Blockchain-empowered federated learning based solutions for Internet of Things security, privacy, and performance

Panagiotis Papadopoulos

Supervisor(s): Mauro Conti, Chhagan Lal

EEMCS, Delft University of Technology, The Netherlands

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
January 28, 2023

Name of the student: Panagiotis Papadopoulos
Final project course: CSE3000 Research Project
Thesis committee: Mauro Conti, Chhagan Lal, Jorge Martinez Castaneda

Abstract

The Internet of Things (IoT) is a rapidly growing technology that connects millions of devices together. However, as more devices connect, the importance of ensuring security, privacy, and performance becomes paramount. Training performance is affected without proper protocols in place, and devices can get compromised. This research focuses on how to enhance IoT security, privacy and performance using federated learning and blockchain. We will first identify the current challenges concerning those three metrics for IoT. Then, we will introduce federated learning and blockchain and explore how to integrate them. Next, we will address a set of related work performed on the field through a series of surveys. Keeping those surveys in mind, we present and compare several novel solutions for various IoT applications that attempt to provide solutions to enhance IoT security. We complete this study by discussing the potential of those solutions, as well as their challenges and then we highlight possible directions for future research in this booming field.

1 Introduction

The Internet of Things (IoT) has evolved into a network of billions of devices around the world that are now connected with each other, all collecting and sharing data [1]. IoT not only provides services for information transfer, analysis, and communications, but also allows for independent operation, coordination, and interaction with other systems [2]. The IoT has changed the lifestyle of human beings, and our society is moving towards always-connected systems. However, the broad spectrum of beneficial IoT applications is accompanied by a wide set of malicious applications too [3]. IoT applications are able to spread at such unprecedented rates, partially due to their strict constraints on resources, power, and storage. Such constraints allow for vast growth, but also hinder security, privacy, and performance.

In the spirit of future-proofing the IoT at the dawn of the 5th and 6th generation of cellular networks [4] [5], different technologies have been proposed. A promising solution is seen in federated learning. While traditional machine learning approaches need to utilize a centralized server to aggregate the training data of all the nodes [6], the training process in federated learning is a distributed process. It allows training models without exchanging raw data while performing all operations on the IoT devices [7]. Another potential solution is the use of blockchain technology, which gained unprecedented popularity after the introduction of Bitcoin [8]. Blockchain can maintain an immutable log of transactions happening in a network and promises to help achieve true decentralization, security, and auditability.

The recent literature shows that both blockchain and federated learning can individually address issues in IoT, but it is also interesting, from an academic perspective, to assess them together. This is also the question this research paper attempts to answer: *"Blockchain-empowered federated learning based solutions for Internet of Things security, privacy, and performance"*. We will approach this main research question by posing and answering some sub-questions in the coming sections of this report:

- What are the main security, privacy, and performance issues that IoT is facing? [Section 2]
- Concerning security, privacy, and performance metrics, how can Federated Learning with Blockchain be used in IoT? [Sections 2 & 3]
- Identify the set of solutions where it is beneficial to use blockchain-empowered federated learning to improve one or more metrics related to security, privacy, and performance metrics. [Section 4]
- What are some possible challenges that arise with federated learning and blockchain solutions? [Section 5]
- Based on the challenges studied in the previous question, what are some possible ways future research can take? [Section 5]

This research contributes to the existing study of the topic by diving deep into the research related to the use of federated learning and blockchain in IoT, analyzing the state of the art, identifying challenges, and exploring future research directions. The rest of this paper is structured as follows: First, in section 2, we present some background on the current issues that IoT faces, and we explain federated learning and blockchain. In section 3, we study surveys that concern the application of federated learning and blockchain for IoT, considering security, privacy, and performance. Section 4 is the core of this research where we study a set of proposed solutions that use blockchain and federated learning to improve those factors. Section 5 discusses our findings, addresses the remaining challenges, and mentions possible ways research can take in the future. In section 6 we draw some conclusions, and lastly, in section 7, we reflect on the ethical aspects of this research and consider its reproducibility and integrity.

2 Background

In this section, we will briefly explain topics that are important for the research performed, as it is important for the readers to get familiar with the terminology first. We will first take a look at IoT's advancements today. Then, we will explore blockchain. What is it, how does it work, and how can it be applied to IoT. Lastly, we will cover federated learning and its novelties. We will look into its architecture, and how can it be applied in the context of IoT to enhance its security, privacy, and performance.

2.1 IoT Security, Privacy, and Performance Concerns

The term suggested in 1999 by Kevin Ashton, an MIT associate, the *Internet of Things* is a blooming field that grabs the attention of the scientific community, as well as the general public. What started as a network of objects with radio-frequency identification (RFID) technology [9], has evolved into a global network of interconnected devices that mostly

act independently to monitor, sense, and report. Applications of it include but are not limited to, smart devices, home applications [10], and healthcare devices [11]. The undeniable benefits proposed by the advancement of IoT cannot be denied, however, they are followed by a series of flaws, concerning security, privacy, as well as performance.

We can look at IoT security through a series of prisms. Let us present some specific attacks and types of attacks that can exploit vulnerabilities of IoT nodes and networks.

