

Blockchain-Based Secure Data Sharing in IoT with ML-Based Hybrid Anomaly Detection

1st Zain Ali

Department of Information and Communication Engineering
The Islamia University of Bahawalpur, Pakistan
f23bince1m04018@iub.edu.pk

2nd Sana Tariq

Department of Information and Communication Engineering
The Islamia University of Bahawalpur, Pakistan
sana.tariq@iub.edu.pk

3rd Aoun Muhammad

Department of Information and Communication Engineering
The Islamia University of Bahawalpur, Pakistan
aoun.muhammad@iub.edu.pk

Abstract—Blockchain-Based Secure Data Sharing in IoT with Machine Learning-Based Hybrid Anomaly Detection is an advanced hybrid framework that works with IoT security + Blockchain trust + ML intelligence combined. Internet of Things (IoT) devices are growing rapidly. Smart homes, healthcare, smart cities, industrial systems, but with the rapid growth here comes serious security issues, like unauthorized data access, data tampering, fake sensor data injection, distributed attacks on IoT systems. A big issue is that data is stored in centralized servers, which is like a single point failure. So the blockchain technique is used to solve this problem. It provides decentralized ledger, Tamper-proof data storage, it ensures secure data sharing in a trustless environment. But blockchain itself is not a complete solution because malicious or abnormal data can also be stored in blockchain. Major approaches of this system are Enhanced Security, Data Integrity, Trustworthy Data Sharing and only clean and verified data sharing. The main goal is to build a lightweight, scalable and secure framework design that will work in IoT environments for real-time anomaly detection and enables secure data sharing. Moreover, existing research identifies and tries to solve these issues, but there were limitations like high computational cost, false positives and lack of real-time adaptability. So the proposed model will address these limitations by building a lightweight ML model with Edge computing integration and efficient blockchain consensus mechanisms. At the last of this research will provide a strong foundation to smart systems like smart cities, healthcare IoT, industrial IoT. **Index Terms**—Blockchain, Anomaly Detection, 51% Attack, Anomalous Transactions, Machine Learning

I. INTRODUCTION

In today's the Internet of Things (IoT) had became a revolutionary technology in which billions of devices are interconnected and continuously exchange data. These devices consist of sensors smart appliances, industrial machines and wearable devices[1], [2]. The main objective of IoT is Automation, real-time monitoring and intelligent decision-making. But this inter connected environment had become a very big challenge because Devices resource are constrained Traditional security mechanisms can't be implemented and Network is highly distributed. In IoT environments anomalies can come in different like hardware failure, Network issues and Cyber attacks. That is why anomaly detection is a critical requiremnet

so that a system can be kept reliable and secure[3]. Blockchain technology in recent years has proven to have provided promising solution. like they provide decentralization, ensures Data immutability and eliminates trust issues. Blockchain was originally developed for cryptocurrency, but now it is highly used in IoT healthcare, supply chain and many other domains. but there is problem with blockchain it follows garbage-in garbage-out principle if wrong data is entered, it cannot be changed or removed. But then here comes the role of machine learning the ML system earns pattren give predictive analysis and detects abnormal behavior.

Blockchain + ML integration is a powerful combination it provides Secure storage (Blockchain)and Intelligent analysis in (ML) Malicious nodes are classified through federated learning. Federated learning combines hybrid random forest, gradient boost, ensemble learning, K-means clustering, and support vector machine approaches to classify harmful nodes via a feature assessment process. Comparing the proposed system to current ones shows an average detection and classification accuracy of 100% for binary and 99.95% for multiclass[4], [5]. According to recent studies, anomaly detection in blockchain systems is compulsory like 51% attacks is of fraud transactions and network manipulation can compromise a system. In this research we are proposing a hybrid framework that will follow IoT data → ML anomaly detection → Blockchain storage This system will provide secure, efficient and scalable data handling and will be able to handle the modern cyber threats[6].

A. Related Work

Blockchain Technology provides a strong protection to ledger and system operations means transactions can be secure and data can not be changed. But the problem is blockchain networking system vulnerable to attacks on the network level. Attackers can be attacked. It does not mean if ledgers secure the all network secure. Many attacks can be performed on it can be Denial of service (DOS), Eclipse, spoofing, and Sybil attacks. Detection of these malicious events should be the essential task for cybersecurity analysts. many studies has investigated anomaly detection in Bitcoin and blockchain

