



## Security of federated learning with IoT systems: Issues, limitations, challenges, and solutions



Jean-Paul A. Yaacoub<sup>a</sup>, Hassan N. Noura<sup>a,\*</sup>, Ola Salman<sup>b</sup>

<sup>a</sup> Univ. Bourgogne Franche-Comté (UBFC), FEMTO-ST Institute, France

<sup>b</sup> American University of Beirut, Electrical and Computer Engineering Department, Beirut, 1107 2020, Lebanon

### ARTICLE INFO

#### Index terms:

Federated learning  
Federated learning threats and attacks  
Federated learning security solutions  
Internet of things  
Machine learning security

### ABSTRACT

Federated Learning (FL, or Collaborative Learning (CL)) has surely gained a reputation for not only building Machine Learning (ML) models that rely on distributed datasets, but also for starting to play a key role in security and privacy solutions to protect sensitive data and information from a variety of ML-related attacks. This made it an ideal choice for emerging networks such as Internet of Things (IoT) systems, especially with its state-of-the-art algorithms that focus on their practical use over IoT networks, despite the presence of resource-constrained devices. However, the heterogeneous nature of the current devices and models in complex IoT networks has seriously hindered the FL training process's ability to perform well. Thus, rendering it almost unsuitable for direct deployment over IoT networks despite ongoing efforts to tackle this issue and overcome this challenging obstacle. As a result, the main characteristics of FL in the IoT from both security and privacy aspects are presented in this study. We broaden our research to investigate and analyze cutting-edge FL algorithms, models, and protocols, with a focus on their efficacy and practical application across IoT networks and systems alike. This is followed by a comparative analysis of the recently available protection solutions for FL that can be based on cryptographic and non-cryptographic solutions over heterogeneous, dynamic IoT networks. Moreover, the proposed work provides a list of suggestions and recommendations that can be applied to enhance the effectiveness of the adoption of FL and to achieve higher robustness against attacks, especially in heterogeneous dynamic IoT networks and in the presence of resource-constrained devices.

### 1. Introduction

As billions of IoT devices are currently in use, with many more to be deployed in the near future, the growth of IoT devices has led to the generation of voluminous data that also contains clients' private information. Thus, leading to higher network overhead, communication, and storage costs, while also causing mixed privacy concerns [1]. In fact, the IoT is now in every aspect of everyone's life and is deployed across lots of domains such as healthcare, industry, smart grids, and robotics, especially with the presence of intelligent automated applications, devices, and services that tend to be more and more Artificial Intelligence (AI)-based and empowered [2]. However, AI demands centralized data to be collected and processed, which is not an easy task due to scalability issues, resource-constrained devices, and power consumption problems [3]. As a result, Federated Learning (FL) was adopted as a distributed and adaptive collaborative AI training approach to sort this issue and offer a higher degree of user-level privacy without the need for any data-sharing

operations. Instead, in FL, it is achieved by sending a copy of untrained ML models to all clients in a given network. However, the IoT's interconnected devices are heterogeneous, resource-constrained, and spread across different geographical locations with little control. This would certainly cause connectivity issues, mainly due to limited bandwidth and resources.

#### 1.1. Problem formulation

However, the main issue is that the implementation of FL in IoT systems makes them prone and vulnerable to potential adversary cyber-attacks such as model inversion and membership inference attacks. Moreover, IoT applications are closely related to sensitive services, especially since they handle sensitive information about users. The main challenge in the IoT domain is preserving the user's privacy without degrading the security level. Therefore, a set of these cyber-attacks can have drastic consequences, especially in sensitive systems such as

\* Corresponding author.

E-mail address: [hassan.noura@univ-fcomte.fr](mailto:hassan.noura@univ-fcomte.fr) (H.N. Noura).

military or medical ones, which would hinder the wider deployment of smart and automated IoT applications. Therefore, new security solutions should be introduced to detect and prevent them. Existing solutions can be divided into two main classes: cryptographic and non-cryptographic. Homomorphic encryption and multi-party computation can be considered cryptographic solutions. On the other hand, an example of a non-cryptographic solution is differential privacy. Either way, appropriate security and privacy solutions for FL in IoT systems should include minimum computations and require minimal resources, especially on the IoT devices side.

### 1.2. Contribution

This paper contributes by adding the following:

- **Discussing:** client/server-related security and privacy issues on local and central servers and highlighting them.
- **Presenting:** the list of all possible FL-related attacks that can target the IoT and suggest suitable security measures.
- **Analyzing** all the available security measures, dividing them between cryptographic and non-cryptographic solutions, and discussing them.
- **Presenting:** a list of all the possible FL-related vulnerabilities that can target the IoT domain and suggesting suitable countermeasures to mitigate these threats.
- **Proposing:** a framework, suggestions, recommendations, and the lessons learned from all this ongoing work.

### 1.3. Organization

This paper is divided into nine sections (see Fig. 1) along with the introduction and is presented as follows: In section II, the FL-IoT background is presented, while discussing the relation between FL and IoT, presenting the FL types, and mentioning the FL-IoT data and application services. In section III, the main FL-IoT challenges are highlighted, while discussing its future opportunities. FL-related security attacks are discussed and analysed in section IV, including attack categories and types. The source of the FL's vulnerability is discussed in section V, while in section VI, the FL-related solutions and countermeasures are analysed. All the learnt lessons are presented in section VII, while our suggestions and recommendations for future research FL directions are presented in section VIII, especially cryptographic and non-cryptographic measures, as well as policies and management. In section IX, we conclude our work.

## 2. Background and preliminaries

The IoT nature with its intelligent AI-empowered applications and services managed to cover key domains in our daily life including industry, medicine, agriculture/farming, smart cities, smart homes, smart transportation, smart autonomous vehicles, robots and modular robots (i.e Unmanned Aerial Vehicles, Unmanned Ground Vehicles, and Unmanned Underwater Vehicles) [4,5]. Unlike traditional AI techniques that rely on the centralized collection and processing of data, FL, as a distributed collaborative AI approach, enabled many intelligent IoT applications by training distributed IoT devices without sharing data (see Fig. 2) while achieving user privacy-preserving and protection. Moreover, FL offers a wide range of IoT services, including preserving data localisation [3], demystifying hidden IoT data patterns, collecting/analysing massive data volumes, real-time IoT data sharing [6], data caching/offloading, enhancing smart/logical real-time decision-making for IoT, attack and anomaly detection, user/data privacy, and IoT security [7]. This will increase operational efficiency, and performance accuracy, and will surely reduce costs. In fact, before proceeding any further, it is important to know how the whole client-server communication takes place in an FL system. The following Fig. 3 is used to highlight how this communication takes place, based on the model already presented in Ref. [8]. On the other hand, an abbreviation list is presented in Table 1.

### 2.1. Federated learning & IoT

IoT systems involve a large number of connected devices that are distributed (decentralized) throughout the network and are connected to the internet. They generate vast amounts of different types of data and can be considered a decentralized data collection system. By integrating FL into the IoT systems, there is a potential to enhance their efficiency, security, and privacy preservation, which will ultimately result in improving both performance and accuracy.

In terms of efficiency, FL can improve the performance of IoT systems as the computational power of IoT end-devices can be utilized to train a local lightweight ML model on data generated by it or by its neighbour's sensors and/or other IoT devices. Another enhancement that can be ensured by integrating FL into the IoT is in situations where the network's connectivity is limited, hence taking advantage of the computational power of edge/fog devices that are located closer to the data source and can perform computations (training) on the data before sending it to the cloud. Thus, local training can be performed also on edge devices, which can ensure a significant gain in saving bandwidth and reducing latency.

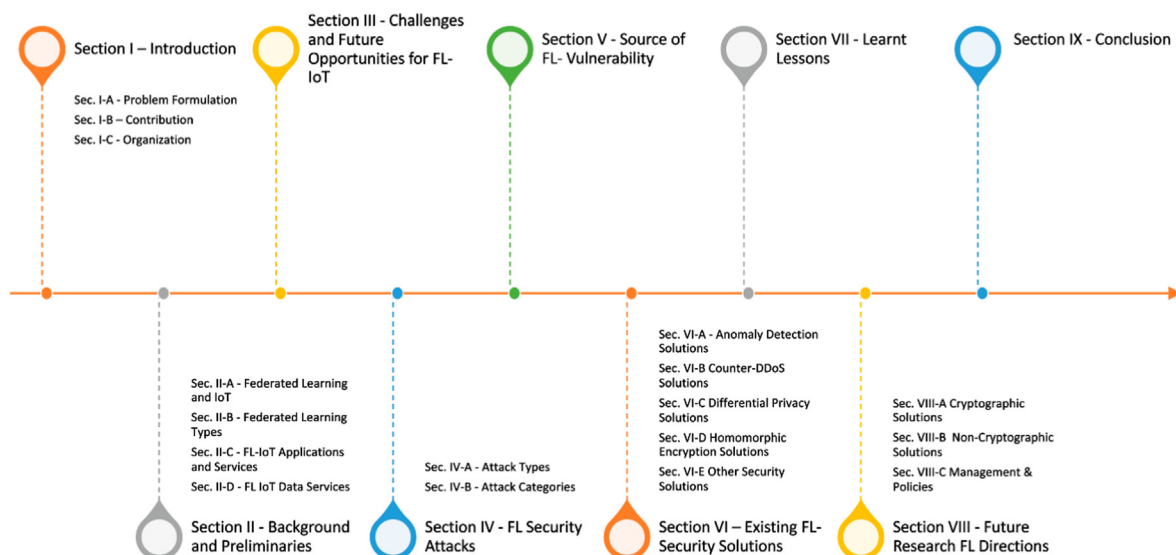


Fig. 1. Structure of the presented survey.

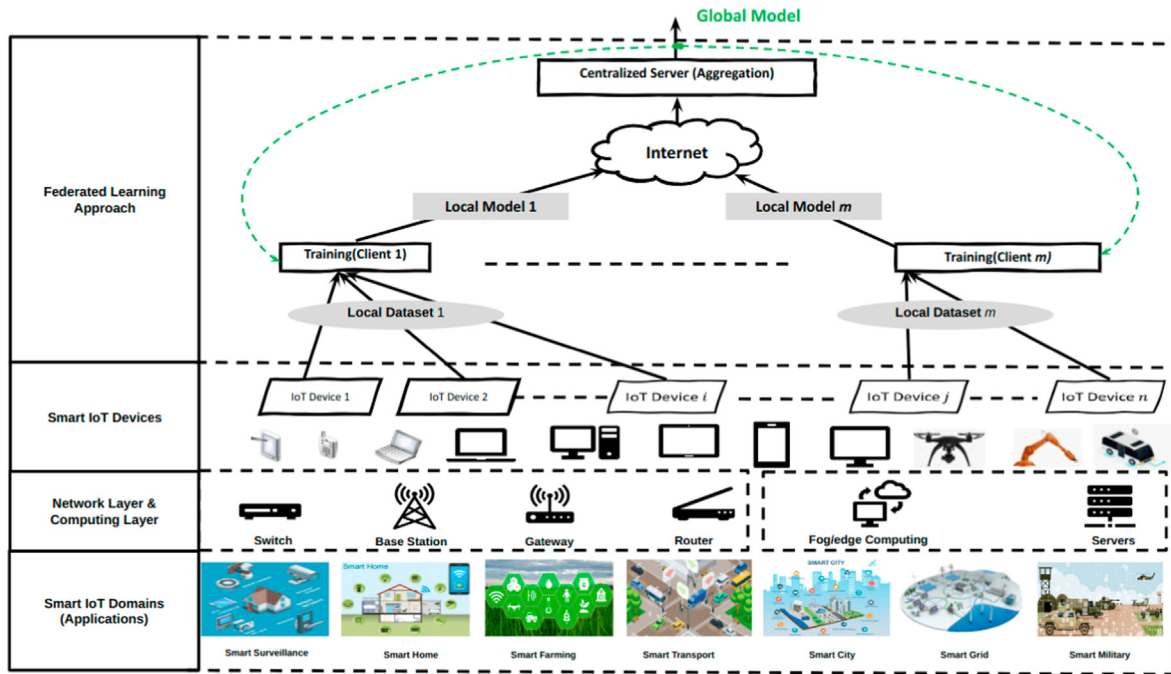


Fig. 2. Relation between Federated Learning and IoT devices, which form local datasets that are used during the training at edge/fog or end IoT devices.

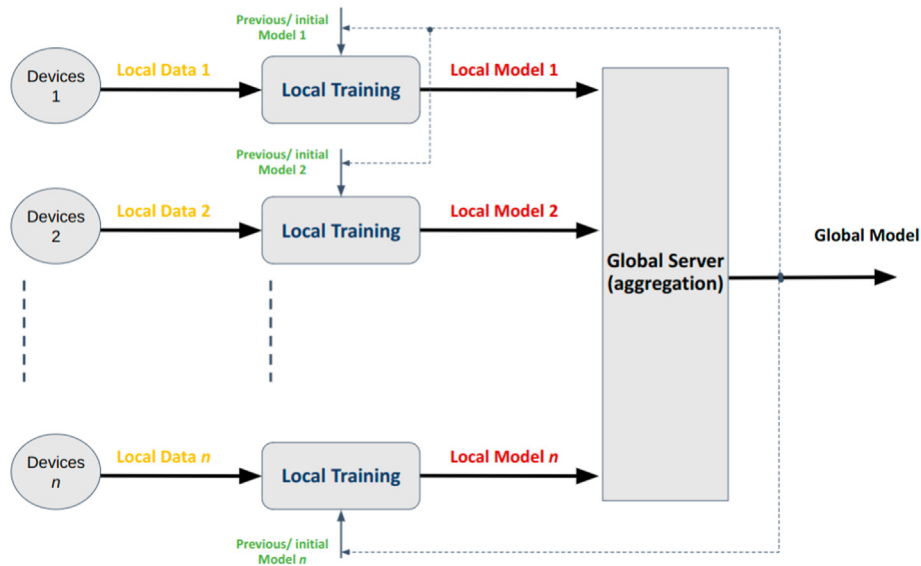


Fig. 3. FL-based Client-Server communications system. Devices can be IoT devices or edge/fog devices.

This ensures the reduction of the amount of data that needs to be transmitted to the cloud (only the local training model). Moreover, another property can be ensured, which is collaborative learning as multiple devices (end IoT devices and/or edge devices as listed previously) from different locations and with different capabilities will work together to construct an accurate and precise global model as it is based on a diverse set of data sources.

In terms of security and data privacy, a set of IoT applications will often involve the communication of sensitive data, which makes them suffer from the concern of security and privacy issues. These issues can be solved by integrating FL with IoT systems since the data persists in the IoT end device(s) and/or edge device(s) and by ensuring secure communication between IoT/edge devices and servers. This will consequently reduce the risk of data breaches and unauthorized access to sensitive information as the data remains decentralized, and each device

can locally train its local model based on the data that it collects. The advantage of FL is that the global model is constructed without the need for the data to leave the IoT/edge devices.

In this section, the different types of FL are presented first. After this, a set of FL-IoT applications and services are described. Then, the different FL-IoT data services are discussed.

### 2.2. Federated learning types

FL allows several models to be trained and aggregated before reaching the final model. To achieve that, FL is divided into 5 main types with different aspects, which are further detailed in Ref. [75], and are presented and discussed as follows:

**Table 1**  
List of abbreviation.

Abbreviation	Explanation
AI	Artificial Intelligence
FL	Federated Learning
CL	Collaborative Learning
ML	Machine Learning
IoT	Internet of Things
DL	Deep Learning
P2P	Peer-to-Peer
IIoT	Industrial IoT
CPS	Cyber-Physical Systems
SA	Smart Agriculture
DLTs	Distributed Ledger Technologies
IID	Independent and Identically Distributed
TL	Transfer Learning
FTL	Federated Transfer Learning
HFL	Horizontal Federated Learning
VFL	Vertical Federated Learning
CSFL	Cross-Silo FL
GAN	Generative Adversarial Networks
DDoS	Distributed Denial of Service
MitM	Man-in-the-Middle
LEGATO	Layerwise Gradient AggregatTiOn
FLPC	Federated Learning Parameter Compression
FDI	False Data Injection
SE	State Estimation
IDS	Intrusion Detection System
IPS	Intrusion Prevention Systems
KPCA	Kernel Principal Component Analysis
LF	Label Flipping
DBA	Distributed Backdoor Attack
FLIP	Federated LearnIng Provable
MLOps	Machine Learning Operations
DTL	Deep Transfer Learning
SGD	Stochastic Gradient Descent
API	Application Programming Interface
SQLIA	SQL-Injection Attack
DNNs	Deep Neural Networks
DP	Differential Privacy
DDos	Distributed Denial of service
FMTL	Federated Multi-Task Learning
FD	Federated Distillation
STD	Sanitized Training Data
MTL	Multi-Task Learning
KD	Knowledge Distillation
MTD	Moving Target Defense
ZKP	Zero-Knowledge Proofs
TEE	Trusted Execution Environment
TAs	Trusted Applications
AD	Anomaly Detection
FEDTIMEDIS	FEDerated TIME DIStributed
LSTM	Long Short-Term Memory
CNN	Convolutional Neural Network
MGVN	Mixed Gaussian Variational self-encoding Networks
AMCNN	Attention Mechanism-based Convolutional Neural Network
SCADA	Supervisory Control and Data Acquisition
IMA	Iterative Model Averaging
GRU	Gated Recurrent Units
MSA	Model Shuffle Attack
LDP	Local Differential Privacy
FFL	Fragmented Federated Learning
HE	Homomorphic Encryption
FHE	Fully Homomorphic Encryption
MPC	Multi-Party Computation
AES	Advanced Encryption Standard
FedDRL	Federated Deep Reinforcement Learning

- **Data Partitioning:** includes the datasets of different clients along with the degree of these features' similarity. In fact, data partitioning can be divided into three main types [76].
  - **Horizontal Data Partitioning:** occurs when the clients' datasets have the same features but with the least sample space intersection.
  - **Vertical Data Partitioning:** occurs when the client datasets are exposed to the same sample space but with different feature spaces.

**Table 2**  
Server-based Attacks and countermeasures.

- **FL Client/Agent-based Attacks:** are usually attacks being carried out either accidentally or deliberately by one or several FL clients/agents, and are presented below (see Table 3) as follows:

Target	Attack Type	Exploit	Outcome	Security Measures
Server-based	Server-Side GAN-Based Attack	Data Privacy	Achieves an invisible attack [9]	Early detection [10]
	DDoS	Server/Data Availability	Clients unable to connect	Early DDoS detection and mitigation
	Eavesdropping Attack	Data privacy	Extracts users data	Enhanced channel security
	Man-In-The-Middle Attack	Data integrity	Replaces packets with malicious ones	Enhanced lightweight encryption

- **Hybrid Data Partitioning:** combines both horizontal and vertical data partitioning but is rarely used.
- **ML Models:** The choice of each homogeneous/heterogeneous model is not uniform. This means that it varies depending on the problem to solve and the dataset. However, Deep Learning (DL) models and tree-based models (Random Forest, Xgboost) are the most commonly adopted models [77].
- **Privacy Mechanisms:** rely on the adoption of FL to overcome privacy and privacy-related issues, most importantly preventing the leakage of clients' data and information. In order to achieve that, FL relies on two techniques, which include differential privacy and other cryptographic methods.
  - **Differential Privacy:** protects privacy by adding random noise to data to mask the gradients. However, adding noise comes at the cost of affecting the model's accuracy.
  - **Cryptographic Methods:** adopts data encryption techniques, most commonly homomorphic encryption and/or secure multi-party computation, to protect the clients data before securely transmitting it on the server. Though these solutions offer a higher level of privacy protection, except that they are computationally expensive.
- **FL System Architecture:** is divided into two main types. However, despite this division, the functioning of the FL systems remains unchanged, except for the client-server communication [78].
  - **Centralized FL Systems:** allow a separate model called the global model to “behave” like a server, where all parameter updates are conducted. Moreover, model learning can either be synchronous or asynchronous.
  - **Decentralized FL Systems:** allow clients to change turns by taking on the role of a server, where a client can randomly retrieve an epoch to make updates on the global model and communicate it to other clients. In fact, decentralized FL systems consist of three main types, which are: Peer-to-Peer (P2P), graph, and blockchain.
- **Scale of Federation:** the scale of federation is divided between two main categories [79] as seen in Fig. 4:
  - **Cross-Device Federated Learning:** is often associated with organizations, where despite having a small client number, they tend to have a large computational power.
  - **Cross-Silo Federated Learning:** is often associated with mobiles, where there are a huge number of clients but with a small computational power.

**Table 3**

Client/Agent-based Attacks and countermeasures.

- **FL Data-based Attacks:** are usually attacks being carried out to intercept, manipulate, or modify the intercepted and hijacked data, which can be done passively via eavesdropping or actively via man-in-the-middle attacks. As a result, several FL data-based attacks are presented below (see Table 4) as follows:

Target	Attack Type	Exploit	Outcome	Security Measures
Client/ Agent-based	Client-Side GAN-Based Attack	User privacy	Exploits the training process's real-time nature [11]	Federated Learning Parameter Compression (FLPC) [12], or anomaly detection algorithm [13]
	Byzantine Attack	Data integrity	Causes convergence problems [14]	Layerwise Gradient AggregatTiOn (LEGATO) [15] or DiverseFL [16]
	Explicit Boosting Attack	Data integrity/user privacy	Evades classification and boosts the local malicious update [17]	Online anomaly detection algorithm [18], Deep Learning (DL) based method [19], and a novel anomaly-based Intrusion Detection System [20]
	Foolsgold Attack	Data integrity/FL security	Uses fake identities to break FL security and authenticity	CONTRA [21] or the removal of malicious nodes from the training environment [22]
	Sybil Attack	Data integrity/FL security	Simulates dummy participant accounts to target the FL [23]	Multi-Armed Bandit for Federated Learning (MABRFL) [24] and Anomaly detection [25]
	Backdoor Insertion Attack	Data integrity/system accuracy	Targets the training data	FLAME [26], universal model-agnostic defense technique (Moat) [27] and Feedback-based Federated Learning (BAFFLE) [28]
	Label Flipping Attack	Data integrity to learn triggers in inputs	Has access to training data poison it and permute labels	Novel defense approach [29] and Kernel Principal Component Analysis (KPCA) and K-mean clustering [30]

**Table 4**

Data-based Attacks and countermeasures.

