




Please cite the Published Version

Hussain, Altaf, Akbar, Wajahat, Hussain, Tariq , Bashir, Ali Kashif , Dabel, Maryam M. Al , Ali, Farman and Yang, Bailin (2024) Ensuring Zero Trust IoT Data Privacy: Differential Privacy in Blockchain using Federated Learning. IEEE Transactions on Consumer Electronics. ISSN 0098-3063

DOI: <https://doi.org/10.1109/TCE.2024.3444824>

Publisher: Institute of Electrical and Electronics Engineers (IEEE)

Version: Accepted Version

Downloaded from: <https://e-space.mmu.ac.uk/635682/>

Usage rights:  In Copyright

Additional Information: © 2024 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Enquiries:

If you have questions about this document, contact openresearch@mmu.ac.uk. Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from <https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines>)

Ensuring Zero Trust IoT Data Privacy: Differential Privacy in Blockchain using Federated Learning

Altat Hussain, Wajahat Akbar, Tariq Hussain, Ali Kashif Bashir, Maryam M. Al Dabel, Farman Ali*, Bailin Yang*

Abstract—In the increasingly digitized world, the privacy and security of sensitive data shared via IoT devices are paramount. Traditional privacy-preserving methods like k -anonymity and l -diversity are becoming outdated due to technological advancements. In addition, data owners often worry about misuse and unauthorized access to their personal information. To address this, we propose a secure data-sharing framework that uses local differential privacy (LDP) within a permissioned blockchain, enhanced by federated learning (FL) in a zero-trust environment. To further protect sensitive data shared by IoT devices, we use the Interplanetary File System (IPFS) and cryptographic hash functions to create unique digital fingerprints for files. We mainly evaluate our system based on latency, throughput, privacy accuracy, and transaction efficiency, comparing the performance to a benchmark model. The experimental results show that the proposed system outperforms its counterpart in terms of latency, throughput, and transaction efficiency. The proposed model achieved a lower average latency of 4.0 seconds compared to the benchmark model's 5.3 seconds. In terms of throughput, the proposed model achieved a higher throughput of 10.53 TPS (transactions per second) compared to the benchmark model's 8 TPS. Furthermore, the proposed system achieves 85% accuracy, whereas the counterpart achieves only 49%.

Index Terms—Blockchain, Differential Privacy, Federated Learning, Internet of Things, Zero Trust Security.

I. INTRODUCTION

THE Smart Home Systems (SHS) that use the Internet of Things (IoT) have become very popular recently. The concept of SHS is also applied to create IoT applications

This work was supported by the "Pioneer" and "Leading Goose" R D. Program of Zhejiang Province (2023C01150), also This research was supported by Zhejiang Provincial Natural Science Foundation of China under Grant No.LD24F020003. Altat Hussain and Tariq Hussain contributed equally to this work and are the first co-authors.

Altat Hussain Department of Computer Science and BI Khushal Khan Khattak University, Karak, Pakistan. (e-mail: altat.hussain@kku.edu.pk).

Wajahat Akbar is from the School of Electronic and Control Engineering at Chang'an University, Xi'an, China (e-mail: wajahatakbar32@gmail.com)

Tariq Hussain and Bailin Yang are at the School of Computer Science and Technology and School of Mathematics and Statistics, Zhejiang Gongshang University, Hangzhou 310018, China; (e-mail: uom.tariq@gmail.com).

Ali Kashif Bashir is from the Department of Computing and Mathematics, Manchester Metropolitan University, UK, and Woxsen School of Business, Woxsen University, Hyderabad, India 502 345, and Centre for Research Impact& Outcome, Chitkara University Institute of Engineering and Technology. Chitkara University, Rajpura, 140401, Punjab, India (e-mail: Dr.alikashif.b@ieee.org)

Maryam M. Al Dabel is from the Department of Computer Science and Engineering, College of Computer Science and Engineering, University of Hafr Al Batin, Saudi Arabia; (e-mail: maldabel@uhb.edu.sa)

Farman Ali is from the Department of Applied AI, School of Convergence, College of Computing and Informatics, Sungkyunkwan University, Seoul 03063, South Korea

Correspondence: (ybl@zjgsu.edu.cn; Farman0977@g.skku.edu)

in various areas such as smart cities, agriculture, healthcare services, etc. [1]. The rapid increase in IoT devices results in generating vast amounts of data (i.e., Big Data) every fraction of a second. These data contain sensitive information about the owners, which requires security and privacy [2], [3]. The term "privacy" refers to the notion that an individual's data will be treated discreetly or that access to the data will require authorization. "security" refers to the ability to protect sensitive data from both eavesdroppers and intruders [4], [5].

Privacy-Preserving Data Sharing (PPDS) methods address risks of re-identification of data owners or revealing sensitive information [6]. PPDS includes various techniques such as data masking [7], encryption [8], k -anonymization [9], [10], and l -diversity [11] that meet privacy requirements [12]. However, these techniques have limitations such as background knowledge, homogeneity, and inference attacks, while FL poisoning attacks present a significant obstacle to achieving data privacy [13]. To address these limitations of the existing PPDS methods, a robust "Zero Trust Architecture (ZTA)" is needed that guarantees a secure mechanism for the transfer of data without risking the re-identification of data owners' sensitive information. The ZTA is a practical implementation emphasizing trust as a vulnerability rather than a fundamental component in network security [14]. ZTA is based on segmenting networks into microcores and perimeters. It suggests that everything, even inside the perimeters of a network, is untrusted instead of building a trusted domain around the network. Thus, it promotes the "never trust, always verify" principle within enterprise networks as well [15].

Additionally, the current methods used to anonymize data often depend on trusting the people who hold the data or the companies they hire to do it for them. This could be problematic if these parties are semi-honest and could potentially cause harm [16], [17]. Likewise, if data controllers receive data owners' datasets with inadequate PPDS techniques, they could engage in malicious activities, posing as attackers with harmful intentions such as stealing sensitive data or misusing data owners' sensitive information [18].

Blockchain technology has recently emerged as a critical solution for safeguarding sensitive data, providing a decentralized method for managing records of sensitive data, ensuring authentication, preventing tampering, and facilitating secure data sharing [19]. Through its resilience, blockchain empowers data owners to manage their data. Differential privacy (DP) offers another mathematical framework for privacy-preserving without disclosing sensitive information [20]. DP quantifies the privacy loss associated with specific data analysis and

achieves this by introducing controlled or randomized noise while maintaining the desired level of privacy. LDP, a type of DP, preserves privacy at the individual data point level rather than the aggregate level, ensuring that no individual data point can be linked to a specific user [21].

FL presents another significant advancement in PPDS techniques, particularly in the domain of machine learning [22]. FL allows model training to be conducted across multiple decentralized devices or servers without the need to centralize the sensitive data [23]. Instead, individual devices or nodes compute model updates on their local data and only share encrypted or aggregated information with a central server or among themselves. This approach minimizes the risk of exposing raw data to third parties, thereby enhancing privacy protection [23], [24]. By leveraging FL in conjunction with LDP and blockchain technology, we can ensure not only the confidentiality of sensitive information but also maintain control over data ownership and usage rights. The integration of FL further strengthens the resilience of blockchain-based systems by enabling collaborative model training while preserving the privacy of individual data points. To address the identified issues and limitations, we propose a mechanism that combines blockchain technology with LDP and FL to achieve robust data privacy and security for sharing sensitive data. This research is motivated by the challenges outlined in [10], detailed below in the "Motivation" section.

A. Motivation

In this section, we carefully examined the problems with the existing methods of sharing data owners' data. Due to pressing privacy concerns, there is a fundamental lack of trust in any data controller or entity involved in this process. As illustrated in Fig. 1, sharing sensitive data directly with controllers without applying any PPDS technique by data owners could potentially be viewed as adversaries since controllers may not be trusted to handle sensitive data. This ambiguity renders that the data controllers may be semi-trust actors in this scenario. Moreover, The existing scheme [10] has implemented k -anonymity on the data owners' sensitive data to reduce the risk of de-identification. But With k -anonymized data, there remains the possibility of an attribution disclosure attack due to a lack of trust for data controllers, as shown in Fig. 1. There is no sense in sharing sensitive information with controllers without any privacy whatsoever, and data owners will never want to do so without any protection. In a nutshell, they consistently prioritize the utmost privacy when sharing data with controllers in a zero-trust environment. Fig. 1, along with Fig. 2, serves to elucidate the attacker models, shedding light on how adversaries can potentially engage in malicious activities, especially in the context of FL.