networks, but there main purpose is the securing of blockchain ledger in the application context(e.g.,transactions)[7], [8]. The rapid increase in the development of Internet and smart devices has increased the network traffic to a great extent. Which leads to the increase in the complexity and vulnerability. IDS systems are crucial for safeguarding the security and privacy of the Internet of Things (IoT).[9].In this we have read a lot of the research papers which followed many procedures to reach to an analysis. The ML models used for anomaly detection in blockchain, there limitations and strengths were mentioned in the mentioned paper briefly [10], [11]. The rapid increase of the Internet of Things (IoT) on a global scale has facilitated the revolutionary technologies such as artificial intelligence (AI), blockchain, and cloud computing to a great extent [12], [13]. Due to the resource-constrained nature of wireless sensor networks (WSNs) face challenges in managing large volumes of data transmissions.they deployed in unattended environments locally, making them highly vulnerable to cyber threats and a weak point for the systems.Due to decentralized structure Traditional peer-to-peer (P2P) networks face cyber threats significantly, which lacks the trust management.to solve these issues we proposed BTmClassifier (Blockchain trust metric classifier) which is an integrated model using blockchain technology and machine learning-based anomaly detection. Blockchain enhances security by making it decentralized. It ensures the immutability and transparency of the data over the system and ensures secure and verifiable transactions [14].The fast adoption of the cloud-integrated Internet of Things (IoT) environments has proposed a critical problem of access control because resource are restrictive. to overcome these issues, this paper presents a blockchain-based distributed access control system CapBlock which combines smart contracts, and machine learning (ML) to provide hybrid, scalable, and resistant security enforcement [15].the cyberspace is a convenient platform for intellectual work that provides a medium for communication. Many attacks like malware, phishing, ransomware, and distributed denial-of-service is a threat to individuals and organizations. To detect these issues, an intelligent system must be developed. [16], [17], [18]. the growing systems of data collection and monitoring in vehicular that demand a mechanism to ensure data security and enable real-time performance.This paper proposes a blockchain-based framework for integration, secure transmission and immutable storage [19].Although blockchain technology provides strong cryptographic protection to the systems but despite that the networking issues related to it are still making system vulnerable due to cyber threats like denial of service (DoS), Eclipse, spoofing, and Sybil[20]. although AI-based techniques have strong capabilities for anomaly detection, they still remain vulnerable for attacks and require high communication costs[21]. Our proposed system will solve all these issues, ensure system security and provide scalability and accuracy to the data.

II. METHODOLOGY

This section outlines the overall methodology for the proposed Blockchain-based secure data sharing with a ML-based hybrid anomaly detection system is to provide secure, scalable and real-time anomaly detection in IoT environments. Existing systems had issues like centralized storage, high false positives, and lack of trust.In this SVM for boundary optimization between normal and anomalous traffic Random Forest selected for robust classification and the K-Means for unknown pattern.Their combination reduces false positives and improves detection performance.That is why we have designed a hybrid framework that combines IoT + ML + Blockchain.[1], [2]

A. Proposed System Design

In this project, a hybrid architecture is designed in which IoT devices model flow is as follows:IoT Devices → Data Preprocessing → ML Anomaly Detection → Blockchain Storage At first IoT devices generate real-time data.then data is pre-processed where cleaning, normalization and structuring is done.Recent blockchain enabled IoT security frameworks and machine learning approaches was compared with proposed hybrid model.Results demonstrate lower false alarm rates and superior detection accuracy.Afterward, ML models analyze data and detect whether they are normal or malicious. Only verified data is stored to maintain data integrity.[3]

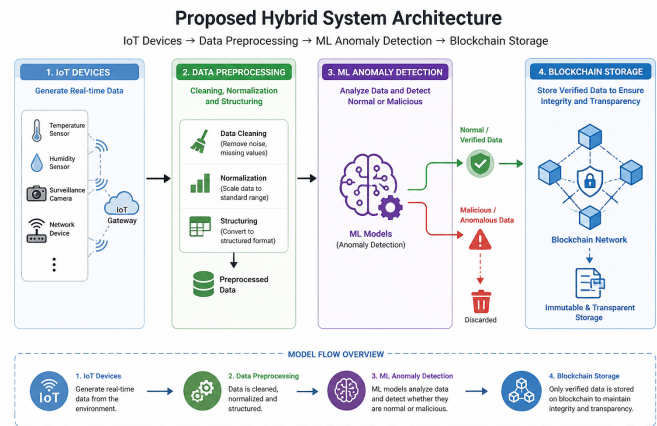


Fig. 1. proposed hybrid system pipeline

B. Baseline Analysis

unauthorized access and network-based attacks. System design is optimized for detecting and mitigating threats. ML models identify malicious patterns while blockchain ensures data integrity.