- **Coding-based Attacks:** the following list includes all the possible coding-related attacks that target the IoT systems and is discussed and presented below (see Table 5) as follows:

Target	Attack Type	Exploit	Outcome	Security Measures
Data-based	Clean-Label Attack	Data privacy, integrity	Avoids changing the input data, and craft a poisoned training data instead [31]	Ensemble-based Nested Training technique [32]
	Data Poisoning Attack	System performance, data integrity	Leads the system to behave in a way that is advantageous to the attacker	Reasonable supply-chain checks, anomaly detection
	Model Poisoning Attack	System performance, data integrity	Poisoning the global model [33]	Sparsefed [34] or MLGuard [35]
	Data Tampering - Modification Attack	Data privacy, integrity, availability	Creates a feature collision effect that merges two dataset classes to fool the ML model	Dynamic Redundant Path Selection (DRPS) [36,37]
	Free-Riding Attack	Data privacy, integrity	Collects the final model, mostly by inserting dummy updates without training the model	Viceroy [38] or P2P Straightforward Protocol (P2PSP) [39]
	Generative Adversarial Networks Attack	Data privacy, integrity	Intercepts training samples via inference, before poisoning the training data	CycleGAN-based [40] or FlowGAN [41]
	Evasion Attack	Data privacy, integrity	Exploits weak spots and vulnerabilities [42]	Ensemble learning approach [43] or region-based classification [44]
	Distributed Backdoor Attack	Data privacy, integrity	Decomposes a global trigger pattern into separate local ones to achieve more stealth and persistence	DeepSight [45] or Federated Learning Provable defense framework [46]
	Model Inversion Attack	Data privacy, integrity	Gets sensitive data from the training set	Black-box model inversion attack [47]
	Membership Inference Attack	Data privacy, integrity	Applies the trained model to specific inputs and evaluating the outcome	Differential Privacy, or purification framework [48]
	Online Adversarial Attack	Data privacy, integrity	Manipulates the model's learning via sending false data	MLOps [49], or secure federated learning with Transformer [50]
	Transfer Learning Attack	Data privacy, integrity	Exploits the pre-trained process that is based on a larger data set	Deep Transfer Learning [51]
	Adversarial Machine Learning Attack	Data privacy, system performance	Small variations in the model data inputs are found and exploited	Data Operations [52] and Machine Learning Operations [53]
	Model Stealing Attack	Data privacy, integrity	The learned model is accurately copied by the attacker	High-performance Deep Neural Networks [54] or PRADA [55]
	Gradient Leakage and Gradient Manipulation Attacks	Data privacy, system accuracy	Aims to steal and recover private, sensitive or confidential training data, and overall accuracy from the shared gradients	BadBatch [56]

2.3. FL-IoT applications and services

The important lessons learned from this review of the FL-IoT services and applications are also highlighted, while also being represented in Fig. 5. We then provide an extensive survey of the use of FL in various key IoT applications, such as:

2.4. FL IoT data services

FL has the potential to cover a wider range of IoT domains while also achieving a wider range of IoT data services, which are presented below as follows:

- **Data production:** occurs frequently and on a daily basis with ongoing tasks and processes, making it readily available for both frequent and efficient access.

- **Data sharing:** makes data sharing available and securely transmitted from an FL server to local clients' devices. In the event that there is a large volume of data, it can be compressed and transmitted to avoid network congestion and bottlenecks.
- **Data offloading:** relies on complementary IoT network technologies to ensure that the data has been delivered to the right FL server, often to reduce bandwidth usage.
- **Data caching:** allows high-speed data storage that stores data subsets to ensure that data is transmitted at a faster pace at each future operation.
- **Data storage:** relies on IoT storage devices such as magnetic, optical, or mechanical media to preserve digital data and information for both ongoing and future operations on FL servers and among their inter-connected clients.

**Table 5**  
Coding-based Attacks and countermeasures.

• **FL Client-Server-based Attacks:** communication attacks also occur, resulting in the total exposure of FL-IoT communications. As a result, these main attacks are presented below (see Table 6) with their suitable countermeasures as follows:

Target	Attack Type	Exploit	Outcome	Security Measures
Coding-based Attacks	SQL Attack	User/Data privacy, integrity	Retrieves or modifies the submitted strings to gain an administrative privilege, extract details, modify databases or compromise users accounts	Uses fully automated technique or the Knuth-Morris-Pratt (KMP) [57]
	Trojanned Model Attack	Data privacy, integrity	Targets pre-trained model pool to execute an arbitrary code	Watermarked data [58]
	Infected Model Attack	Data privacy, integrity	Causes abnormal and suspicious behaviours to the existing model	Prevent third-party interventions and constantly check supply chains
	Adversarial Perturbation Attack	Data privacy, integrity	Logically/physically modifying objects that are used as ML system inputs	Robust training models, advanced authentication measures and notification alerts [59]
	Decision Boundary Detection Attack	Data privacy, integrity	identifies the IDS limitations or exploitable vulnerabilities	Advanced intrusion detection/prevention systems or honeypots
	Textual Training Data Extraction Attack	Data privacy, integrity	Generates text containing verbatim portions of the data it was trained on	Application refactoring, implementing rate limits and authentication, and adding differential privacy [60]
	Inference by Covariance Attack	user privacy, data integrity	Infers the individual user behaviour on a given system, and the system's response	Identifying firm and clear restrictions, recommendations, and applying rate limits
Approximate Copy and High-Fidelity Copy Attack	Data privacy, integrity	Creates a near-perfect/approximate copy to ensure a higher and accurate level of replication	Differential Privacy (DP) OR noise addition [61]	

**Table 6**  
Client-Server-based Attacks and countermeasures.

Target	Attack Type	Exploit	Outcome	Security Measures
Client-Server-based	Local Storage Attack	Data integrity	Extracts all the stored data and inject it with malicious data loads via JavaScript	Lightweight integrity protection [62]
	Cross-Site Flashing Attack	Data privacy	Mitigates and discovers flash applications	FlashOver counter XSS [63]
	Cross-Site Scripting Attack	Data integrity, privacy, availability	Allows the end user to spoof or modify web page contents	XSS-GUARD [64] and XSS-SAFE [65]
	Cascading Style Sheets Injection Attack	User privacy, Data integrity	Can be an XSS, User Interface (UI) or data modification/extraction attack	Injected style sheets [66] or SIACHEN [67]
	Client-Side Resource Manipulation Attack	Data integrity, availability	Exploits and controls the URL that links to the other web page resources	MLPXSS [68], Pulse Connect Secure and Virtual Web Application Firewall [69]
	Cross-origin Resource Sharing	Data integrity, privacy, availability	Causes cross-origin attacks like cross-site request forgery (CSRF)	XSS detection mechanism integrated with HTML5 and CORS [70] or Application Cache [71]
	Session Fixation Attack	User privacy, Data integrity	Intercepts and retrieves the HTTP server's session state information	Serene [72], authentication-based scheme [73] or automated session fixation vulnerability detection [74]

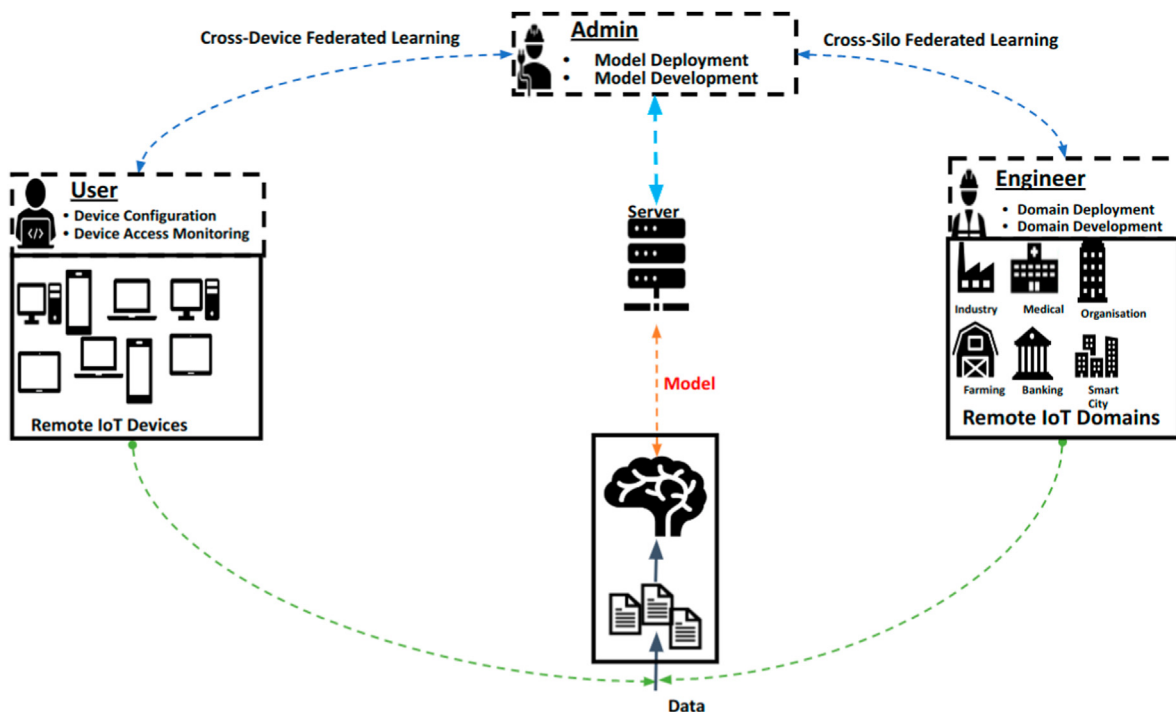


Fig. 4. IoT between cross-Silo and cross-device federated learning.



Fig. 5. A set of IoT Applications that can benefit from the Federated Learning approach.

- **Smart Healthcare:** FL allows hospitals to benefit from the available datasets of non-affiliated hospitals without having the data being centralized in a single location to overcome critical data-related issues such as privacy, security, and access rights [80].
- **Smart Industry and Manufacturing:** FL can be used to reach operational industrial systems and intelligent Industrial IoT (IIoT) applications that require centralized data collection and processing, by coordinating IIoT machines and devices to perform centralized AI training at the network edge with respect to data privacy, security, and confidentiality [1]. Simultaneously, FL models are also being used to develop predictive equipment maintenance labels, especially for users with privacy concerns [81]. Thus, ensuring the fusion of Federated Learning and the Industrial Internet of Things, while introducing the new notion of FL-IIoT [82, 83].
- **Cyber-Physical Systems (CPS):** are relying on FL to develop trustworthy smart services that rely on the dynamic and distributed nature of edge computing, such as smart connected vehicles, and smart logistics, with a high amount of real-time data being produced. A solution called FengHuoLun, was presented in Ref. [84] to implement smart services with MLs trained in a trusted Federated Learning framework, with CPS behaviours being monitored.
- **Smart Cities:** FL plays a key role in processing smart cities, especially with the development of AI and big data, by working on solving privacy and security issues while maintaining waste management, transportation, communications, traffic congestion, traffic lights, and logistics, etc [85,86].
- **Smart Agriculture (SA):** is relying on low-cost and low-energy-consuming sensors and devices to enhance both quantitative and qualitative agricultural production while addressing big data security and privacy violation by presenting a solution called PEFL [87]. PEFL is a deep Privacy-Encoding-based FL framework that uses a perturbation-based encoding and long short-term memory-auto-encoder technique to achieve the intended privacy and identify attacks.
- **Autonomous Robotics and Unmanned Vehicles:** the adoption of FL and Distributed Ledger Technologies (DLTs) to ensure low-latency offloading and real-time collaboration of distributed devices with an advanced autonomy degree and intelligent autonomous systems in a safe, secure, and robust manner [88]. This includes real-time traffic information, decision-making, collision avoidance, and self-driving vehicle prediction, as well as DL in perception, privacy-preserving, control, and other tasks [89].
- **Smart Transportation:** FL is now being implemented to tackle various issues and challenges that surround the smart transportation domain, such as communication delays, calculation processing, data privacy, equipment mobility, smart logistics, resource and system transportation [85].
- **Smart Communication:** FL algorithms are potentially close to becoming the leading 6G enablers, due to their built accurate models based on large decentralized and heterogeneous datasets on resource-constrained devices [90]. The choice of FL is due to its ability to be coordinated by both centralized and distributed nodes.
- **Digital Forensics and Threat Detection:** are now more reliant on supervised ML approaches including FL for inferring system and network anomalies. The main aim is to ensure the interpretability of a model's decision-making process, such as the interpretable federated transformer log learning model presented in Ref. [91] for threat detection, especially for IoT systems and devices [92,93].
- **Ethical Hacking:** may well integrate the FL as a new potential solution to allow the maintenance of IoT users, clients, servers, devices, systems, and (big) data privacy and to avoid any privacy-related attacks and prevent any data exploitation or/and manipulation via AI-based and privacy-preserving solutions [94].
- **Law Enforcement:** started relying on the FL as a promising solution that improves the anti-financial-crime, anti-money laundering, and countering the financing of terrorism processes to achieve enhanced and accurate ML pattern identification and predictive power, without relying on any data sharing or compromising data privacy and security [95]. This can improve the models' accuracy (reduced false positive and false negative rates) with a reduced overall operational cost.
- **Military:** seems to be interested in FL, especially since the army is using centralized ML approaches to train models that rely on servers that host trained models to make predictions. Such approaches require remarkable direct investments in data annotation and tactical model training that are less expensive, complex, and time-consuming [96]. Hence, the reliance on enhanced FL to keep local storage on edge devices to ensure that the exchange of military data is done on tactical central servers in a coordinated, secure, and private real-time manner with the least possible bandwidth consumption and latency [97].

### 3. Challenges and future opportunities for FL-IoT

FL is a promising technique for IoT systems and can provide several benefits, as discussed previously (privacy preservation, scalability, and reduced data transfer). However, FL with IoT systems suffers from several challenges that are presented in Fig. 6. These challenges must be addressed to fully realize the potential of FL, such as:

- IoT Devices Constraints:** FL can be computationally expensive, and IoT devices with limited computation and resources may not be able to participate in the local training model. Therefore, developing lightweight-efficient FL techniques, especially for the training process, can reduce the required training computation, resources such as energy or memory consumption, and the size of communicated data while maintaining high model accuracy and precision. This can ensure that IoT devices with limited computation resources can still contribute to the training process.
- Heterogeneity of IoT Devices and Data:** Federated Learning involves training models on data from multiple IoT devices that may have different hardware configurations and data distributions. Addressing the heterogeneity of devices and data is a major research challenge in the FL approach.
- Model Updates:** IoT devices may have different amounts of data, and some devices may not be available for training at all times. Developing efficient updating model techniques for a distributed environment with less computation, communication, and resources is required.
- Fairness:** Data distributions might be imbalanced or biased, which introduces fairness issues in the trained models that will impose a problem in the global model. Addressing fairness concerns in FL is mandatory and developing a lightweight fairness solution is necessary to be able to apply it with limited IoT devices.
- Optimization Algorithms:** Trained local models at IoT/edge devices (decentralized) rely on optimization algorithms. However, existing optimization algorithms may not be suitable for IoT due to the IoT device constraints (computation and resource) in addition to the distributed nature of the data. Developing new optimization algorithms that can effectively train models on limited IoT devices and decentralized data that preserve high accuracy, precision, and convergence rates.
- Communication Overhead:** FL involves communication between IoT/edge devices and the central server, which can be computationally expensive and time-consuming. Developing efficient compression and quantization solutions is critical to reducing the size of communicated data and consequently reducing the communication delay overhead. This is essential to better responding to IoT devices' constraints in terms of energy, computation, and memory.
- Non-Independent and Identically Distributed (IID) Data:** FL assumes that the data on each source IoT/edge device is IID. However, in many real-world IoT scenarios, the data on each device can be non-IID because it is not sampled from the same distribution. This can occur for a variety of reasons, such as different device types, geographic locations, user demographics, or even temporal variations in data collection. This will pose a hard challenge to FL since traditional ML solutions assume that the training data is IID. However, when the training data is non-IID, the traditional algorithms may not work well, as it will lead to poor model performance (accuracy and convergence rates). Therefore, to fix the issue of non-IID data, we need to develop a lightweight solution that can be based on clustering, meta-learning, or a model personalization technique that takes into account the underlying data distribution and the heterogeneity of the data.
- Generalization:** FL is often used to train models based on a set of specific devices, which cannot be considered representative of the population as a whole. The same techniques that can solve the non-IID issue can also help solve the generalization issue. This solution should be efficient and robust to ensure that locally trained models can be applied to new devices and populations with high accuracy and precision.
- Federated Transfer Learning (FTL):** Transfer Learning (TL) is a popular machine learning technique that involves reusing efficient pre-trained models to construct new efficient ones. FTL is an emerging research direction that involves transferring knowledge across devices in a federated setting, which means it combines TL and FL techniques by using the pre-trained model from a source domain to initialize the models on the devices in the target domain. This can improve the convergence speed and generalization performance of the model. Developing an efficient FTL solution that can find a suitable pre-trained model that is relevant to the target domain and balancing the trade-off between preserving privacy and model

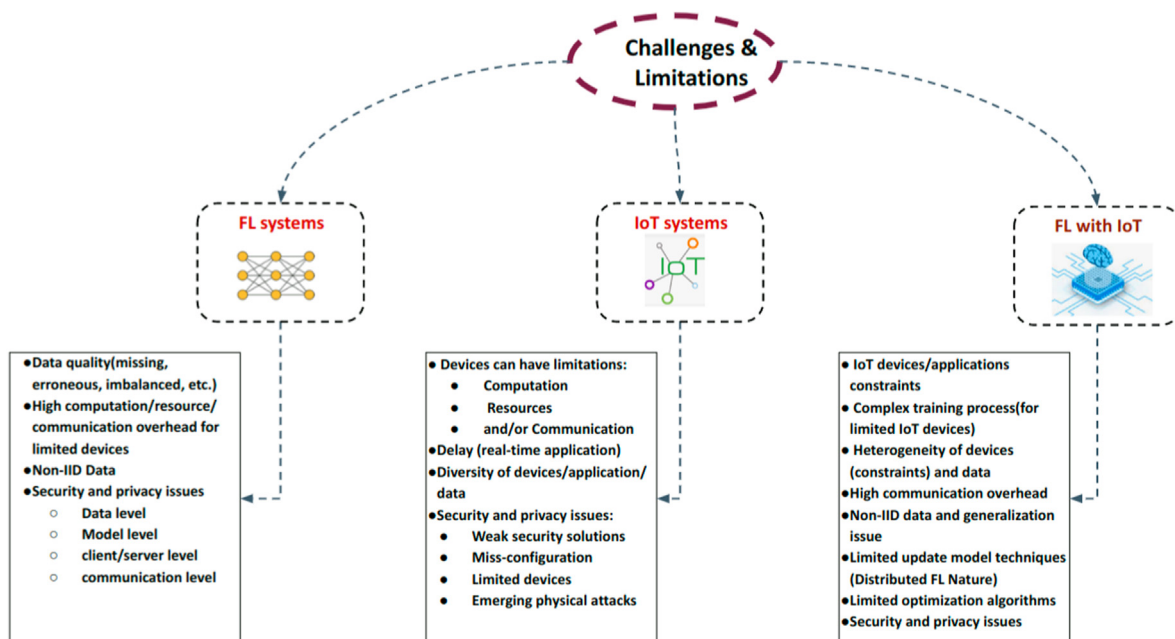


Fig. 6. Existing IoT and/or FL challenges and limitations.

accuracy is mandatory to improve the performance and scalability of Federated Learning. This point can be considered one of the most important contributions required in the field of FL.

- **Explainability:** The trained local models are based on decentralized data, and the model cannot explain the model's predictions. This challenge refers to the difficulty of understanding and interpreting the decision-making process of the trained local model without having access to the individual data on each device. This issue is critical in several applications, especially those involving sensitive or high-stakes decisions, such as healthcare or finance, where it is essential to understand how the model arrived at a decision. Therefore, we need to develop new, robust, and reliable extraction methods that can provide explainable models (explain the decisions made by trained models) without compromising privacy.
- **Standardization:** Hence, FL is still an emerging field, and there is a lack of standardization in terms of protocols, frameworks, and evaluation metrics. Thus, developing standardized protocols and frameworks for FL to facilitate its adoption in different IoT domains can be considered a principal point.
- **Security and Privacy-Preserving Techniques:** FL involves communication between devices and the central server, which creates security and privacy risks at different components (device, server, and network), as we present in the next section. We have to develop lightweight security and privacy-preserving solutions that require the minimum possible overhead in computation, resources, and communication to better respond to the constraints of IoT devices. In addition, these solutions should preserve the model's performance.