- *Attacks against confidentiality and authentication*: Also known as interception attacks, they allow unauthorized access to IoT nodes by intercepting communications. This can be done, for example, in the form of eavesdropping attacks, or node identity theft. Eavesdropping is a type of attack in which a third party manages to listen to communications between nodes. An attacker can gain access to identifying information of the node and replicate it to enter the network. Such attacks lead to a loss of trust in the network, as nodes do not have a way to identify of the sender.
- *Attacks against data integrity*: This type of attack aims to decay data, or to fabricate poisoned data. False data injection attacks are considered one of the most threatening cyber attacks for smart grids [12]. As the false data in the network is used in the set for training and testing the model, the accuracy is decreased, and decision-making is affected which can lead to power outages, in the example of smart grids, or even the loss of life [13].
- *Attacks against availability*: The most common attack against availability in IoT is denial of service. Many IoT devices lack basic security protocols. They very often contain backdoors and manufacturers do not enforce security standards. They also have easily exploitable passwords, that remain mostly unchanged by the owners, and, on top of those, they are continuously connected, therefore susceptible to attacks at any point in time.

2.2 Blockchain

Blockchain, first introduced through Bitcoin in 2008 [14], is a tamper-proof distributed ledger of transactions [15]. Unlike standard data storing techniques, such as databases, that store data in centralized servers, blockchain is a peer-to-peer network, where everyone stores the data locally.

In short, a blockchain consists of a set of blocks that are linked together in a linked-list-alike manner. The first block is called the genesis block and all remaining blocks append after it. Exactly as in a linked list, the blocks are divided into a block header, which contains information about the previous block, and a block body that holds the data of the block [16].

We will now discuss some key features of blockchain that will help us understand its importance:

- Traceability: Every operation is recorded in the ledger, which is available to all participants.
- High Availability: Participants can join and leave the system at any time, as there will always be other participants available that ensure the operation of the blockchain.
- Decentralization: In some blockchain implementations a central authority is eliminated, and in others, it has minimal participation.
- Persistency: Every transaction is verified and stored in every participant's copy of the ledger, therefore making it computationally very expensive to alter.

Blockchain's key features are what make it a great fit for developing IoT Security. Using them as foundations, we can reason about the potential benefits of using blockchain in IoT networks. Blockchain is a standardized and distributed way of storing and accessing data, that can bridge the gap between devices and increase interoperability. It is also resilient, as it is a distributed technology with numerous copies of its set of data, and it eliminates the single point of failure concerns of centralized architectures. Security is also a potential benefit of the use of blockchains, as they are immutable and traceable ledgers. They are resilient against data modification attacks due to their constant data verification.

2.3 Federated Learning

Federated Learning has emerged as a distributed approach to train machine learning models by coordinating multiple devices with a central server without sharing raw data [17].

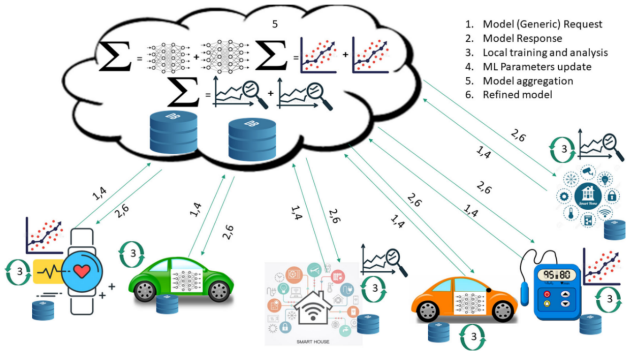

Figure 1: Federated Learning in IoT [18]

The general process of applying federated learning in IoT can be viewed in Figure 1. A shared global model is trained under the coordination of a central server, from a federation of participating devices. The clients first receive the global model and perform machine learning tasks using their resources and local data. Then, the clients upload their new parameter values to the server which aggregates them in order to train the global model. One of the most used algorithms to aggregate the nodes' local models is Federated Averaging [19]. In the last step, the server sends the refined model back to the clients.

Federated Learning is able to offer unique benefits over traditional machine learning approaches. It does not share raw data over the network, as those are not needed to train the global model, therefore ensuring data privacy. Furthermore, federated learning allows for a more decentralized approach to machine learning, as it can train models across multiple devices. Lastly, federated learning facilitates access to het-

erogeneous data, as its clients are spread over a large area [20].

2.4 Integrating Blockchain with Federated Learning

There are many paths we can take to argue why blockchain should be used together with federated learning. In this paper, we can discuss it in the context of using blockchain to mitigate some major federated learning issues. Despite its benefits over traditional machine learning approaches, federated learning does not come without any setbacks. We can identify some major issues and propose why blockchain has the potential to mitigate them.

Federated learning may be a distributed approach to machine learning, but it still requires a central authority to generate the global model and aggregate the individual results of the participating nodes. Such an architecture is prone to man-in-the-middle attacks and presents a single point of failure. Transactions in a blockchain can be recorded without the need for a central authority, and validity is maintained through constant verification. This decentralized nature of blockchain networks makes them resistant to tampering and can help solve the above-mentioned issue of federated learning. Furthermore, in federated learning, there is no reward mechanism in place. Therefore, nodes with more resources, that can contribute more to the model training, have no incentive to participate more actively. Blockchain can be used to track the computational performance of each participating client, and to reward the clients for their contributions using a digital token for example.

A generalized high-level architecture of blockchain-enabled federated learning can be seen in Figure 2. We use some nodes as participants to train the model, and some nodes as miners, that are responsible for maintaining the ledger.

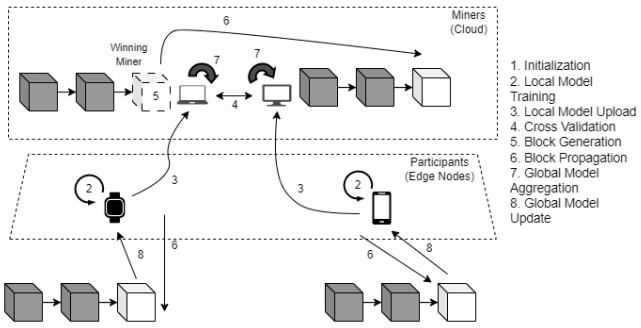

Figure 2: Integration Architecture of blockchain-enabled FL

3 Related Research

As discussed earlier, federated learning and blockchain can bring unique new solutions to battle IoT vulnerabilities, and improve performance. We will, now, provide a review of several recent surveys on the topic of blockchain-enhanced federated learning. We will also provide a table that summarizes the contributions of the surveys.