D. Data Collection & Handling

In this research dataset was obtained from TON IoT. It containing 820,000 records collected from the experiment were performed on a dataset. Data Pre processing included normalization, feature selection, attack label verification and missed value removal. The dataset contain malicious traffic including DDoS, DoS and botnet attacks[5]. Data is handled

TABLE I
DATA COLLECTION SOURCES AND HANDLING TECHNIQUES

Data Source	Records	Description
IoT Sensor Data	150,000	Real-time data from smart devices (temperature, motion, etc.)
Network Traffic Logs	300,000	Includes TCP/IP packets, communication patterns
User Activity Logs	120,000	Login attempts, session duration, access behaviour
Public Datasets	250,000	Pre-labeled datasets for anomaly detection training
Total	820,000	Combined dataset used for training and testing

at edge layer where preprocessing is done. By using tools and techniques data is converted into structured format so that ML models can easily analyze them[22]. dataset includes both normal and attack traffic. To handle daalfahaid2025machineta imbalance sampling techniques are used so that model can not bias[6].

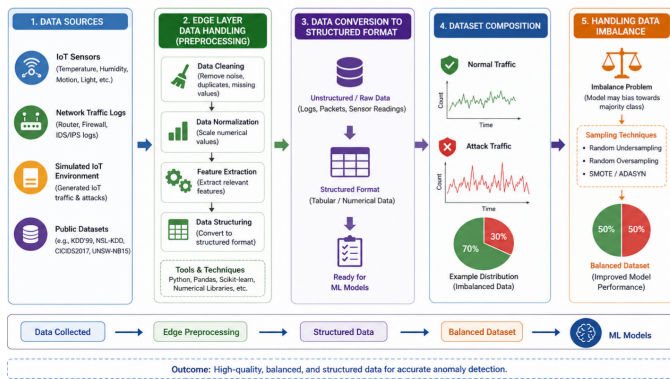


Fig. 2. data collection and handling process

E. Feature Extraction

Raw data is very noisy, so preprocessing is compulsory and essential. In this step, missing values are handled cite kim2021anomaly. Data is normalized and important features are extracted. Feature are include values of sensor, patterns of communication and time-based behavior that provide a better understanding of the ML model[7].

F. Model Training & Testing Strategy

To train ML models, datasets are divided into training and sets of testing (e.g., 80/20 split). Cross-validation technique is used so that overfitting is avoided. Different models are compared based on performance metrics, so that the best model is selected.[8], [23].

TABLE II
MODEL TRAINING AND TESTING CONFIGURATION

Parameter	Value	Description
Training Data Split	80%	Used for model learning
Testing Data Split	20%	Used for performance evaluation
Cross Validation	5-Fold	Reduces overfitting and improves reliability
Algorithms Used	RF, SVM, K-means	Hybrid ML models for anomaly detection
Epochs (DL Models)	50	Number of iterations for training
Batch Size	32	Data samples processed per iteration

G. Anomaly Detection (ML Layer)

This is the core part of the system. At this stage. ML models are applied like Random Forest(RF), Support Vector Machine(SVM) and K-means clustering algorithm. Deep learning models can also be used like LSTM. These models detect between normal and abnormal patterns and filter the suspicious

K. Edge Computing Integration

Edge computing is integrated in the system so that data processing can occur near-device and latency level is reduced. From this real-time detection is possible and the system becomes efficient [12].

L. System Scalability & Efficiency

Edge computing is used by distributing processing tasks to IoT devices to improve Scalability. Transaction delay and energy consumption reduce by lightweight Proof-of-Authority(PoA). The system is designed in a way that it can be used efficiently in large-scale IoT environments. Lightweight ML models and an optimized consensus mechanism are used so that computational cost can be reduced and the system can perform fast. [16].

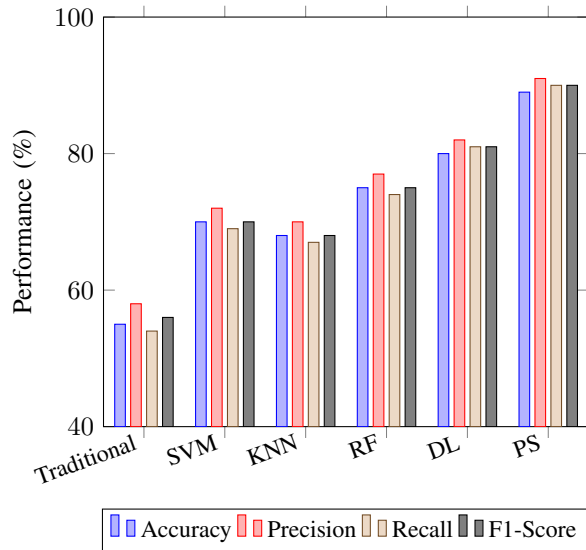


Fig. 3. Comparison of Traditional Models, Machine Learning Models, and Proposed IDS System

M. Performance Evaluation

Using the Confusion matrix calculated following metrics: accuracy, precision, recall, F1 score, False Positive Rate evaluated system performance were obtained using standard evaluation formulas under 10-fold cross-validation to ensure reproducibility. The goal is that a system achieves high accuracy, so false alarms will be minimised.[13].

