

Blockchain-based cybersecurity : A predictive privacy and security of cyber-physical systems

N. Keerthana *

Faculty of Computer Applications

Dr. M. G. R. Educational and Research Institute (Deemed to be University)

Maduravoyal

Chennai 600095

Tamil Nadu

India

Hayder M. A. Ghanimi [†]

Department of Information Technology

College of Science

University of Warith Al-Anbiyaa

Karbala 5751

Iraq

and

Department of Computer Science

College of Computer Science and Information Technology

University of Kerbala

Karbala 5751

Iraq

P. V. V. S. Srinivas [§]

Department of Computer Science and Engineering

Koneru Lakshmaiah Education Foundation

Vaddeswaram 522502

Andhra Pradesh

India

* E-mail: keerthana.mca@drmgrdu.ac.in (Corresponding Author)

[†] E-mail: hayder.alghanami@uowa.edu.iq

[§] E-mail: cnu.pvvs@kluniversity.in

Naveen Reddy Pendli [‡]
Department of Computer Science
University of South Alabama
AL 36688
United States of America

Firas Tayseer Ayasrah [@]
College of Education, Humanities and Science
Al Ain University
Al Ain, 64141
United Arab Emirates

M. Anto Bennet [#]
Department of Electronics and Communication Engineering
Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology
Chennai 600062
Tamil Nadu
India

Sudhakar Sengan ^{\$}
Department of Computer Science and Engineering
PSN College of Engineering and Technology
Tirunelveli 627152
Tamil Nadu
India

Pankaj Dadheech [^]
Department of Computer Science and Engineering
Swami Keshvanand Institute of Technology, Management & Gramothan (SKIT)
Jaipur 302017
Rajasthan
India

Abstract

Suspicious activities in the cyber-crime world are increasing, so modern security measures of Cyber-Physical Systems (CPS) are required. This proposed model for Cybersecurity is known as the Predictive Data Security System and incorporates the use

[‡] E-mail: np1821@jagmail.southalabama.edu

[@] E-mail: firmas.ayasrah@aaau.ac.ae

[#] E-mail: drmantobenet@veltech.edu.in

^{\$} E-mail: sudhasengan@gmail.com

[^] E-mail: pankajdadheech777@gmail.com

of Blockchain (PDSS + BC). Using BC's fixed ledger, this PDSS + BC objective incorporates predictive analytics intending to prevent risks to security. The PDSS integrates a BC for data storage with an ML-based Threat Prediction Model (TPM). In the computer simulation, researchers test the performance of the recommended model against the standard CPS security measures and validate its effectiveness. The empirical results proposed that the recommended PDSS + BC enhances data quality and Anomaly Detection (AD), underscoring the importance of BC in contemporary cybersecurity frameworks. With an optimal Detection Rate (DR) of 91.56% and a minimum End-to-End Delay (EED) of 765 ms, the reference for PDSS + BC is an impressive combination. When accommodated against current state-of-the-art methods, the proposed PDSS + BC performs significantly all around.

Subject Classification: 93E10, 94A13.

Keywords: *Predictive data security, Blockchain, Cybersecurity, Anomaly detection, Machine learning, Data integrity.*

1. Introduction

Higher-than-standard cybersecurity measures are necessary due to the increasing prevalence of threats, including data breaches, phishing, and ransomware [1]. The core component of predictive analytics, Machine Learning (ML), identifies patterns and outliers in past data that may indicate potential attacks. These algorithms may detect future vulnerabilities and attacks by analyzing the behavior of Cyber-Physical Systems (CPS) and past events [2]. Data storage and processing system security is a critical issue due to the significant trust in these computational models for data. The distributed ledger system known as Blockchain (BC) addresses the most critical problem of data integrity by rendering it challenging to manipulate recorded data. The data in a distributed ledger—the technological name for a BC—is shared over a network of nodes as opposed to being centralized on a central server, which makes it highly challenging for hacking [3]. There is a permanent record of all data and transaction inputs because they are encrypted and connected.

Because of this, the data used by the Threat Prediction Models (TPMs) can remain unchanged, which enhances the accuracy of Anomaly Detection (AD). The recommended PDSS could address several cybersecurity problems by combining predictive analytics with BC. To begin with, the system utilizes the BC ledger to secure the data, thereby improving data accuracy and enabling more accurate predictions. Predictive models are dependent on accurate data input, so no one other than authorized users must tamper with the data in any way. The PDSS addresses AD.