FL is vulnerable to several types of attacks, compromising its privacy and security guarantees. Data poisoning attacks involve malicious participants injecting false data to skew the model performance, as shown in Fig. 2. Model poisoning attacks are more sophisticated, where adversaries manipulate model updates to degrade or control the final aggregated model. Inference attacks seek to deduce sensitive information

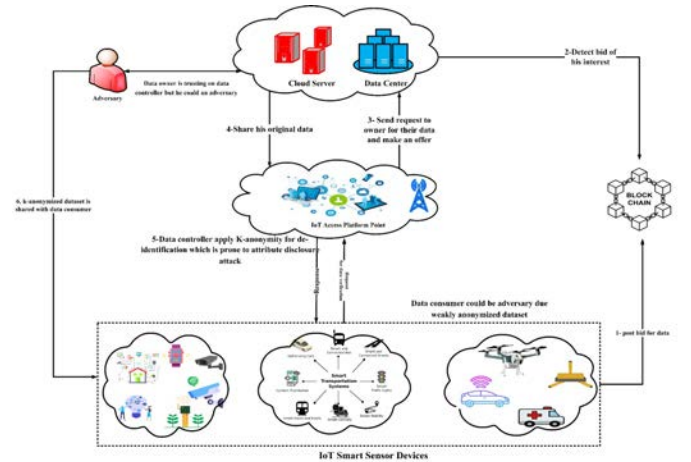


Fig. 1: Attacker Model

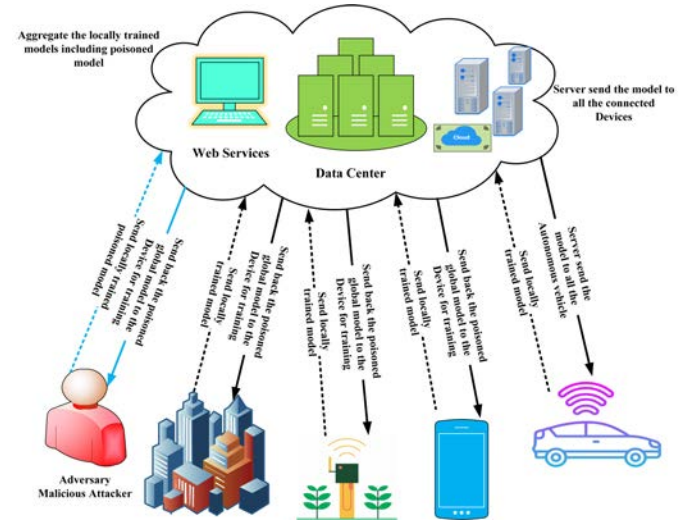


Fig. 2: Poisoning Attack in FL

about the training data from the shared model updates. To solve FL issues, the proposed work used blockchain technology to prevent data poisoning, model poisoning, and inference attacks. Authenticating all entities through blockchain ensures their identities are verified. The main contributions are of the proposed model are:

- To enhance data privacy and security within IoT networks, a novel approach is introduced that combines LDP, Permissioned blockchain, and alongside FL in a zero-trust environment that will protect against poisoning, background knowledge, and inference attacks.
- The registration process on the permissioned blockchain will provide an extra layer of security by verifying the identities of all the entities.
- The experimental results demonstrate higher transaction efficiency, throughput, and low latency and outperforms with respect to its counterpart in other performance metrics.

TABLE I: Notations

Symbols	Description
P2P	Peer-to-peer
PoA	Proof-of-Authority
TPS	Transactions per second
A	Stochastic algorithm operating on a domain \mathcal{D} and yielding results in the co-domain \mathcal{R} .
\mathcal{D}	A dataset.
\mathcal{R}	Codomain in which the algorithm A yields results.
ϵ	Non-negative privacy parameter in DP. Smaller values correspond to stronger privacy guarantees.
D'	A dataset that differs from D by a single data point
O	A possible outcome of the randomized algorithm A
$\Pr[A(D) = O]$	Probability that the algorithm A applied to dataset D produces outcome O .
$P_D(O)$	Notation for $\Pr[A(D) = O]$, the probability of outcome O given dataset D .
$P_{D'}(O)$	Notation for $\Pr[A(D') = O]$, the probability of outcome O given dataset D' .
e	The base of the natural logarithm is approximately equal to 2.71828.
M	Global model
n	Number of decentralized clients in federated learning.
$\mathcal{L}(M, D_i)$	Loss function associated with client i in federated learning.
M^*	Optimal global model parameters in federated learning, minimizing the universal loss function.
Local Loss _{i}	Loss term for client i in federated learning.
D_i	Local dataset of client i
\mathcal{P}	List of Data Owners
\mathcal{C}	List of Data Controllers
\mathcal{S}	List of Data Consumers
\mathcal{B}	Permissioned Blockchain
\mathcal{I}	Interplanetary File System (IPFS)
\mathcal{G}	Initial Global Model
e	An entity (could be a Data Owner, Data Controller, or Data Consumer)
s	List of Data Consumers
c	List of Data Controllers
p	List of Data Owners
θ_p	Updated model parameters from Data Owner p
\mathcal{G}^*	Final updated Global Model

II. RELATED WORK

IoT devices generate high volumes of data, so network security goes beyond conventional measures like firewalls and access control. In dynamic IoT environments, real-time monitoring, data handling and storage, access management, and security are all challenges [25]. The authors of [8473444] propose a zero-trust model that enables hierarchical mining based on IoT. The authors describe IoT infrastructure as a zero-trust model. To address this issue, Amatista has been introduced as a blockchain-based middleware. The relentless advancement of technology has woven its threads of innovation into every corner of the world where human beings exist, bestowing us with improved features and applications that have transformed how we navigate the world. Now, this transformation is more palpable than in the realm of healthcare [26], [27]. With the advancement of technology like 5G [28], the healthcare sector has profoundly evolved, catalyzed by these technological strides. Among the most noteworthy shifts is the digitization of medical records, a repository of sensitive patient information that has transitioned from paper to electronic realms on the network [29]. While this transition has streamlined many aspects of healthcare, it has also unveiled new avenues for digital privacy and security breaches. In a world where medical records encompass deeply personal and private data, sharing such intimate information electronically has spurred the emergence of concerns [30], [31]. With advancements in technology, the challenge of safeguarding the sanctity of medical records has become more pronounced, especially in light of their historical attractiveness to data thieves. This confluence of digitization and data vulnerability casts an even greater imperative on fortifying the security and privacy of patients' information [32].

Encryption has emerged as a cornerstone of data security strategies, ensuring the confidentiality of patient information by rendering it unreadable without the appropriate decryption key. An encryption technique is used in [33]. However, implementing encryption techniques can introduce processing overhead and potential obstacles to key management and distribution problems.

Role-based Access Control (RBAC) has been used in [34] to restrict data access based on users' roles, effectively minimizing the risk of unauthorized data exposure. While RBAC offers a structured approach to access management, its complexity can lead to misconfigurations and challenges in accommodating evolving access requirements.

In technological advancement, blockchain technology has garnered attention for its potential to ensure data immutability and transparency, which are critical for maintaining trustworthy audit trails in healthcare environments. Researchers in [20] used blockchain. Yet, its high computational demands and scalability issues have raised concerns about its practicality for large-scale healthcare systems.

Another technique, the Anomaly Detection System, is used by [35] to rescue the data in healthcare. Anomaly Detection System has proven valuable in identifying unusual patterns in data access, thereby avoiding the detection of unauthorized activities. However, striking a balance between accurate detection and a low rate of false positives remains an ongoing challenge. Moreover, these systems may struggle to identify new or previously unseen attack patterns.

Bio-metric authentication is used in [36], relying on unique physical traits for robust user verification, significantly advancing healthcare data security. However, concerns regarding bio-metric data compromise and the potential for false acceptance

TABLE II: Summary of Techniques, Achievements and Limitations

References	Techniques	Achievements	Limitations
[33]	Encryption	Data confidentiality; Security against breaches	Processing overhead; Key management
[34]	Role-based Access Control	Access restriction; Role-based data handling	Implementation complexity; Access management
[20]	Blockchain Technology	Data immutability; Transparent audit trails	Computational requirements; Scalability
[35]	Anomaly Detection Systems	Unusual activity detection; Intrusion prevention	False positives/negatives; New attack patterns
[36]	Biometric Authentication	Strong user authentication; Non-repudiation	Biometric data compromise; False acceptance
[37]	Homomorphic Encryption	Privacy during processing; Encrypted computation	Computational intensity; Expertise requirement
[38]	Tokenization	Reduced data exposure; Limited sensitive data	Token management; Insider attacks
[39]	Firewalls and IDS	Network traffic control; Threat monitoring	Limited against advanced attacks; False results
[40]	Data Masking	Data obfuscation; Consistent masked data	Reversible masking risks; Data consistency
[41]	K-anonymization	Privacy enhancement; Re-identification defense	Privacy-utility trade-off; Re-identification risk
[42]	T-closeness	Enhanced privacy distribution; Data distortion	Data distortion; T-closeness trade-off
[43]	L-diversity	Attribute disclosure mitigation; Diversity	Data distortion; Privacy-quality trade-off

or rejection warrant careful consideration.