In fact, these challenges can be considered critical for future FL research directions. They should be addressed to fully realize the potential of FL in various IoT domains, such as healthcare and finance. However, current FL security and privacy issues still persist especially over IoT domains, such as data engineers being unable to access raw user data to clean it, identify missing values, and identify the data points that the model will be trained on. Another key issue includes the user's IoT device, especially during the training process, including heterogeneity, storage, computational ability, power consumption, connectivity issues, and communication bottleneck, which keeps the produced data local on each device [98]. Therefore, as part of future directions, the focus should be on detecting and preventing poisoning and data injection attacks [99]. This requires the need to train a model using device-generated data to support communication and reduce the number of communication rounds while sending small model updates to avoid connectivity issues. Another focus should be on the accuracy of both attack detection and prevention operations without risking both user/data privacy and security, especially as the FL is being heavily adopted into the IoT domain (i.e. mainly healthcare, smart cities, and smart grids) [100,101]. Besides, the focus of this work is related to the security and privacy of FL systems with IoT, and consequently, the next sections will describe these topics. Furthermore, let us indicate that in Section VII, we present more details about the challenges and future opportunities that maintain both the security and privacy preservation of FL within the IoT.

#### 4. FL security attacks

FL aims to address the problems of data governance and privacy by collaboratively training algorithms without exchanging any data. This can be done using secure aggregation to maintain private model updates. However, the gradient inversion attack (or input recovery from gradient) still remains a serious security and privacy-preserving threat [102]. FL also suffers from data labeling issues, as they require their training to be supervised, especially when encountering sensitive and heterogeneous data. The importance is to enable collaboration across multiple IoT systems, networks, and organizations using FL. The aim of this paper is to investigate both privacy and security threats, vulnerabilities, and attacks that target the whole FL execution process, including its data distribution

among clients, based on three main categories: Horizontal Federated Learning (HFL), Vertical Federated Learning (VFL), and FTL [103,104]. To ensure FL's security, it is important to secure communications to avoid Byzantine attacks that prevent FL models from converging [105], poisoning attacks that disrupt their training process [106], as well as a backdoor, Generative Adversarial Networks (GAN) and inference-based attacks that target FL's privacy [107]. This also includes attacks that target user/data confidentiality, integrity, availability, authentication, authorization, and accountability [108]. Despite the advantages that the FL brings via ML into the IoT, except that this made it vulnerable to a variety of different attack types that aim to exploit the vulnerabilities and weaknesses of the FL to achieve higher accurate and devastating attacks either overtly or covertly. These attacks are further explained and presented in Refs. [109,110]. As a result, this list of attacks is divided and presented as follows:

##### 4.1. Attack types

In this paper, the attack types are presented depending on the attacker's objective, aim, goal, and motives which can be part of cyber-crimes, cyber-warfare (i.e. sabotage and espionage), hacktivism, or cyber-terrorism/insurgency [111] mostly conducted by violent extremist organizations and groups [112]. In Ref. [113], Beseny et al. discussed terrorists' internet activity (cyber-espionage, online propaganda, or/and fake news) especially on the Dark Web was discussed, while their impact on healthcare was discussed by Tin et al. in Ref. [114]. In Ref. [115], Ghelani et al. discussed cyber-crime activities, especially against banking systems, and presented a banking system model to achieve higher intruder detection.

However, this paper is only concerned with the attack types, which are presented in Fig. 7 and described below as follows:

##### 4.2. Attack categories

In this subsection, the attacks are divided into four main categories, including FL server-based, client-based, client-server (communication) based, and data-based attacks. In each category, a list of specific attacks is also presented and explained to show how these attacks exploit and target them. This is summarized in Fig. 8.

- **FL Server-Side GAN-Based Attack:** aside from the traditional attack type, additional improvements were made to conduct additional tasks during the GAN training process, to enhance the generated samples' quality, without compromising the collaborative learning process nor modifying the shared model. Hence achieving an invisible attack [9]. To mitigate this problem, an early detection method that is lightweight, non-intrusive, and uses characteristics of gradient updates from participants to discover potential GAN-based privacy attackers [10].
- **Distributed Denial of Service (DDoS) Attack:** can usually go undetected targeting the FL servers and network traffic, as well as the resource allocation of the FL systems [116]. As a result, clients are unable to connect to servers and vice versa.
- **FL Communication-based Attacks:** are usually attacks being carried out during communication (passive or active attacks). A set of these main attacks are presented as follows:
  - **Eavesdropping Attack:** takes place when an attacker tries to eavesdrop to extract data via a badly or non-secure communication channel(s). Another attempt can be based on taking over a client's weak security to extract the needed data [117]. This attack is difficult to detect due to the attacker's passive monitoring or re-encryption of the hijacked communication(s).
  - **Man-In-The-Middle (MitM) Attack:** occurs when an attacker intercepts the exchanged model updates among the FL and the

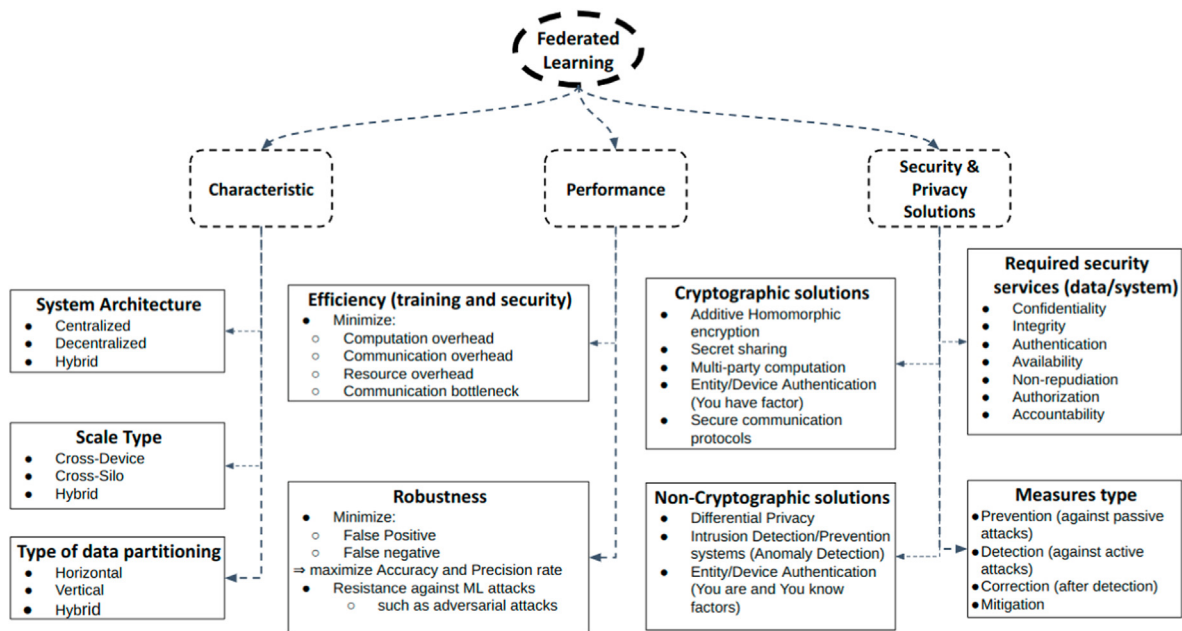


Fig. 7. Proposed federated learning taxonomies.

- **Random Attacks:** occur at random time intervals with the aim to reduce the effectiveness and the accuracy of the FL models.
- **Targeted Attacks:** can also occur at random times but are more specific regarding their objectives and goals as they induce the FL model to output the specified target.
- **Concentrated Attacks:** are attacks that target the FL-based (local/global model), client/server devices, or the data in transit, in order to break in, hijack or intercept incoming/outgoing information and data transmissions. The aim is to target security or/and privacy.
- **Sporadic Attacks:** are random attacks that target the FL-based domain. This often can be as a testing attack to check for any vulnerability or weakness that can be exploited, or as a diversion attack to conduct the initially intended attack on another key part of FL.
- **Separate Attacks:** can target the same FL but at different time frames, and these attacks can be related or not. In fact, they often occur when attacks are not connected together but rather might be part of a collective attack that attempts to target more than one part of FL until some breakthrough is achieved and the main objective is attained.
- **Combined/Joint Attacks:** or cascading attacks, aim to send waves and waves of attacks against a given FL, most often targeting its availability. This attack can range from minutes to weeks in some cases, if not more.

participants before replacing them with malicious updates [118]. Such attacks usually occur whenever there is interference with real networks or via the creation of fake networks controlled by the MitM.

- **Client-Side GAN-Based Attack:** exploits the training process's real-time nature, which occurs when an adversary trains a GAN to generate prototypical samples of the targeted training set to compromise the training set owner's privacy [11]. To mitigate this threat, a defense method called Federated Learning Parameter Compression (FLPC) was presented to ensure a higher and more effective privacy protection [12]. Another scheme was also presented to detect the GAN-based information leakage attack in FL using an anomalous detection algorithm [13].
- **Byzantine Attack:** targets the FL's building block, whenever a client either maliciously or accidentally makes defective updates on the server, which originate from a software bug, error, or exploit (i.e. backdoor), or submits to the server data that is not compatible and also updates it. Thus, causing convergence problems [14]. To mitigate this attack, an aggregation algorithm called Layerwise Gradient AggregatTiOn (LEGATO) was presented to mitigate the adverse effects of Byzantine input [15]. Another method is the DiverseFL, which mitigates the byzantine behaviours in FL [16].
- **Stealthy Boosting Attack:** occurs when in addition to boosting malicious updates, more terms can be added to the learning target by the malicious client, to target the accuracy and validation loss checking controls, as well as to obtain update magnitudes' statistics. This attack was demonstrated in Ref. [17]. **Single Attack** is another stealthy boosting attack type that occurs when a non-colluding malicious attacker causes the model to miss-classify a chosen input set with a

higher level of confidence [23]. In fact, it can also take the form of an **Explicit Boosting** attack, which occurs when a malicious agent mimics a benign agent to overcome the scaling effect to evade having the desired classification outcomes nullified after explicitly boosting the local malicious update [17]. To mitigate the Stealthy Boosting attack, several solutions were presented. For example, an online anomaly detection algorithm was presented to detect and identify stealthy attack vectors with detection thresholds to reduce false positives and increase true positive rates. Thus, achieving the balance between the minimum attack magnitude and detection thresholds [18]. A DL-based method to accurately and precisely detect stealthy False Data Injection (FDI) attacks against the electric power grid's State Estimation (SE) [19]. A novel anomaly-based Intrusion Detection System (IDS) that timely detects and mitigates emerging stealthy DDoS attack types was also presented, even when DDoS attacks are of a low size per source [20].

- **Foolsgold Attack:** occurs when malicious clients introduce several fake identities to send falsified updates to the central server to break both security and authenticity of the FL environment [119]. To mitigate this attack, a defense scheme called CONTRA was presented to overcome poisoning attacks including foolsgold, label-flipping, and backdoor attacks [21]. A poisoning attack mitigation technique was also presented, based on the removal of malicious nodes from the training environment and preventing denial of service attacks [22].
- **Sybil Attack:** occurs when an attacker simulates several dummy participant accounts or selects participants that are previously compromised to conduct advanced attacks against FL [23]. As a result, a Multi-Armed Bandit for Federated Learning (MABRFL) was presented as an adaptive client selection strategy to choose honest

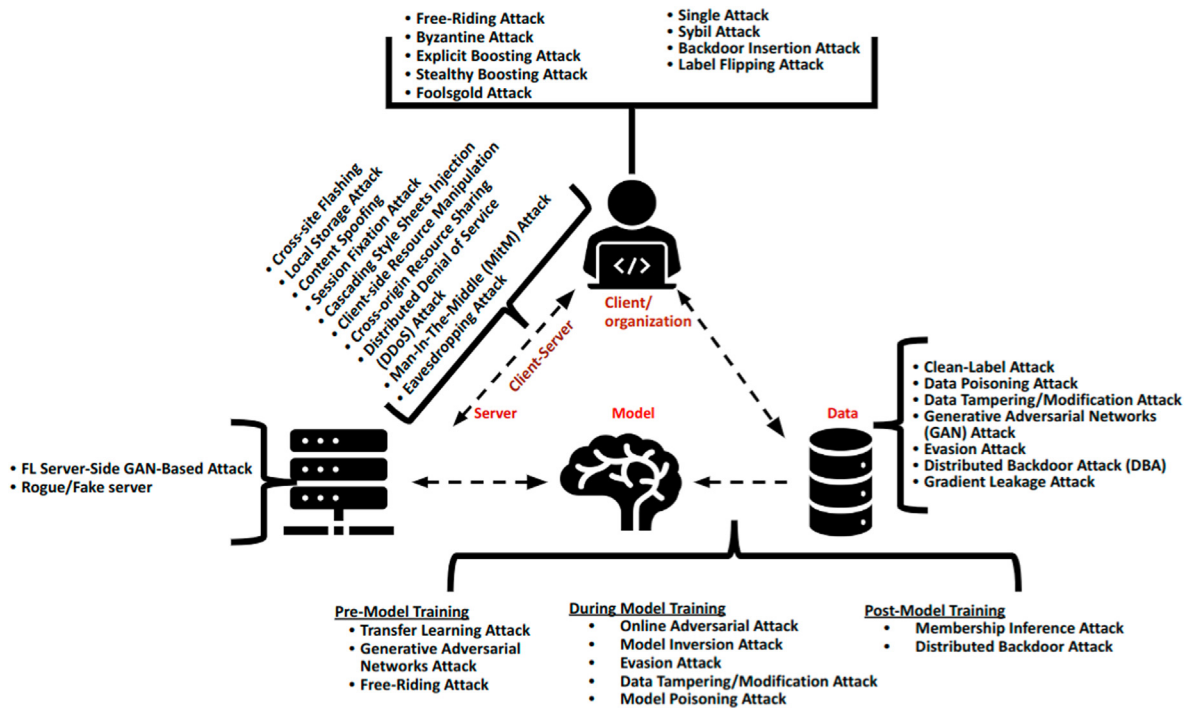


Fig. 8. Main attacks against FL's IoT System.

• **FL Server-based Attacks:** are usually attacks being carried out against or via the servers of the FL and these main attacks are presented below (see Table 2) as follows:

- clients that are more likely to contribute high-quality updates and to identify malicious updates from Sybil and non-Sybil attacks [24]. A defense approach that effectively mitigates the impact of Sybil attacks on model convergence was also presented based on anomaly detection [25].
- **Backdoor Insertion Attack:** targets the training data via manipulation by adding stamps or watermarks to enable models to learn triggers in inputs while preserving the accuracy of clean data. Thus, this attack type is harder to detect [120]. Solutions to mitigate backdoor insertion were also presented. A solution named FLAME was introduced as a defense against backdoor attacks based on detecting and filtering out malicious model updates and on estimating the sufficient amount of noise to be injected to ensure the elimination of backdoors [26]. A universal model-agnostic defense technique (Moat) was also presented to mitigate different poisoning attacks in Federated Learning including backdoor insertions [27]. Lastly, a backdoor detection method called Feedback-based Federated Learning (BAFFLE) was also presented as a novel defense mechanism to secure FL against backdoor attacks with a very high detection rate [28].
- **Label Flipping (LF) Attack:** occurs when attackers who have access to training data poison it and permute labels while keeping the features intact to avoid any detection and fool tangent models [121]. A novel defense approach to mitigate LF attacks was presented, which analyses the resulting clusters before performing model aggregation with higher overall accuracy and lower attack success rates [29]. Another improved defense strategy that emphasizes employing Kernel Principal Component Analysis (KPCA) and K-mean clustering was presented to defend against data-poisoning and Label Flipping (LF) attacks [30].
- **Clean-Label Attack:** occurs when an adversary assumes that the intercepted data is certified and belongs to the same class to avoid changing the input data, and craft a poisoned training data instead that looks similar to the original non-corrupted data [31]. This was achieved using the dog-vs-fish classification task which achieved a

- success rate of 100% in Ref. [122]. Ho et al. presented an ensemble-based Nested Training technique to remove most of the poisoned samples from a poisoned training dataset, while recovering the model's accuracy of malware detection up to 93.2% [32].
- **Data Poisoning Attack:** a data poisoning backdoor attack on a classifier includes adding particular elements to the training data of a system to make it behave in a way that is advantageous to the attacker. It is also known as a subset of model poisoning [109,123], which targets the vulnerability of the ML algorithm while trying to incorporate malicious data points in the training phase to achieve the highest classification error. A similar data poisoning method was presented in Ref. [124]. For example, a face recognition authentication system can be tricked to classify someone wearing a certain pair of glasses as the user “Bob” while acting properly in other situations. Reasonable supply-chain checks should be applied to minimize these threats to the training data. Additionally, the training data cannot be altered by third parties. In fact, poisoning attacks can be targeted: where the global model misclassifies the chosen samples set to an attacker-chosen target while reducing the model's performance impact on the main task, or untargeted: which degrades the overall model performance [125].
- **Model Poisoning Attack:** usually targets a large-scale FL product with multiple clients to modify the updated model ahead of sending it to the central server to aggregate it, leading the global model to be poisoned easily [33]. In fact, it induces adversarial learning targets in the global model, especially since there's no data to save the exchange between the FL server and clients. This allows the adversary to alter it to achieve their malicious objective and intents [105]. This method is more effective than data poisoning since a single manipulation can compromise the global model [17]. A novel optimization-based model stealthy poisoning attack with a high success rate was also presented in Ref. [126]. This can also result in a **model Stealing** attack which occurs when the learned model is accurately copied by the attacker. This copy can be used for a variety of purposes, including analyzing the target system's decision boundaries, recreating the

- behaviour of the target system, or simply reducing costs if the target system is a commercial service. To mitigate this attack, several solutions were presented including: SparseFed, which is a mitigating model with sparsification against poisoning attacks in federated learning [34]. Another solution includes the MLGuard, which employs a lightweight secret-sharing scheme and a novel poisoning attack mitigation technique [35].
- **Data Tampering/Modification Attack:** aims to change or/and alter the training dataset to create a feature collision effect that merges two dataset classes to fool the ML model and consequently misclassify the targeted class [127]. A Dynamic Redundant Path Selection (DRPS) was presented to detect and mitigate data tampering attacks on Cooperative Adaptive Cruise Control (CACC) systems while reducing the number of network paths during tampering detection [36]. Another similar Dynamic Redundant Path Selection (DRPS) solution was presented to detect and mitigate data tampering attacks in networked control systems [37].
  - **Free-Riding Attack:** includes the intentional dissimulating participation in the FL process to collect the final model, mostly by inserting dummy updates without training the model with local data [128]. A novel defense mechanism named “Viceroy” was presented as a new federated optimization algorithm to mitigate free-riding attacks [38], while a P2P Straightforward Protocol (P2PSP) was presented using Shamir’s Secret Sharing (SSS) and the use of Trusted Peers (TPs) to overcome them [39].
  - **Generative Adversarial Networks (GAN) Attack:** is similar to inference attacks, except that it is used to intercept training samples via inference, before using them to poison the training data. This method has been demonstrated in Ref. [11]. To overcome GAN issues, a novel CycleGAN-based architecture based on DDoS detection to generate legal malicious traffic was presented [40]. Additionally, a dynamic traffic camouflaging technique, called FlowGAN was also presented to automatically learn and dynamically morph the on-going traffic flows [41].
  - **Evasion Attack:** often occurs at testing time to manipulate the data input, which would result in an error in the machine learning system. Unlike poisoning attacks, it does not alter the system’s behaviour but rather exploits weak spots and vulnerabilities [42]. It can be divided into three types: white-box (i.e full knowledge of the FL system), grey-box (i.e partial knowledge of the FL system), or black-box (i.e no knowledge about the FL system). To mitigate this attack, an ensemble learning approach based on static and dynamic features extracted from Android malware applications was presented to detect and mitigate adversarial evasion attacks [43]. Another method included a region-based classification that mitigates evasion attacks without affecting the classification’s accuracy [44].
  - **Distributed Backdoor Attack (DBA):** is a novel attack developed by Ref. [129] to exploit the distributed nature of FL by decomposing a global trigger pattern into separate local ones to achieve more stealth and persistence. Thus, adding non-homogeneous data distribution increases the false positives, adding more bias in the FL. As a result, DeepSight was presented as a novel model filtering approach for mitigating backdoor attacks and identifying suspicious model updates [45], while a Federated Learning Provable defense framework (FLIP) was introduced to train the local model on generated backdoor triggers that can cause misclassification and data poisoning attacks [46].
  - **Model Inversion Attack:** by providing inputs and noting the matching output, an attacker can use a model inversion attack to get sensitive data from the training set. Be aware that Model Inversion only creates a composite representing the “essence” of a class, which is likely to be valuable to an attacker, rather than a real instance of the training set. It predicts if the data sample is a member of the target model or not, where the attacker aims to infer information about the data sample from the prediction results. This includes the reconstruction of sensitive attributes or samples. In fact in Ref. [47], Yang et al. presented a black-box model inversion attack that leverages the attacker to build an inversion model to reconstruct the original input sample with high accuracy.
  - **Membership Inference Attack:** occurs when an attacker queries a trained machine learning model to check if a specific example was contained in the model’s training dataset or not [130]. Therefore, the main goal of this attack is to determine whether a particular input was part of the model’s training set by just applying the trained model to specific inputs and evaluating the outcome. For example, we might wish to look through the training set to see whether a certain hospital discharge record is included, to see if a certain person has had treatment for a certain medical condition, or to see if a certain financial transaction is included in the training set for fraud detection. The most common solution is the LDP, except that it does not prevent Label Inference Attacks (LIAs) [131]. A purification framework to defend against inference attacks by purifying the confidence score vectors predicted by the target classifier was also presented in Ref. [48].
  - **Online Adversarial Attack:** occurs the model is learning online from a new continuous data stream, which allows the attacker to manipulate the model’s learning via sending false data [132]. This can be sorted by adopting either a robust data ingestion process or using data versioning [133] in the Machine Learning Operations (MLOps) [49]. To solve the false data injection attack (FDIA) in smart grids, Chen et al. combined secure federated learning (using Paillier cryptosystem) with Transformer to train a detection model while preserving the privacy of all the local training data [50]. Experimental results show that this method protects data privacy and reduces communication overhead.
  - **Transfer Learning Attack:** occurs when a backdoor is created by exploiting the pre-trained process that is based on a larger data set, which the majority of models rely on [134]. To mitigate this threat, transfer models should often be retrained against custom datasets and object functions should be constantly updated, before being further tuned. Another method included the introduction of a novel Deep Transfer Learning (DTL) method that allows learning from data collected from multiple IoT devices while improving the accuracy in detecting IoT-related attacks [51].
  - **Adversarial Machine Learning Attack:** occurs when small variations in the model data inputs are found and exploited to redirect and result in having undesired model outputs [135,136]. To mitigate this attack, a robust Data Operations (DataOps) and MLOps solution can be adopted as a set of practices that aims to reliably and efficiently deploy and maintain machine learning models [53] and to improve quality, speed, and collaboration and promote a culture of continuous improvement [52].
  - \* **Shadow Training:** the attacker builds an attack model that can infer membership by training local models on data that is similar to the target model’s data distribution.
  - \* **Data Transfer:** is similar to the Shadow Training technique, where the attacker trains a local model, which does not need to have the same distribution as the target model, but rather exploits the relation between outputs to infer membership.
  - \* **Threshold:** the attacker uses statistical measurements to infer membership directly, such as maximum confidence and entropy.
  - \* **Likelihood Ratio Attack:** this approach uses different thresholds to account for different input losses to increase accuracy.
  - **Gradient Leakage and Gradient Manipulation Attacks:** Gradient Leakage occurs when an attacker uses the gradient information uploaded by clients with the aim to steal and recover private, sensitive or confidential training data from the shared gradients [137]. Gradient Manipulation occurs when the local model gradients are manipulated to compromise the global model performance, with the aim of targeting the overall accuracy [120]. A novel practical Stochastic Gradient Descent (SGD) based poisoning attack approach

named BadBatch was presented in Ref. [56], to find bad batches that overfit the global model by manipulating local training batches that are provided by compromised clients in each local FL training.

