Anomaly Detection in IIoT Transactions using Machine Learning: A Lightweight Blockchain-based Approach

Mayar Ibrahim Hasan Okfie
Department of Information Technology, College of Computer and Information Sciences, Majmaah University, Saudi Arabia
441103734@s.mu.edu.sa (corresponding author)

Shailendra Mishra
Department of Information Technology, College of Computer and Information Sciences, Majmaah University, Saudi Arabia
s.mishra@mu.edu.sa

Received: 29 March 2024 | Revised: 16 April 2024 | Accepted: 25 April 2024

 DOI: https://doi.org/10.48084/etasr.7384

ABSTRACT

The integration of secure message authentication systems within the Industrial Internet of Things (IIoT) is paramount for safeguarding sensitive transactions. This paper introduces a Lightweight Blockchain-based Message Authentication System, utilizing k-means clustering and isolation forest machine learning techniques. With a focus on the Bitcoin Transaction Network (BTN) as a reference, this study aims to identify anomalies in IIoT transactions and achieve a high level of accuracy. The feature selection coupled with isolation forest achieved a remarkable accuracy of 92.90%. However, the trade-off between precision and recall highlights the ongoing challenge of minimizing false positives while capturing a broad spectrum of potential threats. The system successfully detected 429,713 anomalies, paving the way for deeper exploration into the characteristics of IIoT security threats. The study concludes with a discussion on the limitations and future directions, emphasizing the need for continuous refinement and adaptation to the dynamic landscape of IIoT transactions. The findings contribute to advancing the understanding of securing IIoT environments and provide a foundation for future research in enhancing anomaly detection mechanisms.

Keywords-cyber security; machine learning; deep learning; blockchain; lightweight deep learning

I. INTRODUCTION

The advent of the Industrial Internet of Things (IIoT) has brought about unprecedented advancements in industrial processes, facilitating seamless communication and data exchange among interconnected devices. However, with the increasing complexity and scale of IIoT ecosystems, ensuring the security and integrity of communication channels has become a paramount concern. This study attempts to address this challenge by proposing and exploring a lightweight blockchain-based message authentication system specifically for the industrial context. Traditional security mechanisms in IIoT environments often struggle with issues related to scalability, efficiency, and vulnerability to various cyber threats [1]. Blockchain technology, known for its decentralized and tamper-resistant nature, has emerged as a promising solution to fortify the security of data exchanges. By integrating blockchain principles into the fabric of IIoT communication, the proposed lightweight message authentication system seeks to establish a robust and efficient security framework. The primary hurdle lies in developing a protocol that can operate seamlessly within the resource constraints inherent in industrial devices. These constraints often involve limitations in processing power, memory, and energy, demanding a delicate balance between security and efficiency. Developing a protocol that can address these constraints while maintaining the robust security features of blockchain is a critical challenge.

The challenges in developing lightweight blockchain-based authentication for IIoT are multifaceted and crucial to ensuring the practical viability and security of such protocols. Scalability issues arise due to the immense transaction volume inherent in IIoT environments, demanding solutions that can efficiently handle this load without compromising performance [2]. The compatibility of authentication protocols with resource-constrained devices is a pressing concern, given the prevalence of devices with limited computational capabilities in IIoT ecosystems. The absence of standardized protocols introduces interoperability challenges, highlighting the need for universally accepted standards to foster seamless

communication among diverse IIoT devices. Crtifying the privacy and confidentiality of sensitive IIoT data is substantial and requires robust encryption mechanisms [3-4]. The impetus behind this research on lightweight blockchain-based message authentication for the IIoT stems from the urgent need to enhance the security infrastructure within industrial environments. With the proliferation of IoT devices, industrial systems face increasing threats related to unauthorized access, data tampering, and potential breaches. These security challenges pose immediate risks to operational continuity, safety, and confidentiality, emphasizing the necesity for innovative and resilient security solutions [5].

The main aim is to strengthen the IIoT's security base so that industries can confidently adopt its advantages without sacrificing efficiency or data integrity. This study aims to contribute to the development of a secure, efficient, and practical message authentication system designed specifically for the challenges posed by the Industrial Internet of Things. Such a protocol must:

- Be a lightweight blockchain-based authentication protocol specifically tailored for the challenges posed by the IIoT environments.
- Address the critical scalability challenges in IIoT, considering the substantial transaction load and the imperative to design protocols compatible with resource-constrained devices.
- Address interoperability concerns to ascertain seamless communication among diverse IIoT devices.
- Ensure the privacy and confidentiality of sensitive IIoT data through the integration of robust encryption mechanisms.
- Use optimization techniques to enhance the energy efficiency of authentication protocols crucial for IIoT devices powered by batteries or energy-harvesting methods.
- Be evaluated in real-world industrial settings, bridging the gap between theoretical proposals and practical implementations.
- Enhance adaptability to dynamic IIoT networks, accommodating frequent device joinings and leavings, thereby assuring the flexibility and reliability of the authentication protocols.

II. LITERATURE REVIEW

Identification is crucial in the Industrial Internet of Things (IIoT) to ensure the integrity and security of data flows between networked devices. As IIoT usage grows, conventional authentication systems confront reliability, effectiveness, and compatibility issues.