The researcher also uses the homomorphic encryption technique in [37], allowing computations on encrypted data without decryption, which holds promise for preserving privacy during data processing. Nevertheless, its computational intensity can lead to slower processing speeds, and successful implementation often necessitates specialized expertise.

Tokenization [38] has been employed to replace sensitive data with tokens, thereby limiting the exposure of valuable information within the system. While effective token management is crucial, it's important to note that tokenization might not comprehensively address all security concerns, such as insider threats.

Firewalls and Intrusion Detection Systems (IDS) are used in [39]; they serve as essential safeguards by monitoring and controlling network traffic. However, their efficacy against advanced or zero-day attacks is limited, and the occurrence of false positives and negatives can impact their reliability.

Data masking is used in [40], which involves obfuscating original data while retaining its format and offers a practical approach for testing and analysis. However, reversible masking techniques can potentially result in data leaks, and ensuring the consistency of masked data presents an additional challenge.

K-anonymization is used in [41] for enhancing privacy by generalizing data to a level where individuals cannot be uniquely identified. However, balancing privacy and data utility can be complex, and re-identification attacks remain a major concern.

T-closeness in [42] Guarantees that the distribution of sensitive information within an equivalence class closely mirrors the distribution in the entire data set. However, achieving t-closeness might necessitate significant data distortion.

L-diversity introduces diversity within anonymity groups in [43] to mitigate the risk of attribute disclosure. However, achieving l-diversity can result in increased data distortion and a trade-off between privacy and data quality.

Table II represents existing schemes, their achievements, and limitations.

III. PRELIMINARIES

This section provides an in-depth overview of the proposed method, encompassing its core functionality, definitions, and communication model. To bolster security within the demanding landscape of IoT networks, we present a crucial

safeguarding approach. This method is designed to fortify the security framework, ensuring its robustness and resilience. The following are some key concepts that will be used in the proposed method, and they are necessary to understand their importance in the data-sharing process.

A. Differential Privacy

DP stands as a foundational concept in the domain of data privacy, providing a quantifiable measure of the level of privacy safeguards provided by a specific data analysis or query; its primary function is to ensure that the incorporation or omission of an individual's data from the data set will not significantly modify the results of the analysis. In mathematical terms, its definition can be articulated as follows: A stochastic algorithm denoted as A , operating on a domain \mathcal{D} and yielding results in the codomain \mathcal{R} , guarantees ϵ -DP if, for any pair of data set D and D' that vary only in the information of a single individual and for all measurable sets $O \subseteq \mathcal{R}$,

$$\frac{\Pr[A(D) = O]}{\Pr[A(D') = O]} \leq e^\epsilon.$$

Where the probability is taken over the randomness of the algorithm A .

In this definition, ϵ is a non-negative privacy parameter. Smaller values of ϵ correspond to stronger privacy guarantees, with $\epsilon = 0$ being perfect privacy (no information leakage).

Theorem 1. (LDP): Let $\epsilon > 0$ be a privacy parameter, and let D be a dataset. A randomized algorithm \mathcal{A} is said to satisfy ϵ -Local DP if, for all possible outcomes O and for all datasets D and D' that differ in a single data point, we have:

$$\frac{\Pr[\mathcal{A}(D) = O]}{\Pr[\mathcal{A}(D') = O]} \leq e^\epsilon.$$

Proof. Let D and D' be datasets that differ in exactly one data point. We need to show that for any possible outcome O of the randomized algorithm \mathcal{A} , the following inequality holds:

$$\frac{\Pr[\mathcal{A}(D) = O]}{\Pr[\mathcal{A}(D') = O]} \leq e^\epsilon.$$

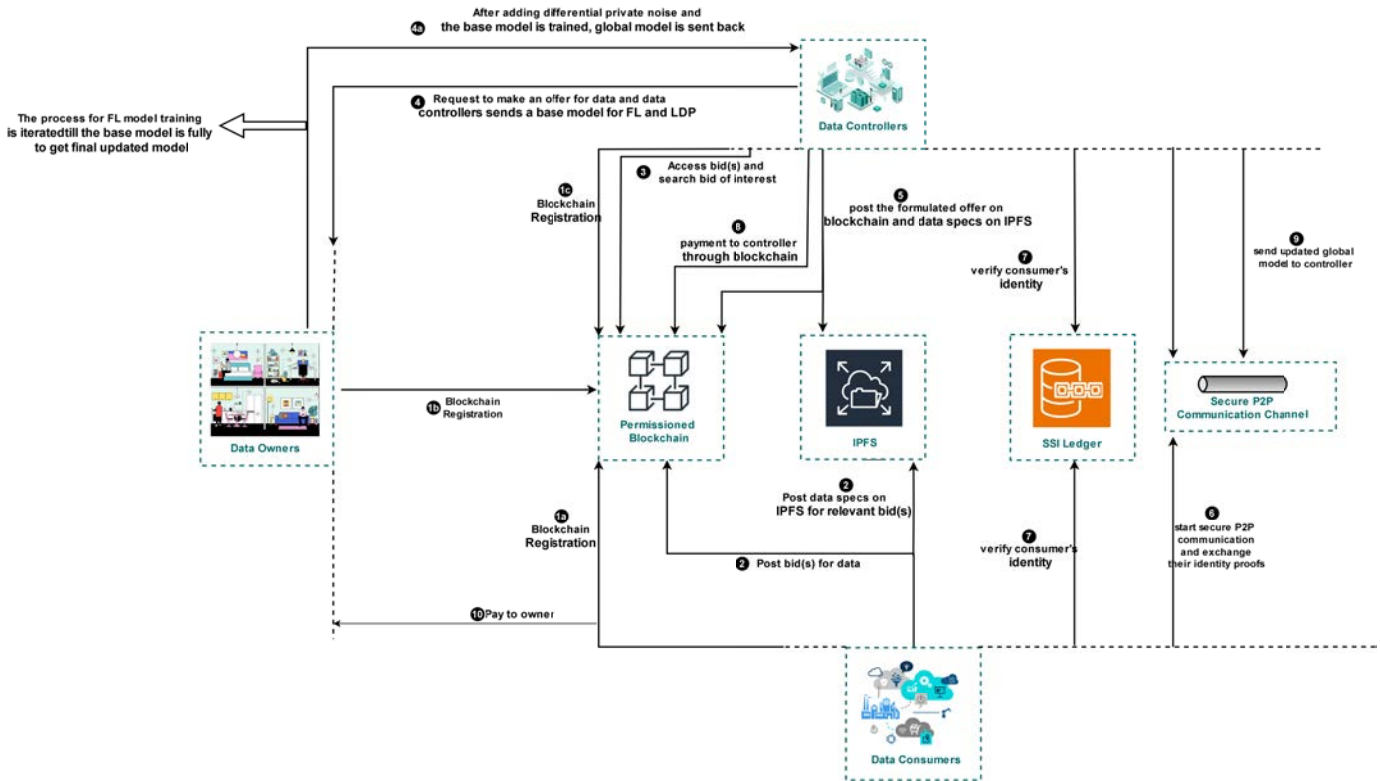


Fig. 4: Proposed model for secure IoT data sharing

ii. *Blockchain bid posting and IPFS data storage:*

A company or researchers (data consumers) submit a bid on the blockchain and upload a dataset specifications file to IPFS. Bids may include links to specification files, bid expiry dates, payment amounts, and tags (e.g., smart home energy usage). A data set requirement file encompasses information about the data set schema, such as the type of device, energy consumption, and timestamps.

iii. *Accessing Bids on the Blockchain:* The data controller (e.g., service providers) accesses the data set specifications and identifies all bids of interest posted by data consumers.

iv. *Requesting Data Acquisition to make an Offer:* Data controllers interconnect with data owners to get their datasets. The data controller sends a base model for training on local data. Data owners train this model, add differential private noise, and return updated parameters. This process iterates until the model is fully trained and a final updated global model is created.

v. *Posting offer on Blockchain and Data Set Specifications on IPFS:* Once the data controller has received the most updated version of the global models, they formulate an offer and post it on the blockchain. The offer includes a link to the data specs file on IPFS. The data controller also reveals the specifications file on IPFS, which includes the data controller's publicly accessible decentralized identities (DIDs). Data consumers acknowledge the offer and access the data specs file uploaded on IPFS by the data controller.

vi. *Exchange of Identity Proofs:* Both parties establish

a secure P2P communication channel to exchange information and identity proofs.

vii. *Confirming Identity:* The identities of both parties (i.e., the controller and consumer) are verified using the SSI ledger.

viii. *Payment to the Controller:* After agreeing to the terms of the offer, the data utilized launches a P-to-P protected channel and makes the required payments to the data controller.

ix. *Sharing the updated Global Model:* Upon successful verification, the final version of the updated model is shared with the data consumer via the secure channel.

x. *Compensation to data owners:* After completing the process of anonymized data sharing, the data controller compensates the specific data owner.