- **SQL Attack:** which targets ML applications, especially when SQL query fragments are included in the attacker's data to retrieve or modify the submitted strings, which make up a part of the submitted query to the server either to gain an administrative privilege, extract details, modify databases or compromise users accounts. Several recommendations were presented in Ref. [138] such as using a parameterized Application Programming Interface (API), and a lookup mechanism to identify unparameterized objects, as well as avoiding the use of dynamic SQL with strings within the database to execute or create queries. This can be done using a fully automated technique that detects, prevents, and reports SQL-Injection Attack (SQLIA), while also capturing all malicious SQL queries [139], or using Knuth-Morris-Pratt (KMP) string matching algorithm to match user's input string with the stored pattern of the injection string to detect any malicious code [57].
- **Trojanned Model Attack:** allows the attacker to upload a malicious model to an already pre-trained model pool to execute an arbitrary code whenever it is loaded or retrained for inference. The use of watermarked data could be a good solution which can protect the ownership and originality of audio data [58] and improve the authenticity, integrity, and safety of data [140].
- **Infected Model Attack:** occurs when a malicious code or/and a modified model are added to the existing model to cause abnormal and suspicious behaviours. Recommendations such as preventing third parties from modifying training data is highly advised, while a supply chain is checked on the in-use training data.
- **Adversarial Perturbation Attack:** occurs when an attacker tries to create an input that generates a desired output. This attack type can take three main forms:
  - 1) **Chosen Class:** when the target is caused to classify a given input as the chosen class.
  - 2) **Excluded Class:** when the classifier is caused to return anything than the specified class.
  - 3) **Perturbed Value:** when the predictor is caused to return a biased value in their choice of direction. This attack can also take a "Physical Realm" by physically modifying objects that are used as ML system inputs. This can be mitigated by adopting robust training models, advanced authentication measures, and notification alerts in case of suspicious activities [59].
- **Decision Boundary Detection Attack:** occurs when the attackers identified the limitations or exploitable vulnerabilities of intrusion detection systems for future bypassing or attacks.
- **Textual Training Data Extraction Attack:** occurs when the model is prompted to generate text containing verbatim portions of the data it was trained on. Mitigation methods include application refactoring to remove sensitive data items, implementing rate limits and authentication, and adding differential privacy or/and a resistant training regime to defend against ML-based privacy attacks [60].
- **Inference by Covariance Attack:** allows the attacker to infer the individual user behaviour on a given system, and the system's response to the user via ML output observation which can be done either passively or actively. This can be mitigated by identifying firm and clear restrictions and recommendations, as well as applying rate limits.
- **Model Stealing Attack:** occurs when an accurate copy of the trained model is obtained to reproduce the targeted system's behaviour to perform "white-box" data extraction. Solutions to mitigate this threat include the High-performance Deep Neural Networks (DNNs) solution called PRADA that perturbs predictions targeted at poisoning the training objective of the attacker [54], and offers protection against DNN Model Stealing Attacks that ensures an effective detection of DNN model extraction attacks [55].
- **Approximate Copy and High-Fidelity Copy by Inference Attack:** Approximate Copy attack occurs when an approximate copy of the targeted model is obtained via input crafting, output observation, and local model training to replicate such behaviours. High-Fidelity Copy is similar to the "Approximate Copy by Inference" attack, with the main difference, being that a high-fidelity copy of the target model is obtained to create a near-perfect copy to ensure a higher and more accurate level of replication.
- **Local Storage Attack:** also known as web storage or offline storage attack, which enables JavaScript sites and applications to store and access data without any expiration date. As a result, the cross-site scripting attack is able to extract all the stored data and inject it with malicious data loads via JavaScript. This can be mitigated using lightweight integrity protection which offers web storage-driven content caching using a generated and pre-determined checksum value [62].
- **Cross-Site Flashing Attack:** occurs via the exploits of the vulnerabilities found in the browsers where the flash applications are often embedded at. One solution to mitigate and discover these vulnerabilities is the FlashOver [63], which is the first fully automated XSS attack discovery system, capable of statically and dynamically analyzing and identifying real-life Flash vulnerabilities.
- **Cross-Site Scripting (XSS) Attack:** occurs when a malicious code/script is injected by the client-side into websites that are displayed by a vulnerable browser. This can be mitigated using the pattern filtering approach [141] or using server-side approaches such as XSS-GUARD, which is a prevention mechanism against XSS attacks on the server side [64] and XSS-SAFE, which is an automated framework that detects and mitigates XSS attacks on the server-side [65].
- **Cascading Style Sheets (CSS) Injection Attack:** injects an arbitrary CSS code into a website, before being rendered in the end user's browser. Depending on the CSS payload, it can be an XSS, User Interface (UI) or data modification/extraction attack. This can be mitigated using injected style sheets that identify a user's installed extensions [66], or SIACHEN, which is based on a white-list and browser-enforced security policy language that mitigates XSS attacks [67].
- **Client-Side Resource Manipulation Attack:** results in an XSS attack by exploiting and controlling the URL that links to the other web page resources. This can result in a **Content Spoofing** attack which is a web security vulnerability that allows the end user to spoof or modify (via injection) the content found on the web page. This can also be exploited using the compounded SQL injection or content spoofing as two different website injection vulnerabilities [142]. This can be mitigated using MLPXSS [68], or a novel configuration using Pulse Secure, Pulse Connect Secure, and Virtual Web Application Firewall [69].
- **Cross-origin Resource Sharing (CORS) Attack:** can cause cross-origin attacks like cross-site request forgery (CSRF). This can be mitigated using an XSS detection mechanism integrated with HTML5 and CORS properties (which detects XSS attacks using browser extensions) [70], or by using an Application Cache (which identifies Cross-origin Resource Status) [71].
- **Session Fixation Attack:** or a Session Fixation Vulnerability (SFV), is a remote code execution attack that exploits software types that are designed with features of the web-server session, mainly to intercept and retrieve the HTTP server's session state information. This can be mitigated using Serene, which offers self-reliant client-side protection against session fixation [72], an authentication-based scheme, which enhances the authenticity of the client on the server [73], or an automated session fixation vulnerability detection, which offers an optimum automatic detection of session management vulnerabilities [74].

### 5. Source of FL vulnerability

Despite the great advantages that the FL offers, except that its introduction into the IoT field left it prone to a variety of exploitable vulnerabilities [108], which are presented in terms of security and privacy in Fig. 9, and also in terms of efficiency, effectiveness and robustness. In this paper, the main vulnerabilities are presented and further summarized in Table 7 as follows:

### 6. Existing FL-security solutions

Due to the increasing and persistent number of attacks against the FL-IoT, which differ in their nature and type, different security measures are highly required to mitigate this threat and ensure early accurate detection and mitigation of these attacks using cryptographic and non-cryptographic measures (see Fig. 7). This will help ensure that the attacks are accurately detected in their early phase and effectively dealt with to avoid these attacks from further expanding.

The introduction of FL into the Internet of Things (IoT)-related

domains (i.e medical [152], communication [153,154], robotics [4,5], industrial [155], etc) has surely introduced new advantages and sophisticated methods to improve the IoT domain. In fact in Ref. [156], Ma et al. investigated FL-based privacy and security issues along with some protective security measures and stated that privacy protection should be made at the client or server side, while security must be maintained at the system level. Unfortunately, FL suffers from different security and privacy issues that require the adoption of several security and privacy preservation techniques to make it more resistant to these attacks. In this work, we divide these techniques into two main classes: cryptographic techniques and non-cryptographic techniques. Examples of cryptographic techniques include homomorphic encryption and/or authentication to secure the aggregation at the aggregator server(s), or by employing non-homomorphic cryptographic algorithms to secure the communication between IoT/edge devices and server(s), in addition to the device/server authentication using the “you have” factor (cryptographic key). On the other hand, examples of non-cryptographic solutions include differential privacy, and anomaly detection/prevention solutions, as well as the “you are/you know” device authentication

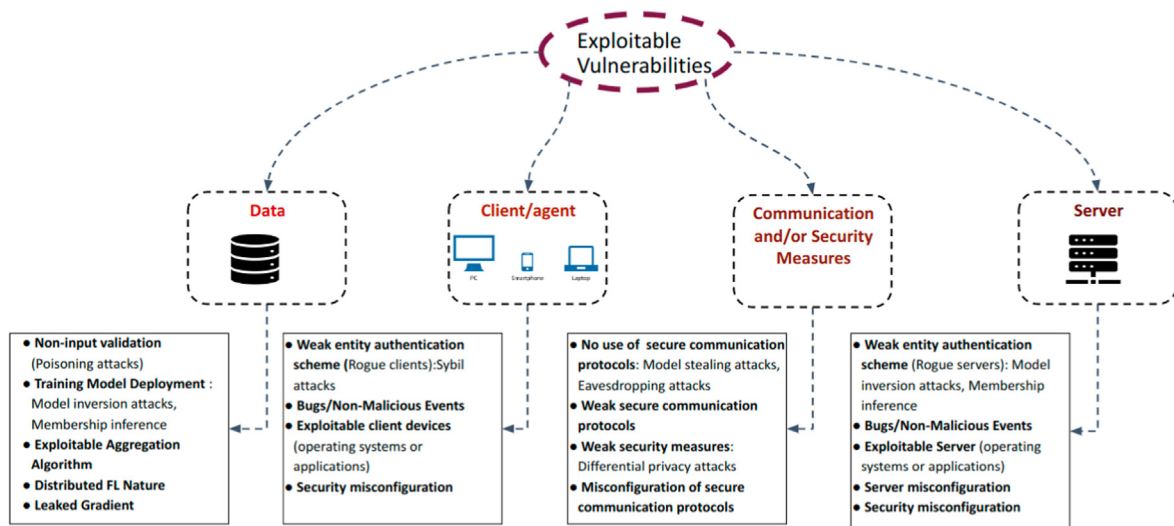


Fig. 9. Possible exploitable FL vulnerabilities and their impact on security, privacy, and performance goals.

- **Communication:** is a key factor for the FL process to achieve convergence after thousands of rounds of communication, except that if it is done on a non-secure channel, it is prone to various open vulnerabilities, mostly related to eavesdroppers that aim to intercept the exchanged models between different clients and the final FL model to replace them with malicious models instead. This issue can be sorted using homomorphic encryption to create the clients' data by exchanging model updates between FL clients and the server. Moreover, it also allows the server to perform computation on encrypted model updates without decryption [143].
- **Leaked Gradient:** despite data not being shared in the training process. An attacker can reveal important information that helps with the reconstruction of raw data either partially or almost fully [144]. As a result, any minor leaked gradient update can reveal sensitive information about the local data, making Byzantine attacks more feasible and harder to detect especially after masquerading as benign participants. As a result, private information may be leaked during gradient transmission [77].
- **Rogue Clients:** since in FL, clients are classified as a critical component, this makes them capable of viewing the states of the intermediate global model and also contributing with updates. This allows clients to either maliciously or accidentally tamper with the training process and corrupt it [77].
- **Exploitable Server(s):** due to the FL's architecture being mostly based on a cluster of either physical-based or cloud-based servers [145], which leaves them exposed to the same type of cloud computing attacks including the Distributed Denial of service (DDoS) attack [146].
- **Exploitable Aggregation Algorithms:** since the inspection for anomalies cannot be done for clients' data and the training pipelines due to privacy constraints, the need for aggregators as critical lines of defense is a must [147]. However, if the anomaly detection mechanisms are not incorporated within the aggregation algorithm, then the correction and maintenance of a robust aggregation algorithm will result in a failure that will expose the whole algorithm, especially for heterogeneous networks [99].
- **Bugs and Non-Malicious Events:** as already known, any low bandwidth or limited computational power can result in the discarding of clients, loss of data, damaging the accuracy of the training process, and leading to a lower-quality and different FL model than the original or intended one. As a result, clients or adversaries can submit false data, drop out of the session, cause data pipeline bugs with a lower ability to detect them, affect the compressed communication [148] or distort servers (i.e noise) [149].
- **Distributed FL Nature:** the nature of distributed FL training, in particular, may result in the creation of distributed backdoors that can be exploited by one or more adversaries at the same time, resulting in a well-coordinated and sophisticated attack. Coordination between past, present, and future participants is also achieved to conduct attacks against current and future global model updates, such as the novel colluding attack, which is called the “Distributed Backdoor Attack” (DBA) [129]. As a result, the robustness of the collaborative training cannot be guaranteed [150].
- **Training Model Deployment:** The deployment of the training model following the assessment of its quality and approval is prone to numerous risks, such as the corruption of that process to cause inference-time vulnerabilities or white-box evasion attacks. This will surely result in low accuracy and robustness, as well as a high form of adversarial noise mostly caused by foreseen [59] and unforeseen adversarial attacks [151].

**Table 7**  
The Main FL-based Vulnerabilities and their impact on IoT's main goals.

Exploitable Vulnerability	FL Components			Targeted Goals					
	Data	Server	Client	Confidentiality	Integrity	Privacy	Availability	Accuracy	Security
Leaked Gradient	✓	–	✓	✓	X	✓	X	✓	X
Rogue Clients	–	–	✓	✓	✓	✓	✓	✓	✓
Exploitable Server	–	✓	–	X	✓	✓	✓	X	✓
Distributed FL Nature	✓	–	–	X	X	✓	✓	✓	✓
Bugs/Accidents	✓	✓	–	X	X	X	✓	✓	✓
Exploitable Aggregation Algorithm	✓	✓	–	X	X	✓	✓	X	✓
Training Model Deployment	✓	–	–	X	X	✓	✓	✓	X

factor.

However, as a set of IoT devices is limited in different terms (i.e. computation, resource, and communication), lightweight security solutions will be ideally required especially for resource-constrained devices. As a result, four main solution types were introduced to tackle this issue and are presented below as follows.

### 6.1. Anomaly detection solutions

The machine learning concept has been heavily introduced into the IoT-related domain and revolutionized to be more active and widely adopted in all IoT-related domains such as medicine and industry. Anomaly Detection (AD) is a proactive solution that detects malicious updates, data, and model poisoning attacks with higher accuracy to prevent them from impacting the FL system [157] such as detecting the malicious data. Jebreel et Ferrer analysed attacks against FL and realized that neurons in the DL's last layer model are related to label-flipping and backdoor attacks. As a result, FL-Defender was presented as an attack detection solution to combat FL-targeted attacks [125]. Experimental results show its effectiveness against label-flipping and backdoor attacks while achieving the lowest attack success rates and maintaining the main task accuracy. However, a major privacy concern has emerged surrounding user data issues. This is mainly due to anomalous data often deviating from a large number of normal data points, which negatively impacts the whole IoT system. As a result, several AD-based solutions were presented and summarized in Table 8. Zhao et al. presented the Multi-Task Deep Neural Network in Federated Learning (MT-DNN-FL) to simultaneously conduct network anomaly detection, Virtual Private Networks (VPN) such as Tor for traffic recognition, and traffic classification task [158]. Experimental results show that the presented approach achieves good detection or classification performance across different tasks. Future work will focus on optimizing the DNN structure without affecting the training performance. Mothukuri et al. presented an

**Table 8**  
A set of recent Anomaly Detection Solutions.

Year	Author(s)	Reference	Type	Description
2019	Zhao et al.	[158]	MT-DNN-FL	Conducts network anomaly detection, VPN (Tor) traffic recognition and classification tasks
2021	Mothukuri et al.	[159]	FL-based AD approach	Recognizes the IoT network intrusions in a proactive manner
2021	Gupta et al.	[160]	FEDTIMEDIS	Mitigates FL threats models
2021	Chen et al.	[161]	CNN-LSTM	Accurately detects anomalies, reduces communication costs, and improves their efficiency
2022	Wu et al.	[162]	FL-MGVN	Overcomes low detection accuracy and high false alarm rates
2022	Liu et al.	[163]	AMCNN-LSTM	Senses time series data in Industrial IoT in an accurate and timely manner

FL-based anomaly detection approach that recognizes the IoT network intrusions in a proactive manner, using decentralized on-device data [159]. Experimental results demonstrated that it overcomes traditional non-FL machine learning versions in terms of the accuracy rate to detect (i.e identify and classify) attacks, and in terms of securing user data privacy. Gupta et al. introduced a hierarchical FL threat model with specific threat scenarios for centralized AD, and a privacy-preserving model to mitigate these threats [160]. As a result, the FEDerated TIME DIStributed (FEDTIMEDIS) Long Short-Term Memory (LSTM) approach was used to train the AD model. Experimental results show that this approach enables high accuracy and easier collaboration.

Chen et al. presented a novel, communication-efficient FL-based deep anomaly detection framework, called the Convolutional Neural Network (CNN)-Long Short Term Memory (CNN-LSTM) model to accurately detect anomalies [161]. Moreover, a gradient compression mechanism was also used to reduce communication costs and improve communication efficiency by 50% without losing accuracy. Numerical results show that CNN-LSTM achieves the highest accuracy on all datasets, while significantly improving communication efficiency and reducing communication costs. Wu et al. presented an anomaly detection classification model that incorporates both FL and Mixed Gaussian Variational self-encoding Networks (MGVN) to form the FL-MGVN [162] to overcome low detection accuracy and high false alarm rates. Experimental results demonstrated that FL-MGVN has higher recognition performance and classification accuracy than other methods. Liu et al. presented the Attention Mechanism-based Convolutional Neural Network (AMCNN-LSTM) model, as a communication-efficient on-device FL-based deep anomaly detection framework that senses time series data in Industrial IoT in an accurate and timely manner [163]. Experimental results show that the AMCNN-LSTM model achieves high accuracy and that the introduced gradient compression mechanism can compress the gradient by 300 times without losing accuracy while reducing the communication overhead by 50%.

### 6.2. Counter-Distributed Denial of Service (DDoS) solutions

DDoS attacks result in the system's degradation or total failure. Additionally, there is ample opportunity for an attacker to harm a client, server, or company in the FL system by simply using too many resources.

As DDoS attacks are increasing against the FL-linked IoT domains, effective counter-DDoS solutions are highly needed. For this reason, several solutions were presented by different authors to detect and deter this type of threats and attacks, and are presented and summarized in Table 9 as follows: Ahakonye et al. presented a federated-learning orchestration that secures the Maritime Supervisory Control and Data Acquisition (SCADA) from DDoS strikes, while also offering both attack detection and mitigation [164]. Li et al. introduced FLEAM as a Federated Learning Empowered Architecture to Mitigate DDoS in industrial IoT. FLEAM is based on an FL-based Iterative Model Averaging (IMA)--Gated Recurrent Units (GRU) called IMA-GRU detection protocol to train a global optimized model, using distributed datasets to remove both data and communication constraints [165]. Experimental results show that the mitigation response time was 72%, lower with an accuracy of 98%.

