Journal of Nonlinear Analysis and Optimization
Vol. 15, Issue. 1 : 2024
ISSN : **1906-9685**

Journal of Nonlinear Analysis and Optimization : Theory & Applications

Editors-in-Chief :

ENHANCING CYBER THREAT HUNTING IN BLOCKCHAIN-BASED IIOT NETWORKS WITH BLOCK HUNTER FEDERATED LEARNING

#1GARIGE ANIL KUMAR, PG Scholar,
#2B.VEERA PRATHAP, Assistant Professor,
#3R. ADI NARAYANA, Assistant Professor,
Department of Computer Science Engineering,
KLR COLLEGE OF ENGINEERING AND TECHNOLOGY, PALONCHA.

ABSTRACT:

Blockchain-based solutions are becoming increasingly popular in a variety of businesses as a means of improving data security. A major application of blockchain technology in the Industrial Internet of Things (IIoT) is a chain-based network. The use of IIoT devices has grown dramatically in today's digital age, particularly in the development of smart factories. Despite its many remarkable qualities, blockchain remains vulnerable to attacks. To protect networks and systems against unanticipated attacks, it is critical to detect abnormalities in blockchain-powered Industrial Internet of Things (IIoT) networks running in smart factories. This study develops a threat hunting system called Block Hunter to autonomously detect attacks in Industrial Internet of Things (IIoT) networks using blockchain technology. The technology accomplishes this through federated learning (FL). To detect abnormalities, Block Hunter combines many machine learning models in a federated environment, as well as a cluster-based structure. Block Hunter is the first federated threat hunting methodology in IIoT networks that can detect aberrant behavior while protecting privacy, according to our current knowledge. Our findings show that the Block Hunter is effective at recognizing anomalous behaviors with high accuracy and minimum data usage.

1.INTRODUCTION

Blockchain data security qualities, such as immutability and tamper-resistance, make it a powerful tool in a variety of areas, including finance, healthcare, military, and networking, thanks to its technical trajectory. The development of Industrial Internet of Things (IIOT) devices is causing the worldwide landscape to become increasingly intelligent and networked; factories, in particular, are gradually improving their intelligence and efficiency as technology advances is defined as a subset of the Internet of Things (IOT) framework. In terms of security requirements, IIOT and IoT differ. The IIOT not only improves consumer convenience and usability, but it also aims to increase industrial safety and efficiency. IIOT devices are typically used in business-to-business (B2B) settings, whereas IoT devices are primarily used in business-to-consumer (B2C) scenarios. This would result in a distinct adversary profile for IIOT networks as opposed to IOT networks, which favor device-to-device interactions. IIOT networks provide a complete framework that supports a wide range of applications and enables us to meet the needs of customers, particularly in the context of smart manufacturing Blockchain technology has gained extensive use in IIOT-based networks, including linked drones, healthcare systems, smart factories, smart homes/buildings, and smart communities, due to its multiple benefits The suggested framework could be used in different IIOT scenarios, although the primary focus of this research is the security of block chain-based IIOT networks in smart factories In today's smart factories, several devices are connected to public networks, and intelligent systems enable a wide range of operations, including temperature monitoring systems, Internet-capable lighting, IP cameras, and IP phones.

These gadgets have the ability to deliver crucial safety services as well as store private and sensitive information. The key worry that will develop as the proliferation of IIOT devices in smart factories is the secure storage, collecting, and sharing of data. As a result, personal, industrial, and essential information are at danger in this scenario. Block chain technology can ensure the availability of communication backbones and data integrity both inside and outside of smart factories through strong authentication. However, privacy and security concerns remain significant barriers in the field of IIoT The probability of fraudulent conduct occurring within blockchain-based networks is a major worry.

Despite its high efficacy, blockchain technology is still vulnerable to cyber assaults. The weaknesses of this block chain network have been exposed, for example, by a 51% cyber attack on Ethereum Classic and three successive attacks in August 2020 resulting in the theft of more than $5 million in bitcoin. Smart factories must ensure the confidentiality of user data during transmission, use, and storage. Stolen data is vulnerable to manipulation by dishonest individuals with malevolent intent who want to access, edit, or use the data. Statistically, these assaults can be classified as anomalous occurrences, demonstrating a significant departure from typical behavior Threat hunting programs and system protection against unauthorized access rely heavily on the detection of out-of-the-ordinary occurrences, which are autonomously identified and filtered.

