A metaheuristic algorithm based hybrid enhanced blockchain and trust management system for heterogeneous IOT systems

Pushpa R ✉

Professor, Department of Computer Science and Engineering,
Akshaya Institute of Technology, Tumkur
pushpa.aitcse@gmail.com

Abstract

Heterogeneous IoT systems provide consumers a variety of data collecting, connectivity, and integration options. This technology is intended to allow many kinds of devices to communicate and exchange data in order to build a more efficient and effective network. This study uses a Logistic regression model to offer a Hybrid Trust & Block chain-based Framework for Heterogeneous IoT System. This framework is intended to provide effective data exchange and connectivity across various IoT devices and infrastructures. Trust levels are also used to identify malicious nodes in a heterogeneous IOT system. This data is then used to build a logistic regression model that can identify malicious nodes. The suggested system's performance is tested using performance metrics (sensitivity, accuracy, and specificity), and the results display that the suggested technique can correctly detect malicious nodes. The suggested system's accuracy was determined to be much greater than that of the current systems. The results showed the proposed framework beats cutting-edge techniques in terms of security.

Key Words: *Heterogeneous IoT systems, Logistic regression, IoT devices, Hybrid Trust, Blockchain-based Framework.*

1. INTRODUCTION

The Internet of Things (IoT) has gained popularity during the last several years. The Internet of Things (IoT) is a network of physical devices (sensors, controllers, and other gadgets) that are linked to the web and may share information with one another. The term "heterogeneous IoT" refers to the networking of many IoT device types on a single network. The issue of protecting such a network becomes more complex as the number of devices rises and the kinds of devices become more diversified [1]. The difficulty with heterogeneous IoT is ensuring a secure, stable, and robust connection between

all devices, regardless of their hardware, software, or protocol. However, the complexity of IoT systems has grown dramatically, necessitating a more secure and dependable infrastructure to operate such systems. In answer to this demand, a heterogeneous IoT system based on trust and block chain technology has emerged as a possible option [2].

Blockchain technology is a system of decentralized ledgers that provides a secure and impermeable way to store and manage data. This technology was developed by Bitcoin. It is made up of a distributed network of computers, which are referred to as nodes that use cryptography to verify and record transactions. This technology is especially well- suited for IoT systems since it allows safe communication and data exchange between the different devices and also provides secure data storage. Additionally, it provides secure data storage. Although the idea of a heterogeneous Internet of Things (IoT) system that makes use of block chain technology is still in its infancy, it is rapidly gaining support [3]. This particular kind of system is an Internet of Things system, and it is made up of a wide number of various components, including sensors, actuators, and other devices. The system is intended to allow for encrypted data sharing and communication between the various devices, as well as the provision of an encrypted and immutable ledger for the data that is being stored [4].

Improved safety, scalability and dependability are among the primary advantages afforded by using a heterogeneous Internet of Things system that makes use of block chain technology. The use of this technology allows devices to securely connect with one another and exchange data without depending on a single centralized server, which lowers the probability of data being compromised [5]. The data may also be retrieved from other nodes in the network thanks to the distributed nature of the block chain technology, which allows the system to be more durable in the event of a breakdown. Last but not least, the immutable ledger of the data that is maintained on the block chain guarantees that the data is immutable, making it far more difficult for hostile actors to alter the data [6]. Many of the authors discussed about heterogeneous IOT using different methods and block chain technology. Some of them are discussed below.

In this research, [7] advised employing block chain technology as the backbone of a security framework in order to combine Internet of Things devices that had varying resource limits under one system. Ethereum may be used to develop a secure system that is resistant to Denial-of- Service attacks; it can also be used to manage device trust; it can store encryption keys; it can store encrypted data; and it can store both. [8] provided a heterogeneous Internet of Things

architecture that was composed of four layers and featured apps, networking, cloud computing, and sensors. The protection of confidentiality in large-scale heterogeneous Internet of things and the movement of huge volumes of data are two issues that the heterogeneous internet of things (IoT) will have to deal with in the future, and this research offered various possible solutions to these issues. [9] Some early research was presented on an architecture for managing heterogeneous Internet of Things systems that makes use of block chain technology. The author then proceeded to develop a framework for the management of a large-scale heterogeneous Internet of Things system. [10] A conceptual model is presented in this study with the purpose of identifying the essential factors that influence the adoption of block chain technology in IoT. Based on the findings, it seems that factors relating to data have a significant impact on the adoption of block chain technology in the internet of things as well as the willingness to use it.