**Table 9**  
A set of recent Counter-DDoS Solutions.

Year	Author(s)	Reference	Type	Description
2021	Li et al.	[165]	FLEAM	Trains a global optimized model, uses distributed datasets to remove data/communication constraints
2021	Zhang et al.	[166]	FLDDoS	Improves accuracy and reduces the communication rounds number
2022	Doriguzzi et al.	[167]	FLAD	Trains feed-forward neural networks for DDoS attack detection
2022	Zainudin et al.	[168]	FL-based DDoS classification	Preserves the privacy of sensitive industrial network data traffic
2022	Lv et al.	[169]	FLDDoS	Counters DDoS attacks and mines its traffic data features

Zhang et al. presented FLDDoS, as DDoS Attack Detection Model based on Federated Learning used on local models that learn their client's data without them sharing it [166]. Moreover, a hierarchical aggregation algorithm and a data re-sampling method were also presented to improve the accuracy of their solution by 4% and reduce the communication rounds number by 40%.

Doriguzzi et al. presented FLAD, as an adaptive Federated Learning Approach to DDoS attack Detection that trains feed-forward neural networks for DDoS attack detection and assigns more computation to members with profiles that are hard to learn, without sharing any test data to monitor the trained model's performance [167]. Results demonstrated that FLAD outperforms the original FL algorithm in terms of accuracy and convergence time by up to 80%. Zainudin et al. presented an FL-based DDoS classification that preserves the privacy of sensitive industrial network data traffic [168]. Experimental results show an accuracy of 98.84% and a loss of 0.061. Lv et al. combined the FL and neural network to present a FLDDoS system to counter DDoS attacks, while also designing a CNN model and a data pre-processing algorithm to mine DDoS traffic data features [169]. Experimental results show that the DDoS detection accuracy is 99% and the multi-class classification is 90%.

Furthermore, we need to implement rate-limiting, in addition to authentication and auditing to identify and stop DDoS attempts as they happen in order to minimize their impact.

### 6.3. Differential privacy solutions

DP is used to defend the FL against privacy attacks while offering a strong defense mechanism against data poisoning attacks. This is often achieved by adding statistical noise in each update to confuse the attackers and distort the eavesdroppers [170]. Despite the FL's emergence as a machine learning technique that safeguards the client's personal information, except that it is still vulnerable to various privacy and security attacks. Therefore, to preserve this privacy, the authors presented several differential privacy-preserving solutions, which are also summarized in Table 10. Yang et al. presented a novel model poisoning attack in differential privacy-based federated learning called Model Shuffle Attack (MSA) to shuffle and scale the model parameters. Experimental results show that it can significantly degrade the global model performance while guaranteeing stealthiness [171]. Choudhury et al. introduced a federated learning framework with two levels of privacy protection, to learn a global model from distributed health data that is locally held at different sites [172]. The experimental results demonstrated the feasibility and effectiveness of the FL framework while offering an elevated level of privacy to maintain the global model's utility. Truex et al. presented LDP-Fed as a novel form of federated learning with a local Differential Privacy system to respond to challenges related to model accuracy, privacy preservation, and system capabilities [173].

**Table 10**  
A set of recent Differential Privacy Solutions.

Year	Author(s)	Reference	Type	Description
2019	Choudhury et al.	[172]	FL framework	Learns a global model from distributed health data
2020	Truex et al.	[173]	LDP-Fed	Responds to challenges related to model accuracy, privacy preservation and system capabilities
2020	Sun et al.	[174]	LDP mechanism	Improves privacy/utility trade-off
	Wei et al.	2020 [175]	DP framework	Effectively prevents information leakage
2021	Hossain et al.	[176]	DeSMP	Shows how a stealthy and persistent model poisoning attack exploits the differential noise
2021	Lian et al.	[177]	COFEL	Reduces the communication time and enhances the privacy protection
2021	Girgis et al.	[178]	CLDP-SGD	Overcomes convergence, communication and privacy problems and achieves a trade-off
2022	Zhang et al.	[179]	PEMFL	Protects the privacy information of the industrial agents
2022	Wang et al.	[180]	Three-plane framework	Secures the Cross-Silo FL using the LDP mechanism
2022	Firdaus et al.	[61]	Secure FL framework	Tackles various FL-related attack types via noise addition

Following an analytical comparison, the LDP-Fed is said to have achieved higher accuracy and system features. Sun et al. presented a novel design of the Local Differential Privacy (LDP) mechanism for federated learning, while also presenting an adaptive range setting for improving the privacy/utility trade-off [174]. A parameter shuffling mechanism was also introduced to mitigate the degradation of privacy due to the high data dimension. Empirical studies show that this LDP-FL achieves higher performance than previous related works. Wei et al. presented a novel framework based on Differential Privacy (DP), where artificial noise (Noising before model Aggregation FL (NbAFL)) is added at the client's side before aggregating it, to effectively prevent information leakage [175]. Extensive simulation results confirm the achievement of a trade-off between convergence performance and privacy protection levels.

Hossain et al. developed an unprecedented DP-exploited Stealthy Model Poisoning (DeSMP) attack for FL models, to show that a stealthy and persistent model poisoning attack can be conducted to exploit the differential noise [176]. The empirical analysis has shown the effectiveness of the presented model. Moreover, a Reinforcement Learning (RL)-based novel defense strategy was also introduced against such attacks. Lian et al. presented COFEL as a novel Communication-efficient and optimized Federated Learning system to reduce communication time and enhance privacy protection using a local differential privacy mechanism [177]. The results show that the accuracy is 22.8% higher than that of CMFL, and the training time is reduced by 20%–48% to achieve an 85% accuracy when compared to traditional FL and CMFL. Girgis et al. presented a distributed Communication-efficient and Local Differentially Private Stochastic Gradient Descent (CLDP-SGD) algorithm to overcome convergence, communication, and privacy problems and ensure that a trade-off is achieved [178]. Preliminary experimental results showed that this algorithm matches the lower bounds on the centralized private distributed Empirical Risk Minimization (ERM).

Zhang et al. presented PEMFL as a Privacy-Enhanced Momentum Federated Learning framework to protect the privacy information of the industrial agents from being inferred by their share parameters [179]. PEMFL combines the Differential Privacy (DP) with Momentum FL (MFL)

and chaos-based encryption method to preserve the privacy information and encrypt the weight parameters of local models. The experimental results have demonstrated the excellence of PEMFL's performance in terms of accuracy and privacy security. Wang et al. presented a three-plane framework for privacy-preserving to secure the Cross-Silo FL using the Local Differential Privacy (LDP) mechanism to provide strong data privacy protection while preserving its high utility [180]. It can also be deployed as an “auxiliary module” to enhance privacy in existing FL systems. Experimental results demonstrated the effectiveness of the presented framework. Firdaus et al. a secure FL framework by empowering blockchain and using the DP to tackle various FL-related attack types such as membership inference attacks via noise addition [61]. Therefore, creating randomized privacy protection. Simulation results showed that this framework offers several advantages for FL privacy protection.

#### 6.4. Homomorphic encryption solutions

The adoption of machine learning in the IoT has surely been beneficial and of great value. However, this led to the rise of privacy and security key concerns. As a result, the demands for homomorphic encryption with federated learning solutions to protect sensitive data have drastically increased. This paper lists the main recent solutions presented by various authors and summarized in Table 11. Jebreel et al. presented Fragmented Federated Learning (FFL) to tackle the accuracy-privacy-security conflict [181]. A lightweight protocol was also presented to allow users to privately exchange and mix encrypted fragments of their updates. Moreover, to achieve security and build trust among users, a reputation-based defense tailored for FFL was designed. The aim was to allow the server to correctly reconstruct a global model from the received mixed updates without any accuracy loss. Experimental results prove the effectiveness of the presented solution. Zhang et al. presented BatchCrypt, as a system solution for Cross-Silo FL (CSFL) [143]. BatchCrypt is used to perform batch encryption and to minimize the encryption and communication overhead caused by Homomorphic Encryption (HE) with an accuracy loss that is less than 1%. Tian et al. provided a secure linear aggregation protocol using a decentralized threshold additive homomorphic encryption for federated learning to protect the users’

**Table 11**  
A set of recent Homomorphic Encryption Solutions.

Year	Author(s)	Reference	Type	Description
2020	Zhang et al.	[143]	BatchCrypt	Perform batch encryption and minimizes the encryption and communication overhead
2021	Tian et al.	[182]	Secure linear aggregation protocol	Protects users' private inputs and reduces communication and computation costs
2021	Jiang et al.	[183]	FLASHE	Secures the model aggregation process in CSFL
2021	Stripelis et al.	[184]	FHE Framework	Used for privacy-preserving biomedical data analysis
2021	Madi et al.	[185]	Secure FL framework	Discussed the use of DP, HE and MPC
2022	Ma et al.	[186]	xMK-CKKS	Protects sensitive data, prevents data leakage, and increases bandwidth demands
2022	Park et al.	[187]	PPFL	Allows the centralized server to aggregate encrypted local model parameters
2022	Kurniawan et al.	[188]	Privacy-preserving scheme	Protects sensitive information and user data privacy

private inputs while reducing communication and computation costs [182]. This protocol is based on Decentralized Threshold Additive Homomorphic Encryption (DTAHE) schemes. Jiang et al. presented Federated Learning Additively Symmetric Homomorphic Encryption (FLASHE) to secure the model aggregation process in Cross-Silo Federated Learning (CSFL) [183]. Simulation results show that FLASHE significantly outperforms other FL models in terms of performance while maintaining the same security level.

Stripelis et al. presented a framework to secure FL using Fully Homomorphic Encryption (FHE) for privacy-preserving biomedical data analysis [184]. Experimental results showed that there was no degradation in performance under encryption. Madi et al. discussed different cryptographic solutions to offer a secure FL framework while investigating the use of DP, HE, and Multi-Party Computation (MPC) [185]. Based on the conducted testing, it was revealed that HE has a better performance with lower bandwidth usage, while MPC can be used in situations where HE cannot be applied.

Ma et al. presented a multi-key homomorphic encryption protocol to design a novel privacy-preserving federated learning (xMK-CKKS) scheme to protect sensitive data by preventing its leakage, while increasing bandwidth demands [186]. The experimental evaluation demonstrated that the model accuracy is still preserved against traditional federated learning, while the energy consumption and computational cost are both reduced.

Park et al. [187] presented a Privacy-Preserving Federated Learning (PPFL) algorithm that uses a Homomorphic Encryption (HE) scheme at the centralized server to conduct arithmetic operations on ciphertexts. This will permit aggregating encrypted local model parameters without decrypting them. Another privacy-preserving scheme that uses HE is presented in Ref. [188] to protect sensitive information and user data privacy for active learning in the context of federated learning. The experimental result showed that this scheme preserved the privacy of active learning with no gradient loss in training model accuracy.

#### 6.5. Other solutions

The aim of this subsection is to present the other FL active countermeasures that are capable of not only detecting but also mitigating the presented attacks above. Here, these main countermeasures are presented as follows:

- **Sanitized Training Data (STD):** is used as an anomaly detector to filter out suspicious training data points. It was initially presented in Ref. [119]. This will secure the permanent erasure of sensitive data from the datasets beyond recovery, even when using forensics tools. It can be applied on large datasets with sensitive information and to clear end-of-life electronic IoT devices, which can be done via physical destruction or by using anti-forensics tools [189].
- **Federated Distillation (FD):** is used to only transfer model outputs with much smaller dimensions than the model sizes, relying on two foundational algorithms named Co-Distillation (CD) and Knowledge Distillation (KD) [190,191]. For example, using regular KD can mitigate the weaknesses of the parameter-averaging FL algorithms, while possibly introducing a trade-off [192].
- **Federated Multi-Task Learning (FMTL):** allows simultaneous model learning for several tasks suitable for FL's statistical and system challenges amongst unbalanced data [193]. MTL can be categorized into two main groups: the first group is based on how the relations are caught amongst tasks, while the second group is based on assuming that the tasks relations are not known beforehand and can be directly learned from the data [194].
- **Moving Target Defense (MTD):** introduces a continuous and dynamic unfolding attack surface across the FL system dimensions to confuse the attackers to further complicate their attacks [195], by constantly moving the FL system's components in a randomized manner [196]. MTD aims to create an asymmetric uncertainty for

cyber threats by reversing attacks by the defenders against the attackers and is enabled by technical trends such as resilient techniques, secure virtualization, workload migration, redundant network connectivity, and real-time compilers [197].

- **Zero-Knowledge Proofs (ZKP):** are used as cryptographic primitives to verify a statement between a proving party and a verifying party without sharing nor revealing any underlying data. This is done to ensure that all the submitted model updates are defined by rules and properties that they must respect to avoid model corruption and FL-related attacks [198]. ZKP is used to verify the statement's validity without revealing the statement nor its content. ZKP relies on algorithms that take some data as input and return 'true' or 'false' as output.
- **Robust Aggregation Methods (RAM):** are applied to sustain communication problems, clients dropouts, and malicious model training and model error updates, all while ensuring FL's aggregation algorithms security [199]. This approach is also agnostic to the corruption level while outperforming classical aggregation approaches in terms of robustness with low corruption levels [200].
- **Trusted Execution Environment (TEE):** is a high-level trusted ecosystem that executes attested and verified codes to establish a digital trust by securing connected FL devices against the injection of false training results by enabling an isolated, cryptographic electronic structure and allowing an end-to-end security [201]. Thus, ensuring that data is securely stored and processed while achieving confidentiality, authenticity, privacy, system integrity properties, and data access rights [202,203]. In fact, TEE applications are known as Trusted Applications (TAs).
- **Batch Normalization (Batch Norm):** is used to train artificial neural networks to become more fast and stable through the input layer normalization via re-centering and re-scaling [204], as well as to make attacks against local data more challenging, such as the privacy-preserving solution called FedBN, presented by Li et al. in Ref. [205]. In fact, Batch normalization also achieves a much higher learning rate with fewer initialization issues as described in Ref. [206]. Batch normalization is a technique used to improve the training of neural networks by normalizing the inputs (i.e convergence speed and generalization performance of models). Based on this, in the context of FL, batch normalization can solve the non-IID and generalization issues, in addition to being an effective countermeasure to mitigate a set of security issues since it reduces the impact of individual client data (malicious ones) on the model in case the training is done by using multiple data clients. Consequently, this ensures that the trained model is more representative of the overall distribution of data, which leads to solving the discussed issues of non-IID and generalization. Moreover, by normalizing the inputs to each layer, batch normalization helps to ensure that the model is more resilient to overfitting on individual client data (that might have malicious targets), while to making it more difficult for attackers to extract sensitive information from the trained model.

## 7. Learnt lessons

In this section, based on the presented solutions above and the issues surrounding the FL-IoT domain, several learned lessons will be presented and discussed to ensure a higher level of understanding of the security and privacy issue of the integration of FL into the IoT domain, while maintaining security and preserving privacy.

A federated learning approach is essential to building smart IoT systems, but unfortunately, it suffers from different security flaws and privacy vulnerabilities, while introducing countermeasures can have a high impact on performance and efficiency. However, maintaining both privacy and security of the FL-IoT domain is essential to maintain a safe and secure IoT environment [207]. Therefore, suggestions and recommendations are essential to achieving this purpose. However, it is also important to focus on the lessons learned from previous events to avoid

them from reoccurring and to take much more precautions using a predictive approach. As a result, these learned lessons are presented below as follows:

- Attacks can be realized at different levels: client (workers), server, data, or during communication.
- Attacks can also be passive (eavesdropping to collect models and try to recover the original data) or active (modifying models to become malicious).
- If workers have been compromised by any traditional techniques, attacks can be indirect. This means that workers should be protected against existing attacks.
- The presence of malicious models that can be loaded on workers' devices or centralized servers introduces new and dangerous vulnerabilities. This will necessitate the development of novel, effective resource-constrained countermeasures.
- Security solutions against all listed attacks cannot be based on only cryptographic solutions; we need to employ also non-cryptographic solutions such as differential privacy, honeypots, intrusion detection, and anomaly detection systems.
- A new kind of encryption algorithm is required for the FL process, which should be homomorphic, ensures the additive property, and also achieves secret sharing.
- Existing homomorphic encryption algorithms follow the asymmetric class, as they introduce an important overhead in terms of computation and storage.
- A trade-off between privacy level and model performance exists, and selecting the optimal threshold(s) is a difficult task.

## 8. Suggestions & recommendations for future research FL directions

Based on the already discussed attacks and solutions, this paper presents a framework that highlights the main attacks and defensive security measures that can be adopted and integrated into the FL domain, especially in each of its layers, which are data, server, and client/agent, as seen in Fig. 10. Therefore, to detect/prevent them and correct their damage, in this section, several suggestions and recommendations for future research directions in maintaining both security and privacy of FL with IoT systems will be proposed, and developing them will ensure a high level of security and preserve privacy for FL with IoT systems. Hence, security and privacy-preserving techniques can be based on cryptographic and non-cryptographic measures. We classify the future directions into three classes: cryptographic, non-cryptographic, and management enhancement (Fig. 11). These research topics are discussed in the following.

### 8.1. Cryptographic measures

From our cryptographic viewpoint, attacks on FL systems can be passive or active, as with any other traditional system. Therefore, existing traditional security measures should be introduced (especially against communication attacks), but with several modifications/propositions that are listed in the following:

- 1) **Lightweight Symmetric Additive cryptographic Algorithms:** local or global models should be encrypted to prevent membership inference, inverse model, and model-stealing attacks. To ensure a secure accurate aggregation process, additive homomorphic message encryption and/or authentication cryptographic algorithms are required. Unfortunately, existing HE solutions are based on asymmetric cryptographic algorithms that introduce an important overhead in all terms and especially in communication overhead as the size of ciphertext is greater than the plaintext. A trade-off between security/privacy and performance exists. According to the constraints of IoT devices, the future direction should ensure that the designed

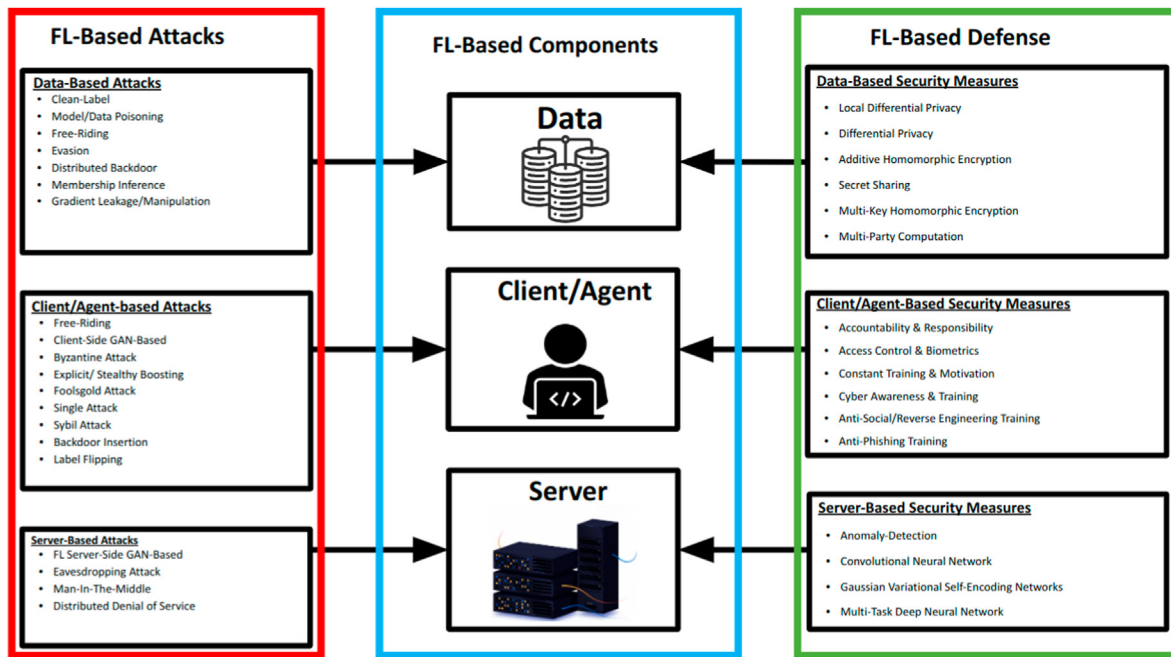


Fig. 10. Possible attacks and countermeasures for the FL system with IoT systems.

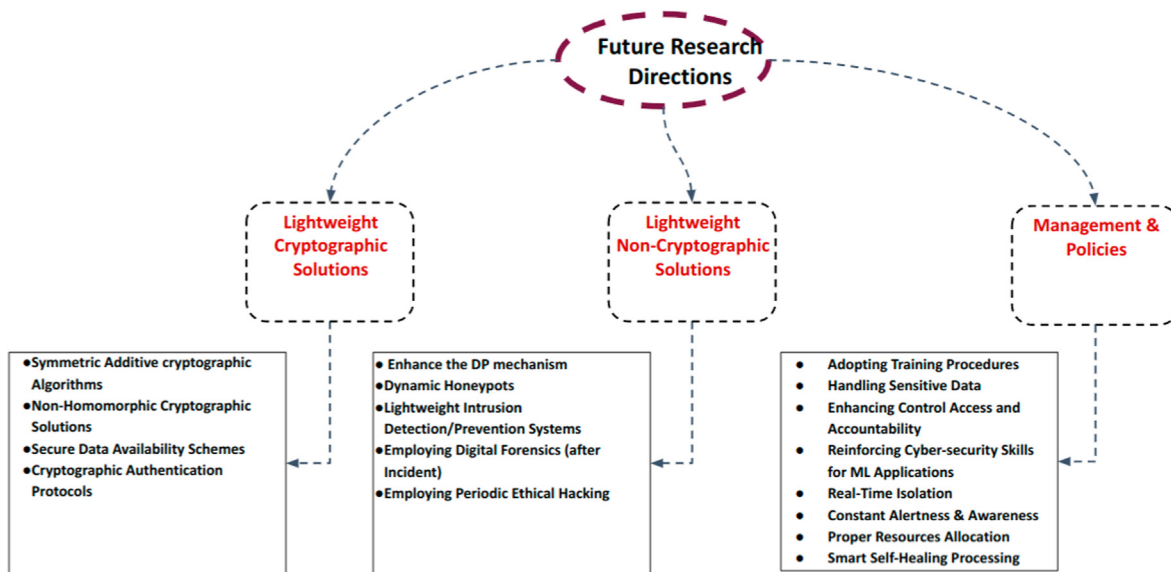


Fig. 11. Proposed future research directions that are divided into three main classes: cryptographic, non-cryptographic, and Management & Policies.

homomorphic schemes are based on the symmetric approach to strike a good balance between security and performance. Thus, the new proposed homomorphic cryptographic algorithms require preserving a higher level of security and privacy with low computation, resource, and communication overhead, especially on the worker (client or organization) side that might use limited IoT devices [208]. This will consequently ensure a lower delay overhead.