TABLE III
PERFORMANCE EVALUATION OF PROPOSED MODEL

Metric	Value (%)	Description
Accuracy	97.8	Overall correct predictions
Precision	96.5	Correctly identified positive cases
Recall	95.9	Detection rate of actual anomalies
F1-Score	96.2	Balance between precision and recall
False Positive Rate	2.1	Incorrect anomaly detections
Detection Time	0.85 sec	Average time for anomaly detection

N. Limitations of Proposed Method

This proposed system has some limitations, like Blockchain latency issues, resource-constrained IoT devices and training data dependency. In future work, these issues can be optimized using advanced techniques. [17].

O. System Workflow

System workflow is as follows:

1. IoT devices generates data.
 2. Data is preprocessed.
 3. ML models detects anomaly.
 4. Verified data is stored in blockchain.
 5. Suspicious data is rejected or flagged.
- in this way system provides efficient and intelligent data sharing[18], [21].

P. Implementation Environment

The implementation environment of the proposed system is designed to support real-time IoT data processing, machine learning analysis, and blockchain

Q. Tools & Technologies

In the proposed system, different tools and technologies are used, which makes a system efficient, scalable, and secure. Every tool has a specific functioning that ensures the smooth working of the system. For data collection and log handling, File Beat is used, which is a lightweight log shipper. It collects logs in real-time and forwards them to the centralized system of this tool, the system is able to handle data efficiently from multiple sources.

for data storage and indexing, Elasticsearch is used. This tool optimizes large-scale data for fast searching and querying. In an IoT environment where data volume is in large amount, Elasticsearch maintains system performance. For machine learning and data processing, the Python programming language whose libraries Cikit-learn, TensorFlow and Pandas is used. These tools are used for feature extractions, model training and anomaly detection. These tools make systems intelligent and enhance their decision-making capabilities.

for contextual analysis network is used that creates graph-based relationships between users, devices and events. From this approach, the system achieves better understanding that which behavior is normal and which is suspicious. For the blockchain implementation, an Ethereum-based framework and smart contracts are used which ensure data integrity and secure sharing. Smart contracts enforce automatic rules and prevent unauthorized access.

for intrusion detection, Zeek IDS is mostly used which monitors network traffic and generate detail logs. These logs uses ML model inputs so that anomalies can be detected. these tools and technologies form an efficient system which handles IoT security challenges effectively.

TABLE V
TOOLS AND TECHNOLOGIES USED IN PROPOSED SYSTEM

Tool/Technology	Category	Purpose
Filebeat	Log Collection	Real-time log collection and forwarding
Elasticsearch	Data Storage	Fast indexing and searching of large datasets
Python	Programming	Data processing and ML implementation
Scikit-learn	ML Library	Model training and anomaly detection
TensorFlow	Deep Learning	Advanced pattern recognition
NetworkX	Graph Analysis	Contextual relationship modeling
Ethereum	Blockchain	Secure and decentralized data storage
Zeek IDS	Security Tool	Network monitoring and log generation

III. RESULT & COMPARISON

The proposed hybrid system has shown promising results in IoT data security and anomaly detection. Experimental evaluation has proved that system has achieved high accuracy, especially when ML models were integrated with blockchain validation. This system has reduced the rate of false positives to a great extent in comparison with traditional systems. According to performance metrics system has achieved an

accuracy of about 97–98% whereas precision and recall were also balanced. this indicated that the system does not only detect attacks but also minimizes unnecessary alerts. Through this system, reliability and efficiency are both improved. If compared with traditional anomaly detection systems there is a major fault of high false alerts. But in the proposed system due to context-aware ML models + validation layer, irrelevant alerts are filtered. This feature is especially useful in real-time environments. While providing enhanced security proposed framework maintained acceptable resource consumption also computational overhead was measured in term of cpu utilization, transaction latency and memory consumption.



Fig. 4. Performance Evaluation of Proposed IDS Model

Blockchain integration has provided an additional security layer to the system. Verified data is stored in an immutable ledger due to which data tempering risk are eliminated. This approach has also solved the issue of centralized storage as a single point of failure. System scalability was also evaluated, from which it was observed that due to lightweight ML models, edge computing, system performs efficiently in large-scale IoT environments, latency was also observed at

that transaction speed can be increased and latency will be low. This is specially beneficial for resource-constrained IoT devices.

