Hindawi
Security and Communication Networks
Volume 2021, Article ID 9430132, 10 pages
https://doi.org/10.1155/2021/9430132

Research Article

BCEAD: A Blockchain-Empowered Ensemble Anomaly Detection for Wireless Sensor Network via Isolation Forest

Xiong Yang[1,2] **Yuling Chen**[1] **Xiaobin Qian**[3] **Tao Li**[1] **and Xiao Lv**[4]

[1]*State Key Laboratory of Public Big Data, College of Computer Science and Technology, Guizhou University, Guiyang, China*
[2]*Guangxi Key Laboratory of Cryptography and Information Security, Guilin, China*
[3]*Guizhou CoVision Science & Technology Co., Ltd., Guiyang, China*
[4]*Guizhou Shuanhui Big Data Industry Development Co., Ltd., Guiyang, China*

Correspondence should be addressed to Yuling Chen; ylchen3@gzu.edu.cn

Received 27 September 2021; Revised 18 October 2021; Accepted 27 October 2021; Published 10 November 2021

Academic Editor: Weizhi Meng

The distributed deployment of wireless sensor networks (WSNs) makes the network more convenient, but it also causes more hidden security hazards that are difficult to be solved. For example, the unprotected deployment of sensors makes distributed anomaly detection systems for WSNs more vulnerable to internal attacks, and the limited computing resources of WSNs hinder the construction of a trusted environment. In recent years, the widely observed blockchain technology has shown the potential to strengthen the security of the Internet of Things. Therefore, we propose a blockchain-based ensemble anomaly detection (BCEAD), which stores the model of a typical anomaly detection algorithm (isolated forest) in the blockchain for distributed anomaly detection in WSNs. By constructing a suitable block structure and consensus mechanism, the global model for detection can iteratively update to enhance detection performance. Moreover, the blockchain guarantees the trust environment of the network, making the detection algorithm resistant to internal attacks. Finally, compared with similar schemes, in terms of performance, cost, etc., the results prove that BCEAD performs better.

1. Introduction

In recent years, the booming Internet of Things is revolutionizing the world. As its supporting technology, wireless sensor networks have also received extensive attention [1, 2]. WSNs are a multihop self-organizing network formed by many sensor nodes deployed in the monitoring area to communicate. It gets rid of the limitation of the cable, realizes the wireless communication of the network, and has a wide range of application scenarios. However, due to the backwardness of WSNs security technology, various security issues limit the practical application of WSNs [3, 4].

Various security technologies and strategies have emerged for protecting network security. Intrusion detection is a classic network security technology [5]. Early intrusion detection systems (IDS) mostly utilize misuse detection. Misuse detection record the attacks by a signature database, then judge an intrusion with the events or data matching the signatures. However, misuse detection is not practical enough because it cannot detect unrecorded attacks. Nowadays, anomaly detection has been more widely used [6] with the development of machine learning. Anomaly detection comes from the statistical community [7]. It establishes a standard model and judges the events or data that do not match with the model as an intrusion. Although anomaly detection requires some model training time and produces a higher false alarm rate, it can detect new unknown intrusions. The performance of anomaly detection will continue to increase and make outstanding contributions to protecting network security with the optimization of modeling algorithms in anomaly detection.

The structure of intrusion detection systems has become richer for stronger practicability and applicability. For example, the proposal of distributed intrusion detection systems (DIDSs) eases the pressure of detecting heterogeneous networks. The DIDS is similar to ensemble learning [8], and the system builds

multiple detection models in a large-scale network. Therefore, the DIDS not only balances network energy consumption but also improves detection performance. However, distributed anomaly detection must ensure trust between nodes to prevent internal attacks. It is the prerequisite for its further application.

In recent years, the emergence of blockchain technology has pointed out a path worth trying to solve the distributed trust problem in the Internet of Things environment. Blockchain is a peer-to-peer distributed network with features such as non-tempera, decentralization, transparency, and system autonomy [9], which can effectively enhance device security and network collaboration in the Internet of Things. Nowadays, some distributed intrusion detection based on blockchain has been proposed [10]. The system packs the detection results or alarms into blocks and shares them in the network to manage trust among domains. However, due to the limitation of the data shared by the blockchain, the above scheme still has its limitations.