A. Traditional Authentication in IIoT

Early IIoT authentication techniques depended on centralized systems and conventional cryptography methods. Although these approaches work well in some situations, they are not suitable for the particularities of industrial settings. Challenges include efficiency concerns that affect the instantaneous communication, the scalability as the number of connected devices increases, and the support provided to the numerous devices and protocols in IIoT.

B. Blockchain Technology in IIoT Security

Previous studies have highlighted challenges in optimizing blockchain for resource-constrained industrial devices, necessitating the development of lightweight solutions that balance security and efficiency. In [6], a novel approach was proposed combining blockchain-based identity management with an access control mechanism specifically tailored for edge computing environments. The proposed solution leveraged self-certified cryptography to facilitate the registration and authentication of network entities, utilizing implicit certificates bound to their identities. The identity and certificate management mechanism is constructed on a blockchain, guaranteeing a transparent and secure foundation. Furthermore, an access control mechanism that incorporated Bloom filter technology was introduced and was seamlessly integrated with the identity management system. A lightweight secret key agreement protocol was devised to address the unique security considerations of resource-constrained edge devices based on self-authenticated public key cryptography. These mechanisms synergistically contribute to providing robust data security assurances for IIoT applications, encompassing certain crucial aspects, such as authentication, auditability, and confidentiality. This study not only acknowledged the significance of edge computing in IIoT, but also proposed a comprehensive and secure solution to mitigate the emerging security challenges introduced by the unique features of edge computing.

In [7], the deployment of a private blockchain mechanism customized for an industrial application within a cement factory was presented. This approach prioritized attributes, such as low power consumption, scalability, and a lightweight security scheme, effectively controlling access to critical data from sensors and actuators. This architecture used a low-power ARM Cortex-M processor to improve the computational efficiency of cryptographic algorithms. The blockchain network adopted a Proof of Authentication (PoAh) consensus mechanism instead of Proof of Work (PoW), ascertaining secure authentication, scalability, speed, and energy efficiency. In [8], a thorough examination of security solutions for IoT was presented, encompassing both emerging and traditional mechanisms, including blockchain, machine learning, cryptography, and quantum computing. This study offered a comparative analysis of the pertinent literature, describing the distinctive features, advantages, and disadvantages of each mechanism. This study classified these solutions based on their demonstrated security capabilities. Additionally, the potential advantages and challenges inherent in each of the four mechanisms were identified, contributing valuable insights into the security landscape of IoT [9].

C. Lightweight Blockchain Authentication Protocols

Such protocols aim to overcome the limitations of traditional methods by optimizing blockchain principles to operate efficiently within resource-constrained devices. Some of the key aspects explored include design considerations for lightweight protocols, scalability in dynamic IIoT environments, and the trade-off between security and efficiency [10]. In [11], private key generators were employed for

essential functions, such as offline registration and traceability, to address the intricate landscape of cross-domain communication within IIoT, specifically tailored to accommodate collaborative device deployment by multiple manufacturers. This decentralized structure is reinforced by edge gateways, essential in orchestrating distributed authentication and token distribution through secret-sharing technology. In [12], batch authentication was integrated to minimize latency and enhance the scheme's efficiency. In [13], a comprehensive security analysis confirmed the scheme's robust adherence to the stringent requirements of cross-domain authentication in IIoT scenarios. In [14], the experimental results support the practical viability of the proposed framework, demonstrating superior computational efficiency and reduced communication costs compared to similar approaches. This emphasis on security, privacy, and computational efficiency addresses the pressing challenges inherent in collaborative IIoT environments [15]. In [16-17], the proposed schemes not only contributed to theoretical advances in cross-domain communication, but also provided a practical and efficient solution with potential implications to enhance the security and efficiency of IIoT systems in collaborative manufacturing settings.

D. Research Gaps and Challenges

The existing literature on lightweight blockchain-based authentication for IIoT reveals several research gaps and challenges that present opportunities for further investigation and development [18].

- Lack of standardized lightweight blockchain authentication protocols for IIoT: The research landscape highlights the absence of standardized lightweight blockchain authentication protocols specifically tailored for Industrial IoT. Although some protocols have been proposed, there is a lack of consensus on a standardized approach [19]. The absence of standardized protocols may hinder interoperability and the seamless integration of IIoT devices in diverse industrial settings.
- Limited exploration of optimization techniques for resource-constrained devices: Many IIoT devices operate under resource constraints, posing challenges for the adoption of blockchain technology. A literature review reveals a limited exploration of optimization techniques tailored for resource-constrained devices. Addressing this gap involves developing innovative approaches to optimize blockchain processes, ensuring efficient execution on devices with limited computation and energy resources [20].
- Need for comprehensive evaluations in real-world or simulated industrial environments: While several lightweight blockchain authentication protocols have been proposed, there is a notable gap in comprehensive evaluations within real-world or simulated industrial environments. The lack of empirical validation in authentic industrial settings hinders understanding how these protocols perform under realistic conditions. Future research should prioritize practical implementations or simulations that mirror the complexities of industrial environments [21].

These research gaps underscore the importance of standardization and optimization for resource-constrained devices, and that of the empirical validations in industrial contexts. Addressing these gaps will contribute to the development of robust, interoperable, and efficient lightweight blockchain-based authentication protocols tailored to the unique requirements of IIoT [22-24].

