# Privacy Protection and Performance Enhancement in IoT Applications using Blockchain and Machine Learning Techniques

**Jeroen Janssen**[1]

**Supervisor(s): Mauro Conti**[1]**, Chhagan Lal**[1]

[1]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 29, 2023

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

## Abstract

**The Internet of Things (IoT) is producing significant amounts of data. Protecting this data from adversaries is therefore a prominent field of research. This paper conducts a review of the current state-of-the-art in the field of IoT integrated with Blockchain (BC) and Machine Learning (ML). The review focuses on the use of privacy and performance metrics to evaluate the effectiveness of these latest solutions. We first provide an overview of related works to have an understanding of what has been done. Then, we present five important privacy and performance metrics that have been used in the review. We then provide a detailed evaluation of state-of-the-art solutions. Finally, we identify and present open problems that need to be addressed in the form of future research directions. This paper provides valuable insights for researchers interested in improving privacy and performance in IoT applications, and opportunities for future research.**

***Keywords*—Internet of Things; Blockchain; Machine Learning; Data Privacy; Performance**

## 1 Introduction

The Internet of Things (IoT) is revolutionizing the IT sector and is stated to be the next most significant advancement since the invention of the Internet [1]. As such, IoT is expected to have a deep economic, commercial and social impact on our lives in the future [2]. It is projected that the IoT market will grow to approximately 75 billion devices by the year 2025 [3], resulting in five times more devices in the IoT market than in 2017.

Due to the growing IoT market, there is a significant increase in the data generated by IoT. The data is considered to be privacy-sensitive [4], since generated data in IoT (directly or indirectly) reflects people's behaviour, interests, lifestyle, and so forth. As such, protecting data from privacy violations is a necessity in the field of IoT.

Recent literature has shown that the combination of Blockchain (BC) and Machine Learning (ML) can improve the privacy of IoT data. Kumar et al. [5] presented a Blockchain-enabled Privacy-Preserving Access Control System (BPACS) for IoT data sharing and accessing data using ML, attempting to improve data integrity and solve data privacy challenges. Also, Kumar et al. [6] presented an intelligent BC framework that integrates BC with ML techniques to protect privacy in IoT. This research is promising and is a step in the right direction.

In this paper, we investigate how the integration of BC and ML impacts privacy and performance in IoT applications. We will do so by answering the following research question: *"How do the combination of machine learning and blockchain impact privacy and performance in IoT data management?"*. The final answer will provide valuable insight to those who want to implement a privacy secure and well performing IoT application. We will first describe the current privacy- and performance-related issues in IoT data management by performing a literature survey of related work in section 2.3. Next, we compare the current state-of-the-art work on this topic in section 3 and critically present their findings. Finally, we propose the design of a technique that addresses privacy- and performance-related issues in the current state-of-the-art techniques. Concretely, this paper provides a survey on the latest advancements in the field of IoT where BC and ML techniques are used.

The rest of this paper is structured as follows. In Section 2, we discuss the background information for BC and ML integrations in IoT data management. Additionally, list the related work to this research, and present their findings and potential limitations. In Section 3, we discuss the method used in this research along with the metrics we will use for review. Review of the state-of-the-art is done in section 3.2. A discussion about the state-of-the-art review along with future research directions is done in Section 4. Furthermore, in this section, we briefly mention in what way future research should be designed. In section 5, we will briefly discuss responsible research and ethics. Finally, a conclusion is given in Section 6.

## 2 Background

It is useful for a reader to understand the related topics to this research field of study. Therefore, we will discuss the background of relevant topics in the following sections. The core of this research is about IoT data management, so we will discuss its background in section 2.1. Next, we consider BC and ML integration techniques as they have the potential to both improve privacy and performance. Therefore, we will provide the background of such integration techniques in section 2.2. Lastly, in section 2.3 we will present related work done in the IoT with BC and ML techniques domain.

### 2.1 IoT data management

IoT describes a group of physical objects that have sensors, processing capabilities and other technologies in order to share data amongst a network [7]. This data is shared between the IoT devices by the use of their networking capabilities. Although the definition IoT has the word *internet* in its name, the devices are not necessarily connected to the internet. It might be that these devices are only connected to a local network, which is not accessible by the outside world.