Therefore, we propose a blockchain-based ensemble anomaly detection scheme. The scheme stores the model of a typical anomaly detection algorithm (isolated forest) in the blockchain network and performs distributed anomaly detection in the WSNs. By constructing a suitable block structure and consensus mechanism, the global model for detection can be iteratively updated to enhance the detection performance. Moreover, the blockchain guarantees the trust environment of the network, making the detection algorithm resistant to internal attacks. Finally, compared with similar schemes in terms of performance, cost, security, etc., the results prove that BCEAD performs better.

1.1. Contributions.

(1) We propose an ensemble anomaly detection network structure for WSN. The structure has three levels, and high-performance nodes in the sink layers bear most of the storage and computing overhead required by the solution. Multiple sink nodes will independently detect anomalies in the local network based on the global anomaly detection model and optimize the model by submitting some parameters. This network structure is suitable for resource-constrained and heterogeneous WSNs and can be equipped with multiple types of anomaly detection algorithms to ensure network security.

(2) We design an anomaly detection model stored and updated by the blockchain. The detection node filters a batch of isolated trees stored in the blockchain to form an isolated forest for detection. In addition, new isolated trees are continuously generated and published to the blockchain, so that the isolated forest model, which is utilized to detect, is constantly updated. In this scheme, the detection model keeps dynamically updated with the environment, which enhances the detection performance while maintaining security.

2. Related Works

2.1. Anomaly Detection Structure. Intrusion detection is an important part of network security protection, and it is a network security technology that actively protects its network and system from illegal attacks. An intrusion detection system [11] usually consists of data collectors, data analyzers, alarm modules, and other parts. Traditional intrusion detection systems begin to feel weak with the expansion of the network scale and the complexity of information. Therefore, distributed intrusion detection systems and collaborative intrusion detection systems have been proposed. They not only analyze system logs but also analyze network traffic and introduce a distributed data collection mechanism into the structure. Subsequently, the anomaly detection structure for network attacks has gradually changed from local and centralized to a distributed structure. Specifically, the new detection system also deploys multiple data analyzers. Data analyzers in each network domain can communicate with each other and share detection models and strategies.

Distributed or collaborative anomaly detection has better detection performance for large-scale heterogeneous networks. However, the distributed anomaly detection has to consider the issue of trust among nodes. For example, the system may crash if a detection node in the detection system falsely sends alarms or maliciously updates the global detection model. Therefore, it is vital to build a trusted environment that can prevent nodes from internal attacks. In addition, distributed detection faces privacy issues. During the detection process, a large amount of data is collected and uploaded by the agent, which exposes the system to hidden dangers and threats of data leakage.

2.2. Anomaly Detection Method. Anomaly detection is the identification of events or observations that do not match the expected pattern. In different scenarios, anomalies are also called outliers, noises, deviations, etc. Isolation forest is a type of typical anomaly detection algorithm [12, 13], which distinguishes abnormal data with "few" and "different" characteristics. Compared with other classic classification methods, isolated forests consume fewer computing resources and can still maintain good performance when processing large amounts of high-dimensional data.

The isolated forest algorithm continuously divides the data by isolated trees, calculates the isolation score according to the height of the data point in the tree, and judges the anomaly according to the average isolated score. Suppose T is a node of an isolated tree, then T may be a leaf node, or an intermediate node with a decision threshold β and two child nodes (T_l, T_r). Assuming a dataset $X = \{x_1, ..., x_n\}$, each data x is a d-dimensional vector. By continuously selecting attributes $q \in d$ randomly and the decision threshold $\beta \in q$ to continuously classify $X' \subseteq X$, an isolated tree can be established. The isolated forest classifies the data by constructing a large number of random isolated trees. In general, data points classified into leaf nodes earlier may be more suitable for the definition of anomalies. Therefore, the

isolation forest can quantify the degree of the anomaly of the data by the average tree length path of the data points.