- 2) **Lightweight Non-Homomorphic Cryptographic Solutions:** aside from maintaining a secure aggregation, we also have to ensure a secure connection between the end entities (IoT devices, client/organization) and servers. These solutions will prevent the interception, modification, and even deletion of transmitted local or global models. Thus, lightweight cryptographic algorithms are required to ensure data confidentiality/integrity, privacy, source authentication, and

data availability. A set of recent lightweight cryptographic algorithms for IoT systems can be the base of this direction such as [209–211].

- 3) **Lightweight Secure Data Availability Schemes:** ensuring the communication availability of the model and updating the model is mandatory to prevent several availability attacks. These solutions should be based on the symmetric keyed secret-sharing variants. The recent work in Ref. [212] can be the base of this direction. In addition, a secure model backup should be also addressed as indicated in Ref. [213].
- 4) **Lightweight Cryptographic Authentication Protocols:** one step in this direction is to define a lightweight symmetric authentication protocol that is capable of striking a good balance between security and performance [214,215].

## 8.2. Non-cryptographic measures

In this part, we list a set of possible future directions for non-cryptographic security and privacy solutions that can be based on ML solutions, honeypot, ethical hacking, and digital forensics techniques as indicated in the following:

- 1) **Enhanced DP Mechanisms:** currently, different techniques for privacy-preserving have been presented such as DP that can provide a defense against privacy attacks but selecting the optimal parameters is a hard task. However, using this countermeasure might decrease the model's performance in terms of accuracy and precision. A trade-off between privacy and performance exists. Moreover, recent statistical techniques can be used to break DP at a certain level. We need to update/modify the DP mechanism to be able to prevent the recent DP attacks without introducing overhead in terms of computation and resources.
- 2) **Dynamic Honeypots:** FL heavily relies on the advantage of the decentralized private deep training to achieve models with real-data results [216]. Introducing honeypots at the client's end or at the centralized server with the dynamic variable selection of vulnerabilities to ensure a higher level of protected deception technology that will be capable of increasing the detection level with a higher level of interaction. Thus, achieving a faster and more accurate ability to collect information about the attacker and analyzing it in a real-time manner.
- 3) **Lightweight Intrusion Detection/Prevention Systems:** Lightweight IDS should also be integrated into FL, but more precisely with adversarial and malware detection solutions, not only to detect malware signs, especially based on the anomaly, signature, and behaviour but also to be able to distinguish and detect polymorphic and zero-day exploits. Unfortunately, in FL systems, attackers exploit the fact that the client/server will load the model. The technique consists of adding malicious code into ML models. This means that if the client/server loads it, its corresponding malicious payload will be executed. Therefore, we need to introduce an anti-malware solution for ML model checking into IDS, especially for FL systems, to detect malicious ML models before loading them at the worker or server side. On the other hand, a prime example in this direction is the Federated Blending-based IDS (F-BIDS) solution to further protect the privacy of existing ML-based IDS presented in Ref. [217]. Moreover, Anomaly detection techniques of IDS in the context of IoT-FL can use lightweight ML models to ensure a higher and faster detection rate with higher accuracy and performance. Besides, Intrusion Prevention Systems (IPS) can take both automated and precise actions after detecting an intrusion to mitigate its corresponding threats in a real-time manner. Using the ML-reinforcement learning approach can be the future of IPS direction.
- 4) **Employing Digital Forensics After Incident:** the adoption of digital forensics into the FL domain is mandatory to preserve its effectiveness and security, as digital forensics techniques allow the detection of the source and the attack vectors [218], especially since the FL system can be compromised at client devices, centralized server, or at network levels [92,93], in addition to the existing vulnerabilities of ML and DL that can also be exploited.
- 5) **Employing Periodic Ethical Hacking:** with FL models being prone to different types of attacks at different levels (clients/servers, network devices), it is important to rely on ethical hacking techniques and tools [94] periodically to detect possible vulnerabilities and recommend suitable countermeasures to ensure a higher level of security and privacy protection. Such a technique can surely identify whether the selected client or server devices are vulnerable to attacks and can be compromised to inform them as their data samples are used during the model training, and consequently, malicious damage can be realized, in addition to preserving the privacy of the training datasets and reducing the information leakage.

## 8.3. Management & policies

- 1) **Adopting Training Procedures:** it is challenging to be detailed about the adoption of mitigation techniques for adversarial perturbation attacks since they will depend on the circumstances. In this case, the attacker tries to create an input that will produce the desired result. This type of attack has been extensively investigated, where an attacker can use it against a global model in a broad range of situations depending on its goals and restrictions. Implementing a training procedure that produces models that are resistant to adversarial perturbation, together with authentication and rate-limiting to make these attacks traceable and slower, can offer some potential solutions.
- 2) **Handling Sensitive Data:** to mitigate against training data extraction, sensitive data should not be present in the local training dataset. This means the application on the client's side should be refactored to remove/anonymize any sensitive feature of data from the local training set if possible
- 3) **Enhancing Control Access and Accountability:** the security of the entire FL-system's devices must be tested to detect and fix any exploitable vulnerability or security gap. Therefore, it is important to motivate all entities, boost confidence, and discourage bad actors to reduce the internal threat and create a safe and trustworthy environment by improving the accountability methods using deterrence policies and limiting the access of different users (i.e. by using the attribute control access scheme). At the same time, maintains users' responsibility and the ability to report any suspicious activities as early as possible. This also requires enhanced policies, especially for models that rely on pre-defined datasets without them being checked and examined before deploying them for testing.
- 4) **Reinforcing Cyber-security Skills for ML Applications:** this means that we require more certified ML-based operators to create, deploy, and test models that can ensure better security (for example the use of batch normalization) in a given IoT system and its components. For example, Li et al. presented the Federated Deep Reinforcement Learning (FedDRL) scheme for ultrashort-term wind power forecasting to improve its accuracy, using the Deep Deterministic Policy Gradient (DDPG) algorithm [219]. This solution also aims to protect users' privacy and reduce communication overhead. Simulation results show that this method has good robustness with the ability to obtain at least 88.34% of network load gain.
- 5) **Real-Time Isolation:** includes the implementation of the FL-based ML mechanisms that instantly disconnect or turn off the central server and/or isolate the suspicious device once a malicious event is detected. This can ensure that the compromised device will not be controlled remotely by an adversary and that the other devices, clients, and servers will not be compromised.
- 6) **Constant Alertness & Awareness:** is required to monitor both users' and devices/systems' behaviours from any suspicious activities being conducted, to locate, identify, and mitigate any possible malicious events from occurring or further expanding.
- 7) **Proper Resources Allocation:** budget spending must be well-defined and organized to ensure that the right cyber-security practices are correctly and properly adopted, while policies are enforced.
- 8) **Smart Self-Healing Processing:** must be adopted in the design phase or added at a later stage of development to ensure that post-trained models can overcome a variety of attacks in a "smart" manner that allows them to recover and re-operate normally by identifying the affected device and isolating it.

## 9. Conclusion

Federated learning has proven itself to be a new AI-based technology that has been introduced into the IoT domain using decentralized learning to improve model accuracy while preserving user/data privacy against a variety of threats and attacks. Thus, being able to extend the machine learning advantages is beneficial for IoT domains, especially for

domains that rely on real-time data transmission and contain sensitive data. However, despite the remarkably unique advantages that the FL offers, both security and data privacy remain as persisting key issues. As a result, this paper provides a comprehensive study on both security and privacy threats, limitations, and challenges, while also presenting and analyzing recent solutions to mitigate their impact and reduce the likelihood of their occurrence. Several taxonomies were also proposed for this purpose to explain the different threats, attacks, and countermeasures for each component in FL-IoT. However, limitations of IoT devices require defining new lightweight security and privacy solutions for the FL-IoT, of which we classify as either cryptographic or non-cryptographic ones. These solutions should reduce the required computation, communication, and resource overhead. Additionally, for cryptographic solutions, symmetric homomorphic cryptographic algorithms are required instead of asymmetric ones, since they require a high overhead. In parallel, efficient secret-sharing variants are also required to ensure data availability. As for non-cryptographic purposes, existing anomaly detection should be adapted to detect/prevent a set of specific ML attacks such as adversarial attacks. Moreover, correction countermeasures should also be adopted. Thus, this work aims to indicate that new lightweight cryptographic and non-cryptographic solutions are required to maintain the security aspect and preserve data privacy, which are the main functionality of IoT devices. These solutions should be considered primarily especially for the resource-constrained IoT devices, as they require maintaining a high level of compatibility and accuracy (reducing false positives and false negatives). Finally, this work offers new perspectives, ideas, and insights toward achieving a much more secure and safe FL-based environment that can be adopted in a risk-free IoT environment. We hope that in the future, following our conducted in-depth research and investigation, the presented suggestions and recommendations are taken into consideration, applied, and further expanded to achieve a safer and more secure FL-based IoT environment.

#### Declaration of competing interest

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

#### References

- [1] Dinh C. Nguyen, Ming Ding, Pubudu N. Pathirana, Aruna Seneviratne, Jun Li, H. Vincent Poor, Federated learning for internet of things: a comprehensive survey, *IEEE Communications Surveys & Tutorials* 23 (3) (2021) 1622–1658.
- [2] Tuo Zhang, Lei Gao, Chaoyang He, Mi Zhang, Bhaskar Krishnamachari, A. Salman Avestimehr, Federated learning for the internet of things: applications, challenges, and opportunities, *IEEE Internet of Things Magazine* 5 (1) (2022) 24–29.
- [3] Mehreen Tahir, Muhammad Intizar Ali, On the performance of federated learning algorithms for iot, *IoT* 3 (2) (2022) 273–284.
- [4] Jean-Paul A. Yaacoub, Hassan N. Noura, Ola Salman, Chehab Ali, Robotics cyber security: vulnerabilities, attacks, countermeasures, and recommendations, *Int. J. Inf. Secur.* 21 (2021) 1–44.
- [5] Jean-Paul Yaacoub, Noura Hassan, Ola Salman, Chehab Ali, Security analysis of drones systems: attacks, limitations, and recommendations, *Internet of Things* 11 (2020), 100218.
- [6] Pablo García Santaclara, Ana Fernández Vilas, Rebeca P Díaz Redondo, Prototype of deployment of federated learning with iot devices, in: *Proceedings of the 19th ACM International Symposium on Performance Evaluation of Wireless Ad Hoc, Sensor, & Ubiquitous Networks*, 2022, pp. 9–16.
- [7] Madumitha Venkatasubramanian, Arash Habibi Lashkari, Saqib Hakak, *IoT Malware Analysis Using Federated Learning: A Comprehensive Survey*, IEEE Access, 2023.
- [8] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, y Arcas Blaise Aguera, Communication-efficient learning of deep networks from decentralized data, in: *Artificial Intelligence and Statistics*, PMLR, 2017, pp. 1273–1282.
- [9] Zhibo Wang, Mengkai Song, Zhifei Zhang, Yang Song, Qian Wang, Hairong Qi, Beyond inferring class representatives: user-level privacy leakage from federated learning, in: *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*, IEEE, 2019, pp. 2512–2520.
- [10] Yayuan Xiong, Fengyuan Xu, Sheng Zhong, Detecting gan-based privacy attack in distributed learning, in: *ICC 2020-2020 IEEE International Conference on Communications (ICC)*, IEEE, 2020, pp. 1–6.
- [11] Briland Hitaj, Giuseppe Ateniese, Fernando Perez-Cruz, Deep models under the gan: information leakage from collaborative deep learning, in: *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, 2017, pp. 603–618.
- [12] Hongbo Cao, Yongsheng Zhu, Yuange Ren, Bin Wang, Mingqing Hu, Wanqi Wang, Wei Wang, Prevention of gan-based privacy inferring attacks towards federated learning, in: *Collaborative Computing: Networking, Applications and Worksharing: 18th EAI International Conference, CollaborateCom 2022, Hangzhou, China, October 15-16, 2022, Proceedings, Part II*, Springer, 2023, pp. 39–54.
- [13] Jianxiong Lai, Xiuli Huang, Xianzhou Gao, Xia Chang, Jingyu Hua, Gan-based information leakage attack detection in federated learning, *Secur. Commun. Network.* (2022) 2022.
- [14] Peva Blanchard, El Mahdi El Mhamdi, Rachid Guerraoui, Julien Stainer, Machine learning with adversaries: byzantine tolerant gradient descent, *Adv. Neural Inf. Process. Syst.* 30 (2017).
- [15] Kamala Varma, Yi Zhou, Nathalie Baracaldo, Ali Anwar Legato, A layerwise gradient aggregation algorithm for mitigating byzantine attacks in federated learning, in: *2021 IEEE 14th International Conference on Cloud Computing (CLOUD)*, IEEE, 2021, pp. 272–277.
- [16] Saurav Prakash, Amir Salman Avestimehr, Mitigating Byzantine Attacks in Federated Learning, 2020 *arXiv preprint arXiv:2010.07541*.
- [17] Arjun Nitin Bhagoji, Supriyo Chakraborty, Prateek Mittal, Seraphin Calo, Analyzing federated learning through an adversarial lens, in: *International Conference on Machine Learning*, PMLR, 2019, pp. 634–643.
- [18] Aditya Ashok, Manimaran Govindarasu, Venkataramana Ajarapu, Online detection of stealthy false data injection attacks in power system state estimation, *IEEE Trans. Smart Grid* 9 (3) (2016) 1636–1646.
- [19] Mohammad Ashrafuzzaman, Yacine Chakhchoukh, Ananth A. Jillepalli, Predrag T. Tosic, Daniel Conte de Leon, Frederick T. Sheldon, Brian K. Johnson, Detecting stealthy false data injection attacks in power grids using deep learning, in: *2018 14th International Wireless Communications & Mobile Computing Conference (IWCMC)*, IEEE, 2018, pp. 219–225.
- [20] Keval Doshi, Yasin Yilmaz, Suleyman Uludag, Timely detection and mitigation of stealthy ddos attacks via iot networks, *IEEE Trans. Dependable Secure Comput.* 18 (5) (2021) 2164–2176.
- [21] Sana Awan, Bo Luo, Fengjun Li, Contra: defending against poisoning attacks in federated learning, in: *Computer Security—ESORICS 2021: 26th European Symposium on Research in Computer Security*, Darmstadt, Germany, October 4–8, 2021, *Proceedings, Part 1* 26, Springer, 2021, pp. 455–475.
- [22] Aashma Uprety, Danda B. Rawat, Mitigating poisoning attack in federated learning, in: *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*, IEEE, 2021, pp. 1–7.
- [23] Yi Liu, Jialiang Peng, Jiawen Kang, Abdullah M. Ilyasu, Dusit Niyato, Ahmed A. Abd El-Latif, A secure federated learning framework for 5g networks, *IEEE Wireless Commun.* 27 (4) (2020) 24–31.
- [24] Wei Wan, Shengshan Hu, Jianrong Lu, Leo Yu Zhang, Hai Jin, Yuan Yuan He, Shielding Federated Learning: Robust Aggregation with Adaptive Client Selection, 2022 *arXiv preprint arXiv:2204.13256*.
- [25] Yupeng Jiang, *Sybil attacks on Differential Privacy Based Federated Learning*, PhD Thesis, Macquarie University, 2022.
- [26] Thien Duc Nguyen, Phillip Rieger, Roberta De Viti, Huili Chen, Björn B. Brandenburg, Hossein Yalame, Helen Möllering, Hossein Fereidooni, Samuel Marchal, Markus Miettinen, et al., {FLAME}: taming backdoors in federated learning, in: *31st USENIX Security Symposium*, vol. 22, USENIX Security, 2022, pp. 1415–1432.
- [27] Arpan Manna, Harsh Kasyap, Somanath Tripathy, Moat: model agnostic defense against targeted poisoning attacks in federated learning, in: *Information and Communications Security: 23rd International Conference, ICICS 2021, Chongqing, China, November 19-21, 2021, Proceedings, Part I* 23, Springer, 2021, pp. 38–55.
- [28] Sebastien Andreina, Giorgia Azzurra Marson, Helen Möllering, Ghassan Karame, Baffle: backdoor detection via feedback-based federated learning, in: *2021 IEEE 41st International Conference on Distributed Computing Systems (ICDCS)*, IEEE, 2021, pp. 852–863.
- [29] Najeeb Moharram Jebreel, Josep Domingo-Ferrer, David Sánchez, Alberto Blanco-Justicia, Defending against the Label-Flipping Attack in Federated Learning, 2022 *arXiv preprint arXiv:2207.01982*.
- [30] Dongcheng Li, W Eric Wong, Wei Wang, Yao Yao, Matthew Chau, Detection and mitigation of label-flipping attacks in federated learning systems with kpc and k-means, in: *2021 8th International Conference on Dependable Systems and Their Applications (DSA)*, IEEE, 2021, pp. 551–559.
- [31] Tolpegin Vale, Stacey Truex, Mehmet Emre Gursory, Ling Liu, Data poisoning attacks against federated learning systems, in: *European Symposium on Research in Computer Security*, Springer, 2020, pp. 480–501.
- [32] Samson Ho, Achyut Reddy, Sridhar Venkatesan, Rauf Izmailov, Ritu Chadha, Alina Oprea, Data sanitization approach to mitigate clean-label attacks against malware detection systems, in: *MILCOM 2022-2022 IEEE Military Communications Conference (MILCOM)*, IEEE, 2022, pp. 993–998.
- [33] Minghong Fang, Xiaoyu Cao, Jinyuan Jia, Neil Gong, Local model poisoning attacks to {Byzantine-Robust} federated learning, in: *29th USENIX Security Symposium, USENIX Security 20*, 2020, pp. 1605–1622.
- [34] Ashwinee Panda, Saeed Mahloujifar, Arjun Nitin Bhagoji, Supriyo Chakraborty, Prateek Mittal, Sparsefed: mitigating model poisoning attacks in federated learning with sparsification, in: *International Conference on Artificial Intelligence and Statistics*, PMLR, 2022, pp. 7587–7624.