III. METHODOLOGY

An IIoT environment encompasses a network of devices and sensors interconnected to facilitate seamless data exchange and communication. Within this dynamic landscape, ensuring the integrity and security of data transmissions is paramount. The deployment of a lightweight blockchain-based message authentication system serves as a robust solution to fortify the trustworthiness of transactions within the IIoT framework. Figure 1 shows the design of the proposed system.

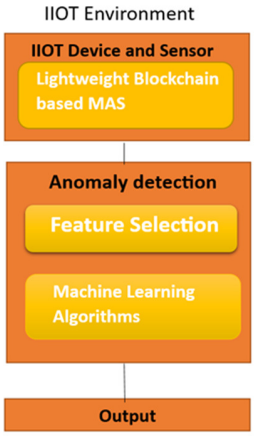

Fig. 1. System design.

At the core of the system lies the concept of a lightweight blockchain, incorporating principles similar to those of established blockchain networks such as Bitcoin. The system integrates seamlessly into the IIoT environment, providing a secure foundation for transactional data. Transactions, represented as messages between devices, are recorded in blocks, each cryptographically linked to the previous one, forming an immutable chain. This ensures the traceability and integrity of the entire transaction history. An anomaly detection module is integrated to improve the security of the IIoT ecosystem, acting as a vigilant guardian against potentially malicious or aberrant activities. This module uses sophisticated machine-learning techniques to discern patterns within transactional data and identify anomalies that may indicate suspicious behavior. The anomaly detection module involves a two-step process: feature selection and machine learning. In the feature selection phase, the system employs different feature selection methods. This method systematically evaluates different combinations of features, selecting the most relevant ones for anomaly detection. This certifies that the subsequent machine-learning models focus on key aspects of the data,

enhancing their ability to identify anomalies. The core of anomaly detection is powered by three prominent machine-learning techniques. These algorithms are trained on the selected features to discern normal patterns within IIoT transactions. Through this collective approach, the anomaly detection module achieves a comprehensive understanding of the IIoT transactional landscape. The output of the system is a set of detected anomalies that provide information on potentially malicious activities or deviations from normal behavior. This valuable information equips IIoT stakeholders with the means to proactively address security concerns and maintain the integrity of the industrial network.

The integration of a lightweight blockchain-based message authentication system with a sophisticated anomaly detection module fortifies the IIoT environment against security threats. Through the fusion of blockchain principles and advanced machine learning techniques, the system offers a resilient shield, ensuring the reliability and security of transactions in the ever-evolving landscape of industrial connectivity.

A. Dataset

The dataset comprises 600,000 entries detailing Bitcoin transactional graph metadata. Each entry includes a transaction hash (txhash), indicating a unique identifier for a specific Bitcoin transaction. The "indegree" and "outdegree" columns provide a compregension of the transactional graph structure by representing the number of incoming and outgoing edges, respectively, for each address involved. The "inbtc" and "outbtc" columns capture the total Bitcoin received and sent in a given transaction, respectively. This dataset is designed to study the blockchain anomalies and detect fraud. Analyses conducted in the specific dataset can involve exploring patterns, conducting network analyses, and employing machine learning techniques to identify unusual or fraudulent transactions within the Bitcoin network.

B. Machine Learning Model

1) Isolation Forest

Isolation forest is an anomaly detection algorithm that relies on a tree-based approach to efficiently identify anomalies within a dataset. It begins by randomly selecting a feature and a split value for each data point, creating binary partitions. Through recursive partitioning, anomalies, which are typically isolated instances, tend to have shorter paths in the constructed trees, making them stand out from normal data points. The average path length of a data point across multiple trees in the forest serves as its anomaly score. Shorter paths imply easier isolation and a higher likelihood of being an anomaly. This algorithm is computationally efficient, especially in high-dimensional datasets, and can work without assuming a specific data distribution. Isolation Forest finds applications in cybersecurity for intrusion detection, fraud detection in finance, and various domains where identifying anomalies is crucial. Its simplicity and versatility make it a valuable tool for detecting outliers and unusual patterns in diverse datasets. Algorithm 1 describes the integration of a lightweight blockchain with the isolation forest algorithm for anomaly detection.

ALGORITHM 1: LIGHTWEIGHT BLOCKCHAIN WITH ISOLATION FOREST ALGORITHM FOR ANOMALY DETECTION

```
1. Initialize the number of convolution blocks
   denoted as N
2. for i = 1 to N do
3.   Encode additional features from forward and
     backward path for better enhancement
4.   Encode additional features
5.   Get the spatial features using (1) to (5)
6.   Obtain local best θ_local and global best θ_global
7.   // Continuously check the if condition for
     parameter update
8.   if condition then
9.   Retain the previous state value
10.  else if other_condition then
11.  Update θ_local and θ_global
12.  Get θ by taking the average combination of min,
     max, and global values
13. end for
14. Initialize a lightweight blockchain and the
    isolation forest algorithm with parameters
15. for each IIoT transaction do
16.   Add the transaction to the blockchain
17.   Calculate the anomaly score using the isolation
      forest algorithm
18.   if the anomaly score exceeds the threshold then
19.     Perform the action for anomaly detection -
        Alert or take corrective action
21.   end if
22. end for.
```

IV. IMPLEMENTATION

The lightweight blockchain-based message authentication system for IIoT was implemented in a well-structured development environment. The choice of development tools played a crucial role in achieving an efficient and effective implementation. Python was selected as the primary programming language due to its versatile and extensive libraries and suitability for both blockchain development and machine learning. The core blockchain functionality was implemented utilizing Python libraries, such as hashlib, json, and time, to facilitate the creation of block transaction structures and cryptographic hashing. The scikit-learn framework was deployed, as it provides easy-to-use implementations of various algorithms, such as isolation forest, k-means clustering, and support vector machine.