The primary goal of this article is to identify suspicious users and transactions in a smart factory-specific block chain-based IIOT network. In this situation, anomaly serves as a catch-all for suspicious activity. Machine Learning (ML) systems can detect abnormalities and attacks on the blockchain, allowing them to discover outliers and patterns. Deep neural networks are seen as a promising solution for anomaly detection due to their ability to autonomously gain representations from the data on which they are trained. Nevertheless, ML and deep learning-based anomaly detection systems are not without challenges. These approaches face hurdles due to a paucity of training data and privacy considerations.

The detection of block chain anomalies is a complicated task. The training procedure takes longer not only because each block must be sent to a central server, but also because the model requires new block data during the testing phase. Furthermore, the regular update of machine learning models to detect abnormalities and respond to emerging threats might expose the models to intentional degradation via causation or data poisoning assaults directed by malicious actors. Adversaries may purposefully transmit modified payloads in order to avoid anomaly detection. A novel and practical approach would be to use Federated Learning (FL) models for anomaly detection while preserving

A summary of the paper's main contributions follows:

Blockchain data security features such as immutability and tamper-resistance, enabled by technological advancements, make it an effective tool for a wide range of industries, including networking, healthcare, finance, and the military. The world is becoming more intelligent and networked as a result of the proliferation of Industrial Internet of Things (IIOT) devices; factories, in particular, are becoming more intelligent and efficient as technology advances [1]. IIOT refers to a subset of the IoT architecture. IIOT and IoT have different security requirements. The IIOT aims to increase industrial productivity and safety while also improving consumer usability and convenience. IIOT devices are typically used in business-to-business (B2B) settings, whereas IoT devices are primarily used in business-to-consumer (B2C) scenarios. As a result, IIOT networks would have a distinct adversary profile from IOT networks, which prioritize device-to-device interactions.

IIOT networks enable a comprehensive infrastructure that supports a wide range of applications and allows us to meet customer demands, particularly in the context of intelligent manufacturing Because of its numerous advantages, blockchain technology has been widely adopted by IIOT-based networks such as connected drones, healthcare systems, smart factories, smart homes/buildings, and smart communitiesThe proposed framework could be useful in a wide range of IIOT scenarios, but the primary goal of this research is to secure block chain-based IIOT networks in smart factories In today's smart factories, many devices are connected to public networks, and intelligent systems provide a variety of functions, including Internet-accessible IP phones, IP cameras, temperature monitoring systems, and lights. These devices can protect sensitive and private data while also providing essential safety functions.The primary concern with the widespread use of IIOT devices in smart factories will be how to securely collect, store, and share data.

This situation puts critical, industrial, and personal information at risk. Block chain technology, with strong authentication, can ensure the availability of communication backbones and data integrity both inside and outside of smart factories. However, there are still numerous challenges in the IIoT space related to privacy and security concerns.One of the primary concerns is the possibility of fraudulent activity occurring in blockchain-based networks. Despite its tremendous efficacy, blockchain technology remains vulnerable to cyberattacks. For example, a 51% cyberattack on Ethereum Classic and three separate attacks in August 2020 [5], which resulted in the theft of more than $5 million in bitcoin, have exposed flaws in this block chain network. Ensuring the confidentiality of user data during transmission, use, and storage is critical in smart factoriesData that has been stolen is vulnerable to manipulation by dishonest individuals who want to access, alter, or use the information for malicious purposes. Based on statistical analysis, these attacks are classified as anomalous events because they deviate significantly from the norm.

The detection and filtering of anomalous events is an important part of threat hunting programs and system security against unauthorized access.The main goal of this article is to identify questionable users and transactions in a block chain-based IIOT network designed for smart factories. In this context, anomaly refers to a wide range of questionable activities. Outliers and patterns can be identified using Machine Learning (ML) algorithms that detect anomalies and attacks on blockchain. Because deep neural networks can learn representations from the data on which they are trained, they appear to be a promising solution for anomaly detection . However, there are several challenges with anomaly detection systems based on deep learning and machine learning. The lack of training data, as well as privacy concerns, pose challenges to these methods . Block chain anomaly detection is a difficult task [8]. In addition to the time required to send each block to a central server, the training process is extended because the model requires new block data for testing . Furthermore, the frequent updating of machine learning models to detect anomalies and respond to new threats may expose the models to deliberate deterioration by hostile actors who use data poisoning or causation. Adversaries may send altered payloads on purpose to avoid anomaly detection. Federated Learning (FL) models would be an innovative and useful method for anomaly detection while maintaining.