Blockchain-enabled Federated Learning: A Survey [21] The authors of this survey first cover the basics of federated learning and blockchain very well, allowing the readers to get familiar with the argumentation. The authors build up on the benefits of federated learning to show its potential and provide three key drawbacks of it. Then, they describe blockchain technology, and why can it be beneficial. That way, the authors pave the way for argumentation towards a solution that combines the two technologies. They use the benefits of blockchain to propose solutions to the key challenges of federated learning.

They then provide a detailed listing of leading research on blockchain-enhanced federated learning. They provide the readers with specific examples of studies that focus on security and privacy, as those two terms are often, wrongfully, used interchangeably in classifications of IoT vulnerabilities. Lastly, the authors take a critical look at open issues of that still need to be addressed, such as consensus algorithms, attack-proof performance, privacy preservation during model sharing, and some trade-offs that need to be made. Although learning performance is the primary factor to consider in federated learning systems, other parameters such as convergence efficiency, privacy protection, or energy consumption need to be considered. The authors list them as trade-offs against learning performance, listing possible ways of improvement for each pair of parameters.

This survey can be noticeably valuable when studying methods of combining federated learning with blockchain technology. The authors first present how federated learning can be used, but also what are its limiting factors, and then attempt to show how those issues can be mitigated with the assistance of blockchain, by providing examples of studies. On top of that, they try to cover security, privacy, and performance individually, and help the reader have a clear view of their respective issues, but also which of them optimise certain solutions. All in all, in this survey the authors use federated learning as a basis for approaching IoT vulnerabilities and utilize blockchain features to optimize the federated learning architectures.

Blockchain-empowered Federated Learning: Challenges, Solutions, and Future Directions [22] The authors of this survey also start by defining federated learning and blockchain, and, like the previous survey, this one puts blockchain forward as an enabling factor for more well-rounded federated learning, where it tries to address three key challenges of federated learning. This survey, too, focuses on incentive mechanisms and model security but differs on the third issue and studies system heterogeneity instead. The authors present a variety of solutions proposed for each issue, including both with and without the use of blockchain. The authors then present some IoT applications where blockchain-enabled federated learning can be used: industrial internet, intelligent transportation, smart healthcare, and wireless networks. They list key concerns for each of them and describe how the proposed research can address those concerns. Then, the authors take the above-mentioned application-specific solutions and classify them into three classes, based on the level of integration of blockchain with

Survey	Authors	Summary	Year
Blockchain-enabled Federated Learning: A Survey [21]	Qu, Uddin, Gan, Xiang, Gao, and Yearwood	The survey presents blockchain-enabled federated learning as a proposal to mitigate federated learning's limitations, in terms of security and privacy.	2022
Blockchain-empowered Federated Learning: Challenges, Solutions, and Future Directions [22]	Zhu, Cao, Saxena, Jiang, and Ferradi	This survey studies blockchain's potential to mitigate key IoT challenges of federated learning. It presents various system models, emerging applications, design challenges and proposed solutions.	2022
Recent Advances on Federated Learning for Cybersecurity for Internet of Things [18]	Ghimire and Rawat	This survey focuses on IoT security using federated learning and security for federated learning itself, through a detailed presentation of applications, while also considering performance through a set of metrics.	2022
Federated Learning for Internet of Things: A Comprehensive Survey [20]	Nguyen, Ding, Pathirana, Seneviratne, Li, and Poor,	The survey illustrates federated learning as an enabler for a wide range of IoT services and a listing of the use of federated learning in different IoT applications.	2021

Table 1: Summary of related surveys

federated learning, namely coupled, decoupled, and overlapped. They compare the advantages and disadvantages of each system model and conclude that there is no best overall solution, and models should be chosen based on the application scenario.

The authors continue by listing additional challenges in blockchain-enabled federated learning and proposed solutions. According to the authors, there is still a need for refinement of the incentive mechanisms, more appropriate selection of clients, more efficient consensus mechanisms, and better model security and data privacy. Lastly, the authors list some unresolved problems in blockchain-enabled federated learning. Based on the investigation of existing research, the authors suggest that there are still performance defects, difficulty in choosing learning parameters, the flexibility of incentives mechanisms and member selection, and privacy and security levels.

This survey does not just list practical applications and technical solutions but dives deep into systematically studying the potential of using blockchain for federated learning, approaching three significant challenges of federated learning and showing how blockchain can be used to mitigate them. They utilize specific IoT applications to bridge into listing the different integration ways, which also acts as an additional metric to inform the readers of the popularity of the models, with decoupled being the most widely used one. Another valuable point of this survey is the explicit mention of privacy, security, and performance issues, as well as proposing solutions. A potential downside of this paper is the absence of key benefits of federated learning and blockchain which would help the readers better understand the motivation behind moving towards blockchain-enabled federated learning. Furthermore, the survey lacks an overview of the challenges and solutions classified with the three proposed federated learning challenges. Such an overview would give the readers a better high-level picture, as the survey is built around those challenges as well.