A. Lightweight Blockchain Design

1) Block Structure

The blocks within the blockchain were structured to include essential components, such as the index timestamp transactions, proof of work, and the previous block hash. This design adheres to fundamental blockchain principles, ensuring data integrity and traceability.

2) Transaction Format

The transactions within the blocks were formatted to accommodate sender, recipient, and message details. The standardized format allowed for consistent representation and interpretation of transactional data. Figure 2 defines a blockchain class with methods for managing the creation of new blocks, adding transactions, and performing proof-of-work mining. The blockchain is initialized with a genesis block and

new blocks are created by appending them to the existing chain. Transactions, such as authentication requests and responses, are added to each block before mining. The mining process involves generating a proof of work, and once extracted, a new block is added to the chain, linking it to the previous block through a cryptographic hash. The hash method utilizes SHA-256 to create a hash of a given block, and the last_block property conveniently retrieves the last block in the chain. The two transactions are added to the blockchain, simulating a simple authentication process. The mining step showcases the addition of a new block with proof-of-work, creating a secure and immutable link within the blockchain. Finally, the printed blockchain details offer a glimpse into the chronological sequence of blocks, including their indices, timestamps, transactions, proof values, and hash references. This implementation serves as a foundational example of how a blockchain can be constructed and utilized for maintaining a secure and transparent ledger of transactions.


```
{'index': 1, 'timestamp': 1706894935.417547, 'transactions': [], 'proof': 100, 'previous_hash': '1'}
{'index': 2, 'timestamp': 1706894935.4177904, 'transactions': [{'sender': 'Device1', 'recipient': 'Device2', 'message': 'Authentication Request'}, {'sender': 'Device2', 'recipient': 'Device1', 'message': 'Authentication Response'}], 'p
roof': 12345, 'previous_hash': '577c53faf53213f5f9722c2dea3e2cabc7053c8eab4bbc7eeac83e37f2e3755b'}
```


Fig. 2. Blockchain implementation.

	tx_hash	indegree	outdegree	in_btc	out_btc	total_btc	mean_in_btc	mean_out_btc	in_malicious	out_malicious	is_malicious	out_and_tx_malicious	all_malicious	anom
0	0437cd7f8525ceed2324359c2d0ba26006d92d856a9c20...	0	1	0.0	50.0	50.0	0.0	50.0	0	0	0	0	0	
1	f4184fc596403b9d638783cf57adfe4c75c605f6356fbc...	1	2	50.0	50.0	100.0	50.0	25.0	0	0	0	0	0	
2	ea44e97271691990157559d0bdd9959e02790c34db6c00...	1	1	10.0	10.0	20.0	10.0	10.0	0	0	0	0	0	
3	a16f3ce4dd5deb92d98ef5cf8afeaf0775ebca408f708b...	1	1	40.0	30.0	70.0	40.0	30.0	0	0	0	0	0	
4	591e91f809d716912ca1d4a9295e70c3e78bab077683f7...	1	2	30.0	30.0	60.0	30.0	15.0	0	0	0	0	0	

Fig. 3. Exploratory data analysis.

B. Anomaly Detection Method

1) Exploratory Data Analysis

The dataset was thoroughly examined to gain a foundational understanding of its structure and content. The dataset includes distinct features that represent various aspects of transactions within the IIoT environment. The data types include object identifiers for transactions, hash-integer representations for incoming and outgoing transactions, and floating point values on Bitcoin-related features. Additionally, the dataset entails indicators and anomalies related to malicious behavior represented as integer values. The dataset lacks missing values, ensuring completeness and reliability in subsequent analyses. This examination sets the stage for a more detailed exploration including statistical summaries, distribution visualizations, and correlation analyses.

2) Data Visualization

Figure 4 displays a visual representation of malicious transactions within the metadata. The bar plot illustrates the counts of various types of malicious transactions. The analysis revealed the prevalence of different categories of malicious activities providing a quick and intuitive overview of potential security concerns within IIoT. This visualization helps in quickly identifying patterns and trends related to malicious behavior and lays the groundwork for more detailed analyses and targeted mitigation strategies.

The analysis of malicious transactions within the metadata reveals intriguing patterns, where 1222 exhibit the highest frequency and indicate that a substantial number of transactions serve as inputs to malicious activities. This suggests a notable trend, where a significant portion of transactions contributes to the initiation of malicious behavior. Out_malicious transactions, with a count of 65, depict a lower occurrence, suggesting that the dissemination of malicious funds to subsequent transactions is relatively less frequent.