Specifically, given a sample set with φ instances, the average path length of each isolated tree is c,

$$c(\varphi) = \begin{cases} 2H(\varphi - 1) - \dfrac{2(\varphi - 1)}{n} & \varphi > 0 \\ 1 & \varphi = 2 \\ 0 & otherwise \end{cases}, \quad (1)$$

where $H(i)$ is the harmonic number, which can be estimated with $\ln(i) + 0.5772156649$ (Eulerian constant). $c(\varphi)$ is the average $h(x)$ for a given φ. The anomaly score s of instance x is defined as

$$s(x, \varphi) = 2^{-(E(h(x))/c(\varphi))}, \quad (2)$$

where $E(h(x))$ is the average value of $h(x)$ in an isolated tree, which is the average height of the tree. From the results, the score of a sample close to 1 is judged to be abnormal; the score close to 0 is judged to be safe; and if the scores of all samples are 0.5, it means that the sample set has no obvious abnormality. However, the isolation forest algorithm also has shortcomings. Because it belongs to unsupervised learning, the algorithm is more dependent on the quality of the training set. In practical applications, it is necessary to ensure the real-time performance of the training set and continuously train and change the detection model to adapt to environmental changes and ensure detection performance.

2.3. Anomaly Detection and Blockchain. Blockchain is a new decentralized infrastructure and distributed computing paradigm that has gradually emerged with the increasing popularity of digital cryptocurrencies such as Bitcoin. At present, it has been highly valued and widely concerned by government departments, financial institutions, technology companies, and capital markets [14]. Blockchain is often understood as a data structure [15]. The block stores the data composed of Merkle tree and forms a chain through hash pointers, thereby ensuring that the data are difficult to be changed. In addition, the blockchain uses pure mathematical methods to establish trust relationships among distributed nodes to form a decentralized trusted distributed system, which has the characteristics of decentralization, network robustness, security, and credibility. In addition, the blockchain uses pure mathematical methods to establish trust relationships among distributed nodes to form a decentralized trusted distributed system, which has the characteristics of decentralization, network robustness, security, and credibility.

Therefore, the blockchain enables mutual trust between different participants [16] in the Internet of Things environment, which greatly reduces the cost of reshaping or maintaining trust for each node. For the trust problem of distributed detection, blockchain is also a kind of solution worth trying. Some schemes, which combine intrusion detection with blockchain to ensure trusted data sharing in distributed intrusion detection, have been proposed [10, 17, 18]. They store different data, such as detection characteristics, detection alarms, and detection results in the blocks, and publish to the blockchain network to share. Subsequently, a type of anomaly detection framework driven by blockchain on edge intelligence appeared [19]. The framework stores the data features to be analyzed on the blockchain, then the cloud-based detection model reads the data features on the blockchain for anomaly detection and feeds back the detection results. In addition, the framework transfers the overhead pressure for detection to the distributed edge network, which is more suitable for the system structure of the IoT. Therefore, the advantages of low detection delay and global model update brought by distributed detection could be supported by the blockchain. Recently, a scheme for detecting abnormal behavior in social networks [20] has been proposed, which combines isolation forest and blockchain technology. The authors claim that the blockchain can protect the privacy leakage problem in the anomaly detection process. They execute the isolated forest algorithm to detect data anomalies by the smart contracts, then marked and stored the abnormal data on a separate blockchain.

All the above detection schemes utilize the advantages of blockchain to solve a certain degree of security or privacy issues, but each has certain limitations. Specifically, they all store detection-related data on the blockchain. Although each solution has made optimizations to reduce the storage overhead, for example, the blockchain only stores the characteristic value or the hash value of the detection data. However, as the detection cycle lengthens, the blockchain will still face a storage bottleneck, resulting in much loss of detection performance. Therefore, we propose the BCEAD, which is different from the traditional scheme, to solve the storage problem and enhance the detection performance by storing the detection model.

3. A Blockchain-Based Ensemble Anomaly Detection

As shown in Figure 1, there is the multilayer network structure in BCEAD, which consists of the sensing layer, the sink layer, and the blockchain layer. The roles and functions of each layer are explained as follows:

Sensing Layer. The sensing layer contains a large number of low-cost, low-energy, and low-performance sensor nodes. The sensor node collects physical information from the external environment in real-time and converges it to the sink node. The collected data will be processed and used to generate the corresponding feature matrix for subsequent anomaly detection.