2. ANALYSIS AND/OR DESIGNS

A. Blockchain

A blockchain is "an encrypted, decentralized, and ever-growing collection of records, called blocks," which are linked together. Every block contains transaction information, the digital signature of the previous block, and a time stamp. [11], which would then invalidate the network consensus[12]. The most important feature of blockchain technology is its security. With its distributed nature, blockchain technology is virtually impossible to hack or manipulate. This makes it ideal for storing information that needs to remain secure, such as financial data or medical records. The data stored on the blockchain is encrypted, making it virtually impossible to alter or delete. Blockchain technology is also extremely efficient. The blockchain allows for the speedy and safe processing of transactions as well as the quick and simple sharing of data stored there. Because it is capable of being employed to track and keep an eye on the flow of goods and amenities, it is perfect for operations like supply chain management. The information from source network to the destination network is carried out by AODV routing protocol, the terminology and the principal of working is described as presents.

B. AODV (Ad hoc On-Demand Distance Vector routing)

AODV preserves only routes in networks that satisfy a need because it is a reactive heterogeneous IOT system. The next hop for getting to destinations is maintained by AODV in a routing table. A route will ultimately time out if no packets are sent along it. Mostly since it only knows about the nodes that are nearby.

1. Terminology of AODV

The following are the three different sorts of routing communications of AODV.

i. RREQ: Route Request

A node starts the route identification procedure by sending an RREQ multicast message when it intends to send/transmit a packet but is unclear of how to get there. Prior to the message reaching the target node, adjacent nodes pass it to their neighbors while keeping track of the communication's source.

ii. RREP: Route Reply

The RREP of the destination node returned to the original node by means of RREQ. Facilitation networks create forwarding pathways as RREP travels back to the source. Intermediary nodes can respond to an RREQ through an RREP if it is aware of the destinations, enabling routers to join a pre-existing route. Once the RREP comes and the path is defined, source-to-destination communication begins.

iii. ERR: Route Error

AODV has less overhead than proactive protocols (fewer route maintenance messages). When a connection is broken and a route is no longer functioning, messages cannot be forwarded, a node sends an RERR message. Messages are recast. RERR shows an inaccessible location. Message- receiving nodes deactivate the route.

C. Key Generation

In the Advanced Encryption Standard (AES) algorithm, the key size, number of rounds, and key generation time contribute to different aspects of encryption key generation:

1. Key Size:

In AES, the key size may be anything between 128 bits and 256 bits. The greater the key size, the more keys may be used, making brute-force attacks more computationally infeasible. Generally, a larger key size provides stronger security, as it increases the complexity of the encryption process and reduces the likelihood of successful attacks. However, larger key sizes also result in increased computational requirements for both key generation and encryption/ decryption operations [13].

2. Number of Rounds:

AES has a set amount of loops, which are 10, 12, or 14 for variables of 128-bit, 192-bit, or 256-bit lengths, accordingly. The number of rounds determines the number of iterations performed during the encryption process. The security of AES improves as the number of rounds increases, thanks to the algorithm's enhanced diffusion and confusion features, making it more resistant to various attacks. However, an increased number of rounds also leads to additional computational overhead and longer encryption/decryption times.

3. Key Generation Time:

The time needed to produce a valid encryption key in AES is referred to as the key generation time. Random bits are frequently utilized during key creation to generate a safe and unique key. The produced key's quality and unpredictability have a direct impact on the encryption's security. While AES does not define a key generation technique, using a reliable and secure key creation method is critical to ensuring the strength of the produced keys. Because of its symmetric nature, the same key may be used for decryption as well as encryption with AES. As a result, the key generation procedure is typically performed only once, and the generated key is securely transmitted between the originator and recipient.