Recent Advances on Federated Learning for Cybersecurity and Cybersecurity for Federated Learning for Internet of Things [18] This survey focuses on the security aspect of IoT applications, presenting the potential benefits of using federated learning, as well as its limitations. The authors begin with a background and comparison of centralized, distributed, and ultimately federated learning. That way they highlight the high-level value of federated learning for the readers before they discuss its applications.

The core of this survey is a detailed projection of federated learning applications for IoT security. The authors provide a listing of addressed issues, whether the proposed solution utilizes federated learning, and what is the domain of application. They then flip the coin and critically reason about the security of federated learning itself, and present attacks and countermeasures. The authors present the success of Federated Learning as a series of metrics that need to be achieved. They review three performance metrics of federated learning, namely accuracy, latency, and resource constraint and what improvements on them does research suggest. The authors, then, review some data sets and sets of machine learning models and algorithms to give the readers information about the trends in research. They close the survey with open challenges and future research directions. They underline that despite its potential and strong attention from the scientific community, federated learning is still in its infancy and needs further studying before it is fully applicable to IoT environments.

This is a unique survey that covers security and privacy, but also performance in such a way that it explains how the three correlate with each other. In other words, how do performance metrics affect privacy and security. It also mentions the potential of blockchain to address problems of federated learning, such as single points of failure, through some research that utilizes blockchain. However, it does not study blockchain extensively.

Federated Learning for Internet of Things: A Compre-

Papers	Security	Privacy	Performance	BC	Application
IOTFLA [23]	●	●	○	○	Smart Home
FL IDS IoV [24]	●	○	●	○	IoV
RAFeL [25]	●	○	●	○	Malware Detection for IoT
(POSTER) Decentralized FL Anomaly Detection [26]	◐	○	◐	○	General IoT
Blockchain & Federated Learning Distributed Architecture for IoT [27]	●	●	◐	●	General IoT

Table 2: Comparison of research papers studied

hensive Survey [20] The authors begin with an introduction to federated learning. They present some classes of federated learning classified by data partitioning, namely horizontal, vertical, and federated transfer, and by networking structure, namely centralized and decentralized. They also briefly talk about IoT and how promising is the use of federated learning for IoT. The authors then illustrate a holistic discussion of the potential of federated learning in enhancing various IoT services, such as data sharing, data offloading, and localization. They mention the challenges of the traditional approach and then provide an extensive discussion of the use of federated learning in a wide variety of IoT applications, including smart healthcare, smart transportation, UAVs, smart city, and smart industry. The authors identify that federated learning can benefit all those applications, but the solutions proposed have limitations, such as lack of scalability, convergence latency, data loss, and lack of simulations.

This survey covers the potential of federated learning for IoT exhaustively. The authors go into depth to show how many ways federated learning can be used to optimize IoT networks. Although the survey does point out issues in proposed solutions that use federated learning, it does not extensively cover the issues of federated learning itself, to give a more spherical view to the readers.

4 Review and comparison of state-of-the-art solutions

In this section, we will take a deeper look at the current state-of-the-art of security, privacy, and performance-aware solutions for IoT. We will study and review some novel architectures and models proposed, regarding their completeness, strong points, and lacking ends. We will also provide a table of comparison of some key features of the papers, such as the context they cover (security, privacy, performance), use of blockchain, and which application of IoT they are targeting.

IOTFLA: A secure and private smart home architecture [23] This paper proposes a novel smart home architecture for a secure and private IoT network using federated learning. It approaches this problem by breaking it down into several distinct building blocks which solves individually.

The paper first breaks down existing architectures for IoT. There are three classic layers, the perception layer, where the nodes reside, the network layer, where all processing and transmission happens, and application, where data gets presented to the users. Then, the paper describes the most important security and privacy concerns that smart home IoT is facing. This is particularly important to make the distinction between privacy and security issues. The paper mentions, among others, IoT's limited resources, security integration, insecure programming, and IoT networks being wide heterogeneous ecosystems, as major security challenges. It then mentions data confidentiality, data anonymization, and access control as major privacy issues. Lastly, the authors browse and select fitting secure data aggregation protocols.

The paper then presents the proposed solution, first as an overview, and then adding core components, security and privacy protocols, and then explaining the benefits of the novel components, namely federated learning, and a secure data aggregation protocol. The architecture consists of smart devices that sense data, a HUB that acts as the "brain" of the system and controls data flows, an intrusion detection system (IDS) that detects suspicious behaviour, and a database that stores sensitive data and security rules. The HUB also acts as a gateway that links the smart devices to the internet, and the IDS is centralized on the guardian, a dedicated computer.

The paper proposes a solid solution to enhance security and privacy for IoT smart home networks. The authors set the requirements and build up an architecture that step by step can solve the problems mentioned. The use of federated learning can improve latency, transfer costs, and privacy. The authors also reason about the different use case applications that federated learning can help within the context of the smart home. They present both the benefits as well as the current limitations. This is important for a variety of reasons:

- To draw the potential benefits and value gained from the technology in a practical and relatable way.
- To motivate the need for additional research in the field. By highlighting the potential benefits and problems that need to be addressed, it can demonstrate the importance of continuing to explore and improve the technology.
- To guide the design and implementation of using the federated learning approach in a smart home.

However, a noticeable issue in this paper is the lack of

implementation. The authors do share some thoughts regarding application scenarios, but they have not performed a proof-of-concept experimental design to illustrate the validity of the system. All in all, this is a great proposed architecture, with novel items, such as secure data aggregation, but it lacks implementation and testing.

Federated learning based IDS approach for the IoV [24] This paper proposes an SDN-driven IoV architecture utilizing collaborative nodes' trustworthiness. It discusses other IDS implementations, both utilizing the advantages of artificial intelligence and some that have leveraged emerging technologies like blockchain, cloud, or edge computing.