Sink Layer. The sink layer converges the environmental information collected by the sensors and submits it to the base station. Compared with sensor nodes, sink nodes have better computing and storage capabilities. Therefore, the sink node is responsible for most of the

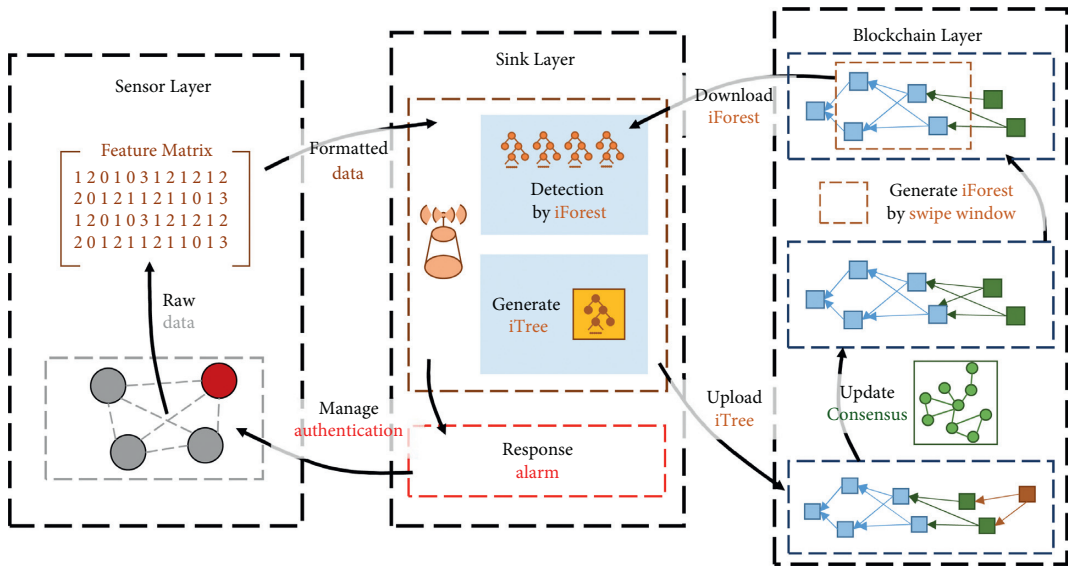

Figure 1: The framework of the BCEAD.

work of BCEAD. In detection, the sink nodes select a batch of the latest isolated trees from the blockchain by the sliding window algorithm. Then, an isolation forest is constructed to detect anomalies for the data submitted by the sensor layer. After detection, the sink node will contribute a new isolated tree to help optimize the global model. Finally, the sink node will respond according to the detection result.

Blockchain Layer. The blockchain layer is mainly responsible for maintaining a safe, usable, and constantly updated global detection model in a distributed network. The sink nodes keep verifying each released block according to the consensus mechanism to ensure that each update for blockchain is benign. The isolated forest, which is a global detection model, will be redundantly stored on the blockchain. Finally, the sliding windows mechanism will screen out suitable blocks to form the isolation forest model for anomaly detection.

3.1. Isolation Blocks. In the detection, the sink node located in the sink layer utilizes the detection model stored in the blockchain layer to detect the data of the local sensor layer. The detection model (isolation forest) is composed of several isolated trees, so a single block in the blockchain is an isolated block containing an isolated tree. Then, the isolated blocks continue to increase, and the blockchain is periodically updated. Therefore, the detection model keeps iterative dynamically.

In BCEAD, the block format includes timeStamp, block ID, the hash value of current and previous blocks, node ID, and isolated tree,

$$Block = [timeStamp \| blockchainI\ D \| hashPre \| hashCur \| NID \| iTree]. \tag{3}$$

An isolated tree contains several nodes, each node contains the decision threshold β and the left node T_l and the right node T_r. The data evaluated by the tree will help the isolated tree to form its structure,

$$iTree = \left[iTree_{left} \| \beta \| iTree_{right}\right]. \tag{4}$$

As shown in Figure 2, the isolation tree $iTree$ is the parameters submitted for each update of the global model. It is also the main content of the block released by the node after each round of consensus. This is the difference between BCEAD and previous solutions: BCEAD does not stores the detect-related data on the blockchain but instead stores the detection model. It avoids data privacy issues and reduces storage and communication overhead. However, the update of the global model can also be consensus because the global nodes can verify whether an isolation tree is appropriate.