The data that is collected on IoT devices is shared, as such other devices in the same IoT network benefit from having access to this data. Oftentimes, this data is transferred through an IoT hub or gateway to enable other connected devices to either analyze this data or take action. A visual representation of the data flow through an IoT system is presented in Fig. 1.

### 2.2 Blockchain-Machine Learning integration

Understanding Blockchain and Machine Learning integrations (BC-ML) begins with understanding what the individual components are. Therefore, we will briefly discuss a definition of both before discussing the integration between them and IoT.

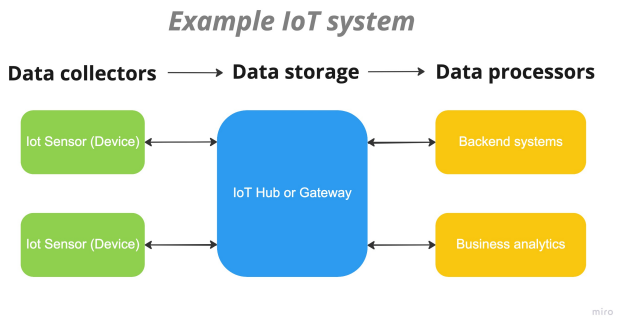


Figure 1: Example data flow through an IoT system

**Blockchain (BC)** is a distributed so-called ledger that is decentralized among a network. Data is recorded in this ledger when a consensus is reached by the majority of nodes in the network. Different consensus algorithms might define a majority in other ways [8]. Data is recorded in an immutable way without the inference of an authority, which is useful functionality for a network such as IoT. Since the BC is distributed among all entities in a network, all entities in the same network have undeniable access to this data. This technology is useful for this research area since it can appropriately address data privacy in IoT applications. On top of that, not only data privacy can be addressed by BC technologies but also user anonymity, with asymmetric cryptography to secure transactions between users [8]. A potential drawback of BC technology is with the fact that data is immutable, if a participant would want to withdraw it's data from the network this would be impossible.

**Machine Learning (ML)** is seen as a subdomain of artificial intelligence (AI). ML algorithms are able to learn from input data. Then, these algorithms can provide a sensible response on never before seen data, based on the previously acquired knowledge [9]. ML is often used in IoT networks to detect adversaries that have joined the network or are attempting to. Detecting adversaries is valuable in order to prevent such entities from doing harm to the network. In order for a ML model to be trained well good data is needed, if this data is not available the ML technology is limited.

**Integration with IoT** of BC and ML techniques can be done in various ways, in order to improve various privacy and performance issues with IoT networks. A few are listed below.

1. IoT applications often have issues with secure storage of data, for which an authority needs to be trusted. BC can provide a secure and decentralized way to store and share data, mitigating data risks in IoT applications.
2. Issues exist with scalability in IoT applications due to protocols that have been implemented [10]. Both BC and ML are scalable technologies. Implementing such protocols through the use of these technologies can mitigate these problems.
3. In a manufacturing environment with IoT applications, downtime of equipment can have high impact [11]. Therefore, predictive maintenance is implemented to prevent downtime. ML models can be trained to improve predictive maintenance when trained on data collected by sensors and devices in the IoT network.

As can be seen, there are both benefits and limitations when integrating BC with ML in IoT. Literature on this topic should mention both the positive and the negative impacts of their proposed solutions.

## 2.3 Related work

For the related work we have selected surveys that consider integrations of BC and ML with IoT. For these surveys, we present what the authors have done, the issues they have identified and present an objective overview of the work the paper has done. We also attempt to provide for the relevant works what is useful for researchers in the same area.

Wu et al. [8] in their paper **Deep reinforcement learning for blockchain in industrial IoT: A survey** focus on the application of BC technology in the Industrial Internet of Things (IIoT) sector. They survey existing literature that proposes solutions for the IoT and examine how these solutions can be adapted for the IIoT. This survey is valuable for researchers in the field of IoT as it provides a deeper understanding of the security and privacy risks of BC from the perspective of ML, which is useful in the design of practical BC solutions for IoT. Lastly, Wu et al. provide a visual representation of how an IoT network with Deep Reinforcement Learning (DRL) and BC should look, we provide this representation in Fig. 2. In their setup device to device (D2D) connections inside of the factory make use of BC for security of data. Devices that collect data can train DRL models and provide these trained models to the network.