With the help of Machine Learning (ML) methods, the framework can detect potential attacks on the business in this example and prevent problems in their tracks. By implementing this preventive test, scientists can reduce the chance of successful attacks [4]. The BC improves the reaction time to newly detected attacks in the natural environment. In the BC, solutions are digitalized with the support of Smart Contracts (SC), which are contracts that perform autonomously. In the case of an AD, the SC has a mandate to initiate predicted behaviors, such as reporting security personnel, implementing preventive measures, or changing the security measures of the CPS. When addressing attacks, having the capacity to respond effectively is essential, particularly in the initial phases when they are less significant. A new invention in cybersecurity, the Predictive Data Security System (PDSS + BC), combines Machine Learning (ML) for prediction with BC security, resulting in a novel model. By linking all these tools collectively, we can enhance the AD and response systems, as well as the data quality that users accept whenever required. The system's adaptability and use might be improved by the scalability of BC and SC functions [5]. AD and learning features will be necessary for the PDSS to remain relevant in the continuously evolving AD environment. The remainder of the article is organized as follows: the summary of predictive analytics for Cybersecurity is discussed in Section 1, the complete survey of background study is discussed in Section 2, the recommended CSP PDSS and CPS based Cybersecurity is detailed in Section 3, the simulation result and analysis are specified in Section 4, and the study is concluded in Section 5.

2. Literature Review

This study reviews the literature to identify key developments in predictive data security CPS and BC. [6] Blockchain-Enabled Data Security (BEDS) utilizes BC technology and advanced cryptographic methods to secure web-based content. In BEDS, BC provides data privacy and accuracy, while digital encryption prevents misconduct and unauthorized changes. Data security is of the highest priority at BEDS because all files are encrypted and maintained on the BC, making hacking impractical. As a result, the effectiveness of cryptographic methods in ensuring data security and maintaining user privacy has become a significant problem. The Privacy-Preserved Threat Intelligence Framework (P2TIF) [7] integrates BC and DL to enhance security and privacy in the IIoT setting. Using a deep variational autoencoder to prevent data prediction attacks,

P2TIF features a scalable BC for secure data transmission. Using data created by ToN-IoT and IoT-Botnet, the model estimates an attention-based Deep Gated Recurrent Neural Network (DGRNN), which is the central component of the AD framework. BC provides, for security reasons, analysis and predictions of medical data and medical conditions, and P2TIF is adaptable and vital when it addresses data and AD. The solution their actions are available with is the III-level system that was addressed in [8]. It requires keeping track of healthcare sensors in real-time, the BC database, and the SC to determine who can access the data. The method [9] describes the application of BC for countering attacks in the context of Industry 4.0. It categorizes several types of cyberattacks and examines the potential of using BC to enhance the security, accuracy, and privacy of data used in intelligent production facilities. Among the discussed protocols are linear structures and Directed Acyclic Graphs (DAGs). To enhance the functionality of BC in industrial applications, the authors recommend further investigation into integrating BC + ML and other modern technologies [9]. [10] proposed the Block Hunter, a Federated Learning (FL) for IDS in BC-based IIoT networks. Block Hunter is a cluster-based architecture that utilizes machine learning (ML) for Anomaly Detection (AD) and privacy. Due to their FL, it is easy to AD that can consume little bandwidth, hence enhancing cybersecurity within IIoT environments [11-15]. Cutting-edge BC is not integrated with predictive analytics tools that can provide real-time AD and response. Further research is needed to integrate these methods for enhanced cybersecurity efficiency.