3.2. Sliding Window. The number of samples (isolated trees) used by the isolation forest will affect the detection performance of the model. The optimal value of the number of samples has been verified as $\varphi = 250$. BCEAD stores the detection model on the blockchain network, and the number of isolated blocks included in the blockchain keeps increasing. Therefore, the detection scheme has to filter the appropriate number of isolated blocks.

As shown in Figure 3, BCEAD sets up a sliding window algorithm to screen isolated trees for the detection model. The sliding window is essentially a fixed-size list, including a

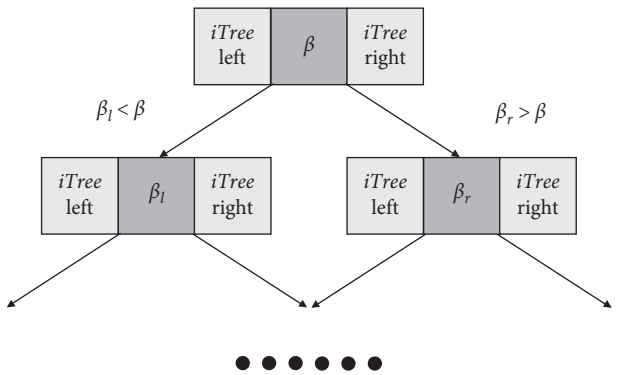

Figure 2: The structure of the isolated tree.

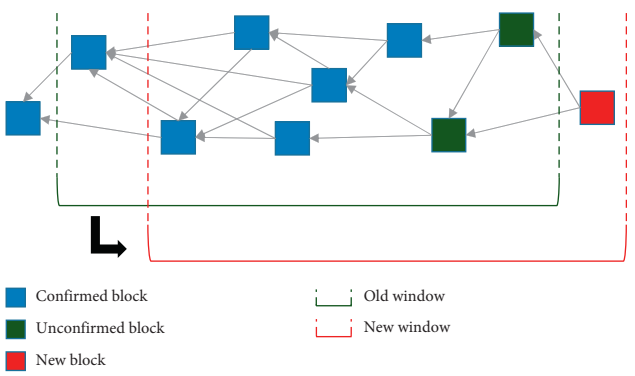

Figure 3: The sliding window for isolation forest.

series of block indexes. With the addition of new blocks, the total length of the blockchain will gradually increase, but the swiping window always traces back from the latest block and includes φ blocks. The update of the swiping window after each round of consensus in the blockchain network, each sink node will get a new slide window when synchronizing the latest blockchain.

Therefore, after detection, the sink node submits an isolated tree trained by the current data. The isolated tree will be packaged into blocks and released to the blockchain network. The isolated tree will be packaged into blocks and released to the blockchain network, and the model generated by each subsequent detection may utilize the isolated tree. This setting ensures the real-time update of the global detection model. Moreover, the training set used to generate the model is the current real-time detection data, which ensures the performance of the detection algorithm.

3.3. Block Forest. In the blockchain network, all nodes store a public ledger locally. During the continuous update of the ledger data, all nodes always communicate to reach consensus and ensure consistency. In BCEAD, the blockchain network uses IOTA (tangle) consensus. Compared with the traditional blockchain, the block in tangle is a single transaction. Each new transaction will verify and quote the previous two, so the network does not need to reach a consensus immediately [21]. Therefore, the network does not require miners, which avoids mining attacks such as in [22, 23]. Users can complete transaction verification by themselves, and the cost of the transaction is only the computational cost of verifying the other two transactions [24]. Therefore, the tangle blockchain can achieve mutual trust between nodes in distributed detection systems. It can both ensure security and considerable performance.

All sink nodes in BCEAD act as participants in the tangle, using isolated trees instead of blockchain transactions, and each block packs an isolated tree separately. In the block releasing, the new isolated tree will be packed into a specified structure block, and quote after the previous two verified blocks on the chain. The specific process is as follows:

(1) *Verify Block.* The sink node visits the tips list (maintains unconfirmed blocks in the network) and randomly selects blocks for verification. The verification will review the format of the block, the node ID of the publishing block, and its reference block. When a block is verified, the sink node puts it into the list, which contains references, and ends the verification work when the list has 2 elements.

(2) *Release Block.* The sink node refers to two blocks in the list, attaches the block ID, its ID, encrypts the block content with its private key, and calculates the hash value of the current block, then broadcasts the block to the blockchain network. Other nodes in the network will review the format and source of the new block, add it to the end of the tangle, and update tips. After several rounds of blockchain updates, subsequent references to a block will prove its credibility.