It then gives a high-level overview of the proposed IDS model, describing the different types of nodes participating. Those involve a cloud server at the top layer, controller nodes (SDN) on base stations, and then vehicles, certification authority nodes, and Roadside Units (RSUs) at the bottom layer. Each SDN controller collects and processes flow information of vehicles in the network and RSUs under its managed zone. Then, the SDN controllers and the cloud server train an IDS model for the network. Finally, the SDN controllers independently monitor the network with the trained model.

Then, the paper defines the metrics used to establish the trust level of the network. The proposed IDS model uses a trust estimator to assess nodes for maliciousness and derives features such as stranger nodes, traffic flow, and node properties to reflect the trust characteristics of IoV nodes. These features serve as input for the classification module which uses Federated learning to evaluate whether the nodes are malicious or not.

This paper presents a novel Intrusion Detection System (IDS) model that combines Federated learning, trust, and Software-Defined Networking (SDN) for efficient model training. The authors also attempt to take performance into account in their model, with the use of SDN controllers to monitor specific areas of the network. On top of that, the authors provide a simulation to evaluate the proposed model, with promising results using CNN (Convolutional Neural Networks).

RAFeL: Robust and Data-Aware Federated Learning Malware Detection [25] This paper proposes a framework that integrates a customized encoding algorithm with a novel Federated Learning-based defence technique for IoT networks.

The framework ensures that the aggregated global model is free of tampered data. The way it achieves that is by combining state-of-the-art FedProx aggregation algorithm [28] with RAPID (Robust and Active Protection with Intelligent Defense). The authors utilize the distribution patterns of IoT devices, which consist of a set of indicative features, to identify potentially manipulated data. In short, if a distribution pattern is anomalous it is a sign of a malicious device, whereas non-tampered devices will show a similar distribution of training data and local model updates. If the Federated Learning server detects a similar distribution, it executes FedProx aggregation. If it detects anomalies, it executes the RAPID defence mechanism. The key challenge is to extract the indicative features that change distribution (and make it anomalous) when a device is tampered with. The authors propose an algorithm that detects and updates indicative features, as well as flagging users as malicious.

Furthermore, the paper proposes a performance-aware bitwise encoding to reduce communication overhead. It takes advantage of splitting the full precision data into chunks and applies a customized encoding technique to each chuck. The motivation behind this technique is that different portions of values (weights, activations, gradients) have different characteristics in machine learning and deep learning, and most significant bits tend to be very sparse.

This paper offers a solution that makes the network more secure by mitigating manipulated data in server aggregation and also improves performance by reducing communication overhead with a customized compression technique, which also ensures security. It is also worth mentioning that the authors validated their framework through an experimental setup, where the encoding technique appears to have a significant contribution in reducing communication cost, and the model shows resilience against data manipulation attacks.

POSTER: Decentralized Federated Learning for IoT Anomaly Detection [26] This paper proposes an anomaly detection method based on decentralized federated learning. The model is based on the principles of decentralization through peer-to-peer communication, and the local training and storing of data that federated learning provides. All this, adds a more efficient decentralization algorithm that allows for non-full peer-to-peer model transmission.

The authors do provide an experimental setup which shows that the method does not lack in performance against centralized federated learning architectures, but it takes some time to catch up due to the fact that decentralized nodes take some time to catch up with the model trained. On the other hand, the setup is just 100 epochs and 8 clients big, so it is not truly representative.

Another key downside of the paper is the lack of background information. The authors do not describe federated learning and the concept of anomaly detection or present its benefits and drawbacks as a method over other security solutions. The paper does not motivate about the main issues IoT is facing, in order to present the proposed model as a solution. The paper also lacks in reviewing the latest related research relevant to the topic. It provides a small overview of centralized and decentralized solutions with a small mention of their drawbacks. Lastly, the novelties of the system design are not clearly described, nor is its application to real-world IoT applications.

Blockchain and Federated Learning-enabled Distributed Architecture for IoT [27] This paper provides a proposal for a novel architecture for secure IoT networks. The authors cover the motivation for the integration of federated learning and blockchain as a complementarity relationship, where blockchain is used to address the issues of federated learning. The authors identify a need for a lightweight authentication scheme, auditable local model updates, a feedback-based reward system, and the importance

of cyber resilience to prepare for and recover from cyber-attacks. These issues are presented as important considerations for the system design.

The system is a distributed multi-layered approach, composed of four modules: local nodes, edge nodes, a blockchain-enabled fog network, and a core distributed cloud. Local nodes use their own data and resources to train the model, edge nodes aggregate and validate the global model, fog network stores the global models permanently and the core cloud handles authentication. Furthermore, the paper proposes a reward scheme to counter free-ride attacks and make the blockchain part more resilient. The system is based on the successful participation of the nodes. In case the participation is not above a set threshold, the node gets penalized.

The paper puts forward an experimental setup to validate the proposed system, using real-world data, but on a relatively small scale. However, the analysis shows that the proposed model is stable and accurate in attack-free environments, and it also succeeds in mitigating attacks, such as poisoning attacks. The reward system is also tested successfully against free-riding attacks.

5 Discussion and Future Research Directions

In this section, we will reflect on the findings of studying state-of-the-art solutions for IoT security. We will compare those findings with what the related surveys suggest and highlight the most interesting elements. Based on the above, we will then discuss where research can focus in the future.