3. EXPERIMENTAL SECTION

In this study, we proposed a hybrid trust and block chain- based framework to ensure security for IOT systems with the help of trust evaluation and key strength mechanism involves several key steps. Firstly, the trust management system provides security features that enable nodes to handle the unpredictable nature of nodes, Transaction management system maintains the ledger of transaction among the nodes. Later on, these trust values are fed to the IDS to detect and remove the miscellaneous nodes from the planned path. And their corresponding details of the final secure path are pushed into transaction management system, and the Hash values of the generated blocks are transferred to the cloud for future reference. Secondly, key strength assessment phase determines the required key size, generates cryptographic keys, and evaluates their strength. The integration of trust and key strength factors guides the decision-making process.

A. Trust Model

Trust is a crucial security feature that helps nodes handle the unpredictable nature of IOT; as a result, trust calculation and maintenance in IOT are challenging. The network's performance will undoubtedly suffer from an unreliable node. Because of this, figuring out every node's trust level has a big influence on how much protection a node can offer for an encrypted connection.

The trust values of the nodes are calculated in IOT while data is transmitted ion is carried out using AODV as follows.

Trust calculation algorithm for AODV

Step 1 PR1=Count the number of packets received at each node.
PS2=Count the number of packets sent by each node
RR1=Count the number of RREQ received at each node.
RS2= Count the number of RREP sent by each node.

Step 2 Calculate the P1 from neighbor node value:

$$P1 = \frac{\text{No. of packet sent}}{\text{No. of packet Received}}$$

Calculate the P2 from neighbor node value:

$$P2 = \frac{\text{No. of Route Replay sent}}{\text{No. of Route request Received}}$$

Step 3 Calculate Trust Value = (P1 * α) + (P2 * β) where α and β = static weighting factor

Step 4 Insert Trust value into Neighbor Table

B. Key Assessment

The evaluation of the key's strength consists of determining the entropy of the cryptographic keys that were produced by using Shannon entropy.

Shannon Entropy Calculation:

The formula for computing the Shannon entropy of a cryptographic key is as follows:

$$\text{Entropy} = \sum_{i=1}^{n} p_i \cdot \log_2 (p_i) \quad (8)$$

where n is the total number of bits in the key and pi is the likelihood that each bit is 1. Shannon entropy is used to calculate the key strength of randomly generated cryptographic keys. The code computes the entropy of the keys for various key sizes (128, 192, and 256 bits) and conducts statistical analysis over a certain number of iterations. The key strength is then calculated by multiplying the likelihood of '1' and '0' bits in the key by the Shannon entropy. This study focuses on an optimization method known as PSO.