The application of a blockchain network for trusted communication among nodes can resist internal attacks because the blockchain network always verifies every message, even it comes from the trusted nodes. Therefore, attackers can only deplete the performance of the global detection model by publishing extreme or malicious isolated trees. Therefore, in the blockchain consensus, the verification includes querying the ID of the promulgator, and the purpose is to prohibit a node from frequently publishing the block. The experimental part in Section 4 proves that a single node publishing block has to set two rounds of consensus cycle cooling time, which can effectively prevent malicious nodes from destroying the model and ensure the benign performance of the detection model.

3.4. Algorithm Description. This paper proposes an isolation forest-based anomaly detection algorithm based on blockchain. The algorithm stores the isolation forest in the blockchain, and each detection will build a model based on the blockchain to detect the network data. The details are shown in Algorithm 1.

In this algorithm, first initialize the sink node SN (line 2). The sensor node runs $sniferPackets$ to capture the external environment information and obtain the unprocessed sensor information Raw_data (line 3). Through the feature extraction function $featureExtractor$, the feature matrix FM is obtained after processing the sensor data (line 4–6).

```
input: log files and data packets of the networks
output: response the detection and update the detection model
(1)  begin:
(2)  SN ← this;
(3)  Raw_data ← snifferPackets();
(4)  for all elements of Raw_data do
(5)     vector = featureExtractor(elements);
(6)     add(FM,bvector)
(7)  end
(8)  iForest = slideWindow(blockChain);
(9)  state = detection(FM, iForest);
(10) if state == False then
(11)    Response();
(12) end
(13) updateBlockchain();
(14) end
```

Algorithm 1: B-iForest.

The sink node uses the algorithm *sli de Win do w* to filter out suitable blocks from the blockChain to form an isolation forest *iForest* (line 8). Then, the sink node detects the feature matrix *FM* according the *iForest* and returns the detection result state (line 9). When the system detects an abnormality, that is, *state* = *False*, use *Response* to respond (line 10–11). Finally, the blockchain will be updated according to the update algorithm *up da teBlcokchain*, which is detailed in Algorithm 2 as follows:

In Algorithm 2, first initialize the number of samples of the detection algorithm, that is, the number of isolated trees φ (line 2). The algorithm *MCMC* filters out the latest and quotable block list *tips* in the chain (line 4). According to the selection algorithm *Select*, the block to be quoted is selected from the list of references (line 6) and is verified (line 7), and the index of the verified block in the *List* (line 8) is recorded until the number of *List*'s elements reaches two (line 5). The isolated tree *iTree* to be submitted into blocks is packed and the two previous blocks that have been verified (line 11) is quoted. Finally, the packaged block to the entire network is broadcasted and the blockchain is updated (line 12).

4. Experiment

4.1. Data Processing. We implement related experiments of the proposed scheme through *Python*3.8. The isolation forest algorithm comes from the machine learning package Scikit-learn by *Python*. All experiments are executed on an x64 Windows10 personal computer using an Intel(R) Core (TM) i5-8500 CPU 3.00 GHz processor. We choose kddcup.data_10percent.gz in the popular KDD CUP′99 dataset in IDS research, that is, 10% of the dataset sampling. We select four typical attack samples with different attributes, $Buffer_overflow$, *po d*, $guess_passws$, and *nmap*, remove the data labels, and mix them with normal samples to generate raw data for simulation. In the simulation process, samples are continuously sampled from the raw data to BCEAD for detection. The changes in the environment are simulated by controlling the abnormal proportions in the samples.

In this section, we conduct evaluation experiments through the following indicators:

Table 1 is the confusion matrix, where TP indicates that the real sample is positive and the prediction is also positive. FP indicates that the real sample is negative, but the prediction is positive. FN indicates that the real sample is positive, but the prediction is negative. TN means that the real sample is negative and the prediction is also negative. Some evaluation indicators, such as accuracy rate, precision rate, recall rate, and F1 value, can be obtained by the confusion matrix.