2. BACKGROUND WORK.

Existing System

Sayadi et al. proposed a technique for detecting abnormalities in bitcoin transactions. A study was conducted to investigate the use of One-Class Support Vector Machines (OCSVM) and K-means to classify outliers based on shared categories and statistical significance. After examining their efforts in regard to the advancement of findings for acknowledgement, they concluded that highly accurate results were possible. Reference describes a method for identifying abnormalities in blockchain-based Internet of Things networks. The procedure of finding unusual behavior on the blockchain was demonstrated. The goal of data collection in branches is to determine the group's ability to detect anomalous activity. A innovative technique was devised to improve the security of blockchains and their associated technology. Encoder-decoder deep learning regression is another way for identifying security flaws in blockchains. This study created an algorithm that analyzes data acquired while monitoring the bitcoin blockchain in order to detect suspicious behavior. This model demonstrates its ability to detect existing attacks by studying the Ethereum network's historical data, as supported by empirical research. According to Chai et al., FL has the potential to improve environmental data interchange and teaching when integrated with a hierarchical blockchain architecture. This architecture can offer substantial benefits for large transport networks. Federated learning (FL) successfully addresses the distribution patterns and privacy concerns of the Internet of Vehicles. To encourage individuals to share their information, a sophisticated framework with a large number of leaders and participants is used to replicate the process of trading. Empirical study has demonstrated that deploying a hierarchical program can improve people's ability to share, learn, and protect themselves from certain forms of hostile attacks. In addition, the authors of present a thorough study of how FL could improve security features and successfully avoid various malware attacks. This article examines some significant challenges and potential solutions that researchers should consider in order to improve the practical application of FL.

Disadvantages

- The current system does not include the Isolation Forest (IF) model, a tree-based strategy for finding anomalies.
- The technique makes no mention of the Cluster-Based Local Outlier Factor.

Proposed System

The detection of abnormal behavior is a critical component in protecting an automated system from unexpected attacks. To discover any possible concerns, each updated data block must be transmitted to a central computer. This serves little purpose and raises privacy concerns. Florida's decisions appear to be a sensible response to the situation. FL is used to confirm the model's validity and to create a universal model capable of detecting anomalies. To include the parameters, we shall send a request containing them to the parameter server. We will update our core model after we have received information about the devices, data, and service providers associated with each smart factory. Implementing a cluster-based design improves resource allocation efficiency in all intelligent factories while also speeding up blockchain transaction processing. Clustering allows the computer to build the underlying network in a hierarchical form.

Advantages

Federation Construction: The local recipient of the model is the cluster, a subset of smart factory members.

Decentralized Training: Upon selecting a cluster of smart factories, its model is updated with local data. Data Model Accumulation: In charge of compiling and combining the data models. Individual data is not transmitted and integrated from the federation to the server. FedAvg, or Model Aggregation: The parameter server computes an improved global model by aggregating model weights

3. MODULES

The Service Provider must enter a valid user name and password to log in to this module. Upon successful login, he can perform several tasks like logging in, training and testing IIOT network datasets, Examine Trained and Examined

View and Authorize

Users The administrator can see a list of all enrolled users in this module. In this, the administrator may see user information such name, email address, and address, and they can also approve people.

Remote User

There are n numbers of users present in this module. Prior to beginning any operations, the user must register. The user's information is saved in the database after they register. Upon successful registration, he must use his permitted user name and password to log in. Following a successful login, the user can perform many tasks like VIEW YOUR PROFILE, PREDICT CYBER THREAT HUNTING TYPE, and REGISTER AND LOGIN.