C. Particle Swam Optimization

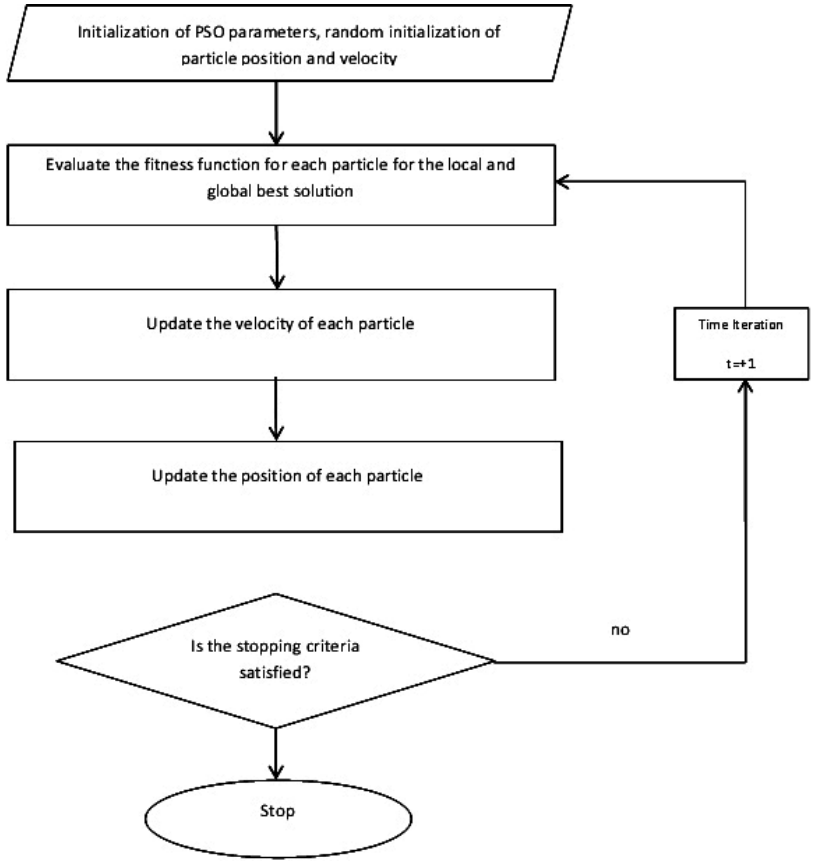
Particle swarm optimization (PSO) is a population-based search method based on the social actions of birds throughout a flock as modeled. The particle swarm concept was developed with the intention of visualizing the graceful and surprising choreographed of a flock of birds while also identifying patterns that govern a bird's capacity for synchronous flight and the ability to quickly change directions by rejoining in an optimal shape. The concept evolved into a straightforward and efficient optimization process from this fundamental objective. [14].

The formula for PSO based feature selection is as follows:

$$x^{t+1}_{id} = x^{t1}_{id} + v^{t+1}_{id}$$

$$v^{t+}_{id}{}^{1} = w * v^{t}_{id} \; c_1 * r_{1i}(p_{id} - x^{t}_{id}) + c_2 * r_{2i} * (p_{gd} - x^{t}_{id})$$

Where d, D is the d^{th} dimension of our search space. In PSO, t is the current iteration, x^{t}_{id} is the particle's current location, and v^{t}_{id} is the particle's current velocity in d in the i^{th} iteration. w is the inertia weight, which allows velocities from the previous iteration to have a controlling influence on the current one, while C_1 and C_2 are the learning rates. r_{1i} and r_{2i} are simply random numbers that are uniformly distributed in the range [0, 1], and p_{id} and p_{gd} are the p_{best} and g_{best}.



A. Logistic Regression

When it comes to ML methods, logistic regression is one that is is not a "black box" since we understand exactly how it operates. Black box models, like extremely deep neural networks (DNN), frequently undergo iterations that are more sophisticated. The outcome of a binomial logistic regression might be either 0 or 1. In the realm of ML, logistic regression is one of the few methods for which the inner workings are not a mystery. Employing black box

models, such as a very deep neural network (DNN), may lead to more fruitful iterations. The result of a binomial logistic regression might be either 0 or 1. Determine if a given event will produce a 1 or a 0 as the following step. Since 1 is as likely as not, we may express the chance that it won't happen as 1-p. Bernoulli distribution is one kind of binomial distribution, and this serves as an illustration of it. Logistic regression uses a modified linear regression model to characterize the issue at hand [16].

$\hat{y} = \beta_0 + \beta_1 x_1 + \ldots\ldots\ldots\ldots + \beta_n x_n$. Sigmoid function examples include $P = 1/\ 1+ e^{-\hat{y}}$. Continuing with this, NN with a sigmoid activation function might be employed to depict logistic regression. In additional disputes, we can calculate the weighted variables by reducing the function of cross-entropy loss generated by the common SGD approach. $l(p, y) = -y \log(p) - (1-y) \log(1-p)$ *where* $y \in [0,1]$.