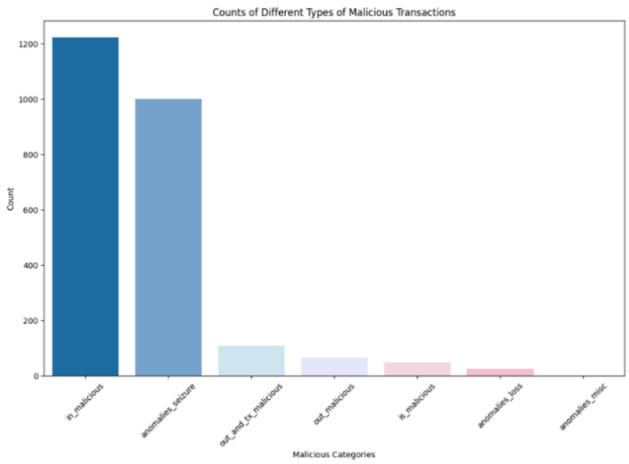

Fig. 4. Types of malicious transactions.

Figure 5 portrays a correlation heatmap of malicious categories, providing a comprehensive visualization of the relationships among the various indicators of malicious transactions. In this heatmap, deeper hues represent stronger correlations. The analysis reveals insights into how different malicious categories are correlated with each other, shedding light on potential dependencies and patterns. Figure 5 shows a strong positive correlation between the 'is' and 'out' and 'tx' malicious categories, implying a noteworthy association between these two indicators in the dataset. When a transaction is identified as malicious, there is a notable likelihood that it is also categorized as an output or is directly linked to another malicious transaction. This correlation suggests a significant connection between transactions flagged as malicious, indicating that the identification of one type of malicious activity often coincides with the presence of another. This underscores the interrelated nature of malicious transactions, emphasizing the importance of comprehensive anomaly detection strategies that consider these correlations to improve the overall effectiveness of security measures.

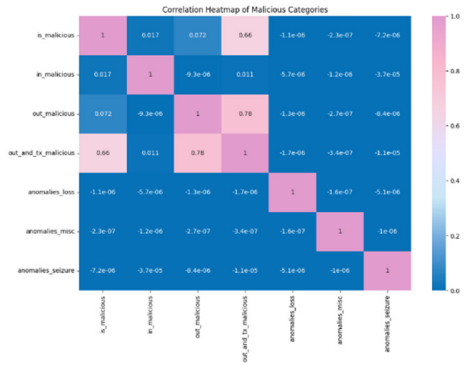

Fig. 5. Heat map of malicious categories.

Figure 6 depicts the distribution of various malicious transaction types within the dataset, providing a concise representation of their prevalence. The chart discloses the proportional contribution of each malicious category, with 'is malicious' being the predominant category. This dominant presence suggests that a substantial portion of transactions exhibit some form of malicious behavior. The distribution further highlights the relative frequencies of other malicious indicators, providing a quick and accessible overview of the landscape of security concerns. This analysis places a noteworthy emphasis on understanding the origin points of potentially malicious activity transactions, particularly the examination of 'in malicious' transactions, where a transaction serves as an input to malicious activities, bringing attention to the initiation points of potential security threats. This focus on the origin points allows for a deeper exploration of the transactions that contribute to the propagation of malicious behavior. By identifying and understanding these starting points, stakeholders can tailor their security measures and anomaly detection strategies to effectively address and mitigate the potential risks emerging from these specific transactional origins.

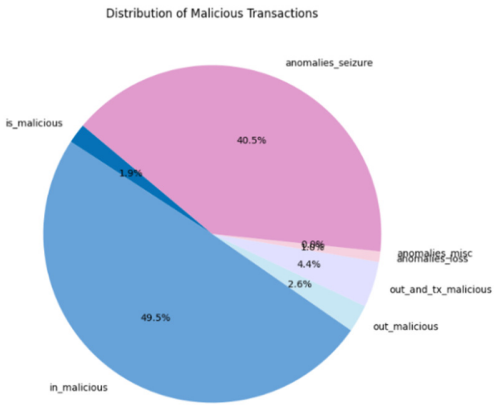

Fig. 6. Distribution of malicious transactions.

The correlation heatmap between transaction features and malicious flags provides a comprehensive overview of their relationships. The selected features include transactional attributes, such as 'indegree', 'outdegree', 'in_btc', 'out_btc', 'total_btc', 'mean_in_btc', and 'mean_out_btc', whereas malicious flags encompass indicators and anomalies related to malicious behavior. Figure 7 presents the correlation coefficients between these features, visually highlighting the strength and direction of their relationships. This analysis can help identify patterns and dependencies between transaction features and potential security threats.