3 .Proposed CPS – Predictive Data Security System for Cybersecurity

The PDSS + BC and data analytics in order to enhance security through attack prevention. First, the CPS generates a BC network and sets the SC to address data security problems and actions. It collects and processes data from multiple sources and then constructs a model using the historical data. This model AD generates attack probabilities for new incoming data. The system incorporates this detail in the BC using the SHA-256 and SHA-3 hybrids to maintain the accuracy of the data and the model. SC is deployed in the event of AD, which could entail predetermined actions, such as alerting security to the area. This integration ensures that all system records and responses are non-alterable and traceable, thereby enabling organizations to better defend against cyber-attacks. Every transaction in the BC is a block containing a hash of block data and the

hash of the previous block. The Hash Function (HF) is responsible for providing data integrity by generating a Hash value for the block. The hash is assessed by the Eq. (1).

$$H_i = \text{Hash}(B_i) = \text{Hash}(H_{i-1}, \text{Data}_i) \quad (1)$$

Where, the block of 'i' is indicated in H_i , the hash of the previous block is given in H_{i-1} , and Data_i is the block 'i' of data. To check the validity of the generated by BC, Eq. (2) is applied; in this, each block's hash should be equal to the stored hash.

$$\text{Verify}(B_i) = \begin{cases} 1 & \text{if } H_i = \text{Hash}(H_{i-1}, \text{Data}_i) \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

Where, $\text{Verify}(B_i)$ returns '1' if the block is valid; '0' if it is tampered with. When training the predictive model, a loss function 'L' is used to evaluate the model performance in Eq. (3).

$$L(\theta) = \frac{1}{N} \sum_{i=1}^N (y_i - f(x_i; \theta))^2 \quad (3)$$

Where, $\theta \rightarrow$ The model parameters, $x_i \rightarrow$ The input features, $y_i \rightarrow$ The actual labels, $f(x_i; \theta) \rightarrow$ The predicted output for input ' x_i ', $N \rightarrow$ The number of data samples. The probability $P(T)$ of an attack as 'T' specified features 'X' is computed using a Logistic Regression (LR) in Eq. (4).

$$P(T|X) = \sigma(w^T X + b) \quad (4)$$

Where, $w \rightarrow$ the vector of weights, $b \rightarrow$ the bias term, $\sigma \rightarrow$ the sigmoid function. SC's condition 'C' selects the response using Eq. (5).

$$C = \begin{cases} 1 & \text{if } P(T|X) > \theta_{\text{threshold}} \\ 0 & \text{Otherwise} \end{cases} \quad (5)$$

Where, $\theta_{\text{threshold}} \rightarrow$ The predefined threshold for attack probability. Whenever condition 'C' is fulfilled, the SC acts 'A', as Eq. (6).

$$A = \text{Execute}_{\text{Action}}(\text{Action}_{\text{Type}}) \quad (6)$$

Where, $\text{Action}_{\text{Type}}$ could be alerting security personnel and updating firewall rules.

For improving security, the SHA-256 and SHA-3 algorithms are used for hashing, as shown in Eq. (7).

$$H_{\text{final}} = \text{SHA3}(\text{SHA256}(B_{\text{data}})) \quad (7)$$

Where the data in the block $\rightarrow B_{data}$, $SHA256(B_{data})$ computes the SHA-256 hash of the block data, SHA3 applies SHA-3 to the result of SHA-256 (Algorithm 1).

Algorithm 1 for PDSS

- Step 1. Function PDSS_Integration (newData, Historical Data, Labels, Threshold)
- Step 2. Initialize BCNetwork()
- Step 3. DeploySC()
- Step 4. Features = PreprocessData (CollectDataFromSources())
- Step 5. Model = TrainModel (Features, Labels)
- Step 6. For Each dataPoint IN historicalData
- Step 7. blockData = PrepareBlockData (dataPoint, Model)
- Step 8. blockHash = ComputeHybridHash (blockData)
- Step 9. StoreBlockInBC (Block Data, Block Hash)
- Step 10. End FOR
- Step 11. AttackProbabilities = PredictAttackProbabilities (newData, Model)
- Step 12. For Each Probability IN attackProbabilities
- Step 13. IF probability > threshold, Then
- Step 14. RecordAttackDataInBC (newData, probability)
- Step 15. IF SCConditionMet() Then; ExecuteSmartContractAction()
- Step 16. End If
- Step 17. End If
- Step 18. End For
- Step 19. Return attack probabilities
- Step 20. End Function
- Step 21. Function ComputeHybridHash (Data)
- Step 22. Return SHA3(SHA256 (Data))
- Step 23. End Function