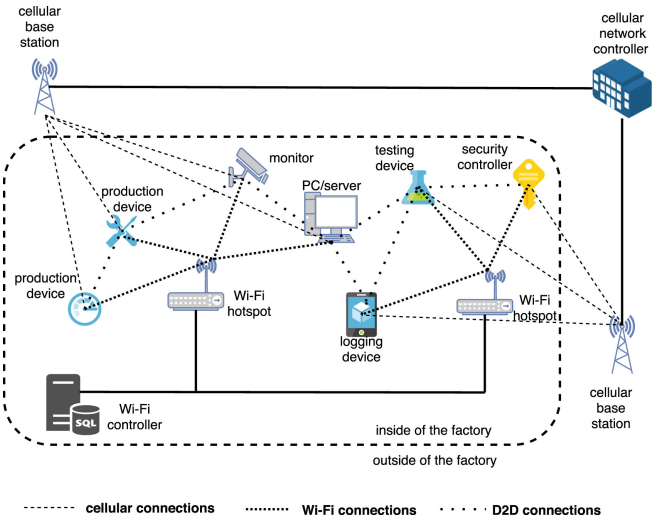


Figure 2: Proposed IIoT network with DRL and BC (Wu, 2021, page 10)

The authors first provide an overview of BC, Deep Learning (DL), and IIoT and then list the main contributions made by recent works in the field where DL is combined with BC for IoT or IIoT solutions. Some useful findings from these main contributions include proposals to improve node selection in BC-driven IoT using DL techniques, which aim to

improve performance. Additionally, the authors list contributions that make use of BC to improve data sharing capabilities and those that aim to improve the performance of DL.

Furthermore, the authors identify potential risks regarding privacy leakage and pseudonymity when analyzing data stored on BC ledgers with DL. They discuss performance issues caused by the consensus mechanisms used by BC and provide an overview of different consensus mechanisms and how DL affects blockchain privacy. They even touch upon data storage in BC, as the way data is stored can affect how well it can be kept private. During these parts the authors do a survey of current state-of-the-art research and address how, if at all, these state-of-the-art can address the threats and problems stated before.

In conclusion, the work done by Wu et al. can be used to improve privacy and performance in IoT networks that use BC for data storage. The potential threats and solutions listed by the authors, along with representative works, are valuable for creating the best version of an IoT network when implementing BC. However, it should be noted that the solutions mainly focus on privacy threats arising from the use of DL to analyze data in a BC IoT network and do not provide much information on improving IoT networks that make use of both DL and BC. Additionally, the structure of the paper could be improved, as the authors do not provide an explanation for the order in which they discuss potential threats and issues.

In their paper **Security and Privacy in IoT using Machine Learning and Blockchain: Threats and Countermeasures**, Waheed et al. [9] state that previous surveys have not looked into both security and privacy in the same paper. The authors first provide a thorough explanation of the methodology used to collect relevant papers for this survey, including search keywords, selection criteria, and search repositories. They then explain the study, which consists of providing a generic classification of IoT threats reported in recent literature, literature reviews of ML algorithms and BC techniques for IoT security and privacy, and highlighting research gaps in this literature. They also provide a taxonomy of the latest security and privacy solutions in IoT using ML and BC, identifying and analyzing the integration of ML and BC that strengthen security and privacy in IoT, and highlighting and discussing existing challenges to ML and BC techniques in IoT security and privacy to suggest future directions.

The work done by Waheed et al. is valuable as they identify threats to both security and privacy, and they conduct a comprehensive analysis of potential solutions. Furthermore, they provide clear tables with threats and solutions that can be used to improve security and privacy in an IoT network. The solutions provided by Waheed et al. are split into three categories: (i) solutions that make use of ML, (ii) solutions that make use of BC, and (iii) solutions that use an integration of ML and BC techniques. This makes their survey strong as they consider solutions not limited to one technique.

However, it should be noted that the authors discuss IoT security more than IoT privacy. Of the selected papers that contributed to this field, 14 out of 17 papers consider IoT security, but only 7 out of these same 17 papers consider IoT privacy. This trend continues with Waheed et al. discussing 4 papers on IoT security that use ML algorithms as opposed to no papers being discussed on IoT privacy that use ML algorithms, even though they presented one survey being published that considered IoT privacy with machine learning. Additionally, this can also be seen when making use of BC techniques, where 9 and 6 papers were selected for IoT security and IoT privacy, respectively. This is not necessarily a bad thing, since this representation of previous works might be an accurate representation of the work done in this field. However, one should consider researching additional works on privacy if one is looking for purely privacy related works. On a positive note, the authors consider existing solutions to the presented threats more equally. This representation of previous work may be an accurate representation of the work done in this field.