Algorithm 1 begins by registering each entity (i.e., data owners (\mathcal{P}), data controllers (\mathcal{C}), and data consumers (\mathcal{S})) on \mathcal{B} (line 1-3). \mathcal{S} then post bids on the \mathcal{B} and upload their data specifications on \mathcal{I} (line 4-6). \mathcal{C} access these bids from the \mathcal{B} and identify their bids of interest (lines 7-9). Subsequently, \mathcal{C} send their initial global model \mathcal{G} to each \mathcal{P} , who trains the model locally in the FL process and adds differential private and sends back the updated model θ_p (line 10-15). The process in lines 10-15 repeats till the model is fully trained for the FL process, and finally, \mathcal{C} receives updated global model \mathcal{G}^* from

\mathcal{P} . \mathcal{C} then formulate an offer based on \mathcal{G}^* , post this offer on \mathcal{B} and upload \mathcal{G}^* dataset specification to \mathcal{I} (line 16-19). If \mathcal{S} accepts an offer, They make a payment to \mathcal{C} and exchange identity proofs via a secure P2P channel and verify identities via SSI ledger (lines 20-26). \mathcal{C} share \mathcal{G}^* with \mathcal{S} via a secure P2P channel. Finally, \mathcal{C} compensates the specific \mathcal{P} for their contribution and successful data-sharing (lines 27-34).

V. IMPLEMENTATION AND RESULTS

This section provides a detailed description of the model's implementation and testing approaches. Using the Remix Ethereum platform, we built a private Ethereum blockchain network to enhance the effectiveness and throughput of blockchain transactions. In addition, the FL and LDP processes were executed using Python. Proof-of-Authority (POA) was the consensus protocol we chose. The simulation has three main functions: registration, offer, and finalization. Finalization is a payable function, whereas offer and registration are non-payable. A particular dataset is registered through register function, which is the first step. Subsequently, offers are generated and responded to the registered bid using the offer function.

We conducted experiments involving multiple arbitrary data owners in the FL context. Our experimentation comprised 70 rounds of transactions, each encompassing the complete trading procedures, including bidding, offers exchanges, finalization, and FL rounds.

Subsequently, we conducted a comprehensive comparative analysis between the proposed scheme and the approach presented by [10]. The findings are presented in the subsequent figures. In our experiments, we accessed the privacy accuracy, transaction latency, throughput, transaction efficiency, correlation, and analysis of variance. The subsequent sections delve into the detailed outcomes of the proposed experiments.

A. Latency

Latency is the time required for a transaction to be submitted, integrated into the blockchain, and validated. It can be calculated using the following equation (2) [26].

$$Latency = Cost/Throughput \quad (2)$$

We measured the transaction latency for 70 rounds of transactions carried out within our proposed model and compared these results with the benchmark model [10] as shown in Fig. 5. The latency assessment is conducted through the system's processor clock, particularly during key bidding phases like registration, offering, and finalization, utilizing the formula specified in equation (2).

Fig. 5 clearly shows that the proposed model outperformed the benchmark in terms of latency. The proposed model achieved a lower average transaction latency of 4.0 seconds, while the benchmark model had a latency of 5.6 seconds. The increase in latency with each subsequent transaction is due to the direct relationship between latency and the number of transactions. As the transaction count grows, the time required to process each transaction increases, consequently elevating the overall latency.

Algorithm 1 : Blockchain-based Secure Data Sharing with LDP and FL

Input:

\mathcal{P} : List of Data Owners
 \mathcal{C} : List of Data Controllers
 \mathcal{S} : List of Data Consumers
 \mathcal{B} : Permissioned Blockchain
 \mathcal{I} : Interplanetary File System (IPFS)
 \mathcal{G} : Initial Global Model

Output:

\mathcal{G}^* : Final updated Global Model

```

1: for each entity  $e \in (\mathcal{P} \cup \mathcal{C} \cup \mathcal{S})$  do
2:   Register  $e$  on  $\mathcal{B}$ 
3: end for
4: for each consumer  $s \in \mathcal{S}$  do
5:    $s$  posts bids on  $\mathcal{B}$  and uploads data specifications on  $\mathcal{I}$ 
6: end for
7: for each controller  $c \in \mathcal{C}$  do
8:    $c$  accesses bids from  $\mathcal{B}$  and identifies bids of interest
9: end for
10: for each controller  $c \in \mathcal{C}$  do
11:   for each owner  $p \in \mathcal{P}$  do
12:      $c$  sends  $\mathcal{G}$  to  $p$ 
13:      $p$  trains  $\mathcal{G}$  locally, applies DP and sends updated
        model parameters  $\theta_p$  back
14:   end for
15: end for
16: for each controller  $c \in \mathcal{C}$  do
17:    $c$  formulates an offer and posts it on  $\mathcal{B}$ 
18:    $c$  uploads updated data set specifications on  $\mathcal{I}$ 
19: end for
20: for each consumer  $s \in \mathcal{S}$  do
21:   if  $s$  accepts the offer then
22:      $s$  makes payment to  $c$  through a secure P2P channel
23:      $c$  and  $s$  exchange identity proofs via a secure P2P
        channel
24:     Identities are verified using the self-sovereign identity
        (SSI) ledger
25:   end if
26: end for
27: for each controller  $c \in \mathcal{C}$  do
28:    $c$  shares the final updated global model  $\mathcal{G}^*$  with  $s$  via
        a secure P2P channel
29: end for
30: for each controller  $c \in \mathcal{C}$  do
31:   for each owner  $p \in \mathcal{P}$  do
32:      $c$  compensates  $p$  after successful data sharing
33:   end for
34: end for
=0

```

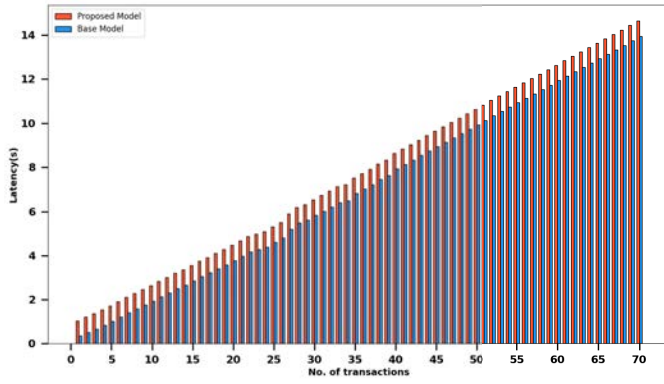


Fig. 5: Latency

B. Throughput

Throughput on a blockchain refers to the rate at which valid transactions are processed per second (tps). It is calculated using the equation (3).

$$Throughput(TP) = execution\ cost/latency \quad (3)$$

To assess the efficiency of our proposed model, we measured TP for key procedures (offer, finalize, register) during the experiment. This throughput was calculated using the formula outlined in Equation (3). To visualize the performance improvement, Fig 6 offers a comprehensive comparison of throughput between our model and benchmark models, all processing 70 transactions.

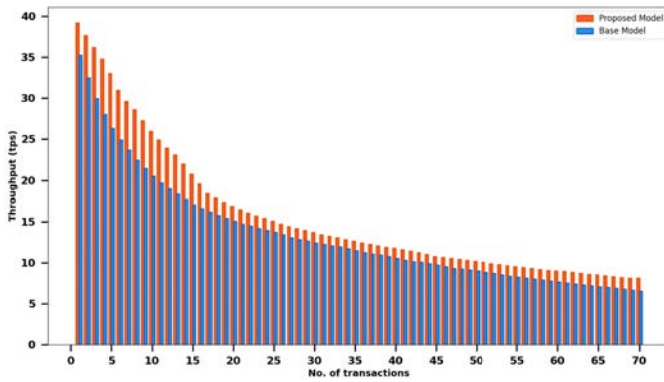


Fig. 6: Throughput

Fig. 6 compares the benchmark model and the proposed model's throughput, clearly showing that the proposed model achieved a higher throughput than its counterpart. The proposed model achieved an average throughput of 10.23 tps, while the benchmark model achieved a throughput of 8 tps. It is important to note that an inverse relationship exists between the number of transactions and throughput; as the number of transactions increases, throughput tends to decrease. Therefore, in each subsequent transaction, throughput decreases in Fig. 6.

C. Accuracy

Accuracy represents the mean of the accurate prediction produced by the model, as shown in equation (4).

$$Accuracy = (validtransactions)/(TotalTransactions) \quad (4)$$

To evaluate the accuracy of the proposed scheme, we experimented with 70 iterations on three different IoT-connected device data. We employed LDP to preserve privacy while calculating the scheme's accuracy for both federated and personal (individual) settings. The results are presented in Fig 8 and 7, respectively, subsequently comparing these results with those obtained by [10]. Personal accuracy signifies the accuracy of an individual model, whereas federated accuracy represents the accuracy achieved when aggregating models from all clients to a global model.