5.1 Discussion on findings

The field of IoT security cannot be considered novel, but it is also not mature yet. Many IoT devices have significant security vulnerabilities, and there are ongoing efforts to improve the security of IoT devices and networks. Federated learning and blockchain, on the other hand, are very new technologies, that are still under active research. While they have shown promising benefits and the potential to address numerous issues related to IoT security, there are still many challenges and limitations that need to be addressed before they can be widely adopted and considered mature.

Federated learning's potential

The surveys studied take different ways of presenting the topic of IoT security, but they also agree on the greater picture. They all put federated learning forward as a promising solution to enhance IoT security, privacy, and performance. Some surveys also discuss the potential of blockchain to mitigate some concerns about federated learning, but we will discuss more on that when we talk about state-of-the-art solutions.

Most surveys also build up the motivation of using federated learning, by listing the challenges of the traditional centralized or decentralized machine learning methods and how federated learning can mitigate them. Some surveys also mention integration techniques, which are very important to highlight the insights and give a high-level picture to the readers. The studied surveys that focus on blockchain-enhanced federated learning [22] [21] follow a similar approach. They present the drawbacks of federated learning as motivation to utilize blockchain technology, with arguments such as centralized processing, lack of incentive mechanism, and robustness. Then, they present a high-level system architecture.

Critical assessment of federated learning and challenges in implementation

It is important to not only illustrate the benefits of using a technology, such as federated learning, but to also take a step back and critically assess the security of the technology itself. Otherwise, the proposed solutions suddenly may create more problems than they solve. Surveys such as [18] promote this, by listing attacks and countermeasures against federated learning, or [21], by listing persisting issues related to blockchain, such as consensus algorithms and privacy preservation. On the contrary, some surveys, such as [20], do not go into depth about the challenges of federated learning and only focus on its benefits and the problems it solves. That way the readers do not get a spherical view of the technology and might create a wrong impression.

A key part of all the surveys studied is a discussion on the challenges that lie ahead on the road of implementing such models and applying them efficiently. A fundamental issue that is covered in all surveys is performance, both in training the model as well as training defence mechanisms. IoT has inherently low resources available, making the distribution of calculations a very hard puzzle. The surveys that mainly focus on federated learning itself mention issues such as poisoning attacks of the aggregation server, resource management, and communication overhead. On the other hand, the surveys that propose blockchain-enhanced federated learning focus on what we could identify as mostly blockchain-related concerns, such as lack of incentive mechanisms, model security, and privacy concerns.

Security, Privacy, and Performance of proposed solutions

After studying several state-of-the-art proposed solutions for various IoT applications and problems, we can now discuss some common issues and key points covered regarding security, privacy, and performance.

The papers studied take security and privacy into account. It appears that they all aim to enhance security, in one way or another, but not all of them explicitly aim for privacy. IOTFLA [23] strongly focuses on privacy. The authors discuss separately about security and privacy issues that IoT is facing and reflect on how the proposed model addresses them.

A fundamental constraint of IoT nodes is the lack of resources which limits the calculations that are possible to be performed on the nodes. This raises the question of whether any architecture or model proposed takes performance into account and whether the overhead calculations affect the overall performance. The authors of RAFeL [25] do take performance into account. They review the performance of their encoding algorithm and make sure that not only it achieves higher compression rates, but it also reduces the communication overhead without impacting the model's performance. In a similar manner, the authors of [27] set as a design requirement to have a "light weighted authentication scheme". Furthermore, the authors of [26] propose a decentralized model to improve security, but also improve the decentralization algorithm for better performance. The IDS for IoV paper [24]

proposes a similar, in essence, model to the poster paper [26], but goes much more into depth. They attempt to improve performance with the use of SDN controllers to monitor specific areas of the network. In other words, they try to spread the costs of calculation over the network. On the other hand, the authors of IOTFLA [23] focus on security and privacy, but not on performance. They mention that they take resource limitations into account, but they do not provide any context on how this is done, and they also assume that the IoT devices in the network have the resources available to perform the necessary calculations.

Use of blockchain in proposed solutions

The surveys studied showed unanimity in proposing federated learning to enhance security, privacy, and performance, but they varied in considering blockchain. The same trend can be seen in the papers studied. The papers discuss challenges with traditional centralized machine learning methods [23] and use them as grounds to propose federated learning, but also discuss federated learning's own issues, such as communication overhead and data poisoning in [25], and use them as design requirements for their model. Papers that propose anomaly detection systems/IDS rather than architectures, such as [24] and [26], do not focus on the drawbacks of traditional methods, but on the benefits of the novel designs. Only one of the papers proposed a blockchain-enhanced federated learning architecture. The authors of [27] use model poisoning attacks, secure aggregation, malicious nodes, and free-ride attacks as motivations to integrate blockchain with federated learning. As this is a vastly broad approach, the authors tried to limit the scope and set the system requirements through a series of unresolved issues, such as auditable local model updates, and contribution-based rewards. It is notable that the authors of [24] mention blockchain's potential to provide safer cooperative learning, but does not consider it in the model. It is clear from the study that most solutions do not utilize blockchain yet. That could be due to the complexity of designing and maintaining purely decentralized solutions using blockchain, or due to scalability constraints. Federated learning involves training models on a large number of decentralized devices, and blockchain networks can have scalability issues when handling a large number of transactions.

5.2 Future Research Directions

After studying surveys and comparing state-of-the-art solutions focused on enhancing IoT security, privacy, and performance using federated learning and blockchain, we can, now, identify some common unresolved issues on this blooming field and propose some future search directions.