Decision tree classifiers

Decision tree classifiers have proven effective in a wide range of applications. The ability to extract descriptive decision-making knowledge from the provided data is their key characteristic. Training sets can be used to create decision trees. The following is the process for creating such a generation based on the set of objects (S), each of which is a member of one of the classes C1, C2,..., Ck:

Step 1: The decision tree for S has a leaf labeled with this class if every item in S is a member of the same class, such as Ci. Step 2. If not, let T be a test with O1, O2,... On as potential results. The test divides S into subsets S1, S2,..., Sn where each object in Si has result Oi for T. This is because each object in S has a single outcome for T. T becomes the decision tree's root, and we create a subsidiary decision tree for each outcome Oi by applying the same process recursively to the set Si.

Gradient boosting

Gradient boosting is a machine learning technology that has applications in a variety of fields, including classification and regression problems. Several subpar prediction models, often decision trees, are used to build the prediction model. The user's material includes two references: and. When the weak learner is a decision tree, gradient-boosted trees are formed, which frequently outperform random forests.Gradient boosted trees, like other boosting approaches, generate models slowly and gradually. Nonetheless, they outperform earlier techniques by allowing for the optimization of any differentiable loss function.

K-Nearest Neighbors (KNN)

- A simple but effective classification algorithm uses a similarity metric to determine categories.
- In contrast with parametric
- Ineffective instruction provides knowledge only after the test case is presented.
- To categorize new data, the K-nearest neighbors are determined using the training data.

Example

- The training set is made up of the k-nearest samples in the feature space.
- Feature spaces are mathematical representations used to express categorical variables that lack a numerical measurement scale.
- Instance-based learning techniques are naturally slow since it may take more time to incorporate examples that closely resemble the input vector for testing or prediction into the training dataset.

Logistic regression Classifiers

The relationship between a set of independent (explanatory) variables and a categorical dependent variable is studied using logistic regression analysis. When the dependent variable has only two possible values, such as 0 and 1, or Yes and No, logistic regression is utilized. When the dependent variable has three or more unique values (married, single, divorced, or widowed), the term multinomial logistic regression is commonly used. The methodology is used in a similar manner, even if the dependent variable's data format differs from multiple regressions. Discriminant analysis and logistic regression are two techniques used to evaluate categorical answer variables. Many statisticians believe that logistic regression is a more adaptable and helpful modeling tool than discriminant analysis in most cases. Unlike discriminant analysis, logistic regression does not make the assumption that the independent variables are uniformly distributed. This program performs multinomial and binary logistic regressions with numerical and categorical independent variables. In addition to the regression equation, there are quality of fit, odds ratios, confidence intervals, probability, and deviance. The diagnostic residual charts and reports, as well as the residual analysis, are complete. An independent variable subset selection search can be used to find the best regression model with the fewest independent variables. It gives ROC curves and confidence intervals for predicted values to aid in the selection of the best classification cutoff point. You can validate your findings by automatically classifying rows that were not included in the research.

Naïve Bayes

The essential notion underlying the supervised learning strategy known as the naive bayes technique is that the presence or absence of one feature within a class has no bearing on the presence or absence of another feature. Nonetheless, it appears to be both useful and potent. It works in a manner similar to other guided learning strategies. Numerous reasons have been proposed in the literature. The explanation based on representation bias will be the primary focus of this course. The naive bayes classifier is a linear classifier, as are logistic regression, linear discriminant analysis, and linear SVM. The learning bias, also known as discrepancies, occurs during the process of estimating the classifier's parameters. The Naive Bayes classifier is commonly used in research, while practitioners looking for practical results are less likely to employ it. The researchers discovered that it is quite accurate in comparison to other systems, that it is very simple to design and execute, that picking its settings is a straightforward procedure, that learning occurs quickly even on very large databases, and so on. However, end customers are unaware of the benefits of this strategy and do not receive a clear and straightforward model. As a result, we convey the learning process's outcomes in an original manner. The understanding and implementation of the classifier are simplified. The theoretical foundations of the naive Bayes classifier are addressed in the first section of this lecture. After that, we test the technique on the Tanagra dataset. When we compare the model's parameters (the findings) to those from other linear techniques, such as logistic regression, linear discriminant analysis, and linear SVM, we get extremely consistent results. This accounts for a significant portion of the method's performance advantage over other solutions. In the second section, the same dataset is processed using the following tools: Orange 2.0b, R 2.9.2, Knime 2.1.1, Weka 3.6.0, and RapidMiner 4.6.0. Above all, we want to understand the outcomes that were found.