- [35] Youssef Khazbak, Tianxiang Tan, Guohong Cao, Mlguard: mitigating poisoning attacks in privacy preserving distributed collaborative learning, in: 2020 29th International Conference on Computer Communications and Networks (ICCCN), IEEE, 2020, pp. 1–9.
- [36] Taisei Awaji, Rintaro Tashima, Ryogo Kubo, Detection and mitigation of data tampering attacks for cooperative acc systems based on c-v2x, in: 2022 IEEE International Conference on Consumer Electronics-Taiwan, IEEE, 2022, pp. 149–150.
- [37] Kento Aida, Kenta Yamada, Ryosuke Hotchi, Ryogo Kubo, Dynamic network path provisioning and selection for the detection and mitigation of data tampering attacks in networked control systems, *IEEE Access* 9 (2021) 147430–147441.
- [38] Cody Lewis, Vijay Varadarajan, Nasimul Noman, Attacks against federated learning defense systems and their mitigation, *J. Mach. Learn. Res.* 24 (30) (2023) 1–50.
- [39] Cristóbal Medina-López, Vicente González-Ruiz, Leocadio G. Casado, On mitigating pollution and free-riding attacks by shamir's secret sharing in fully connected p2p systems, in: 2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC), IEEE, 2017, pp. 711–716.
- [40] Chin-Shiuh Shieh, Thanh-Tuan Nguyen, Wan-Wei Lin, Wei Kuang Lai, Mong-Fong Horng, Denis Miu, Detection of adversarial ddos attacks using symmetric defense generative adversarial networks, *Electronics* 11 (13) (2022) 1977.
- [41] Jie Li, Lu Zhou, Huaxin Li, Yan Lu, Haojin Zhu, Dynamic traffic feature camouflaging via generative adversarial networks, in: 2019 IEEE Conference on Communications and Network Security (CNS), IEEE, 2019, pp. 268–276.
- [42] Battista Biggio, Igino Corona, Davide Maiorca, Blaine Nelson, Nedim Šrđić, Pavel Laskov, Giorgio Giacinto, Fabio Roli, Evasion attacks against machine learning at test time, in: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer, 2013, pp. 387–402.
- [43] Usman Ahmed, Jerry Chun-Wei Lin, Gautam Srivastava, Mitigating adversarial evasion attacks of ransomware using ensemble learning, *Comput. Electr. Eng.* 100 (2022), 107903.
- [44] Xiaoyu Cao, Neil Zhenqiang Gong, Mitigating evasion attacks to deep neural networks via region-based classification, in: Proceedings of the 33rd Annual Computer Security Applications Conference, 2017, pp. 278–287.
- [45] Phillip Rieger, Thien Duc Nguyen, Markus Miettinen, Ahmad-Reza Sadeghi, Deepsight: Mitigating Backdoor Attacks in Federated Learning through Deep Model Inspection, 2022 *arXiv preprint arXiv:2201.00763*.
- [46] Kaiyuan Zhang, Guan hong Tao, Qiuling Xu, Siyuan Cheng, Shengwei An, Yingqi Liu, Shiwei Feng, Guangyu Shen, Pin-Yu Chen, Shiqing Ma, et al., Flip: A Provable Defense Framework for Backdoor Mitigation in Federated Learning, 2022 *arXiv preprint arXiv:2210.12873*.
- [47] Ziqi Yang, Ji yi Zhang, Ee-Chien Chang, Zhenkai Liang, Neural network inversion in adversarial setting via background knowledge alignment, in: Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security, 2019, pp. 225–240.
- [48] Ziqi Yang, Bin Shao, Bohan Xuan, Ee-Chien Chang, Fan Zhang, Defending Model Inversion and Membership Inference Attacks via Prediction Purification, 2020 *arXiv preprint arXiv:2005.03915*.
- [49] Dominik Kreuzberger, Niklas Kühn, Sebastian Hirschl, Machine Learning Operations (Mlops): Overview, Definition, and Architecture, 2022 *arXiv preprint arXiv:2205.02302*.
- [50] Po-Yu Chen, Shusen Yang, Julie A. McCann, Jie Lin, Xinyu Yang, Detection of false data injection attacks in smart-grid systems, *IEEE Commun. Mag.* 53 (2) (2015) 206–213.
- [51] Ly Vu, Quang Uy Nguyen, Diep N. Nguyen, Dinh Thai Hoang, Eryk Dutkiewicz, Deep transfer learning for iot attack detection, *IEEE Access* 8 (2020) 107335–107344.
- [52] Julian Ereth, Dataops-towards a definition, *LWDA* (2018) 104–112, 2191.
- [53] Alla Sridhar, Suman Kalyan Adari, What is mlops?, in: *Beginning MLOps with MLFlow* Springer, 2021, pp. 79–124.
- [54] Tribhuvanesh Orekondy, Bernt Schiele, Mario Fritz, Prediction Poisoning: towards Defenses against Dnn Model Stealing Attacks, 2019 *arXiv preprint arXiv:1906.10908*.
- [55] Mika Juuti, Sebastian Szlyler, Samuel Marchal, N. Asokan, Prada: protecting against dnn model stealing attacks, in: 2019 IEEE European Symposium on Security and Privacy (EuroS&P), IEEE, 2019, pp. 512–527.
- [56] Ilia Shumailov, Zakhar Shumaylov, Dmitry Kazhdan, Yiren Zhao, Nicolas Papernot, Murat A. Erdogdu, Ross J. Anderson, Manipulating sgd with data ordering attacks, *Adv. Neural Inf. Process. Syst.* 34 (2021) 18021–18032.
- [57] Oluwakemi Christiana Abikoye, Abdullahi Abubakar, Ahmed Haruna Dokoro, Oluwatobi Noah Akande, Aderonke Anthonia Kayode, A novel technique to prevent sql injection and cross-site scripting attacks using knuth-morris-pratt string match algorithm, *EURASIP J. Inf. Secur.* (1) (2020) 1–14, 2020.
- [58] Thoriq Bayu Aji, Jangkung Raharjo, Ledy Novamizanti, Mutiarahmi Khairunnisa, Robust audio watermarking via quantization and particle swarm optimization, in: *AIP Conference Proceedings*, vol. 2482, AIP Publishing LLC, 2023, 150003.
- [59] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, Vladu Adrian, Towards Deep Learning Models Resistant to Adversarial Attacks, 2017 *arXiv preprint arXiv:1706.06083*.
- [60] Jinyuan Jia, Neil Zhenqiang Gong, Defending against Machine Learning Based Inference Attacks via Adversarial Examples: Opportunities and Challenges, *Adaptive Autonomous Secure Cyber Systems*, 2020, pp. 23–40.
- [61] Muhammad Firdaus, Harashta Tatimma Larasati, Kyung-Hyune Rhee, A secure federated learning framework using blockchain and differential privacy, in: 2022 IEEE 9th International Conference on Cyber Security and Cloud Computing (CSCloud)/2022 IEEE 8th International Conference on Edge Computing and Scalable Cloud (EdgeCom), IEEE, 2022, pp. 18–23.
- [62] Sebastian Lekies, Martin Johns, Lightweight integrity protection for web storage-driven content caching, in: 6th Workshop on Web, vol. 2, 2012.
- [63] Steven Van Acker, Nick Nikiforakis, Lieven Desmet, Wouter Joosen, Piessens Frank, Flashover: automated discovery of cross-site scripting vulnerabilities in rich internet applications, in: Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security, 2012, pp. 12–13.
- [64] Prithvi Bisht, V.N. Venkatakrishnan, Xss-guard: precise dynamic prevention of cross-site scripting attacks, in: International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment, Springer, 2008, pp. 23–43.
- [65] Shashank Gupta, Brij Bhooshan Gupta, Xss-safe: a server-side approach to detect and mitigate cross-site scripting (xss) attacks in javascript code, *Arabian J. Sci. Eng.* 41 (3) (2016) 897–920.
- [66] Pierre Laperdrix, Oleksii Starov, Quan Chen, Alexandros Kapravelos, Nick Nikiforakis, Fingerprinting in style: detecting browser extensions via injected style sheets, in: 30th USENIX Security Symposium, vol. 21, USENIX Security, 2021, pp. 2507–2524.
- [67] Ashar Javed, Jens Riemer, Jörg Schwenk, Siachen: a fine-grained policy language for the mitigation of cross-site scripting attacks, in: International Conference on Information Security, Springer, 2014, pp. 515–528.
- [68] Fawaz Mahioub Mohammed Mokbal, Dan Wang, Azhar Imran, Jiuchuan Lin, Faheem Akhtar, Xiaoxi Wang, Mlpxss: an integrated xss-based attack detection scheme in web applications using multilayer perception technique, *IEEE Access* 7 (2019) 100567–100580.
- [69] Brian Evan Maher, Sachin Purushottam Joglekar, Jesper Mikael Johansson, Protecting websites from cross-site scripting, *May 12, US Patent 9 (32) (2015) 519*.
- [70] Chih-Hung Wang, Yi-Shauin Zhou, A new cross-site scripting detection mechanism integrated with html5 and cors properties by using browser extensions, in: 2016 International Computer Symposium (ICS), IEEE, 2016, pp. 264–269.
- [71] Sangho Lee, Hyungsub Kim, Jong Kim, Identifying Cross-Origin Resource Status Using Application Cache, *NDSS*, 2015.
- [72] Philippe De Ryck, Nick Nikiforakis, Lieven Desmet, Piessens Frank, Wouter Joosen, Serene: self-reliant client-side protection against session fixation, in: IFIP International Conference on Distributed Applications and Interoperable Systems, Springer, 2012, pp. 59–72.
- [73] Salman Ahmed, Qamar Mahmood, An authentication based scheme for applications using json web token, in: 2019 22nd International Multitopic Conference (INMIC), IEEE, 2019, pp. 1–6.
- [74] Rahul Kumar, Aakash Kumar Goel, Automated session fixation vulnerability detection in web applications using the set-cookie http response header in cookies, in: Proceedings of the 7th International Conference on Security of Information and Networks, 2014, pp. 351–354.
- [75] Chen Zhang, Yu Xie, Hang Bai, Bin Yu, Weihong Li, Yuan Gao, A survey on federated learning, *Knowl. Base Syst.* 216 (2021), 106775.
- [76] Qinbin Li, Zeyi Wen, Zhaomin Wu, Sixu Hu, Naibo Wang, Li Yuan, Xu Liu, Bingsheng He, A Survey on Federated Learning Systems: Vision, Hype and Reality for Data Privacy and Protection, *IEEE Transactions on Knowledge and Data Engineering*, 2021.
- [77] Virajji Mothukuri, Reza M. Parizi, Seyedamin Pouriye, Yan Huang, Dehghantanha Ali, Gautam Srivastava, A survey on security and privacy of federated learning, *Future Generat. Comput. Syst.* 115 (2021) 619–640.
- [78] Jianxin Zhao, Yanhao Feng, Xinyu Chang, Peng Xu, Shilin Li, Chi Harold Liu, Wenke Yu, Jian Tang, Jon Crowcroft, Energy-efficient and fair iot data distribution in decentralised federated learning, *IEEE Transactions on Network Science and Engineering* (2022).
- [79] Qiang Yang, Yang Liu, Yong Cheng, Yan Kang, Tianjian Chen, Yu Han, Federated learning, *Synthesis Lectures on Artificial Intelligence and Machine Learning* 13 (3) (2019) 1–207.
- [80] Dianwen Ng, Xiang Lan, Melissa Min-Szu Yao, Wing P. Chan, Mengling Feng, Federated learning: a collaborative effort to achieve better medical imaging models for individual sites that have small labelled datasets, *Quant. Imag. Med. Surg.* 11 (2) (2021) 852.
- [81] Zeki Murat Çınar, Abubakar Abdussalam Nuhu, Qasim Zeeshan, Orhan Korhan, Mohammed Asmael, Babak Safaei, Machine learning in predictive maintenance towards sustainable smart manufacturing in industry 4.0, *Sustainability* 12 (19) (2020) 8211.
- [82] Quoc-Viet Pham, Kapal Dev, Praveen Kumar Reddy Maddikunta, Thippa Reddy Gadekallu, Thien Huynh-The, et al., Fusion of Federated Learning and Industrial Internet of Things: A Survey, 2021 *arXiv preprint arXiv:2101.00798*.
- [83] Parimala Boopalan, Swarna Priya Ramu, Quoc-Viet Pham, Kapal Dev, Praveen Kumar Reddy Maddikunta, Thippa Reddy Gadekallu, Thien Huynh-The, et al., Fusion of Federated Learning and Industrial Internet of Things: A Survey, *Computer Networks*, 2022, 109048.
- [84] Chong Zhang, Xiao Liu, Xi Zheng, Rui Li, Huai Liu, Fenghuo: a federated learning based edge computing platform for cyber-physical systems, in: 2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), IEEE, 2020, pp. 1–4.
- [85] Zhaoshua Zheng, Yize Zhou, Yilong Sun, Wang Zhang, Boyi Liu, Keqiu Li, Applications of federated learning in smart cities: recent advances, taxonomy, and open challenges, *Connect. Sci.* 34 (1) (2022) 1–28.
- [86] Swarna Priya Ramu, Parimala Boopalan, Quoc-Viet Pham, Praveen Kumar Reddy Maddikunta, Thien Huynh-The, Mamoun Alazab, Thanh Thi Nguyen, Thippa Reddy Gadekallu, Federated learning enabled digital twins for smart cities:

- concepts, recent advances, and future directions, *Sustain. Cities Soc.* 79 (2022), 103663.
- [87] Prabhath Kumar, Govind P. Gupta, Rakesh Tripathi, Pefl: deep privacy-encoding-based federated learning framework for smart agriculture, *IEEE Micro* 42 (1) (2021) 33–40.
- [88] Xianjia Yu, Jorge Peña Queraltó, Jukka Heikkonen, Tomi Westerlund, An Overview of Federated Learning at the Edge and Distributed Ledger Technologies for Robotic and Autonomous Systems, *CoRR*, 2021.
- [89] Xianjia Yu, Jorge Peña Queraltó, Tomi Westerlund, Towards Lifelong Federated Learning in Autonomous Mobile Robots with Continuous Sim-To-Real Transfer, 2022 *arXiv preprint arXiv:2205.15496*.
- [90] Yi Liu, Xingliang Yuan, Zehui Xiong, Jiawen Kang, Xiaofei Wang, Dusit Niyato, Federated learning for 6g communications: challenges, methods, and future directions, *China Communications* 17 (9) (2020) 105–118.
- [91] Gonzalo De La Torre Parra, Luis Selvera, Joseph Khoury, Hector Irizarry, Elias Bou-Harb, Rad Paul, Interpretable federated transformer log learning for cloud threat forensics, in: *Proceedings of the Network and Distributed Systems Security, NDSS Symposium*, 2022.
- [92] Jean-Paul A. Yaacoub, Hassan N. Noura, Ola Salman, Chehab Ali, Advanced Digital Forensics and Anti-digital Forensics for IoT Systems: Techniques, Limitations and Recommendations, *Internet of Things*, 2022, 100544.
- [93] Jean-Paul A. Yaacoub, Hassan N. Noura, Ola Salman, Chehab Ali, Digital Forensics vs. Anti-digital Forensics: Techniques, Limitations and Recommendations, 2021 *arXiv preprint arXiv:2103.17028*.
- [94] Jean-Paul A. Yaacoub, Hassan N. Noura, Ola Salman, Chehab Ali, A Survey on Ethical Hacking: Issues and Challenges, 2021 *arXiv preprint arXiv:2103.15072*.
- [95] Momina Shaheen, Muhammad Shoab Farooq, Tariq Umer, Byung-Seo Kim, Applications of federated learning: taxonomy, challenges, and research trends, *Electronics* 11 (4) (2022) 670.
- [96] **Strategic Plan. Marine Corps Science and Technology Strategic Plan.**
- [97] Shiqiang Wang, Tiffany Tuor, Theodoros Salonidis, Kin K. Leung, Christian Makaya, He Ting, Kevin Chan, Adaptive federated learning in resource constrained edge computing systems, *IEEE J. Sel. Area. Commun.* 37 (6) (2019) 1205–1221.
- [98] Kaiyue Zhang, Xuan Song, Chenhan Zhang, Shui Yu, Challenges and future directions of secure federated learning: a survey, *Front. Comput. Sci.* 16 (1–8) (2022).
- [99] Li Tian, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, Virginia Smith, Federated optimization in heterogeneous networks, *Proceedings of Machine Learning and Systems* 2 (2020) 429–450.
- [100] Wonsuk Oh, Girish N. Nadkarni, Federated learning in health care using structured medical data, *Advances in Kidney Disease and Health* 30 (1) (2023) 4–16.
- [101] Sharnil Pandya, Gautam Srivastava, Rutvij Jhaveri, M. Rajasekhara Babu, Sweta Bhattacharya, Praveen Kumar Reddy Maddikunta, Spyridon Mastorakis, Md Jalil Piran, Thippa Reddy Gadekallu, Federated learning for smart cities: a comprehensive survey, *Sustain. Energy Technol. Assessments* 55 (2023), 102987.
- [102] Hatamizadeh Ali, Hongxu Yin, Pavlo Molchanov, Andriy Myronenko, Wenqi Li, Prerna Dogra, Andrew Feng, Mona G. Flores, Jan Kautz, Daguang Xu, et al., Do gradient inversion attacks make federated learning unsafe? *IEEE Trans. Med. Imag.* (2023).
- [103] Q Li, Z Wen, and B He. **Federated Learning Systems: Vision, Hype and Reality for Data Privacy and Protection.** arxiv 2019. *arXiv preprint arXiv:1907.09693*.
- [104] Muhammad Asad, Moustafa Ahmed, Chao Yu, A critical evaluation of privacy and security threats in federated learning, *Sensors* 20 (24) (2020) 7182.
- [105] Alberto Blanco-Justicia, Josep Domingo-Ferrer, Sergio Martínez, David Sánchez, Adrian Flanagan, Kuan Eik Tan, Achieving security and privacy in federated learning systems: survey, research challenges and future directions, *Eng. Appl. Artif. Intell.* 106 (2021), 104468.
- [106] Junjie Tan, Ying-Chang Liang, Nguyen Cong Luong, Dusit Niyato, Toward smart security enhancement of federated learning networks, *IEEE Network* 35 (1) (2020) 340–347.
- [107] Rémi Gosselin, Loïc Vieu, Faiza Loukil, Alexandre Benoit, Privacy and security in federated learning: a survey, *Appl. Sci.* 12 (19) (2022) 9901.
- [108] Nader Bouacida, Prasant Mohapatra, Vulnerabilities in federated learning, *IEEE Access* 9 (2021) 63229–63249.
- [109] Yi Liu, Neeraj Kumar, Zehui Xiong, Wei Yang Bryan Lim, Jiawen Kang, Dusit Niyato, Communication-efficient federated learning for anomaly detection in industrial internet of things, in: *GLOBECOM 2020-2020 IEEE Global Communications Conference*, IEEE, 2020, pp. 1–6.
- [110] Yi Liu, Lei Xu, Xingliang Yuan, Cong Wang, Bo Li, The Right to Be Forgotten in Federated Learning: an Efficient Realization with Rapid Retraining, 2022 *arXiv preprint arXiv:2203.07320*.
- [111] F Gregory Gause III, The price of order: settling for less in the middle east, *Foreign Aff.* 101 (10) (2022).
- [112] David Waterman, Unbranding: Disenfranchising Terrorism and Disenchanting Terrorists, 2022.
- [113] János Besenyő, Attila Gulyás, Darko Trifunovic, Hezbollah and the internet in the twenty-first century, *Int. J. Intell. Count. Intell.* (2022) 1–17.
- [114] Derrick Tin, Dennis G. Barten, Harald De Cauwer, Luc Jm Mortelmans, Gregory R. Ciottone, Terrorist attacks in western europe: a counter-terrorism medicine analysis, *Prehospital Disaster Med.* 37 (1) (2022) 19–24.
- [115] Diptiben Ghelani, Tan Kian Hua, Surendra Kumar Reddy Koduru, Cyber Security Threats, Vulnerabilities, and Security Solutions Models in Banking, *Authorea Preprints*, 2022.
- [116] Boyang Zhang, Tao Zhang, Zhijian Yu, Ddos detection and prevention based on artificial intelligence techniques, in: *2017 3rd IEEE International Conference on Computer and Communications (ICCC)*, IEEE, 2017, pp. 1276–1280.
- [117] Chuan Xu, Giovanni Neglia, What else is leaked when eavesdropping federated learning?, in: *CCS Workshop Privacy Preserving Machine Learning (PPML)*, 2021.
- [118] Derui Wang, Chaoran Li, Sheng Wen, Surya Nepal, Xiang Yang, Man-in-the-middle attacks against machine learning classifiers via malicious generative models, *IEEE Trans. Dependable Secure Comput.* 18 (5) (2020) 2074–2087.
- [119] Gabriela F. Cretu, Angelos Stavrou, Michael E. Locasto, Salvatore J. Stolfo, Angelos D. Keromytis, Casting out demons: sanitizing training data for anomaly sensors, in: *2008 IEEE Symposium on Security and Privacy (Sp 2008)*, IEEE, 2008, pp. 81–95.
- [120] Eugene Bagdasaryan, Andreas Veit, Yiqing Hua, Deborah Estrin, Vitaly Shmatikov, How to backdoor federated learning, in: *International Conference on Artificial Intelligence and Statistics*, PMLR, 2020, pp. 2938–2948.
- [121] Clement Fung, Chris JM. Yoon, Ivan Beschastnikh, The limitations of federated learning in sybil settings, in: *23rd International Symposium on Research in Attacks, Intrusions and Defenses, RAID*, 2020, pp. 301–316, 2020.
- [122] Mourad Benmalek, Mohamed Ali Benrekia, Yacine Challal, Security of federated learning: attacks, defensive mechanisms, and challenges, *Revue des Sciences et Technologies de l'Information-Série RIA: Rev. Intelligence Artif.* 36 (1) (2022) 49–59.
- [123] Micah Goldblum, Dimitris Tsipras, Chulin Xie, Xinyun Chen, Avi Schwarzschild, Dawn Song, Aleksander Madry, Bo Li, Tom Goldstein, Dataset Security for Machine Learning: Data Poisoning, Backdoor Attacks, and Defenses, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.
- [124] Battista Biggio, Blaine Nelson, Pavel Laskov, Poisoning Attacks against Support Vector Machines, 2012 *arXiv preprint arXiv:1206.6389*.
- [125] Najeeb Moharram Jebreel, Josep Domingo-Ferrer, Fl-defender: combating targeted attacks in federated learning, *Knowl. Base Syst.* (2022), 110178.
- [126] Xingchen Zhou, Ming Xu, Yiming Wu, Ning Zheng, Deep model poisoning attack on federated learning, *Future Internet* 13 (3) (2021) 73.
- [127] Shafahi Ali, W Ronny Huang, Mahyar Najibi, Octavian Suciu, Christoph Studer, Tudor Dumitras, Tom Goldstein, Poison frogs! targeted clean-label poisoning attacks on neural networks, *Adv. Neural Inf. Process. Syst.* 31 (2018).
- [128] Yann Fraboni, Richard Vidal, Marco Lorenzi, Free-rider attacks on model aggregation in federated learning, in: *International Conference on Artificial Intelligence and Statistics*, PMLR, 2021, pp. 1846–1854.
- [129] Chulin Xie, Keli Huang, Pin-Yu Chen, Bo Li, Dba: distributed backdoor attacks against federated learning, in: *International Conference on Learning Representations*, 2019.
- [130] Nicholas Carlini, Steve Chien, Milad Nasr, Shuang Song, Andreas Terzis, and Florian Tramèr. Membership inference attacks from first principles, in: *2022 IEEE Symposium on Security and Privacy (SP)*, IEEE, 2022, pp. 1897–1914.
- [131] Ruihan Wu, Jin Peng Zhou, Kilian Q. Weinberger, Chuan Guo, Does Label Differential Privacy Prevent Label Inference Attacks?, 2022 *arXiv preprint arXiv:2202.12968*.
- [132] Jiaming Zhang, Jitao Sang, Xian Zhao, Xiaowen Huang, Yanfeng Sun, Yongli Hu, Adversarial privacy-preserving filter, in: *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 1423–1431.
- [133] Andjela Mladenovic, Avishek Joey Bose, Hugo Berard, William L. Hamilton, Simon Lacoste-Julien, Pascal Vincent, Gauthier Gidel, Online Adversarial Attacks, 2021 *arXiv preprint arXiv:2103.02014*.
- [134] Shahbaz Rezaei, Xin Liu, A Target-Agnostic Attack on Deep Models: Exploiting Security Vulnerabilities of Transfer Learning, 2019 *arXiv preprint arXiv:1904.04334*.
- [135] Hanjun Dai, Hui Li, Tian Tian, Xin Huang, Lin Wang, Jun Zhu, Le Song, Adversarial attack on graph structured data, in: *International Conference on Machine Learning*, PMLR, 2018, pp. 1115–1124.
- [136] Samuel G. Finlayson, John D. Bowers, Joichi Ito, Jonathan L. Zittrain, Andrew L. Beam, Isaac S. Kohane, Adversarial attacks on medical machine learning, *Science* 363 (6433) (2019) 1287–1289.
- [137] Ligeng Zhu, Zhijian Liu, Song Han, Deep leakage from gradients, *Adv. Neural Inf. Process. Syst.* 32 (2019).
- [138] Maksy Sendiang, Anritsu Polii, Jusuf Mappadang, Minimization of sql injection in scheduling application development, in: *2016 International Conference on Knowledge Creation and Intelligent Computing (KCIC)*, IEEE, 2016, pp. 14–20.
- [139] Kei Wei, Muthusrinivasan Muthuprasanna, Suraj Kothari, Preventing sql injection attacks in stored procedures, in: *Australian Software Engineering Conference (ASWEC'06)*, IEEE, 2006, p. 8.
- [140] Muath AlShaikh, Malek Alzaqebah, Sana Jawarneh, Robust watermarking based on modified pigeon algorithm in dct domain, *Multimed. Tool. Appl.* 82 (2) (2023) 3033–3053.
- [141] Yusof Imran, Al-Sakib Khan Pathan, Preventing persistent cross-site scripting (xss) attack by applying pattern filtering approach, in: *The 5th International Conference on Information and Communication Technology for the Muslim World (ICT4M)*, IEEE, 2014, pp. 1–6.
- [142] Syed Zeeshan Hussain, Nancy Agarwal, Content spoofing via compounded sql injection, in: *International Conference on Intelligent Computing and Communication Technologies*, Springer, 2019, pp. 244–252.
- [143] Chengliang Zhang, Suyi Li, Junzhe Xia, Wei Wang, Feng Yan, Yang Liu, {BatchCrypt}: efficient homomorphic encryption for {Cross-Silo} federated learning, in: *2020 USENIX Annual Technical Conference (USENIX ATC 20)*, 2020, pp. 493–506.
- [144] Yoshinori Aono, Takuya Hayashi, Lihua Wang, Shihō Moriai, et al., Privacy-preserving deep learning via additively homomorphic encryption, *IEEE Trans. Inf. Forensics Secur.* 13 (5) (2017) 1333–1345.