As a result, the multinomial logistic regression is limited to addition new stochastic cells to the stratum. To maintain the result as the distribution of probability, we could substitute the SoftMax activation equation for the logistic function. With SoftMax, the decrease is renewed the log- likelihood loss while doing. [17]. The collected trust values are then used to train a logistic regression model that incorporates Naive Bayes, an ensemble, and a support vector machine. The performance of this suggested model will be assessed by testing at a later stage. The performance of a classifier may be measured using a number of different factors described in the literature. Whose, the formula is shown below in mathematical form.

B. Performance metrics

In terms of the performance metrics represented in the confusion matrix, we may assess the efficacy of a technique by looking Accuracy: The proportion of subjects correctly identified as part of the total no of subjects.

Accuracy =TP+TN/TP+TN + FP + FN (3)

Sensitivity: Recall, also known as sensitivity, is the percentage of true positive indicators that the system accurately recognizes.

Sensitivity=TP/TP+FN. (4)

Precision: The amount of successful predictions may be used as a measure of a viewpoint's accuracy. Another name for this idea is "predictive value."

Precision=TP/TP+FP. (5)

1-Score: A statistic that considers combined accuracy and recall is the F1-score.

F1-score=2*(Precision*Recall/Precision+Recall) (6)

Specificity: The algorithm has classified the negative as specificity appropriately.

Specificity=TN/TN+FP (7)

Whereas, FN= False Negative, FP= False Positive, TN= True Negative, and TP= True Positive.

Depending on the topic, the experimental section should include synthesis procedures, characterization techniques etc and specification of the instruments. If it is simulation, author will have to mention the type simulation software with specifications [3,7]. Write the importance of that software used relevant to research work carried out.

4. RESULTS AND DISCUSSION

Logistic Regression

Our suggested approach properly identified 348 benign nodes as benign and 149 harmful nodes as miscellaneous nodes. There are some misclassifications are also occurred, the proposed model predicts 1 miscellaneous node as normal node, and 2 normal nodes as miscellaneous nodes.

Ensemble Learning

Logistic regression model predicted 147 miscellaneous node as miscellaneous nodes and correctly predicted 342 normal nodes as normal nodes. There are some misclassifications are also occurred, the Logistic regression model predicts 3 miscellaneous nodes as normal node, and 8 normal nodes as miscellaneous nodes.

Naïve Bayes

Logistic regression model predicted 145 miscellaneous node as miscellaneous nodes and correctly predicted 341 normal nodes as normal nodes. There are some misclassifications are also occurred, the Naïve Bayes model predicts 5 miscellaneous nodes as normal node, and 9 normal nodes as miscellaneous nodes.

SVM

Logistic regression model predicted 145 miscellaneous node as miscellaneous nodes and correctly predicted 335 normal nodes as normal nodes. There are some misclassifications are also occurred, the SVM model predicts 5 miscellaneous nodes as normal node, and 15 normal nodes as miscellaneous nodes.

The following graphic depicts an accuracy examination of four algorithms, with the proposed and present approaches along the x-axis and the accuracy value along the y-axis. There is a 98% success rate in group instruction. The accuracy of the Gaussian Navies base is 97%. Logistic regression, the recommended voting classifier, achieves 99% accuracy, while the support vector machine achieves 96% accuracy. Based on comparisons with various techniques, it may be said that logistic regression yields useful results.

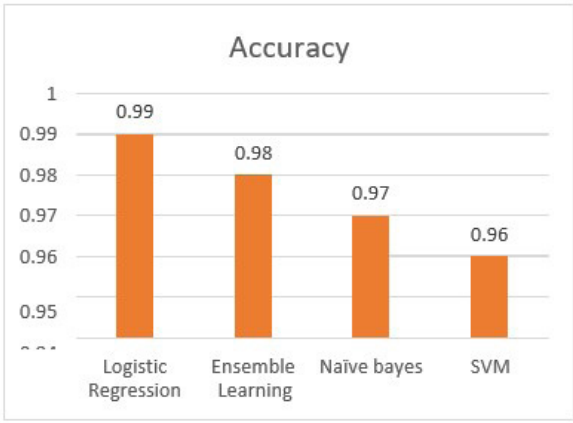


Figure-1 Accuracy study

The figure below depicts a sensitivity analysis of four algorithms, where the suggested solution and current algorithms make up the x-axis and the sensitivity value occupies the y-axis. Almost often, learning in a group requires sensitivity. The Gaussian Navies base has a 96% sensitivity. The accuracy of the recommended voting classifier logistic regression is 99%, while the support vector machine's sensitivity is 95%. One may assume that logistic regression produces successful results when compared to other techniques.