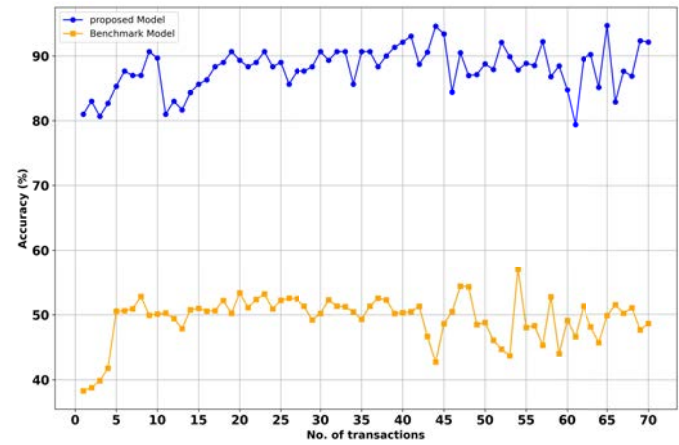


Fig. 7: Base model accuracy vs. proposed model federated accuracy

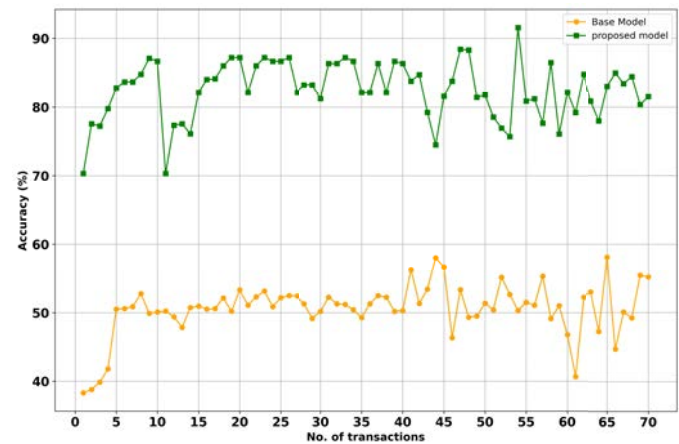


Fig. 8: Base model accuracy vs. proposed model personnel accuracy

The proposed scheme demonstrated an enhanced privacy accuracy of 85% in both personal and federated settings compared to the approach presented by [10], which achieved an accuracy of 49%. The improved accuracy in our proposed models can be attributed to the robust training mechanism

employed in FL and the applied LDP, which enhances data accuracy. Both LDP and FL are new methods that improve accuracy and ensure strong data privacy. The experiments for the proposed model were executed within a Python Anaconda environment to implement FL and LDP.

D. Transaction Efficiency

Another performance metric for the proposed model is transaction efficiency (TE). It can be calculated using the following equation (5).

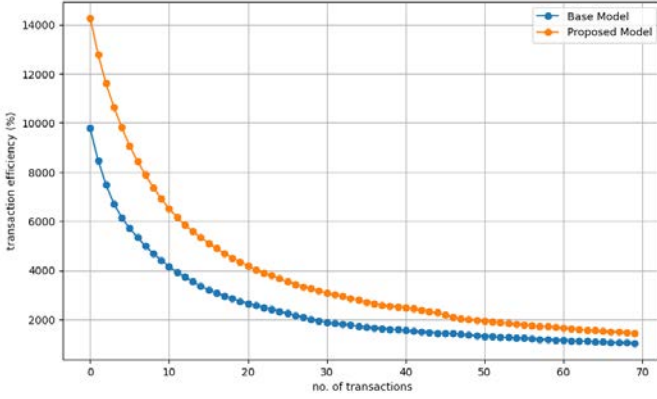


Fig. 9: Throughput Transactions efficiency (TTE)

$$TE = \frac{\text{Total Throughput}}{\text{Resource Consumption}} \quad (5)$$

Throughput is the total number of transactions the system performs in a specific time. Resource consumption refers to the specific resources consumed to perform the total throughput; here, resource consumption is the execution cost in the form of gas used. In the experiment, transaction efficiency is evaluated for 70 rounds of transactions. Fig.9 visually represents the transaction efficiency comparison between the proposed and base models. The visual representation shows that the proposed model has better transaction efficiency than the base model. The successive decrease in efficiency for each subsequent transaction is due to the inverse relationship between transaction efficiency and latency.

E. Performance Consistency of Transaction Throughput

Standard deviation is a statistical measure that quantifies the degree of variation or dispersion within a dataset. It is an essential tool in understanding the consistency of data. In the context of transaction throughput, a lower standard deviation indicates more consistent performance, while a higher standard deviation signifies greater variability. The standard deviation (σ) is calculated using formula in equation (6):

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} \quad (6)$$

Where N is the number of throughput values, x_i represents each throughput value, and μ is the mean throughput.

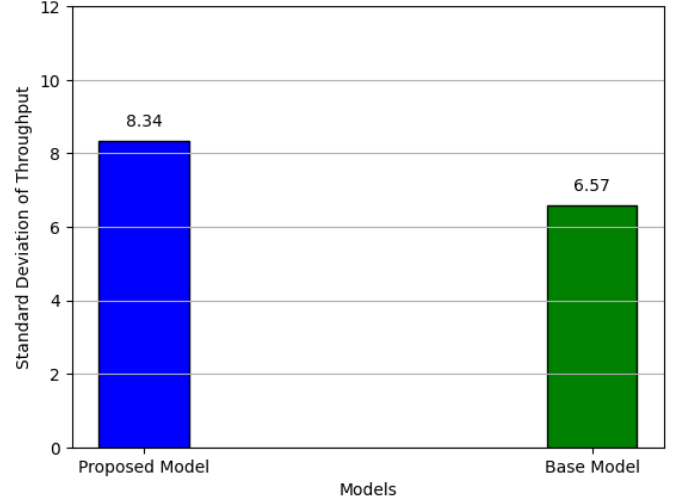


Fig. 10: Consistency of transaction throughput

The performance of the proposed and base models is compared by analyzing the standard deviation of their throughput values. The proposed model shows σ of approximately 8.34, while the base model has a slightly lower σ of about 6.57, as shown in Fig. 10. Although the proposed model exhibits higher variability, it generally achieves a higher throughput, indicating better performance when peak throughput is prioritized. This analysis suggests that the proposed model in terms of achieving a higher transaction throughput.

F. Comparative Analysis of privacy Loss (ϵ)

Privacy Loss (ϵ) is a metric used to quantify the amount of privacy leakage or risk in a system, particularly in the context s involving data privacy and security. A lower (ϵ) value indicates better privacy preservation, as it suggests less information leakage per transaction.

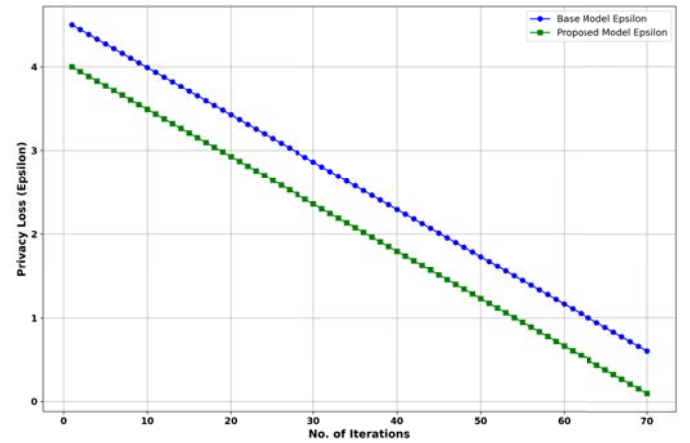


Fig. 11: Analysis for privacy loss (ϵ)

In comparing the performance of the base and proposed models based on their privacy loss (ϵ) values, we observe distinct trends. The base model initially starts with a higher (ϵ) value of around 4.5 and gradually decreases to about 0.6 over

70 iterations, indicating a slower rate of privacy improvement. On the other hand, the proposed model begins with a lower (ϵ) value near 4.0 and decreases more rapidly to approximately 0.1 over the same number of iterations, as shown in Fig. 11. This faster reduction suggests a more effective privacy-preserving mechanism in the proposed model than the base model.

G. Comparing Throughput-Latency Trade-off

In Fig. 12 comparing through and latency trade-offs between the base and the proposed models, we observe clear differences in how they handle the transaction processing. The base model consistently shows lower throughput values at various latency levels compared to the proposed model. This suggests that the proposed model can handle more transactions per unit of time while keeping latency in check better than the base model.