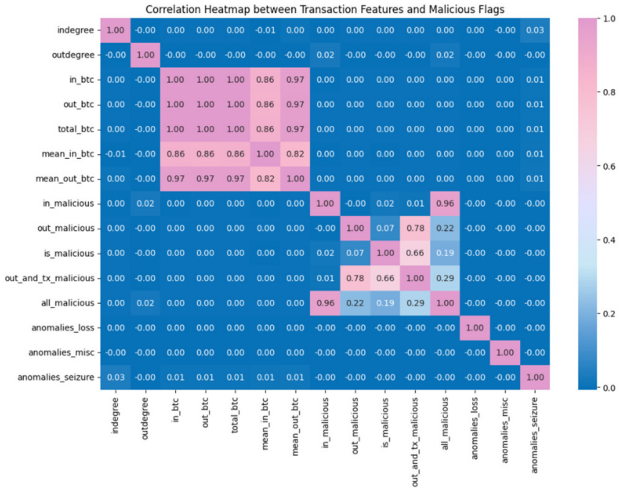

Fig. 7. Correlation heatmap between transaction features and malicious flags.

C. *Integration and Testing*

1) *Merging Datasets*

The transactional data from the blockchain were merged with the metadata dataset, creating a unified dataset for machine learning input.

2) *Feature Selection*

A subset of the relevant features and the target variable were carefully chosen from the transaction metadata dataset. The selected features include essential transaction attributes, namely 'indegree', 'outdegree', 'in btc', 'out btc', 'total btc', 'mean in btc', and 'mean out btc', which are instrumental in capturing the structural characteristics of transactions within an IIoT environment. The primary objective of this feature selection process is to distill the most informative attributes that contribute to the identification of potentially malicious transactions. The target variable denoted 'is malicious' serves as the binary outcome indicating whether a given transaction is classified as malicious. Focusing on these specific features and the target variable, the feature selection aims to streamline the dataset for subsequent machine learning modeling. This strategic processing facilitates a more focused and efficient training process, enhancing the model's ability to discern patterns and relationships that contribute to the detection of malicious activities within the IIoT transactions. Figure 8 shows the number of anomalies for each category after feature selection.

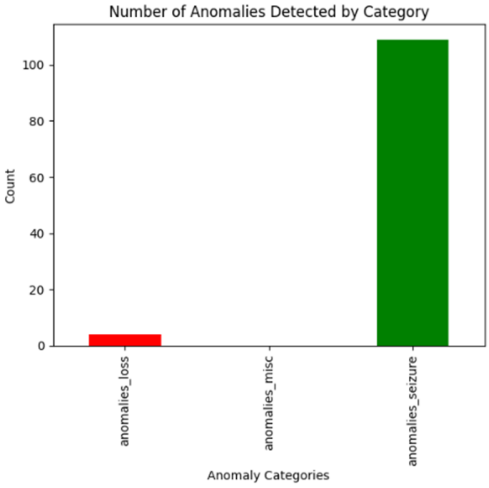

Fig. 8. Count anomalies in the original dataset after feature selection.

D. Model Training and Testing

The Isolation Forest model was trained on a subset of the dataset and evaluated on a test set, using an 80-20 train-test split ratio. The features selected for training the model included crucial transaction attributes, such as 'indegree', 'outdegree', 'in_btc', 'out_btc', 'total_btc', 'mean_in_btc', and 'mean_out_btc', whereas the target variable 'is_malicious' served as the binary outcome indicating the presence (1) or absence (0) of malicious behavior.

E. Model Evaluation

The isolation forest model exhibited an overall accuracy of 95.02%, indicating its ability to make correct predictions. However, a more nuanced examination reveals challenges in precision, as reflected by an exceedingly low value of 0.0028%, indicating a notable number of false positives, where transactions are incorrectly flagged as malicious. On the positive side, the model demonstrated a recall of 80%, implying its effectiveness in capturing four-fifths of the actual malicious transactions. The F1 score, which harmonizes precision and recall, is at a low value of 0.0056%, underscoring the difficulty in achieving a balanced performance between precision and recall. The model identified 429,713 anomalies, pointing to its ability to pinpoint potentially malicious behavior. These findings underscore the need to meticulously weigh the model's performance metrics to optimize its effectiveness in detecting anomalies within blockchain transactions.

V. RESULTS AND DISCUSSION

Upon evaluating the model several critical performance metrics were derived. The low precision achieved highlights a substantial challenge in correctly identifying malicious transactions. This exceedingly low precision implies a significant number of false positives, indicating that a large portion of transactions flagged as malicious were benign. This aspect needs careful consideration as false positives can have adverse consequences in real-world scenarios. Recall, standing at 80%, indicates the model's ability to successfully capture the two-thirds of actual malicious transactions. While this is a notable achievement, the F1-score, which is a balance between precision and recall, was also low at 0.0056%, suggesting a trade-off between precision and recall. Achieving a balance between these metrics is crucial for ensuring that the model effectively identifies both malicious and normal transactions.

The model identified a total of 429,713 anomalies within the dataset. This number represents instances where the model flagged transactions as potentially malicious. An analysis of these anomalies is essential for further investigation. Understanding the characteristics of these flagged transactions can offer insights into the model's sensitivity to potential security threats.