Random Forest

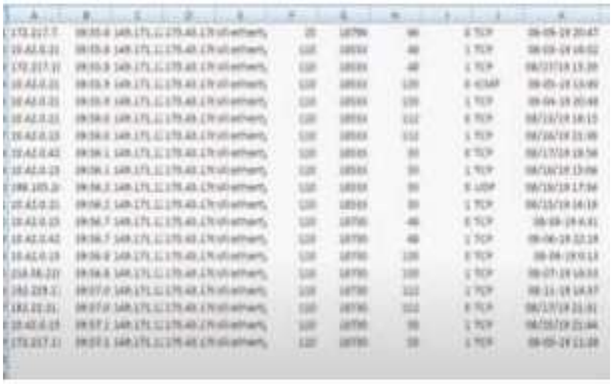
Random forests, also known as random choice forests, are a type of ensemble learning used in training to create a large number of decision trees for tasks such as regression and classification, among others. The random forest algorithm selects the alternative that is selected by the majority of the trees to determine the outcome of the class assignment. Regression tasks use the mean or average forecast produced by each tree as their output. Random choice forests address one of the most common difficulties with decision trees: overfitting in training sets. While decision trees are generally effective, graduated improved trees provide more accuracy. However, the data's fundamental qualities may have an impact on its performance. Tin Kam Ho pioneered the random choice forest method in 1995 [1]. Ho employed the random subspace technique, describing its use as a concrete application of Eugene Kleinberg's "stochastic discrimination" approach to item classification. After developing a variant of the approach, Leo Breiman and Adele Cutler were able to register the trademark "Random Forests" in 2006. Minitab, Inc. is currently the only company owning this trademark.The expansion combines Ho's random feature selection method with Breiman's "bagging" methodology, as well as additional contributions from Amit and Geman. As a result, a set of decision trees is generated that effectively reduce variance. Organizations typically utilize random forests as "blackbox" models because to their low setup costs and ability to give precise forecasts for a wide range of data types.

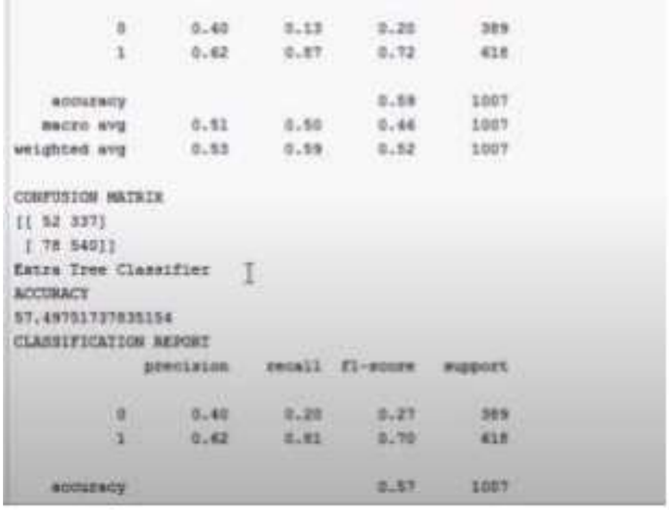
SVM

Discriminant machine learning generates a discriminant function from a training dataset that is identically distributed and independent. In classification tasks, this function may accurately predict new occurrence labels. Unlike generative machine learning techniques, discriminant classification functions assign data points x to one of the many classes in a classification problem without requiring the creation of conditional probability distributions. Discriminant techniques require less computing resources and a smaller training dataset than generative methods. Their major purpose is to identify irregularities in projections. This is especially true when using multidimensional feature spaces to calculate posterior probabilities. Learning a classifier in geometry is the process of determining the equation for a three-dimensional surface that divides the classes of a feature space into groups of the optimal size.