Integration of Blockchain with Federated Learning

Research shows that blockchain is capable of alleviating some key issues of federated learning, but further research is needed. Federated learning is itself a complicated technology to apply, and combining it with blockchain makes the task even more difficult. A possible direction is towards systems that resemble the generic architecture we saw in Figure 2. The authors of [27] followed this generic design but added another layer above with fog nodes. This architecture uses blockchain to verify the local model results and store the global model safely. The authors of [29] also use this approach. Some other ways are to use blockchain to incentivise nodes with a reward scheme, or smart contracts [30].

Simulations and real-world implementation

Another critical factor is to design the model such as it can also be implemented in real-world applications. So far most of the proposed solutions studied at least provide some simulated scenarios to validate the model, but they make assumptions and have not been realized in scale. Federated learning and blockchain are still very new technologies, so heavy testing and simulation are needed too. The provided experimental setups that validate models could be scaled up substantially to simulate conditions closer to the real world and to measure the true efficiency of the models.

Security, Privacy, and Performance

Although federated learning can enhance security and privacy compared to traditional machine learning models, issues persist. IoT nodes remain relatively exposed, as they have limited resources available. It is still easy for a malicious party to gain control of a set of end nodes of an IoT network. This can lead to poisoned data, and backdoor attacks, compromising the model's accuracy. That is why research mostly focuses on anomaly detection. However, even with the best IDS systems in place, the very presence of a central aggregation server will always be a bottleneck to security. Innovative solutions are needed to preferable move to purely decentralized architectures, using blockchain for example. As this is a very complex and difficult operation, the alternative is to develop novel security solutions to shield the central server.

Lastly, one of the biggest challenges that federated learning for IoT has to solve is communication overhead. Huge numbers of participant nodes serving unbalanced data to the aggregation server result in reduced performance. Some possible solutions are better compression algorithms, such as the one discussed in [25] and the introduction of an incentive to filter the quality of the data.

6 Conclusion

In this paper, we conducted research on state-of-the-art of IoT security. Motivated by the ongoing issues of IoT security, we introduced two promising technologies, namely federated learning and blockchain. Using those two technologies, the aim of the research was to study federated learning and explore how blockchain can integrate with federated learning through state-of-the-art solutions. We presented those technologies and studied them through proposed solutions and related surveys. We can conclude that federated learning is a highly promising technology to enhance the security, privacy, and performance of IoT. It offers substantial benefits compared to traditional centralized machine learning models. However, as federated learning itself presents some challenges, blockchain is deemed a very good fit to integrate with federated learning and mitigate those issues. The technologies themselves, as well as their integration, are very popular topics around the scientific community, but there is still a long way to go until they can be applied efficiently in large-scale real-world scenarios.

7 Responsible Research

In this section, we will discuss the scientific integrity of this paper in the context of source gathering, and then its reproducibility as a literature study.

The core of this research has been the literature study and review of different applications of IoT Security that utilize blockchain and federated learning. For that reason, it is of paramount importance to use credible and reliable sources to gather information from. After the initial talks with our supervisor, we decided upon using IEEE Xplore and ACM as our repositories, due to the strict quality control the papers that are published there undergo. After an initial literature study and paper gathering, our supervisor checked the proposed papers and helped us finalize the list, further ensuring the quality of the papers studied. Lastly, the study of the related work was crucial in this research, to form a plan for the comparison of the research papers studied later. That is why a wide selection of surveys on the topic of IoT security and IoT security using federated learning and blockchain were studied. All in all, we believe that the initial careful selection of the work studied and the selection of a plethora of papers allowed us to conduct a solid scientific research.

Despite the fact that this research did not involve any experiments or results gathering, it is still important to consider reproducibility. For that purpose, all the references of the work that was studied to produce this research paper can be found below, in the style of IEEE. References of any diagrams or schematics that were used or studied can also be found below. The inclusion of the sources and the in text citations that link to relevant scientific papers are deemed enough for any future research effort to reproduce the conducted work.

References

[1] K. Shafique, B. A. Khawaja, F. Sabir, S. Qazi, and M. Mustaqim, "Internet of things (iot) for next-generation smart systems: A review of current challenges, future trends and prospects for emerging 5g-iot scenarios," *IEEE Access*, vol. 8, pp. 23 022–23 040, 2020.

[2] J. Lin, W. Yu, N. Zhang, X. Yang, H. Zhang, and W. Zhao, "A survey on internet of things: Architecture, enabling technologies, security and privacy, and applications," *IEEE Internet of Things Journal*, vol. 4, no. 5, pp. 1125–1142, 2017.

[3] I. Butun, P. Österberg, and H. Song, "Security of the internet of things: Vulnerabilities, attacks, and countermeasures," *IEEE Communications Surveys Tutorials*, vol. 22, no. 1, pp. 616–644, 2020.

[4] D. Wang, D. Chen, B. Song, N. Guizani, X. Yu, and X. Du, "From iot to 5g i-iot: The next generation iot-based intelligent algorithms and 5g technologies," *IEEE Communications Magazine*, vol. 56, no. 10, pp. 114–120, 2018.

[5] J. H. Kim, "6g and internet of things: a survey," *Journal of Management Analytics*, vol. 8, no. 2, pp. 316–332, 2021. [Online]. Available: https://doi.org/10.1080/23270012.2021.1882350

[6] J. Liu, J. Huang, Y. Zhou, X. Li, S. Ji, H. Xiong, and D. Dou, "From distributed machine learning to federated learning: a survey," *Knowledge and Information Systems*, vol. 64, no. 4, pp. 885–917, 2022. [Online]. Available: https://doi.org/10.1007/s10115-022-01664-x

[7] A. Imteaj, U. Thakker, S. Wang, J. Li, and M. H. Amini, "A survey on federated learning for resource-constrained iot devices," *IEEE Internet of Things Journal*, vol. 9, no. 1, pp. 1–24, 2022.