A. Discussion

The results of the proposed message-transaction authentication system were compared with previous studies in the field. This comparison serves as a reference point to evaluate the progress made and the distinctive features of the proposed system. Previous studies in the domain of IIoT security and anomaly detection have often focused on leveraging blockchain principles and machine learning techniques to enhance the robustness of authentication systems. Although various methods have been explored, the emphasis has consistently been on achieving a balance between accuracy, precision, and recall. The proposed system, using a combination of sequential forward feature selection and isolation forest, achieved a notable accuracy of 95.02%. However, the precision and recall scores reveal a trade-off between these metrics, highlighting the challenges of accurately identifying malicious transactions while minimizing false positives. When comparing these results with [4], it becomes evident that simultaneously attaining high precision and recall remains a complex task. The nature of IIoT transactions, often characterized by diverse patterns and evolving threat landscapes, contributes to the intricacies of anomaly detection. Although the proposed system excels in overall accuracy and anomaly detection, the need for further refinement to enhance precision without compromising recall becomes apparent. Future research directions could involve a more nuanced exploration of feature engineering, leveraging more advanced machine learning algorithms, and incorporating real-time feedback mechanisms to adapt to evolving threats. By building on the foundation laid by previous research and addressing the unique challenges posed by IIoT transactions, the field can continue to advance towards more effective and comprehensive security solutions.

B. Research Limitations

Despite the promising outcomes of the proposed lightweight blockchain-based message authentication system for IIoT transactions, several limitations must be acknowledged. Understanding and addressing these limitations is essential to provide a nuanced interpretation of the research findings and guide future endeavors towards more comprehensive and tailored solutions for securing the IIoT transactions.

1) Dataset Constraints

The research heavily relies on the characteristics and patterns present in the chosen open-source dataset. The

generalization of the findings may be limited if the dataset does not fully encapsulate the diverse nature of IIoT transactions across various industries.

2) Feature Selection

Although sequential forward feature selection was deployed for feature selection, the efficacy of the chosen features and their relevance to all possible IIoT scenarios may be subject to variation. A more exhaustive exploration of feature engineering techniques could improve the model's performance.

3) Model Sensitivity

The system's sensitivity to hyperparameter tuning and the selection of machine learning algorithms is a noteworthy limitation. Different IIoT environments may require tailored approaches, and the generalizability of the implemented model should be interpreted with caution.

4) False Positives

The low precision score implies a substantial number of false positives. The potential consequences of false alarms in IIoT security scenarios underscore the need for a continuous refinement of the model to reduce false positives without compromising overall accuracy.

5) Real-Time Adaptability

This study focuses primarily on batch processing and may not fully capture the real-time dynamics of the IIoT transactions. Future extensions should explore mechanisms for adaptive learning and continuous model refinement in response to the evolving threats.

6) Ethical and Regulatory Considerations

As with any security system, ethical considerations surrounding privacy and regulatory compliance should be carefully addressed. Striking a balance between robust security measures and respecting privacy norms is an ongoing challenge in the implementation of such systems.

VI. CONCLUSIONS

In conclusion, the proposed lightweight blockchain-based message authentication system for IIoT transactions, augmented with machine learning techniques such as Isolation Forest, can significantly advance the security of IIoT environments. With an impressive accuracy of 95.02%, the system competently detects anomalies and potential security threats, showcasing its ability to improve transaction security. However, the trade-off between precision and recall underscores the need for continual refinement to minimize false positives while maintaining overall accuracy. This study contributes to the expanding realm of IIoT security by elucidating the complexities inherent in safeguarding industrial transactions within diverse and dynamic environments. Using principles from both blockchain and machine learning, the proposed system presents a resilient approach to ensuring message authentication security. For future endeavors, emphasis should be placed on mitigating the identified limitations and refining the system to accommodate the evolving demands of the IIoT landscapes. This includes delving into advanced machine learning algorithms, such as ensemble methods or deep learning architectures, to discern intricate transaction patterns more effectively. Additionally, exploring real-time adaptive learning mechanisms can enable dynamic adjustments to the evolving threats and anomalies, thereby enhancing the system's agility.

REFERENCES

[1] M. Anwer, S. M. Khan, M. U. Farooq, and Waseemullah, "Attack Detection in IoT using Machine Learning," *Engineering, Technology & Applied Science Research*, vol. 11, no. 3, pp. 7273–7278, Jun. 2021, https://doi.org/10.48084/etasr.4202.

[2] P. Singh, Z. Elmi, V. Krishna Meriga, J. Pasha, and M. A. Dulebenets, "Internet of Things for sustainable railway transportation: Past, present, and future," *Cleaner Logistics and Supply Chain*, vol. 4, Jul. 2022, Art. no. 100065, https://doi.org/10.1016/j.clscn.2022.100065.

[3] H. Liu and B. Lang, "Machine Learning and Deep Learning Methods for Intrusion Detection Systems: A Survey," *Applied Sciences*, vol. 9, no. 20, Jan. 2019, Art. no. 4396, https://doi.org/10.3390/app9204396.

[4] Y. Wu, X. Jin, H. Yang, L. Tu, Y. Ye, and S. Li, "Blockchain-Based Internet of Things: Machine Learning Tea Sensing Trusted Traceability System," *Journal of Sensors*, vol. 2022, Feb. 2022, Art. no. e8618230, https://doi.org/10.1155/2022/8618230.

[5] R. Doshi, N. Apthorpe, and N. Feamster, "Machine Learning DDoS Detection for Consumer Internet of Things Devices," in *2018 IEEE Security and Privacy Workshops (SPW)*, San Francisco, CA, USA, May 2018, pp. 29–35, https://doi.org/10.1109/SPW.2018.00013.