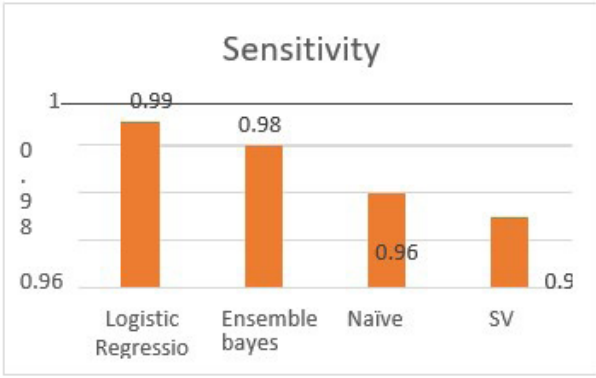


Figure-2 Sensitivity study

The following graph displays a specificity study of 4 algorithms, where the x-axis indicates the suggested and current technique and the y-axis displays the specificity value. A specificity of 97% applies to ensemble learning. A 97% specificity is included in the Gaussian Navies basis. In contrast to the support vector machine, which has a 95% specificity, the proposed voting classifier logistic regression has a 99% specificity. Logistic regression is thought to provide successful results when compared to other techniques.

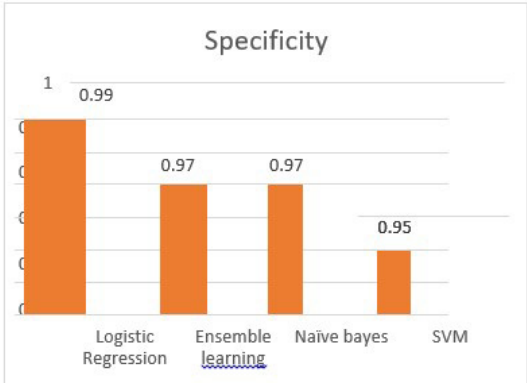


Figure-3 Specificity study

5. CONCLUSIONS

The suggested hybrid trust and blockchain-based architecture for heterogeneous IoT systems has shown promising outcomes in terms of scalability, fault tolerance, and security. The framework may be used to a variety of IoT systems and offers a trust-based foundation for safe data exchange and communication. The framework can successfully analyze data from various IoT devices and identify abnormalities or malicious actions by employing the logistic regression approach. Trust levels are also used to identify malicious nodes in a heterogeneous IOT system. This data is then used to build a logistic regression model that can identify malicious nodes. The suggested method is compared to existing algorithms based on performance metrics to demonstrate its efficiency. For proposed and current algorithms, performance principles like as specificity, sensitivity, and accuracy are compared. The suggested method has a better accuracy score of 99% than the current approach. All suggested algorithm values are greater than current algorithm values, demonstrating that the new method is very efficient. As a result, it is possible to infer that logistic regression is very accurate in detecting harmful behaviors in heterogeneous IOT systems.

ACKNOWLEDGEMENT

I thank the Department of Computer Science and Engineering, Akshaya Institute of Technology, Tumkuru for providing facilities to conduct experiments.

REFERENCES

1. L. Tseng, L. Wong, S. Otoum, M. Aloqaily, and J. Ben Othman, "Blockchain for Managing Heterogeneous InternetofThings: A Perspective Architecture, "*IEEENetw.,*2020, vol.34, no. 1, pp. 16–23, doi:10.1109/MNET.001.1900103.