[8] L. AU Tseng, X. Yao, S. Otoum, M. Aloqaily, and Y. Jararweh, "Blockchain-based database in an iot environment: challenges, opportunities, and analysis," *Cluster Computing*, vol. 23, no. 3, pp. 2151–2165, 2020. [Online]. Available: https://doi.org/10.1007/s10586-020-03138-7

[9] L. AU Shancang, X. Li Da, and S. Zhao, "The internet of things: a survey," *Information Systems Frontiers*, vol. 17, no. 2, pp. 244–259, 2015. [Online]. Available: https://doi.org/10.1007/s10796-014-9492-7

[10] K. Xu, X. Wang, W. Wei, H. Song, and B. Mao, "Toward software defined smart home," *IEEE Communications Magazine*, vol. 54, no. 5, pp. 116–122, 2016.

[11] Y. Zhang, L. Sun, H. Song, and X. Cao, "Ubiquitous wsn for healthcare: Recent advances and future prospects," *IEEE Internet of Things Journal*, vol. 1, no. 4, pp. 311–318, 2014.

[12] Z. Hu, Y. Wang, X. Tian, X. Yang, D. Meng, and R. Fan, "False data injection attacks identification for smart grids," in *2015 Third International Conference on Technological Advances in Electrical, Electronics and Computer Engineering (TAEECE)*, 2015, pp. 139–143.

[13] J. Xin, V. V. Phoha, and A. Salekin, "Combating false data injection attacks on human-centric sensing applications," vol. 6, no. 2, jul 2022. [Online]. Available: https://doi.org/10.1145/3534577

[14] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," *Cryptography Mailing list at https://metzdowd.com*, 03 2009.

[15] C. Natoli, J. Yu, V. Gramoli, and P. Veríssimo, "Deconstructing blockchains: A comprehensive survey on consensus, membership and structure," 08 2019.

[16] L. Xu, N. Shah, L. Chen, N. Diallo, Z. Gao, Y. Lu, and W. Shi, "Enabling the sharing economy: Privacy respecting contract based on public blockchain," in *Proceedings of the ACM Workshop on Blockchain, Cryptocurrencies and Contracts*, ser. BCC '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 15–21. [Online]. Available: https://doi.org/10.1145/3055518.3055527

[17] J. Konečný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, "Federated learning: Strategies for improving communication efficiency," 2016. [Online]. Available: https://arxiv.org/abs/1610.05492

[18] B. Ghimire and D. B. Rawat, "Recent advances on federated learning for cybersecurity and cybersecurity for federated learning for internet of things," *IEEE Internet of Things Journal*, vol. 9, no. 11, pp. 8229–8249, 2022.

[19] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas, "Communication-efficient learning of deep networks from decentralized data," 2016. [Online]. Available: https://arxiv.org/abs/1602.05629

[20] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, and H. Vincent Poor, "Federated learning for internet of things: A comprehensive survey," *IEEE Communications Surveys Tutorials*, vol. 23, no. 3, pp. 1622–1658, 2021.

[21] Y. Qu, M. P. Uddin, C. Gan, Y. Xiang, L. Gao, and J. Yearwood, "Blockchain-enabled federated learning: A survey," vol. 55, no. 4, nov 2022. [Online]. Available: https://doi.org/10.1145/3524104

[22] J. Zhu, J. Cao, D. Saxena, S. Jiang, and H. Ferradi, "Blockchain-empowered federated learning: Challenges, solutions, and future directions," *ACM Comput. Surv.*, nov 2022, just Accepted. [Online]. Available: https://doi.org/10.1145/3570953

[23] U. M. Aïvodji, S. Gambs, and A. Martin, "Iotfla : A secured and privacy-preserving smart home architecture implementing federated learning," in *2019 IEEE Security and Privacy Workshops (SPW)*, 2019, pp. 175–180.

[24] A. Hbaieb, S. Ayed, and L. Chaari, "Federated learning based ids approach for the iov," ser. ARES '22. New York, NY, USA: Association for Computing Machinery, 2022. [Online]. Available: https://doi.org/10.1145/3538969.3544422

[25] S. Shukla, G. Kolhe, H. Homayoun, S. Rafatirad, and S. M. P D, "Rafel - robust and data-aware federated learning-inspired malware detection in internet-of-things (iot) networks," in *Proceedings of the Great Lakes Symposium on VLSI 2022*, ser. GLSVLSI '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 153–157. [Online]. Available: https://doi.org/10.1145/3526241.3530378

[26] Z. Lian and C. Su, "Decentralized federated learning for internet of things anomaly detection," in *Proceedings of the 2022 ACM on Asia Conference on Computer and Communications Security*, ser. ASIA CCS '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 1249–1251. [Online]. Available: https://doi.org/10.1145/3488932.3527285

[27] P. Kumar Sharma, P. Gope, and D. Puthal, "Blockchain and federated learning-enabled distributed secure and privacy-preserving computing architecture for iot network," in *2022 IEEE European Symposium on Security and Privacy Workshops (EuroSPW)*, 2022, pp. 1–9.

[28] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," 2018. [Online]. Available: https://arxiv.org/abs/1812.06127

[29] L. Cui, Y. Qu, G. Xie, D. Zeng, R. Li, S. Shen, and S. Yu, "Security and privacy-enhanced federated learning for anomaly detection in iot infrastructures," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 5, pp. 3492–3500, 2022.

[30] M. Ali, H. Karimipour, and M. Tariq, "Integration of blockchain and federated learning for internet of things: Recent advances and future challenges," *Computers Security*, vol. 108, p. 102355, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167404821001796