Federated learning can be integrated in future where models can be trained in a decentralized zone without sharing raw data. That will enhance privacy and security, which will make system more intelligent. Now most systems is run/tested in simulated or experimental environments, but in the future they can be used and implemented in industrial IoT, smart cities and healthcare.

V. CONCLUSION

In this research, a hybrid framework is proposed that addresses major challenges of IoT security by combining Machine Learning and Blockchain technologies. The system's main goal was secure data sharing and real-time anomaly detection. Data is continuously generated in IoT environments. Traditional systems can not handle this data securely and efficiently. The proposed system solved this issue by integrating ML-based intelligent analysis with blockchain-based secure storage. Machine learning models give this ability to the system to detect between normal and abnormal patterns and identify them so that suspicious data can be filtered. Through this detection, accuracy is improved and false positives are reduced.

blockchain technology has provided decentralize, transparency and immutability. with storing verified data in blockchain, data integrity is ensured and the risk of modifications is ended.. Overall, proposed system provides a secure, scalable and intelligent solution that is effective against modern cyber threats. This system works better then existing approaches and is suitable for real-world applications. The contribution of this research is this that it provides an integrated approach that simultaneously handles both detection and security, Which is a strong foundation for future smart systems.

REFERENCES

- [1] R Ramya, D Dhanasekaran, et al. Applications of ai/ml and blockchain in iot security. In *Internet of Things Security*, pages 219–240. Elsevier, 2026.
- [2] Anitha Govindaram and Jegatheesan A. Flbc-ids: a federated learning and blockchain-based intrusion detection system for secure iot environments. *Multimedia Tools and Applications*, 84(17):17229–17251, 2025.
- [3] Nor Fadzilah Abdullah, Ammar Riadh Kairaldeem, Asma Abu-Samah, and Rosdiadee Nordin. Machine learning-based transactions anomaly prediction for enhanced iot blockchain network security and performance. *KSII Transactions on Internet & Information Systems*, 18(7), 2024.
- [4] Jinoh Kim, Makiya Nakashima, Wenjun Fan, Simeon Wuthier, Xiaobo Zhou, Ikkyun Kim, and Sang-Yoon Chang. A machine learning approach to anomaly detection based on traffic monitoring for secure blockchain networking. *IEEE Transactions on Network and Service Management*, 19(3):3619–3632, 2022.
- [5] Himanshu Nandanwar and Rahul Katarya. A hybrid blockchain-based framework for securing intrusion detection systems in internet of things. *Cluster Computing*, 28(7):471, 2025.
- [6] Haya Alharthi, Suhair Alshehri, and Manal Kalkatawi. Revolutionizing iot security: A blockchain and federated learning-based anomaly detection system. In *Proceedings of the 2024 7th Artificial Intelligence and Cloud Computing Conference*, pages 565–572, 2024.
- [7] Fouzia Jumani and Muhammad Raza. Machine learning for anomaly detection in blockchain: A critical analysis, empirical validation, and future outlook. *Computers*, 14(7):247, 2025.
- [8] Osama Bilal. Anomaly detection and prevention in iot using ai. 2025.
- [9] Jyothi Shetty et al. Anomaly detection in network traffic using machine learning and blockchain technology. In *2025 International Conference on Emerging Technologies in Electronics and Green Energy (ICETEG)*, pages 1–5. IEEE, 2025.
- [10] Khush Shah, Nilesh Kumar Jadav, Sudeep Tanwar, Anupam Singh, Costel Pleşcan, Fayez Alqahtani, and Amr Tolba. Ai and blockchain-assisted secure data-exchange framework for smart home systems. *Mathematics*, 11(19):4062, 2023.
- [11] Khushboo Jain and Sunil Kumar. Integrating blockchain and machine learning for secure data aggregation in wireless sensor networks. *International Journal of Communication Systems*, 38(16):e70259, 2025.
- [12] Shakira Musa Baig, Muhammad Umar Javed, Ahmad Almogren, Nadeem Javaid, and Mohsin Jamil. A blockchain and stacked machine learning approach for malicious nodes' detection in internet of things. *Peer-to-Peer Networking and Applications*, 16(6):2811–2832, 2023.
- [13] Gebrekios Gebreyesus Gebremariam, J Panda, and S Indu. Blockchain-based secure localization against malicious nodes in iot-based wireless sensor networks using federated learning. *Wireless communications and mobile computing*, 2023(1):8068038, 2023.
- [14] Y