2. B. S. Egala, A. K. Pradhan, V. Badarla, and S. P Mohanty, "Fortified-Chain: A Blockchain-Based Framework for Security and Privacy-Assured Internet of Medical Things with Effective Access Control," *IEEE Internet Things J.,* 2021, vol. 8, no. 14, pp. 11717–11731,doi: 10.1109/JIOT.2021.3058946.

3. M. A. Rahman, M. S. Abuludin, L. X. Yuan, M. S. Islam, and A. T. Asyhari, "EduChain: CIA-Compliant Blockchain for Intelligent Cyber Defense of Microservices in Education Industry 4.0, "*IEEE Trans. Ind. Informatics*, 2022, vol. 18, no. 3, pp. 1930–1938, doi: 10.1109/TII.2021.3093475.

4. W. Li and W. Meng, "BCTrustFrame: Enhancing Trust Management via Blockchain and IPFS in 6G Era," *IEEE Netw.,* 2022, vol. 36, no. 4, pp. 120–125, doi: 10.1109/MNET.013.2100768.

5. S. M. Basha, "Inter-Locking Dependency Evaluation Schema based on Block-chain Enabled," no. April 2022.

6. JOHN KENEDI, "Access to Justice for Victims of Ethnic Based Violence in Ethiopia : The Case of Oromia Region BASED VIOLENCE IN ETHIOPIA : THE CASE OF," 2021.

7. K. Yuzik and D. Makaroff, "Blockchain-Based Security for Heterogeneous IoT Systems," *Proc. 30th Annu Int. Conf. Comput. Sci. Softw. Eng.,* pp. 63–72, 2020.

8. T. Qiu, N. Chen, K. Li, M. Atiquzzaman, and W. Zhao "How can heterogeneous internet of things build our future: A survey," IEEE Commun. Surv. Tutorials, doi:10.1109/COMST.2018.2803740. vol. 20, no. 3, pp. 2011–2027, 2018,

9. J. Li, T. Liu, D. Niyato, P. Wang, J. Li, and Z. Han, "Contract-Theoretic Pricing for Security Deposits in Sharded Blockchain with Internet of Things (IoT)," IEEE Internet Things J., vol. 8, no. 12, pp. 10052–10070, 2021, doi: 10.1109/JIOT.2021.3049227.

10. Y. Liu, K. Wang, K. Qian, M. Du, and S. Guo, "Tornado: Enabling Blockchain in Heterogeneous Internet of Things Through a Space-Structured Approach," IEEE Internet Things J., vol. 7, no. 2, pp. 1273–1286, 2020, doi: 10.1109/JIOT.2019.2954128.

11. K. Fanning and D. P. Centers, "B lockchain and Its Coming," pp. 53–57, 2016, doi: 10.1002/jcaf.

12. A. Lakhan, M. Ahmad, M. Bilal, A. Jolfaei, and R. M. Mehmood, "Mobility Aware Blockchain Enabled Offloading and Scheduling in Vehicular

Fog Cloud Computing," IEEE Trans. Intell. Transp. Syst., vol. 22, no. 7, pp. 4212–4223, 2021, doi: 10.1109/TITS.2021.3056461

13. S. Kumar, "Privacy preservation and security challenges : a new frontier multimodal machine learning research September, 2022, doi: 10.1504/IJSNET.2022.10049988.

14. M. A. M. De Oca and T. St, "Université Libre de Bruxelles Frankenstein ' s PSO : An Engineered Composite Particle Swarm Optimization Algorithm," Most, no. March, 2007.

15. J. L. Fernández Martínez and E. García Gonzalo, "The Generalized PSO: A New Door to PSO Evolution," J. Artif. Evol. Appl., vol. 2008, no. 1, pp. 1–15, 2008, doi: 10.1155/2008/861275.

16. V. P. Nguyen, E. M. Ung, A. Krishna, and S. Tham, Lossless compression of topology of 3D Triangulated irregular networks. 2016. doi: 10.1109/ICICS.2015.7459974.

17. A. Field, Methods of Regression An Example Logistic Regression on SPSS Methods Residuals," Options, pp. 1– 4, 2009.