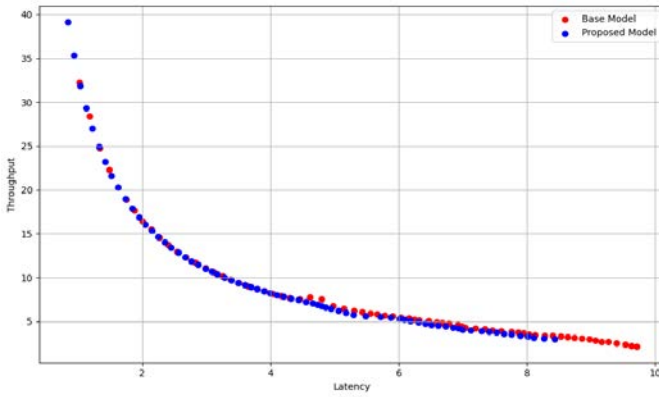


Fig. 12: Throughput-latency Trade-off

H. 95th Percentile Latency Comparison

In performance investigation, the 95th percentile latency is a metric demonstrating the latency value less which 95% of latency amount falls. This measure is crucial for considerate the worst-case performance scenarios that a arrangement strength counter. The bar chart in Fig. 13 compares the 95th percentile latencies of base and proposed models. The base model shows a 95th percentile latency of nearly 9.32 seconds, while the proposed model demonstrates a lower 95th percentile latency of around 7.96 seconds. This specifies that the proposed model not only achieves in terms of average latency but also preserves more reliable performance under high-load circumstances. The lower 95th percentile latency of the proposed model recommends it can holder the widely hold of the transactions more efficiently, making it a more robust excellent for applications with low latency responses.

I. Comparison of Precision, Recall and F1-Score

In the assessment of the machine learning models, precision, recall, and F1-score are crucial to measure the performance widely. Precision is the ratio of the true positive prediction to the total number of predictions, demonstrative the accuracy of the positive prediction, and can be considered using equation (7). Recall is the ratio of the true positive predication to

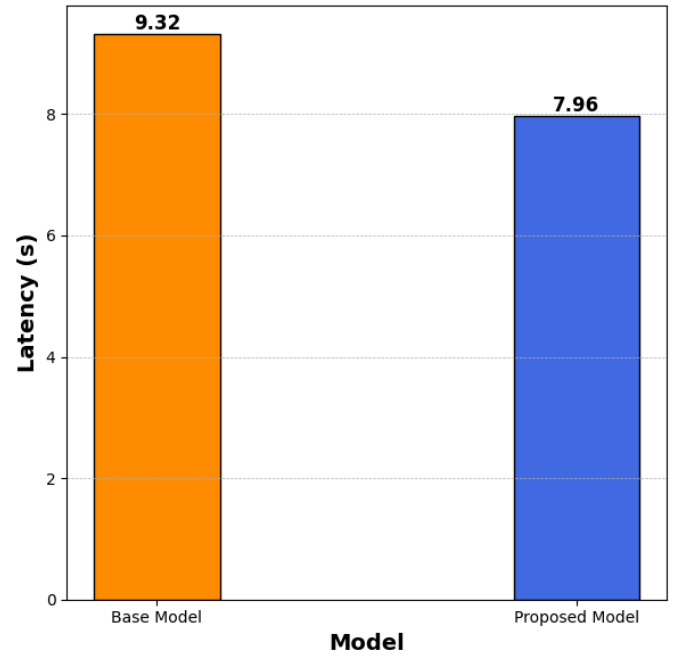


Fig. 13: 95th percentile latency

the total number of actual positives, imitating the model's capability to recognize all the appropriate occurrences as shown in equation (8). The F1-Score is the harmonic mean of precision and recall, utilizing equation (9), providing a balance between the two metrics.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (7)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (8)$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (9)$$

The bar chart in Fig. 14 illustrates the precision, recall, and F1-Score for the Base Model and the Proposed Model. The proposed model outperforms the base model across all three metrics, with a precision of 0.89, a Recall of 0.93, and an F1-Score of 0.91. In contrast, the base model has a precision of 0.80, a recall of 0.83, and an F-Score of 0.82. This shows that the proposed model not only predicts more accurately but also retrieves a higher proportion of the actual positives, achieving a better balance between precision and recall.

J. Linear regression

1) *Correlation*: The correlation coefficient (commonly denoted as r) between variables X and Y in a dataset with n pairs of observations can be calculated using the following equation (10).

$$r = \frac{n \sum_{i=1}^n (X_i Y_i) - (\sum_{i=1}^n X_i) (\sum_{i=1}^n Y_i)}{\sqrt{\left[n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2 \right] \left[n \sum_{i=1}^n Y_i^2 - (\sum_{i=1}^n Y_i)^2 \right]}} \quad (10)$$

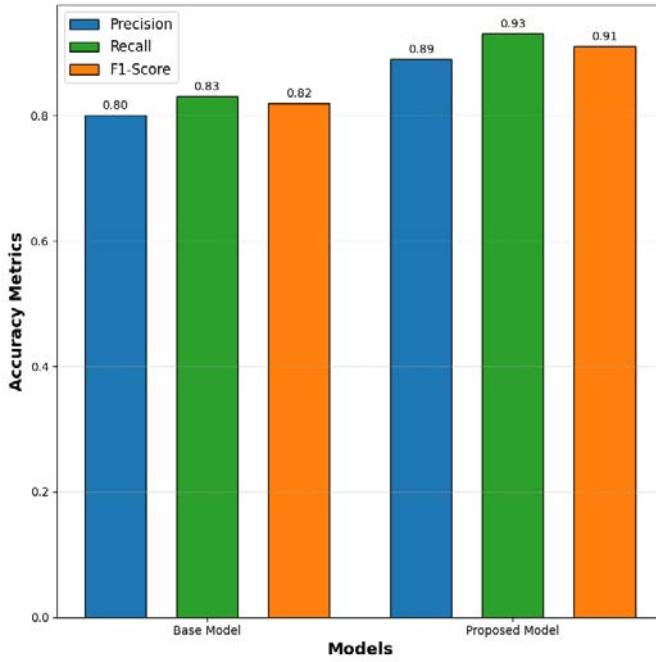


Fig. 14: Comparison of Precision, Recall, and F1-Score

X, Y : Variables for which correlation is being calculated

\sum : The summation symbol.

n : Present the number of data points.

XY : Product of X and Y .

$\sum XY$: Sum of the products of X and Y .

$\sum X, \sum Y$: Sum of X values and Y values, respectively.

2) *Analysis of Variance (ANOVA)*: ANOVA employs the F-statistic to assess the significance of variation between group means relative to variation within groups. One-way ANOVA calculates the F-statistics as in equation (11).

$$F = \frac{MS_{\text{between}}}{MS_{\text{within}}} \quad (11)$$

The general steps for calculating ANOVA involve computing the sum of squares. Degree of freedom and mean square, then use these values to compute the F-statistic.

Throughput Proposed Model: The R2 value indicates that 73% of the variability in the dependent variable, Proposed Model, is explained by the three explanatory variables. P-value of the F-statistic from the ANOVA table and a sig-level of 5%, the explanatory variables provide significantly more explanatory power than a basic mean. This highlights the substantial contribution of the explanatory variables to the model.

Latency Proposed Model: The R-squared value indicates that 49% of the variability in the dependent variable, the Proposed Model, is explained by the three explanatory variables. P-value of the F-statistic in the ANOVA IV and a significance level of 5%, the explanatory variables provide significantly more information than a basic mean.

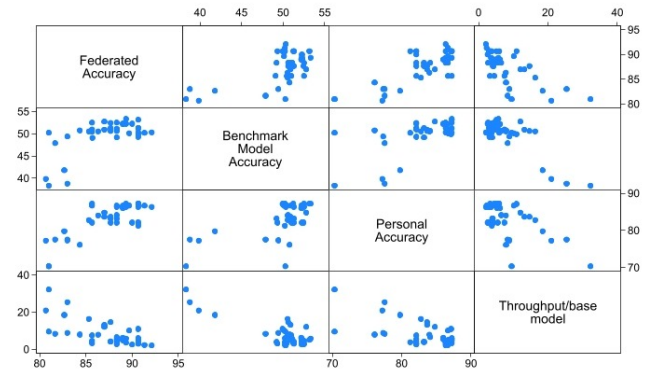


Fig. 15: Throughput-based model correlation with explanatory variables

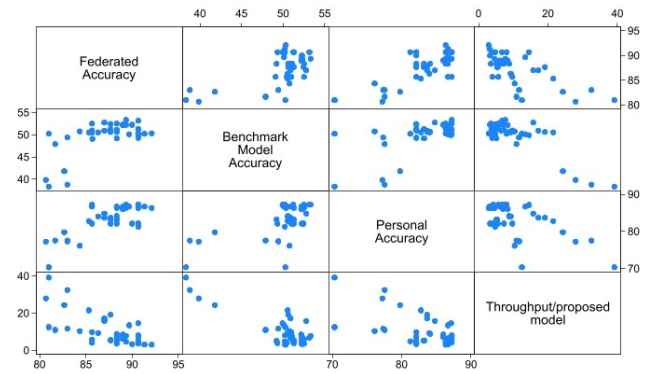


Fig. 16: Throughput proposed model correlation with explanatory variables

Fig. 15 shows the correlation for the throughput of the base model. Fig shows that the relationship of federated accuracy with benchmark model accuracy, personal accuracy, and throughput base model is weakly positive, strongly positive, and weak negative, respectively. The relationship between benchmark model accuracy and personal accuracy and throughput base model is weak and negative. In contrast, the personal accuracy and throughput base models have a strong positive relationship.