## 3 Proposed study

Following the analysis of related work in our field of research, we have composed a list of metrics by which we will review current state-of-the-art solutions. The structure of this section is as follows, we first present our privacy and performance metrics with their explanation in section 3.1. Afterwards, the actual review of the selected solutions is presented in section 3.2 to assess whether these proposed solutions are privacy-preserving and/or performance-enhancing. The review by privacy and performance metrics can be seen in table 1.

### 3.1 Privacy and performance metrics

The metrics used for review of the state-of-the-art are the following five.

**Confidentiality**; defined as preserving authorized restrictions on access and disclosure, including means for protecting personal privacy and proprietary information. In the field of BC-ML, IoT confidentiality is most often achieved by encryption of data. Because the nature of BC makes data accessible and disclosed to everyone in the network.

**Anonymity (Pseudonymity)**; the notion of a person (or entity) not being identifiable by name. In a BC framework, it is required for all transactions to be linked back to a user. Therefore, anonymity is achieved by using pseudonyms. In an open BC, this means that people who join are anonymous. Opposed to this, in a private BC where users have to be admitted, the entity that admits a user knows the identity of this user. This identity is not disclosed to other users on this private BC. Hence, they would have pseudonymity.

**Transparency**; in BC applications, transparency is related to the transactions done on the network and which entity does the computing of the new blocks for the ledger. Transparency in the ML field is about being able to know on what data the model is trained. Such that one can infer why a model predicts its outcomes.

**Consent Management**; in data applications, it is important that a user is in control over who gets access to their personal data. As such, we selected this metric for our reviews. Consent management in BC applications has to do with what data gets recorded on the ledger of the users. The ML aspect of consent management has to do with what data may be used to train and verify a model.

**Performance**; since performance can mean many things, such as in BC the time per block generated or in ML time per epoch during training, we will have this as a more generic metric. A full improvement is achieved when a paper proves improvement through a theoretical substantiation and measuring experiments. If either is missing, we denote that partial improvement is achieved, and if no notion of performance is done in the state-of-the-art research, we will note that no improvement is gained.

### 3.2 State-of-the-art evaluation

The below evaluations of the state-of-the-art are structured as follows. First, we present the problem addressed in the IoT field. Then, the proposed solution along with the steps taken. We furthermore note the method used in the paper to evaluate the proposed solution. Lastly, we present our review with the positive points of the proposed solution and why it is useful for IoT applications. Opposing potential weaknesses or limitations the proposed solution might have.

**Blockchain-based Auditable Privacy-Preserving Data Classification for Internet of Things**, in [12] Zhao et al. state existing issues in IoT data with privacy preservation. Existing approaches usually make use of a designated converter interacting with a semi-honest verifier. Zhao et al. note that for the malicious behaviour of the data center and data processor, this approach is insufficient. They note that it is challenging to design a mechanism which can be against a malicious data center/data processor while guaranteeing privacy of data. It is important for data to remain private in an IoT network, especially if there exists a malicious entity in the network.

To counteract the issues found they propose a blockchain-based auditable Privacy-Preserving Data Classification (PPDC) scheme for IoT. Then, go on to show the correctness of PPDC against malicious data processors/data centers. Zhao et al. also provide a new group signature to improve encryption methods of data and allow calculation of correctness by all entities connected to the IoT network. Data privacy is shown by providing a hardness assumption of calculating private keys by an adversary when the encryption/decryption keys are based on a type-3 asymmetric pairing group. They go on to show correctness by proof.

In this paper a powerful solution for guaranteeing privacy of data in a BC based IoT application. Therefore, people should aim to implement this proposed solution when looking to improve an IoT application by use of BC technology. It should, however, be noted that the system is subject to breaking if the private keys related to the data encryption are leaked. Another limitation of their solution is the amount of transactions needed to store one secured piece of data. Which makes it computational heavy, even though they improved the amount of calculation over the flawed previous solution called CLS [16].