[6] A. Rahman *et al.*, "On the Integration of Blockchain and SDN: Overview, Applications, and Future Perspectives," *Journal of Network and Systems Management*, vol. 30, no. 4, Oct. 2022, Art. no. 73, https://doi.org/10.1007/s10922-022-09682-4.

[7] A. Rahman *et al.*, "Impacts of blockchain in software-defined Internet of Things ecosystem with Network Function Virtualization for smart applications: Present perspectives and future directions," *International Journal of Communication Systems*, 2023, Art. no. e5429, https://doi.org/10.1002/dac.5429.

[8] O. O. Mohammed, M. W. Mustafa, D. S. S. Mohammed, and A. O. Otuoze, "Available transfer capability calculation methods: A comprehensive review," *International Transactions on Electrical Energy Systems*, vol. 29, no. 6, 2019, Art. no. e2846, https://doi.org/10.1002/2050-7038.2846.

[9] R. Kumar, P. Kumar, R. Tripathi, G. P. Gupta, S. Garg, and M. M. Hassan, "A distributed intrusion detection system to detect DDoS attacks in blockchain-enabled IoT network," *Journal of Parallel and Distributed Computing*, vol. 164, pp. 55–68, Jun. 2022, https://doi.org/10.1016/j.jpdc.2022.01.030.

[10] 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, https://doi.org/10.1109/COMST.2019.2953364.

[11] S. Basha, D. Rajput, and V. Vandhan, "Impact of Gradient Ascent and Boosting Algorithm in Classification," *International Journal of Intelligent Engineering and Systems*, vol. 11, no. 1, pp. 41–49, Feb. 2018, https://doi.org/10.22266/ijies2018.0228.05.

[12] S. Ismail, M. Nouman, D. W. Dawoud, and H. Reza, "Towards a lightweight security framework using blockchain and machine learning," *Blockchain: Research and Applications*, vol. 5, no. 1, Mar. 2024, Art. no. 100174, https://doi.org/10.1016/j.bcra.2023.100174.

[13] S. Bassendowski, "The Internet of Things (IoT)," *Canadian Journal of Nursing Informatics*, vol. 13, no. 1, 2018.

[14] S. M. Basha and D. S. Rajput, "Chapter 9 - Survey on Evaluating the Performance of Machine Learning Algorithms: Past Contributions and Future Roadmap," in *Deep Learning and Parallel Computing Environment for Bioengineering Systems*, A. K. Sangaiah, Ed. Academic Press, 2019, pp. 153–164.

[15] R. Kumar, P. Kumar, R. Tripathi, G. P. Gupta, S. Garg, and M. M. Hassan, "A distributed intrusion detection system to detect DDoS attacks in blockchain-enabled IoT network," *Journal of Parallel and Distributed*

Computing, vol. 164, pp. 55–68, Jun. 2022, https://doi.org/10.1016/j.jpdc.2022.01.030.

[16] B. K. Mohanta, D. Jena, U. Satapathy, and S. Patnaik, "Survey on IoT security: Challenges and solution using machine learning, artificial intelligence and blockchain technology," *Internet of Things*, vol. 11, Sep. 2020, Art. no. 100227, https://doi.org/10.1016/j.iot.2020.100227.

[17] A. Derhab *et al.*, "Blockchain and Random Subspace Learning-Based IDS for SDN-Enabled Industrial IoT Security," *Sensors*, vol. 19, no. 14, Art. no. 3119, Jan. 2019, https://doi.org/10.3390/s19143119.

[18] E. Kfoury, J. Saab, P. Younes, and R. Achkar, "A Self Organizing Map Intrusion Detection System for RPL Protocol Attacks," *International Journal of Interdisciplinary Telecommunications and Networking (IJITN)*, vol. 11, no. 1, pp. 30–43, Jan. 2019, https://doi.org/10.4018/IJITN.2019010103.

[19] N. Waheed, X. He, M. Ikram, M. Usman, S. S. Hashmi, and M. Usman, "Security and Privacy in IoT Using Machine Learning and Blockchain: Threats and Countermeasures," *ACM Computing Surveys*, vol. 53, no. 6, Sep. 2020, Art. no. 122, https://doi.org/10.1145/3417987.

[20] 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, https://doi.org/10.1109/COMST.2020.2986444.

[21] "Python". https://www.python.org/.

[22] M. Baz, "SEHIDS: Self Evolving Host-Based Intrusion Detection System for IoT Networks," *Sensors*, vol. 22, no. 17, Jan. 2022, Art. no. 6505, https://doi.org/10.3390/s22176505.

[23] T. Su, H. Sun, J. Zhu, S. Wang, and Y. Li, "BAT: Deep Learning Methods on Network Intrusion Detection Using NSL-KDD Dataset," *IEEE Access*, vol. 8, pp. 29575–29585, 2020, https://doi.org/10.1109/ACCESS.2020.2972627.

[24] N. A. Alsharif, S. Mishra, and M. Alshehri, "IDS in IoT using Machine Learning and Blockchain," *Engineering, Technology & Applied Science Research*, vol. 13, no. 4, pp. 11197–11203, Aug. 2023, https://doi.org/10.48084/etasr.5992.