Fig. 16 shows the correlation for the throughput of the proposed model. The proposed model has weak negative, personal, and benchmark model accuracy, which has a weak positive relation with federated accuracy.

Fig. 17 and Fig. 18 show the correlation between the latency of the base and the proposed model, respectively. In Fig. 17 benchmark model accuracy has a weak negative, while personal accuracy and latency have a strong positive relationship with federated accuracy. The latency base model has a weak negative, and personal accuracy has a strong positive relation with benchmark model accuracy. Latency-based models and personal accuracy have a weak negative relation. In Fig. 18, the benchmark model and personal accuracy have a weak positive, while the latency proposed model has a strong positive relation with federated accuracy. The latency proposed

TABLE III: Correlation matrix for throughput (proposed model)

	Federated Accuracy	Benchmark Model Accuracy	Personal Accuracy	Proposed Model
Federated Accuracy	1	0.632	0.785	-0.690
Benchmark Model Accuracy	0.632	1	0.627	-0.826
Personal Accuracy	0.785	0.627	1	-0.592
Proposed Model	-0.690	-0.826	-0.592	1

TABLE IV: Analysis of variance for throughput (Proposed Model)

Source	DF	Sum of squares	Mean squares	F	Pr > F	p-values	signification codes
Model	3.000	2035.032	678.344	32.734	< 0.0001		***
Error	36.000	746.025	20.723				
Corrected Total	39.000	2781.057					

TABLE V: Correlation matrix for latency (proposed model)

	Federated Accuracy	Benchmark Model Accuracy	Personal Accuracy	Proposed Model
Federated Accuracy	1	0.632	0.785	0.678
Benchmark Model Accuracy	0.632	1	0.627	0.521
Personal Accuracy	0.785	0.627	1	0.475
Proposed Model	0.678	0.521	0.475	1

TABLE VI: Analysis of variance for latency (Proposed Model)

Source	DF	Sum of squares	Mean squares	F	Pr > F	p-values	signification codes
Model	3.000	97.268	32.423	11.545	< 0.0001		***
Error	36.000	101.099	2.808				
Corrected Total	39.000	198.367					

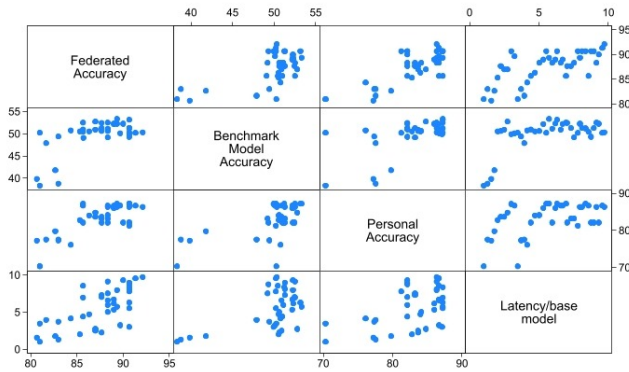


Fig. 17: Latency-based model correlation with explanatory variables

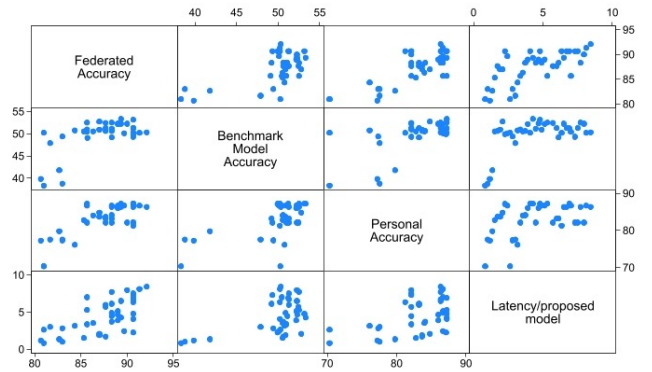


Fig. 18: Latency proposed model correlation with explanatory variables

model has weak negative and personal accuracy and a strong positive relationship with benchmark model accuracy. Latency proposed model and personal accuracy have a weak negative relation.

K. Comparison table

In table VII, we provide a side-by-side comparison of the base scheme, referenced from [10], and our proposed scheme, highlighting key performance metrics and security features.

VI. CONCLUSION AND FUTURE PROSPECTS

This research introduces a resilient, secure, and zero-trust architecture for exchanging sensitive information in IoT net-

works, ensuring the highest level of privacy preservation throughout the process. The approach integrates a robust P2P communication channel to guarantee the safe transmission of sensitive data. FL empowers models to operate locally and precisely on the data owner's side, avoiding the need for central data aggregation. LDP further fortifies privacy by offering robust guarantees to each user during data collection and analysis. In the proposed framework, data owners have exclusive authority over implementing LDP, ensuring robust data privacy protection within a zero-trust architecture for sharing their data sets. Blockchain technology, which fosters a decentralized environment, effectively eliminates the vulnerability of a single point of failure. The blockchain meticulously

TABLE VII: Comparison Table for Base and Proposed Scheme

Achievements	Base Scheme [10]	Proposed Scheme
Throughput (tps)	8	10.53
Latency (s)	5.6	4
Accuracy (%)	49%	85%
Security	Weak	Registration of entities enhances more security
Privacy	Weak privacy technique, background knowledge attack was possible	No privacy threat due to FL and LDP
Registration of entities	No	Yes
Privacy technique	K-anonymity	FL and DP
Security technique	Blockchain without registration of entities	Blockchain with registration of entities
Application of privacy technique	Data controller was applying anonymization technique on trust-based relation	Data owner itself will anonymization technique due to data controller can be an adversary

records the registration of different entities, bids, offers, and prices, while the data set specifications are saved in a decentralized file system, i.e., IPFS. The proposed architecture's fundamental performance metrics revolve around throughput, latency, privacy, accuracy, ANOVAN, and correlation matrix, all of which are essential for ensuring its effectiveness. The results of our comparison metrics indicate the proposed model's significant advantages over the base-scheme model. We intend to explore the integration of quantum cryptography to enhance data security further. Additionally, we plan to investigate the application of advanced machine-learning techniques to improve the accuracy and efficiency of our models. Further studies will also focus on optimizing performance metrics to ensure scalability in larger environments and leveraging generative AI to create synthetic datasets for more robust model training and validation.

ACKNOWLEDGMENT

Altat Hussain and Tariq Hussain contributed equally to this work and are the first co-authors.