- [145] Mikail Mohammed Salim, Shailendra Rathore, Jong Hyuk Park, Distributed denial of service attacks and its defenses in iot: a survey, *J. Supercomput.* 76 (7) (2020) 5320–5363.
- [146] Felix Sattler, Klaus-Robert Müller, Wojciech Samek, Clustered federated learning: model-agnostic distributed multitask optimization under privacy constraints, *IEEE Transact. Neural Networks Learn. Syst.* 32 (8) (2020) 3710–3722.
- [147] Hongyi Wang, Mikhail Yurochkin, Yuekai Sun, Dimitris Papailiopoulos, Yasaman Khazaeni, Federated Learning with Matched Averaging, 2020 *arXiv preprint arXiv:2002.06440*.
- [148] Aritra Dutta, El Houcine Bergou, Ahmed M. Abdelmoniem, Chen-Yu Ho, Atal Narayan Sahu, Marco Canini, Panos Kalnis, On the discrepancy between the theoretical analysis and practical implementations of compressed communication for distributed deep learning, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, 2020, pp. 3817–3824.
- [149] Hang Xu, Chen-Yu Ho, Ahmed M. Abdelmoniem, Aritra Dutta, El Houcine Bergou, Konstantinos Karatsenidis, Marco Canini, Panos Kalnis, Compressed Communication for Distributed Deep Learning: Survey and Quantitative Evaluation, Technical report, 2020.
- [150] Nicola Rieke, Jonny Hancox, Wenqi Li, Fausto Milletari, Holger R. Roth, Shadi Albarqouni, Spyridon Bakas, Mathieu N. Galtier, Bennett A. Landman, Klaus Maier-Hein, et al., The future of digital health with federated learning, *NPJ digital medicine* 3 (1) (2020) 1–7.
- [151] Daniel Kang, Yi Sun, Dan Hendrycks, Tom Brown, Jacob Steinhardt, Testing Robustness against Unforeseen Adversaries, 2019 *arXiv preprint arXiv:1908.08016*.
- [152] Jean-Paul A. Yaacoub, Mohamad Noura, Hassan N. Noura, Ola Salman, Yaacoub Elias, Raphaël Couturier, Chehab Ali, Securing internet of medical things systems: limitations, issues and recommendations, *Future Generat. Comput. Syst.* 105 (2020) 581–606.
- [153] Noura Hassan, Tarif Hatoum, Ola Salman, Jean-Paul Yaacoub, Chehab Ali, Lorawan security survey: issues, threats and possible mitigation techniques, *Internet of Things* 12 (2020), 100303.
- [154] Jean Paul A. Yaacoub, Javier Hernandez Fernandez, Hassan N. Noura, Chehab Ali, Security of power line communication systems: issues, limitations and existing solutions, *Computer Science Review* 39 (2021), 100331.
- [155] Jean-Paul A. Yaacoub, Ola Salman, Hassan N. Noura, Nesrine Kaaniche, Chehab Ali, Mohamad Malli, Cyber-physical systems security: limitations, issues and future trends, *Microprocess. Microsyst.* 77 (2020), 103201.
- [156] Chuan Ma, Jun Li, Ming Ding, Howard H. Yang, Feng Shu, Tony QS. Quek, H. Vincent Poor, On safeguarding privacy and security in the framework of federated learning, *IEEE network* 34 (4) (2020) 242–248.
- [157] Brett Weinger, Jinoh Kim, Alex Sim, Makiya Nakashima, Nour Moustafa, K John Wu, Enhancing iot anomaly detection performance for federated learning, *Digital Communications and Networks* 8 (2022).
- [158] Ying Zhao, Junjun Chen, Di Wu, Jian Teng, Shui Yu, Multi-task network anomaly detection using federated learning, in: *Proceedings of the Tenth International Symposium on Information and Communication Technology*, 2019, pp. 273–279.
- [159] Virajji Muthukuri, Prachi Khare, Reza M. Parizi, Seyedamin Pouriyeh, Dehghantanha Ali, Gautam Srivastava, Federated-learning-based anomaly detection for iot security attacks, *IEEE Internet Things J.* 9 (4) (2021) 2545–2554.
- [160] Deepti Gupta, Olumide Kayode, Smriti Bhatt, Maanak Gupta, Ali Saman Tosun, Hierarchical federated learning based anomaly detection using digital twins for smart healthcare, in: 2021 IEEE 7th International Conference on Collaboration and Internet Computing (CIC), IEEE, 2021, pp. 16–25.
- [161] Mingzhe Chen, Nir Shlezinger, H. Vincent Poor, Yonina C. Eldar, Shuguang Cui, Communication-efficient federated learning, *Proc. Natl. Acad. Sci. USA* 118 (17) (2021), e2024789118.
- [162] Dongmin Wu, Yi Deng, Mingyong Li, Fl-mgvm: federated learning for anomaly detection using mixed Gaussian variational self-encoding network, *Inf. Process. Manag.* 59 (2) (2022), 102839.
- [163] Yi Liu, Sahil Garg, Jiangtian Nie, Yang Zhang, Zehui Xiong, Jiawan Kang, M Shamim Hossain, Deep anomaly detection for time-series data in industrial iot: a communication-efficient on-device federated learning approach, *IEEE Internet Things J.* 8 (8) (2020) 6348–6358.
- [164] Love Allen Chijioke Ahakonye, Cosmas Ifeanyi Nwakanma, Jae Min Lee, and Dong-Seong Kim. Fed-marine: Federated Learning Framework for Ddos Detection and Mitigation in Maritime-Scada Network.
- [165] Jianhua Li, Lingjuan Lyu, Ximeng Liu, Xuyun Zhang, Xixiang Lyu, Fleam: a federated learning empowered architecture to mitigate ddos in industrial iot, *IEEE Trans. Ind. Inf.* 18 (6) (2021) 4059–4068.
- [166] Jiachao Zhang, Peiran Yu, Le Qi, Song Liu, Haiyu Zhang, Jianzhong Zhang, Flddos: ddos attack detection model based on federated learning, in: 2021 IEEE 20th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), IEEE, 2021, pp. 635–642.
- [167] Roberto Doriguzzi-Corin, Domenico Siracusa, in: *Flad: Adaptive Federated Learning for Ddos Attack Detection*, 2022 *arXiv preprint arXiv:2205.06661*.
- [168] Zainudin Ahmad, Rubina Akter, Dong-Seong Kim, Jae-Min Lee, Privacy-preserving Federated Learning-Based Ddos Classification for IiOT Networks, 2022, pp. 504–505.
- [169] Dingyang Lv, Xinyu Cheng, Jinghui Zhang, Wei Zhang, Wei Zhao, Xu He, Ddos attack detection based on cnn and federated learning, in: 2021 Ninth International Conference on Advanced Cloud and Big Data (CBD), IEEE, 2022, pp. 236–241.
- [170] Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, Li Zhang, Deep learning with differential privacy, in: *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 2016, pp. 308–318.
- [171] Ming Yang, Hang Cheng, Fei Chen, Ximeng Liu, Meiqing Wang, Xibin Li, Model Poisoning Attack in Differential Privacy-Based Federated Learning, *Information Sciences*, 2023.
- [172] Olivia Choudhury, Aris Gkoulalas-Divanis, Theodoros Salonidis, Issa Sylla, Yoonyoung Park, Grace Hsu, Amar Das, Differential Privacy-Enabled Federated Learning for Sensitive Health Data, 2019 *arXiv preprint arXiv:1910.02578*.
- [173] Stacey Truex, Ling Liu, Ka-Ho Chow, Mehmet Emre Gursoy, Wenqi Wei, Ldp-fed: federated learning with local differential privacy, in: *Proceedings of the Third ACM International Workshop on Edge Systems, Analytics and Networking*, 2020, pp. 61–66.
- [174] Lichao Sun, Jianwei Qian, Xun Chen, Ldp-fl: Practical Private Aggregation in Federated Learning with Local Differential Privacy, 2020 *arXiv preprint arXiv:2007.15789*.
- [175] Wei Kang, Jun Li, Ming Ding, Chuan Ma, Howard H. Yang, Farhad Farokhi, Shi Jin, Tony QS. Quek, H. Vincent Poor, Federated learning with differential privacy: algorithms and performance analysis, *IEEE Trans. Inf. Forensics Secur.* 15 (2020) 3454–3469.
- [176] Md Tamjid Hossain, Shafkat Islam, Shahriar Badsha, Haoting Shen, Desmp: differential privacy-exploited stealthy model poisoning attacks in federated learning, in: 2021 17th International Conference on Mobility, Sensing and Networking (MSN), IEEE, 2021, pp. 167–174.
- [177] Zhuotao Lian, Weizheng Wang, Chunhua Su, Cofel: communication-efficient and optimized federated learning with local differential privacy, in: *ICC 2021-IEEE International Conference on Communications, IEEE*, 2021, pp. 1–6.
- [178] Antonious Girgis, Deepesh Data, Suhas Diggavi, Kairouz Peter, Ananda Theertha Suresh, Shuffled model of differential privacy in federated learning, in: *International Conference on Artificial Intelligence and Statistics*, PMLR, 2021, pp. 2521–2529.
- [179] Zehui Zhang, Linlin Zhang, Qingdan Li, Kunshu Wang, Ningxin He, Tiegang Gao, Privacy-enhanced momentum federated learning via differential privacy and chaotic system in industrial cyber-physical systems, *ISA Trans.* 128 (2022) 17–31.
- [180] Chen Wang, Xinkui Wu, Gaoyang Liu, Tianping Deng, Kai Peng, Shaohua Wan, Safeguarding cross-silo federated learning with local differential privacy, *Digital Communications and Networks* 8 (4) (2022) 446–454.
- [181] Najeeb Moharram Jebreel, Josep Domingo-Ferrer, Alberto Blanco-Justicia, David Sánchez, Enhanced security and privacy via fragmented federated learning, *IEEE Transact. Neural Networks Learn. Syst.* (2022).
- [182] Haibo Tian, Fangguo Zhang, Yunfeng Shao, Bingshuai Li, Secure Linear Aggregation Using Decentralized Threshold Additive Homomorphic Encryption for Federated Learning, 2021 *arXiv preprint arXiv:2111.10753*.
- [183] Zhifeng Jiang, Wei Wang, Yang Liu, Flashe: Additively Symmetric Homomorphic Encryption for Cross-Silo Federated Learning, 2021 *arXiv preprint arXiv:2109.00675*.
- [184] Dimitris Stripelis, Hamza Saleem, Tanmay Ghai, Nikhil Dhinagar, Umang Gupta, Chrysovalantis Anastasiou, Greg Ver Steeg, Srivatsan Ravi, Muhammad Naveed, Paul M. Thompson, et al., Secure neuroimaging analysis using federated learning with homomorphic encryption, in: 17th International Symposium on Medical Information Processing and Analysis, vol. 12088, SPIE, 2021, pp. 351–359.
- [185] Abbas Madi, Oana Stan, Aurélien Mayoue, Arnaud Grivet-Sébert, Cédric Gouy-Pailler, Sirdey Renaud, A secure federated learning framework using homomorphic encryption and verifiable computing, in: 2021 Reconciling Data Analytics, Automation, Privacy, and Security: A Big Data Challenge (RDAAPS), IEEE, 2021, pp. 1–8.
- [186] Jing Ma, Si-Ahmed Naas, Sigg Stephan, Xixiang Lyu, Privacy-preserving federated learning based on multi-key homomorphic encryption, *Int. J. Intell. Syst.* 37 (2022).
- [187] Jaehyoung Park, Hyuk Lim, Privacy-preserving federated learning using homomorphic encryption, *Appl. Sci.* 12 (2) (2022) 734.
- [188] Hendra Kurmiawan, Masahiro Mambo, Homomorphic encryption-based federated privacy preservation for deep active learning, *Entropy* 24 (11) (2022) 1545.
- [189] Antonio Emanuele Cinà, Kathrin Grosse, Ambra Demontis, Sebastiano Vascon, Zellerger Werner, Bernhard A. Moser, Alina Oprea, Battista Biggio, Marcello Pelillo, Fabio Roli, Wild Patterns Reloaded: A Survey of Machine Learning Security against Training Data Poisoning, 2022 *arXiv preprint arXiv:2205.01992*.
- [190] Hyowoon Seo, Jihong Park, Seungeun Oh, Mehdi Bennis, Seong-Lyun Kim, 16 federated knowledge distillation, *Machine Learning and Wireless Communications* (2022) 457.
- [191] Daliang Li, Junpu Wang, Fedmd: Heterogenous Federated Learning via Model Distillation, 2019 *arXiv preprint arXiv:1910.03581*.
- [192] Alessio Mora, Irene Tenison, Paolo Bellavista, Irina Rish, Knowledge Distillation for Federated Learning: a Practical Guide, 2022 *arXiv preprint arXiv:2211.04742*.
- [193] Rui Hu, Yuanxiong Guo, Hongning Li, Qingqi Pei, Yanmin Gong, Personalized federated learning with differential privacy, *IEEE Internet Things J.* 7 (10) (2020) 9530–9539.
- [194] Virginia Smith, Chao-Kai Chiang, Maziar Sanjabi, S Talwalkar Ameet, Federated multi-task learning, *Adv. Neural Inf. Process. Syst.* 30 (2017).
- [195] Lei Cheng, Hong-Qi Zhang, Jing-Lei Tan, Yu-Chen Zhang, Xiao-Hu Liu, Moving Target Defense Techniques: A Survey, *Security and Communication Networks*, 2018, p. 2018.
- [196] Saikil Sengupta, Ankur Chowdhary, Abdulhakim Sabur, Adel Alshamrani, Dijiang Huang, Subbarao Kambhampati, A survey of moving target defenses for network security, *IEEE Communications Surveys & Tutorials* 22 (3) (2020) 1909–1941.

- [197] Jianjun Zheng, Akbar Siami Namin, A survey on the moving target defense strategies: an architectural perspective, *J. Comput. Sci. Technol.* 34 (2019) 207–233.
- [198] Parno Bryan, Jon Howell, Gentry Craig, Mariana Raykova, Pinocchio: nearly practical verifiable computation, *Commun. ACM* 59 (2) (2016) 103–112.
- [199] Matei Grama, Maria Musat, Luis Muñoz-González, Jonathan Passerat-Palmbach, Daniel Rueckert, Amir Alansary, Robust Aggregation for Adaptive Privacy Preserving Federated Learning in Healthcare, 2020 *arXiv preprint arXiv:2009.08294*.
- [200] Pillutla Krishna, Sham M. Kakade, Zaid Harchaoui, Robust aggregation for federated learning, *IEEE Trans. Signal Process.* 70 (2022) 1142–1154.
- [201] Pramod Subramanyan, Rohit Sinha, Ilia Lebedev, Srinivas Devadas, Sanjit A. Seshia, A formal foundation for secure remote execution of enclaves, in: *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, 2017, pp. 2435–2450.
- [202] Mohamed Sabt, Mohammed Achemlal, Abdelmadjid Bouabdallah, Trusted execution environment: what it is, and what it is not, in: 2015 *IEEE Trustcom/BigDataSE/Ispa*, vol. 1, IEEE, 2015, pp. 57–64.
- [203] Attia Qammar, Jianguo Ding, Huansheng Ning, Federated learning attack surface: taxonomy, cyber defences, challenges, and future directions, *Artif. Intell. Rev.* (2022) 1–38.
- [204] Dinh C. Nguyen, Quoc-Viet Pham, Pubudu N. Pathirana, Ming Ding, Aruna Seneviratne, Zihuai Lin, Octavia Dobre, Won-Joo Hwang, Federated learning for smart healthcare: a survey, *ACM Comput. Surv.* 55 (3) (2022) 1–37.
- [205] Xiaoxiao Li, Meirui Jiang, Xiaofei Zhang, Michael Kamp, Dou Qi, Fedbn: Federated Learning on Non-iid Features via Local Batch Normalization, 2021 *arXiv preprint arXiv:2102.07623*.
- [206] Sergey Ioffe, Christian Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift, in: *International Conference on Machine Learning*, pmlr, 2015, pp. 448–456.
- [207] CSIA NITRD, Iwg: Cybersecurity Game-Change Research and Development Recommendations, 2013.
- [208] Khalil Hariss, Noura Hassan, Acis: lightweight and robust homomorphic block cipher additive scheme, in: 2022 *International Wireless Communications and Mobile Computing (IWCMC)*, IEEE, 2022, pp. 413–418.
- [209] Noura Hassan, Ola Salman, Raphaël Couturier, Chehab Ali, Lesca: lightweight stream cipher algorithm for emerging systems, *Ad Hoc Netw.* 138 (2023), 102999.
- [210] Hassan N. Noura, Ola Salman, Raphaël Couturier, Chehab Ali, A single-pass and one-round message authentication encryption for limited iot devices, *IEEE Internet Things J.* 9 (18) (2022) 17885–17900.
- [211] Hassan N. Noura, Ola Salman, Raphaël Couturier, Chehab Ali, Lorca: lightweight round block and stream cipher algorithms for iot systems, *Vehicular Communications* 34 (2022), 100416.
- [212] Noura Hassan, Mohamad Noura, Ola Salman, Raphael Couturier, Chehab Ali, Efficient & secure image availability and content protection, *Multimed. Tool. Appl.* 79 (2020) 22869–22904.
- [213] Noura Hassan, Ola Salman, Chehab Ali, Raphael Couturier, Preserving data security in distributed fog computing, *Ad Hoc Netw.* 94 (2019), 101937.
- [214] Hassan N. Noura, Reem Melki, Chehab Ali, Secure and lightweight mutual multi-factor authentication for iot communication systems, in: 2019 *IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*, IEEE, 2019, pp. 1–7.
- [215] Reem Melki, Hassan N. Noura, Chehab Ali, Lightweight multi-factor mutual authentication protocol for iot devices, *Int. J. Inf. Secur.* 19 (2020) 679–694.
- [216] Cheng Huang, Jiaxuan Han, Xing Zhang, Jiayong Liu, Automatic Identification of HoneyPot Server Using Machine Learning Techniques, *Security and Communication Networks*, 2019, p. 2019.
- [217] Ons Aouedi, Kandaraj Piamrat F-bids, Federated-blending Based Intrusion Detection System, *Pervasive and Mobile Computing*, 2023, 101750.
- [218] Aine MacDermott, Thar Baker, Shi Qi, Iot forensics: challenges for the iot era, in: 2018 9th *IFIP International Conference on New Technologies, Mobility and Security (NTMS)*, IEEE, 2018, pp. 1–5.
- [219] Li Yang, Ruinong Wang, Yuanzheng Li, Meng Zhang, Chao Long, Wind power forecasting considering data privacy protection: a federated deep reinforcement learning approach, *Appl. Energy* 329 (2023), 120291.