Support Vector Machines (SVM) are a discriminant approach used in convex optimization. SVM solves the problem analytically, consistently producing the same hyperplane optimal value. Machine learning approaches such as genetic algorithms (GAs) and perceptrons are rarely employed for this type of categorization task. The parameters used to start and end a perceptron have a significant impact on the results. Each training set and kernel used to transfer data from the input space to the feature space is associated with a unique set of SVM model parameters. In contrast, the models employed by the GA classifier and perceptron vary with each training cycle. A large number of hyperplanes will successfully achieve the main purpose of perceptrons and Genetic Algorithms (GAs), which is to minimize mistakes during training.

4. RESULTS





```
           0       0.40      0.13      0.20       389
           1       0.62      0.87      0.72       418

    accuracy                           0.58      1007
   macro avg       0.51      0.50      0.46      1007
weighted avg       0.53      0.59      0.52      1007

CONFUSION MATRIX
[[ 52 337]
 [ 78 540]]
Extra Tree Classifier
ACCURACY
57.49751737835154
CLASSIFICATION REPORT
              precision    recall  f1-score   support

           0       0.40      0.20      0.27       389
           1       0.62      0.81      0.70       418

    accuracy                           0.57      1007
```

View Cyber Threat Hunting Type Type Found Ratio Details

Cyber Threat Hunting Type	Ratio
Packet Drop	[illegible]
Packet Hijacking	[illegible]

5. CONCLUSION

This study describes the development of the Block Hunter framework, which employs federated learning to detect anomalies in IIOT smart factories utilizing block chains. Block Hunter reduces resource consumption and increases throughput for block chain-based IIOT network hunting by implementing a cluster-based architecture. In order to identify anomalies, the Block Hunter framework was assessed utilizing an assortment of machine learning methods (NED, IF, CBLOF, K-means, PCA). Additionally, an evaluation was conducted on the Block Hunter's performance with respect to the interval between blocks, miner selection, and block size. Further research utilizing generative adversarial networks (GAN) to construct and execute a framework resembling that of a block hunter would be highly intriguing. Hence, it is prudent to conduct research on the development and implementation of IIOT-integrated blockchain networks employing various consensus algorithms.

REFERENCES

1. J. Wan, J. Li, M. Imran, D. Li, and F. e Amin, “A blockchain-basedsolution for enhancing security and privacy in smart factory,” IEEETransactions on Industrial Informatics, vol. 15,no. 6, pp. 3652–3660,2019.
2. F. Scicchitano, A. Liguori, M. Guarascio, E. Ritacco, and G. Manco,“Blockchain attack discovery via anomaly detection,” ConsiglioNazionale delle Ricerche, Istituto di Calcolo e
3. Reti ad Alte Prestazioni(ICAR), 2019, 2019.
4. Q. Xu, Z. He, Z. Li, M. Xiao, R. S. M. Goh, and Y. Li, “An effective blockchain-based, decentralized application for smart building system management,” in Real-Time Data Analytics for Large Scale Sensor Data.Elsevier, 2020, pp. 157–181.
5. B. Podgorelec, M. Turkanovi´c, and S. Karakatiˇc, “A machine learningbased method for automated blockchain transaction signing includingpersonalized anomaly detection,” Sensors, vol. 20, no. 1, p. 147, 2020.
6. A. Quintal, “Veriblock foundation discloses mess vulnerability in ethereum classic blockchain,” VeriBlock Foundation. [Online]. Available: com/news-releases/veriblock-foundationdiscloses-mess-vulnernability-in-ethereum-classic-blockchain301327998.html
7. [M. Saad, J. Spaulding, L. Njilla, C. Kamhoua, S. Shetty, D. Nyang,and D. Mohaisen, “Exploring the attack surface of blockchain: A comprehensive survey,” IEEE Communications Surveys & Tutorials
8. R. A. Sater and A. B. Hamza, “A federated learning approach to anomaly detection in smart buildings,” arXiv preprint arXiv:2010.10293, 2020.
9. O. Shafiq, “Anomaly detection in blockchain,” Master’s thesis, TampereUniversity, 2019.
10. A. Yazdinejadna, R. M. Parizi, A. Dehghantanha, and H. Karimipour,“Federated learning for drone authentication,” Ad Hoc Networks, p.102574, 2021.

11. D. Preuveneers, V. Rimmer, I. Tsingenopoulos, J. Spooren, W. Joosen,and E. Ilie-Zudor, “Chained anomaly detection models for federated