**Privacy-Preserving Support Vector Machine Training Over Blockchain-Based Encrypted IoT Data in Smart Cities**, Shen et al. presented in [13] a problem with existing Support Vector Machine (SVM) training methods where an implicit assumption was made that data can reliably be collected from multiple data providers while maintaining data privacy. Since this is generally not true in reality they proposed SecureSVM which is designed to be privacy-preserving. It is important for the data used to train an SVM classifier to be privacy preserving. On the contrary, if this is not the case entities would be reluctant to share their data.

The proposed solution employs a public-key cryptosystem in order to encrypt data that needs to remain private. The specific cryptosystem they use is Paillier since it is more efficient than other algorithms (e.g. Goldwasser-Micali, RSA and Rabin). The proposed solution is different than other SVM solutions since they train on encrypted data instead of unencrypted data. Therefore, they use a simpler optimization algorithm named gradient descent. This in order to make it computationally feasible. The data is all stored on a BC in order to provide secure, reliable and tamperproof data sharing.

Evaluation of the proposed SecureSVM is done by comparing precision and recall against traditional SVM on two commonly used sufficiently large datasets. It is shown that both precision and recall remain in an error margin range compared to SVM. This means that the data is now secure but the proposed classifier can still identify trends and make accurate predictions. Lastly, in the evaluation the authors provide performance time of SecureSVM on both datasets and show that the computation time is acceptable.

The proposed solution can be used to train classifiers where the supplied data is encrypted in order to be privacy secure. One major drawback of this work is that both the problem they end up solving and datasets used for evaluation are healthcare related. Therefore, it is not certain that this proposed solution will also perform well on non-healthcare related datasets. This is not consistent with the sector they were considering, namely smart cities. Also, for the performance evaluation they only provide the performance of their proposed solution and not that of the traditional SVM. The au-

Table 1: Review of state-of-the-art by metrics.

| Paper | Confidentiality | Anonymity | Transparency | Consent Management | Performance |
|---|---|---|---|---|---|
| PPDC [12] | ● | ● | ● | ○ | ◐ |
| secureSVM [13] | ● | ◐ | ○ | ◐ | ◐ |
| PPSF [6] | ◐ | ◐ | ○ | ○ | ● |
| BC Federated Learning [14] | ● | ◐ | ○ | ● | ◐ |
| IoT healthcare FL + BC [15] | ◐ | ○ | ○ | ○ | ◐ |

No: ○, Partially: ◐, Yes: ●

thors should have included this in order to make a comparison between the respective performances. The authors make no notion of anonymity of entities in this IoT network. However, since they adopt a BC-based solution at least pseudonymity could be achieved. Lastly, they state that future work needed to be done is generalizing their framework to enable the construction of a wide range of privacy-preserving ML training algorithms.

**PPSF: A Privacy-Preserving and Secure Framework Using Blockchain-Based Machine-Learning for IoT-Driven Smart Cities**, smart cities face challenges such as centralization, security, privacy and scalability with the evolution of IoT. As such, Kumar et al. present a Privacy-Preserving and Secure Framework (PPSF) in [6]. With the proposed PPSF framework the authors aim to improve data privacy and attack detection (security) of the IoT network.

Their solution contains a two-level privacy scheme and an intrusion detection scheme. On top of the PPSF, they propose a Gradient Boosting Anomaly Detector (GBAD) based on the LightGBM utility system. Lastly, they propose a decentralized deployment solution by integrating an InterPlanetary File System (IPFS) with blockchain. Where the actual privacy sensitive data is stored securely on a filesystem outside of the BC ledger (but inside of the IoT network) and only the hash needed to access this data is recorded on the BC ledger.

They compare this model with three well-known ML techniques and show that the model performs significantly better when detecting various attacks on the IoT network. They use two intrusion detection datasets which are commonly used, namely ToN-IoT and BoT-IoT. They evaluate the precision, true positive rate, false positive rate and performance of execution time for various actions.