4. Result and Discussion

Using a predictive data security model that incorporates BC requires a robust hardware and software infrastructure. Hardware states to the computer model that embodies computing resources, including multi-core CPUs, sufficient and efficient RAM, SSD storage, and a high-speed network. The software includes a BC, Hyperledger Fabric, Machine Learning (ML), TensorFlow, and a database for non-BC data (Table 1). The performance is evaluated using End-to-End Delay (EED) and Detection Rate (DR) metrics. The proposed PDSS is compared with past methods, such as BEDS + P2TIF (Fig. 1 and Fig. 2). In a predictive data security model, EED refers to processing and responding to a request. It plays a significant role in assessing the CPS's performance in terms of its response to the incident. EED is defined as the time taken between the moment the request is made and the moment the response is received, Eq. (8).

$$Latency = Time_{response} - Time_{request} \quad (8)$$

The Threat Detection Rate (TDR) measures the CSP's ability to differentiate real attacks from other tasks. It is the proportion of true attacks correctly classified (True Positives (TP)) to the sum of true attacks correctly classified and the sum of true attacks misclassified as noise (False Negatives (TN)). This metric demonstrates the system's capability to achieve AD as described in Eq. (9).

$$TDR = \frac{Number\ of\ TP}{Sum\ of\ (TP+FN)} \quad (9)$$

Table 1 compares the performance metrics—EED and DR—for three predictive data security models: The success of the goals of BEDS, P2TIF, and PDSS. The average EED of BEDS is 1092 *ms*, which is the highest among the three models, indicating that the model's response time is slow. This slower EED could slow down the accurate time AD and the model's efficiency for the DR, resulting in 86% BEDS, which is lower than P2TIF

Table 1
Comparison of Performance

Techniques	EED (ms)	DR (%)
BEDS	1092	86
P2TIF	987	87.67
PDSS	765	91.56

and PDSS. This implies that even using BEDS acceptance reasonable AD, it might not be as efficient as the other models in AD.

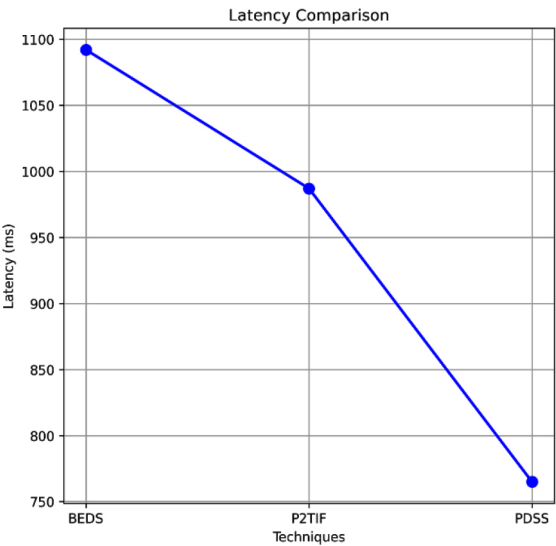


Figure 1
Comparison of EED



Figure 2
Comparison of DR

Compared to BEDS, P2TIF has a lower EED of 987 *ms* and proposes a better trade-off between EED and DR, with a determined DR of 87.67%. PDSS is again the most striking, with the least EED of 765 *ms*, implying that the model responds faster than all the others. Moreover, the DR of PDSS is the highest at 91.56% of the respondents rated their organization's effectiveness in AD as superior to the industry average. PDSS has the lowest latencies and the highest DR, followed by P2TIF and then BEDS (Figure 1 and Figure 2).

5. Conclusion and Future Work

The novel idea of the PDSS + BC for enhancing Cybersecurity. Using ML to predict potential risks and BC to verify the data's authenticity, the PDSS mitigates common weaknesses of conventional CPS. The reviews and assessments indicate positive results in enhancing AD and data security. For future work improvement, more detailed and complex ML, such as Deep Learning (DL), could be implemented to increase the accuracy of AD predictions. Furthermore, studying BC scalability options and SC improvements may enhance the system's efficiency and adjustability in response to the constantly changing cyber-attacks. Incorporating attacks into intelligence feeds and learning mechanisms will be essential to sustain the CPS's capability to adapt to changing AD landscapes.

