26; Springer International Publishing: Berlin/Heidelberg, Germany, 2021; pp. 114–130.
75. Jeong, H.; Son, H.; Lee, S.; Hyun, J.; Chung, T.M. FedCC: Robust Federated Learning against Model Poisoning Attacks. *arXiv* **2022**, arXiv:2212.01976.
76. Lyu, X.; Han, Y.; Wang, W.; Liu, J.; Wang, B.; Liu, J.; Zhang, X. Poisoning with cerberus: Stealthy and colluded backdoor attack against federated learning. In Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2023.
77. Wei, K.; Li, J.; Ding, M.; Ma, C.; Jeon, Y.S.; Poor, H.V. Covert model poisoning against federated learning: Algorithm design and optimization. *IEEE Trans. Dependable Secur. Comput.* **2023**. [CrossRef]
78. Sun, J.; Li, A.; DiValentin, L.; Hassanzadeh, A.; Chen, Y.; Li, H. Fl-wbc: Enhancing robustness against model poisoning attacks in federated learning from a client perspective. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 12613–12624.
79. Mao, Y.; Yuan, X.; Zhao, X.; Zhong, S. Romoa: Ro bust Mo del A ggregation for the Resistance of Federated Learning to Model Poisoning Attacks. In Proceedings of the Computer Security–ESORICS 2021: 26th European Symposium on Research in Computer Security, Darmstadt, Germany, 4–8 October 2021; Springer: Berlin/Heidelberg, Germany, 2021; pp. 476–496.
80. Pillutla, K.; Kakade, S.M.; Harchaoui, Z. Robust Aggregation for Federated Learning. *IEEE Trans. Signal Process.* **2022**, *70*, 1142–1154. [CrossRef]
81. Wang, X.; Liang, Z.; Koe, A.S.V.; Wu, Q.; Zhang, X.; Li, H.; Yang, Q. Secure and efficient parameters aggregation protocol for federated incremental learning and its applications. *Int. J. Intell. Syst.* **2022**, *37*, 4471–4487. [CrossRef]
82. Hao, M.; Li, H.; Xu, G.; Chen, H.; Zhang, T. Efficient, private and robust federated learning. In Proceedings of the Annual Computer Security Applications Conference, Virtual Event, 6–10 December 2021; pp. 45–60.
83. Smith, V.; Chiang, C.K.; Sanjabi, M.; Talwalkar, A.S. Federated multi-task learning. *Adv. Neural Inf. Process. Syst.* **2017**, *30*.
84. Shalev-Shwartz, S.; Zhang, T. Accelerated proximal stochastic dual coordinate ascent for regularized loss minimization. In Proceedings of the International Conference on Machine Learning, Beijing, China, 22–24 June 2014; pp. 64–72.
85. Wassan, S.; Suhail, B.; Mubeen, R.; Raj, B.; Agarwal, U.; Khatri, E.; Gopinathan, S.; Dhiman, G. Gradient Boosting for Health IoT Federated Learning. *Sustainability* **2022**, *14*, 16842. [CrossRef]
86. Feng, J.; Yang, L.T.; Zhu, Q.; Choo, K.K.R. Privacy-preserving tensor decomposition over encrypted data in a federated cloud environment. *IEEE Trans. Dependable Secur. Comput.* **2018**, *17*, 857–868. [CrossRef]
87. Feng, J.; Yang, L.T.; Ren, B.; Zou, D.; Dong, M.; Zhang, S. Tensor recurrent neural network with differential privacy. *IEEE Trans. Comput.* **2023**. [CrossRef]
88. Karras, A.; Karras, C.; Giotopoulos, K.C.; Tsolis, D.; Oikonomou, K.; Sioutas, S. Federated Edge Intelligence and Edge Caching Mechanisms. *Information* **2023**, *14*, 414. [CrossRef]
89. Karras, A.; Karras, C.; Schizas, N.; Avlonitis, M.; Sioutas, S. AutoML with Bayesian Optimizations for Big Data Management. *Information* **2023**, *14*, 223. [CrossRef]