This paper creates a strong framework basis that is useful in IoT applications, and this solution uses both BC and ML techniques. Therefore, if any researchers want to extend this framework, it can easily be done. Since our paper considers privacy and performance and not security, it might seem by Table 1 that this paper doesn't add much. However, this paper did some privacy contributions and made a major performance contribution. It should be noted that their biggest contribution is in the space of attack detection which is outside of the scope for our paper. For the privacy of data, we note a partial solution because the authors stored privacy sensitive data on a secure location but did not look for improvement in the encryption of this data, nor did they provide in which way they encrypt the data in their solution. As a potential limitation of the proposed solution, we note that execution time of actions scale with the amount of nodes in the IoT network. Consequently, as the IoT network grows bigger with more nodes performance takes a hit. Therefore, this proposed framework might be best for private IoT networks where the amount of nodes is limited. Their future work suggests designing a prototype so that they can assess the efficiency of the proposed framework.

**Privacy-preserving Decentralized Learning Framework for Healthcare System**, Kasyap et al. [14] state that medical applications such as clinical trials and drug discovery would not be effective without collaboration of institutions. However, the problem is that it would be at the cost of an individual's privacy. A previous solution to this problem was by enforcing several pacts and compliances to avoid data breaches, and collecting participant's data to a central trusted repository. Due to the COVID pandemic this central repository has proven to be obsolete and the authors state that a design of a distributed and decentralized Collaborative Learning system is needed which could inference knowledge from every data point.

As a potential solution, they state that Google has proposed Federated Learning in order to train models in-place so that data is kept intact to the device [17]. Although this solution is privacy-preserving in nature it is susceptible to inference, poisoning and Sybil attacks. Therefore, they present a BC-based Federated Learning architecture with two layers of participation to improve global model accuracy and guarantee participant privacy. Their solution makes use of BC's channel mechanism in order to train models in parallel and distribute them.

In order to evaluate their proposal, the authors run an experiment on top of a federated testbed called PySyft[1]. Their infrastructure setup simulates federated training in a multi-channel BC. The MedNIST[2] dataset was used for training purposes. The authors show training loss for both a privacy-leaking experiment and privacy-preserving experiment. From the results it is shown that training on privacy-preserving data is harder than on privacy-leaking data. However, over time the trained model on privacy-preserving data is only marginally worse in accuracy and testing loss than the model trained on privacy-leaking data. It should be noted, however, that at inference time the privacy-preserving model performs 35% less accurate than the privacy-leaking model.

The proposed solution achieves privacy of data. Pseudonymity is achieved by having participants ask healthcare instances for access to the network. As such, the identity of the participant is known to the healthcare instance but not to other participants in this network. Since participants can choose to join this network and make their data available we denote this as being able to give consent for usage of data. Even though the authors claim to have improved performance, only partial explanation is given and no metrics are given. They do show that some computations during the training of models are theoretically faster than previous approaches. A major limitation of this model is that it operates under the assumption that honest participants stay honest. The model would break in terms of privacy when a participant starts behaving maliciously. Also, in order for interested parties to implement this technique they should find out if their IoT network can facilitate the proposed solution with the resource constraints of their specific network.

**A framework for privacy-preservation of IoT healthcare data using Federated Learning and blockchain technology** [15] proposed by Singh et al., is a framework aimed at addressing the centralized storage of data often found in smart city IoT applications. Centralized storage of data has some problems when we consider an IoT application, such

[1]https://github.com/OpenMined/PySyft

[2]https://www.dropbox.com/s/5wwskxctvcxiuea/MedNIST.tar.gz

as: security issues, single point of failure and increased latency. Similar to the earlier reviewed state-of-the-art PPSF, this proposal is a framework. Therefore, researchers looking to implement a specific solution should aim to build on top of this proposed framework in order to gain privacy, anonymity and performance benefits.

In order to counteract the problems with centralized storage, an architecture containing BC and Federated Learning (FL) is proposed. With this proposal the authors move the learning to the data such that data does not need to be stored on a centralized form of storage. Also, with their framework Singh et al. aim to improve scalability.

Evaluation is done of overhead and reliability in the network. Since the proposal is purely theoretical, measurements are done by providing the formula's for both the overhead calculations and the reliability calculations. Plots are then presented in which can be seen that the biggest impact in overhead is caused by latency between devices. Reliability is shown to be mainly impacted by the binomial probability distribution.

This research, which proposes a framework for privacy-preserving IoT networks, is particularly valuable for individuals and organizations seeking to develop specific applications utilizing such a framework. The authors present a visual representation of the proposed IoT network, which is beneficial for understanding the design. However, the authors do not provide a detailed implementation of the framework, and the explanations for its privacy-preserving properties are limited. Additionally, the authors do not provide sufficient information on the framework's transparency and consent management capabilities, beyond the fact that it incorporates blockchain technology. As such, it is unclear whether this framework represents an improvement in these areas. Overall, the primary limitation of this proposed solution is its theoretical nature. It is possible that the metrics provided by the authors may not hold up when applied to a real-world IoT application.