References

- [1] S. Ahmadjee, C. Mera-Gómez, R. Bahsoon and R. Kazman, "A study on blockchain architecture design decisions and their security attacks and threats", *ACM Transactions on Software Engineering and Methodology (TOSEM)*, vol. 31, no. 2, p. 1-45, (2022).
- [2] Y. Qi, M. S. Hossain, J. Nie and X. Li, "Privacy-preserving blockchain-based federated learning for traffic flow prediction," *Future Generation Computer Systems*, vol. 117, p. 328-337, (2011).
- [3] Z. Wang, S. Zhang, Y. Zhao, C. Chen and X. Dong, "Risk prediction and credibility detection of network public opinion using blockchain technology," *Technological Forecasting and Social Change*, vol. 187, p. 122177, (2023).

- [4] Z. Rahman, X. Yi and I. Khalil, "Blockchain-based ai-enabled industry 4.0 cps protection against advanced persistent threat," *IEEE Internet of Things Journal*, vol. 10, no. 8, p. 6769-6778, (2022).
- [5] V. Jeevika Tharini, B. Ravi Kumar, P. Sahaya Suganya Princes, K. Sreekanth, B. R. Kumar and S. Sudhakar, "Business Decision-Making Using Hybrid LSTM for Enhanced Operational Efficiency," In *International Conference on Multi-Strategy Learning Environment*, p. 155-166). Singapore: Springer Nature Singapore, (2024).
- [6] A. I. Taloba, A. Elhadad, R. M. A. El-Aziz and O. R. Shahin, "Prediction of data threats over web medium using advanced blockchain-based information security with crypto strategies", *Journal of Ambient Intelligence and Humanized Computing*, p. 1-17, (2021).
- [7] P. Kumar, R. Kumar, G. P. Gupta, R. Tripathi and G. Srivastava, "P2TIF: A blockchain and deep learning framework for privacy-preserved threat intelligence in industrial IoT", *IEEE Transactions on Industrial Informatics*, vol. 18, no. 9, p. 6358-6367, (2022).
- [8] M. Wang, H. Zhang, H. Wu, G. Li and K. Gai, "Blockchain-based secure medical data management and disease prediction," In *Proceedings of the 4th ACM International Symposium on Blockchain and Secure Critical Infrastructure*, pp. 71-82, (2022).
- [9] S. B. ElMamy, H. Mrabet, H. Gharbi, A. Jemai and D. Trentesaux, "A survey on the usage of blockchain technology for cyber-threats in the context of industry 4.0", *Sustainability*, vol. 12, no. 21, p. 9179, (2020).
- [10] A. Yazdinejad, A. Dehghantanha, R. M. Parizi, M. Hammoudeh, H. Karimipour and G. Srivastava, "Block hunter: Federated learning for cyber threat hunting in blockchain-based IIoT networks," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 11, p. 8356-8366, (2022).
- [11] M. A. Ghanimi, Hayder, P. Melgarejo-Bolivar, A. Tumi-Figueroa, S. Ray, D. Pankaj and S. Sudhakar, "Merkle-Damgård hash functions and blockchains: Securing electronic health records," *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 27, no. 2-A, p. 237-248, (2024).
- [12] M. A. Ghanimi, Hayder T. Gopalakrishnan, D. G. Joel Sunny, K. Amarendra, D. Pankaj and S. Sudhakar, "Chebyshev polynomial approximation in CNN for zero-knowledge encrypted data analysis," *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 27, no. 2-A, p. 203-214, (2024).

- [13] P. Vidya Sagar, M. A. Ghanimi Hayder, L. P. Arokia Jesu, Raja, D. Pankaj and S. Sudhakar, "Secure multi-party computation in deep learning: Enhancing privacy in distributed neural networks," *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 27, no. 2-A, p. 249–259, (2024).
- [14] G. R. K. Rao, M. A. Ghanimi, Hayder, V. Ramachandran, D. Al-Qahani, D. Pankaj and S. Sudhakar, "Enhanced security in federated learning by integrating homomorphic encryption for privacy-protected, collaborative model training," *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 27, no. 2-A, p. 361–370, (2024).
- [15] K. R. Jaya, T. Gopalakrishnan, M. Divyapushpalakshmi, K. Amarendra, D. Pankaj and S. Sudhakar, Security and privacy concerns in social networks mathematically modified metaheuristic-based approach," *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 27, no. 2-A, p. 371–382, (2024).

Received November, 2024