REFERENCES

- [1] Tinashe Magara and Yousheng Zhou. Internet of things (iot) of smart homes: Privacy and security. *Journal of Electrical and Computer Engineering*, 2024(1):7716956, 2024.
- [2] Leong Yee Rock, Farzana Parveen Tajudeen, and Yeong Wai Chung. Usage and impact of the internet-of-things-based smart home technology: a quality-of-life perspective. *Universal access in the information society*, 23(1):345–364, 2024.
- [3] Taotao Wang, Shengli Zhang, Qing Yang, and Soung Chang Liew. Account service network: a unified decentralized web 3.0 portal with credible anonymity. *IEEE Network*, 2023.
- [4] Ado Adamou Abba Ari, Olga Kengni Ngangmo, Chafiq Titouna, Ousmane Thiare, Alidou Mohamadou, and Abdelhak Mourad Gueroui. Enabling privacy and security in cloud of things: Architecture, applications, security & privacy challenges. *Applied Computing and Informatics*, 20(1/2):119–141, 2024.
- [5] Ziwei Liu, Ziyu Xu, Xiyu Zheng, Yongxing Zhao, and Jinghua Wang. 3d path planning in threat environment based on fuzzy logic. *Journal of Intelligent & Fuzzy Systems*, (Preprint):1–14, 2024.
- [6] Dezhi Han, HongXu Zhou, Tien-Hsiung Weng, Zhongdai Wu, Bing Han, Kuan-Ching Li, and Al-Sakib Khan Pathan. Lmca: a lightweight anomaly network traffic detection model integrating adjusted mobilenet and coordinate attention mechanism for iot. *Telecommunication Systems*, 84(4):549–564, 2023.
- [7] Chunhuan Ni, Li Shan Cang, Prosanta Gope, and Geyong Min. Data anonymization evaluation for big data and iot environment. *Information Sciences*, 605:381–392, 2022.
- [8] Pejman Panahi, Cüneyt Bayılmış, Unal Çavuşoğlu, and Sezgin Kaçar. Performance evaluation of lightweight encryption algorithms for iot-based applications. *Arabian Journal for Science and Engineering*, 46(4):4015–4037, 2021.
- [9] Junqi Guo, Minghui Yang, and Boxin Wan. A practical privacy-preserving publishing mechanism based on personalized k-anonymity and temporal differential privacy for wearable iot applications. *Symmetry*, 13(6):1043, 2021.
- [10] Doderio JM Rodriguez-Garcia M, Sicilia MA. A privacy-preserving design for sharing demand-driven patient datasets over permissioned blockchains and p2p secure transfer. *PeerJ Computer Science*, 7:e568, 2021.
- [11] Brijesh B Mehta and Udai Pratap Rao. Improved l-diversity: scalable anonymization approach for privacy preserving big data publishing. *Journal of King Saud University-Computer and Information Sciences*, 34(4):1423–1430, 2022.
- [12] Xuefei Yin, Yanming Zhu, and Jiankun Hu. A comprehensive survey of privacy-preserving federated learning: A taxonomy, review, and future directions. *ACM Computing Surveys (CSUR)*, 54(6):1–36, 2021.
- [13] Natarajan Deepa, Quoc-Viet Pham, Dinh C Nguyen, Sweta Bhat-tacharya, B Prabadevi, Thippa Reddy Gadekallu, Praveen Kumar Reddy Maddikunta, Fang Fang, and Pubudu N Pathirana. A survey on blockchain for big data: Approaches, opportunities, and future directions. *Future Generation Computer Systems*, 131:209–226, 2022.
- [14] Eduardo B Fernandez and Andrei Brazhuk. A critical analysis of zero trust architecture (zta). *Computer Standards & Interfaces*, 89:103832, 2024.
- [15] Samia Masood Awan, Muhammad Ajmal Azad, Junaid Arshad, Urooj Waheed, and Tahir Sharif. A blockchain-inspired attribute-based zero-trust access control model for iot. *Information*, 14(2):129, 2023.
- [16] Keping Yu, Liang Tan, Moayad Aloqaily, Hekun Yang, and Yaser Jararweh. Blockchain-enhanced data sharing with traceable and direct revocation in iiot. *IEEE transactions on industrial informatics*, 17(11):7669–7678, 2021.
- [17] Shuxin Shi, Dezhi Han, and Mingming Cui. A multimodal hybrid parallel network intrusion detection model. *Connection Science*, 35(1):2227780, 2023.
- [18] Xuanmei Qin, Yongfeng Huang, Zhen Yang, and Xing Li. A blockchain-based access control scheme with multiple attribute authorities for secure cloud data sharing. *Journal of systems architecture*, 112:101854, 2021.
- [19] Ibrar Yaqoob, Khaled Salah, Raja Jayaraman, and Yousof Al-Hammadi. Blockchain for healthcare data management: opportunities, challenges, and future recommendations. *Neural Computing and Applications*, pages 1–16, 2022.
- [20] Bessem Zaabar, Omar Cheikhrouhou, Faisal Jamil, Meryem Ammi, and Mohamed Abid. Healthblock: A secure blockchain-based healthcare data management system. *Computer Networks*, 200:108500, 2021.
- [21] Ahmed El Oudrhiri and Ahmed Abdelhadi. Differential privacy for deep and federated learning: A survey. *IEEE access*, 10:22359–22380, 2022.
- [22] Chen Zhang, Yu Xie, Hang Bai, Bin Yu, Weihong Li, and Yuan Gao. A survey on federated learning. *Knowledge-Based Systems*, 216:106775, 2021.
- [23] Tuo Zhang, Lei Gao, Chaoyang He, Mi Zhang, Bhaskar Krishnamachari, and A Salman Avestimehr. Federated learning for the internet of things: Applications, challenges, and opportunities. *IEEE Internet of Things Magazine*, 5(1):24–29, 2022.
- [24] Haoyuan Cheng, Tianguang Lu, Ran Hao, Jiamei Li, and Qian Ai. Incentive-based demand response optimization method based on federated learning with a focus on user privacy protection. *Applied Energy*, 358:122570, 2024.
- [25] Barjinder Kaur, Sajjad Dadkhah, Farzaneh Shoeleh, Euclides Carlos Pinto Neto, Pulei Xiong, Shaheer Iqbal, Philippe Lamontagne, Suprio Ray, and Ali A Ghorbani. Internet of things (iot) security dataset evolution: Challenges and future directions. *Internet of Things*, page 100780, 2023.
- [26] Laraib Javed, Adeel Anjum, Bello Musa Yakubu, Majid Iqbal, Syed Atif Moqurrah, and Gautam Srivastava. Sharechain: Blockchain-enabled model for sharing patient data using federated learning and differential privacy. *Expert Systems*, 40(5):e13131, 2023.
- [27] Prayag Tiwari, Abdullah Lakhani, Rutvij H. Jhaveri, and Tor-Morten Grønli. Consumer-centric internet of medical things for cyborg applica-

- tions based on federated reinforcement learning. *IEEE Transactions on Consumer Electronics*, 69(4):756–764, 2023.
- [28] Mohammad Kamrul Hasan, Taher M Ghazal, Rashid A Saeed, Bishwajeet Pandey, Hardik Gohel, Ala'A Eshmawi, Sayed Abdel-Khalek, and Hula Mahmoud Alkhassawneh. A review on security threats, vulnerabilities, and counter measures of 5g enabled internet-of-medical-things. *IET Communications*, 16(5):421–432, 2022.
 - [29] W Nicholson Price and I Glenn Cohen. Privacy in the age of medical big data. *Nature medicine*, 25(1):37–43, 2019.
 - [30] Ken Miyachi and Tim K Mackey. hocbs: A privacy-preserving blockchain framework for healthcare data leveraging an on-chain and off-chain system design. *Information Processing & Management*, 58(3):102535, 2021.
 - [31] Kapal Dev, Chih-Lin , and Sunder Ali Khowaja. Guest editorial dense - data integrity, integration and security issues for consumer data in industry 5.0. *IEEE Transactions on Consumer Electronics*, 69(4):809–812, 2023.
 - [32] Ibrar Yaqoob, Khaled Salah, Raja Jayaraman, and Yousof Al-Hammadi. Blockchain for healthcare data management: opportunities, challenges, and future recommendations. *Neural Computing and Applications*, pages 1–16, 2021.
 - [33] Sangjukta Das and Suyel Namasudra. A novel hybrid encryption method to secure healthcare data in iot-enabled healthcare infrastructure. *Computers and Electrical Engineering*, 101:107991, 2022.
 - [34] Mohammad Fareed and Ali A Yassin. Privacy-preserving multi-factor authentication and role-based access control scheme for the e-healthcare system. *Bulletin of Electrical Engineering and Informatics*, 11(4):2131–2141, 2022.
 - [35] Abdel Mlak Said, Aymen Yahyaoui, and Takoua Abdellatif. Efficient anomaly detection for smart hospital iot systems. *Sensors*, 21(4):1026, 2021.
 - [36] Dhananjay Nigam, Shilp Nirajbhai Patel, PM Raj Vincent, Kathiravan Srinivasan, Sinouvassane Arunmozhi, et al. Biometric authentication for intelligent and privacy-preserving healthcare systems. *Journal of Healthcare Engineering*, 2022, 2022.
 - [37] Li Zhang, Jianbo Xu, Pandi Vijayakumar, Pradip Kumar Sharma, and Uttam Ghosh. Homomorphic encryption-based privacy-preserving federated learning in iot-enabled healthcare system. *IEEE Transactions on Network Science and Engineering*, 2022.
 - [38] Yan Zhuang, Chi-Ren Shyu, Shenda Hong, Pengfei Li, and Luxia Zhang. Self-sovereign identity empowered non-fungible patient tokenization for health information exchange using blockchain technology. *Computers in Biology and Medicine*, 157:106778, 2023.
 - [39] Teri Lenard and Roland Bolboaca. A statefull firewall and intrusion detection system enforced with secure logging for controller area network. In *European Interdisciplinary Cybersecurity Conference*, pages 39–45, 2021.
 - [40] Maida Ahtesham. Bigdata applications in healthcare: Security and privacy challenges. In *International Conference on Digital Technologies and Applications*, pages 231–240. Springer, 2022.
 - [41] Mercedes Rodriguez-Garcia, Miguel-Angel Sicilia, and Juan Manuel Doderio. A privacy-preserving design for sharing demand-driven patient datasets over permissioned blockchains and p2p secure transfer. *PeerJ Computer Science*, 7:e568, 2021.
 - [42] Rajiv Bagai, Eric Weber, and Vikas Thammanna Gowda. Data sanitization for t-closeness over multiple numerical sensitive attributes. 2023.
 - [43] Keiichiro Oishi, Yuichi Sei, Yasuyuki Tahara, and Akihiko Ohsuga. Semantic diversity: Privacy considering distance between values of sensitive attribute. *Computers & Security*, 94:101823, 2020.