## 4 Discussion and Future Work

In the previous section it can be seen that various work is being done in the field of IoT with BC and ML techniques. All solutions aim at improving the privacy in the network, with regards to the data that is shared and collected. Also, an attempt is made to provide anonymity of an individual in the network. Along the way, all solutions have performance improvement where the PPSF proposed by Kumar et al. stand out positively with the most performance gain. Most notably is the PPDC proposed by Zhao et al., because it provides privacy, anonymity and transparency while even gaining performance benefits. In the following list we provide future research directions.

- We note that, apart from PPDC, no state-of-the-art solution is able to provide transparency. Probably, providing transparency is hard since privacy-preserving techniques are inherently non-transparent. There is high correlation between transparency and privacy, most techniques need to make a trade-off between these. Therefore, research should be done into providing transparency in a privacy-preserving way.
- Another thing that stood out during review of state-of-the-art solutions is that when an approach provides full privacy of data, it can most often only provide pseudonymity as well. A common approach in these state-of-the-art solutions is having some form of authority in the network. Nodes that want to join the network first have to make themselves known to the authority. Then, potential adversaries are denied access to the network by the authority. This approach is heavily used since most privacy problems addressed in the state-of-the-art arise when attacks on privacy come from within the network. Future research is suggested to investigate the possibility of providing full anonymity while preserving privacy of data.
- A future research direction that arises from the review of state-of-the-art is the question around consent management. A structure for consent management is only sparsely implemented in the reviewed state-of-the-art solutions. Such a structure would benefit users in a network to control which data is shared and which data is kept private. However, an opt-in strategy of data sharing might negatively impact the performance of trained models. Furthermore, attacks on the network might be harder to detect and/or prevent. Therefore, research needs to be done to find out the impact on performance in an IoT network when only a small portion of participants share their data.

For future research directions it is advised to build on top of a framework that implements BC and ML in the IoT domain. For the reviewed state-of-the-art frameworks PPSF is a promising candidate, it implements anonymity and performance in a better way than the IoT healthcare FL + BC framework reviewed. Furthermore, the PPSF framework provides better reproducibility and has been implemented as opposed to the purely theoretical form of the IoT healthcare FL + BC framework.

## 5 Responsible Research

In this section, we will provide insight into the scientific integrity of this paper. We will do so by both providing the integrity and reproducibility of our research. Integrity can be found in section 5.1 and reproducibility is presented in section 5.2

### 5.1 Integrity

Since this work is a literature study, no experiments have been done. As such, the discussion of data gathering, fabrication, falsification or the trimming of data is irrelevant. The integrity we do provide, is related to the collected papers used in this study. We note that all papers are properly referenced throughout this paper and a references section which adheres to the IEEE xplore reference style is provided. The collection of papers has been done by accessing the IEEE xplore and ACM databases. We decided to use these two databases because of their thorough quality checks before accepting papers. As a last notable attempt in order to guarantee integrity

we first provided the metrics by which we compared the state-of-the-art before the actual comparison. By doing so we prevented altering the metrics in such a way that could put some works in a very positive or negative light by our choosing.

### 5.2 Reproducibility

Reproducibility in the sense of resulting data being processed or an experiment that was conducted is not applicable to this paper. However, we can provide insight into the reproducibility of collecting the related work. Since we do provide references one can easily check these papers if interested. An improvement that could have been done in order to improve reproducibility is providing a systematic manner in which we collected the relevant work, such as the used search queries and inclusion criteria. Then, interested parties could extend upon our research if more recent relevant work is available by the defined gathering criteria.

## 6 Conclusion

In this paper, we reviewed state-of-the-art research in the field of IoT where BC and ML techniques are used to improve privacy and performance. We explored technologies used in latest research and composed metrics to analyse state-of-the–art solutions on the combination of IoT, BC and ML. We note that the integration of BC and ML in IoT can positively impact both privacy and performance. An informational table was constructed for easy comparison of the state-of-the-art solutions. Afterwards, we discussed the reviews and pointed out the works that added the most in this field in terms of privacy and performance. Furthermore, we presented future research directions based on the privacy metrics that were lacking in the reviewed state-of-the-art. Lastly, we concluded that future researchers should try and build on top of a privacy-preserving framework when developing new IoT applications. We believe that researchers will find our work insightful as a guide to their future research on IoT applications to both protect privacy of participants and enhance the performance of the network.

## References


[1] 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.

[2] F. Hussain, R. Hussain, S. A. Hassan, and E. Hossain, "Machine learning in iot security: Current solutions and future challenges," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 3, pp. 1686–1721, 2020.

[3] V. Friedman, "On the edge: Solving the challenges of edge computing in the era of iot," Jun 2020. [Online]. Available: https://www.databank.com/blogs/2018/08/30/solving-edge-computing-challenges-in-era-of-iot/

[4] A. Al-Hasnawi, I. Mohammed, and A. Al-Gburi, "Performance evaluation of the policy enforcement fog module for protecting privacy of iot data," in *2018 IEEE International Conference on Electro/Information Technology (EIT)*, 2018, pp. 0951–0957.

[5] P. M. Kumar, B. Rawal, and J. Gao, "Blockchain-enabled privacy preserving of iot data for sustainable smart cities using machine learning," in *2022 14th International Conference on COMmunication Systems NETworkS (COMSNETS)*, 2022, pp. 1–6.

[6] P. Kumar, R. Kumar, G. Srivastava, G. P. Gupta, R. Tripathi, T. R. Gadekallu, and N. N. Xiong, "Ppsf: A privacy-preserving and secure framework using blockchain-based machine-learning for iot-driven smart cities," *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 3, pp. 2326–2341, 2021.

[7] A. Gillis, "What is the internet of things (iot)?" accessed: 19-12-2022. [Online]. Available: https://www.techtarget.com/iotagenda/definition/Internet-of-Things-IoT

[8] Y. Wu, Z. Wang, Y. Ma, and V. C. Leung, "Deep reinforcement learning for blockchain in industrial iot: A survey," *Computer Networks*, vol. 191, p. 108004, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1389128621001213

[9] N. Waheed, X. He, M. Ikram, M. Usman, and S. Hashmi, "Security and privacy in iot using machine learning and blockchain: Threats and countermeasures," *ACM Computing Surveys*, vol. 53, pp. 1–37, 12 2020.

[10] D. Ottolini, I. Zyrianoff, and C. Kamienski, "Interoperability and scalability trade-offs in open iot platforms," in *2022 IEEE 19th Annual Consumer Communications Networking Conference (CCNC)*, 2022, pp. 1–6.

[11] Y. Liu, W. Yu, T. Dillon, W. Rahayu, and M. Li, "Empowering iot predictive maintenance solutions with ai: A distributed system for manufacturing plant-wide monitoring," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 2, pp. 1345–1354, 2022.

[12] Y. Zhao, X. Yang, Y. Yu, B. Qin, X. Du, and M. Guizani, "Blockchain-based auditable privacy-preserving data classification for internet of things," *IEEE Internet of Things Journal*, vol. 9, no. 4, pp. 2468–2484, 2022.

[13] M. Shen, X. Tang, L. Zhu, X. Du, and M. Guizani, "Privacy-preserving support vector machine training over blockchain-based encrypted iot data in smart cities," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 7702–7712, 2019.

[14] H. Kasyap and S. Tripathy, "Privacy-preserving decentralized learning framework for healthcare system," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 17, no. 2s, jun 2021. [Online]. Available: https://doi.org/10.1145/3426474

[15] S. Singh, S. Rathore, O. Alfarraj, A. Tolba, and B. Yoon, "A framework for privacy-preservation of iot healthcare data using federated learning and blockchain technology," *Future Generation Computer Systems*, vol. 129, pp. 380–388, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167739X21004726

[16] L. Garms and A. Lehmann, "Group signatures with selective linkability," in *Public-Key Cryptography – PKC 2019*, D. Lin and K. Sako, Eds. Cham: Springer International Publishing, 2019, pp. 190–220.

[17] H. B. McMahan, E. Moore, D. Ramage, and B. A. y Arcas, "Federated learning of deep networks using model averaging," *CoRR*, vol. abs/1602.05629, 2016. [Online]. Available: http://arxiv.org/abs/1602.05629