4.1.1. Accuracy. The ratio of the number of correctly classified samples to the total number of samples.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}. \tag{5}$$

4.1.2. Precision. The probability that a sample predicted to be positive is indeed a positive sample.

$$Precision = \frac{TP}{TP + FP}. \tag{6}$$

4.1.3. Recall Rate. The probability of being correctly predicted as a positive sample among all positive samples is

$$Recall = \frac{TP}{TP + FN}. \tag{7}$$

4.1.4. F1 Value. An indicator of comprehensive precision rate and recall rate is

$$F1\_score = \frac{2 \times Recall \times Precision}{Recall + Precision}. \tag{8}$$

input: the iTree given by last detection
output: the newest detection model given by the blockchain
(1) begin:
(2) $\varphi \leftarrow$ this;
(3) //estimatorSize, the size of slideWindow
(4) tips $\leftarrow$ MCMC();
(5) while length(List)< 2 do
(6) preBlock = Select(tips)
(7) if $Verify(preBlock) == True$ then
(8) add (List, preBlock)
(9) end
(10) end
(11) curBlock = publish (List, iTree)
(12) blockchain = Brodcast (curBlock)
(13) end

Algorithm 2: Update Blockchain.

Table 1: Confusion matrix.

Label	Positive	Negative
Positive	True positive (TP)	False positive (FP)
Negative	False negative (FN)	True negative (TN)

4.2. Detection Performance. Jia et al. [25] proposed the connection between model generalization error and individual learners in ensemble learning. The upper limit of the generalization error is

$$PE \leq \frac{\overline{p}\left(1 - S^2\right)}{S^2}. \tag{9}$$

Here, $\overline{p}$ is the average value of all relevancies between every two classifiers, and S denotes the mean intensity of the individual classifier. Equation (9) shows that the generalization performance of the global model is better when the classification of individual classifiers is stronger, and the correlation between the classifiers is smaller. In EAD, isolated trees are generated independently by sink nodes during detection, so the low correlation between isolated trees ensures the detection performance of the global isolation forest.

As shown in Figure 4, some sampling points of the Nmap attack are drawn after being processed by PCA, and BCEAD can effectively distinguish between normal points and abnormal points. The ACU can reach 96.67%, and the F1 value is about 0.6 in the detection. It shows that BCEAD can effectively detect anomalies and distinguish between normal traffic and attack events in the network.

Figure 5 shows the accuracy of BCEAD's detection of various typical attacks. The detection performance of the detection model is maintained at a high level of 94% to 96%, and it has a good detection accuracy rate for all kinds of attacks. Figure 6 shows the false positive rate (FPR) of BCEAD detection of various typical attacks. The detection effect is worse for *Guess_passw d* and *Nmap* attacks, which is consistent with the response of the curve in Figure 5. The

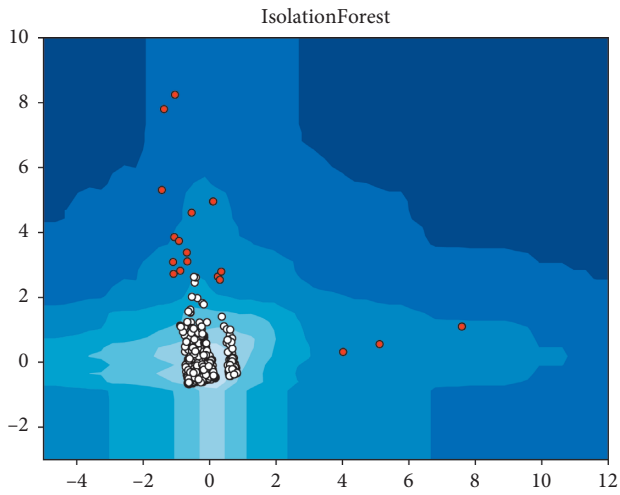

Figure 4: BCEAD detects two-dimensional data after PCA.

difference in the detection performance for different attacks is mainly due to the different launch frequency of various attacks. It confirms that the quality of the training set directly affects the performance of the detection model, and BCEAD, which updates the training data in real-time, can effectively guarantee its detection performance.

Figure 7 shows the $F1$ value changes of the scheme proposed by Liu et al. and BCEAD during a period. In the figure, the abscissa indicates the time by the ratio change of abnormal and normal data in the real-time detection. The blockchain-based anomaly detection proposed by Liu et al. does not have the update of the detection model, so the detection performance of the data is always constant in a

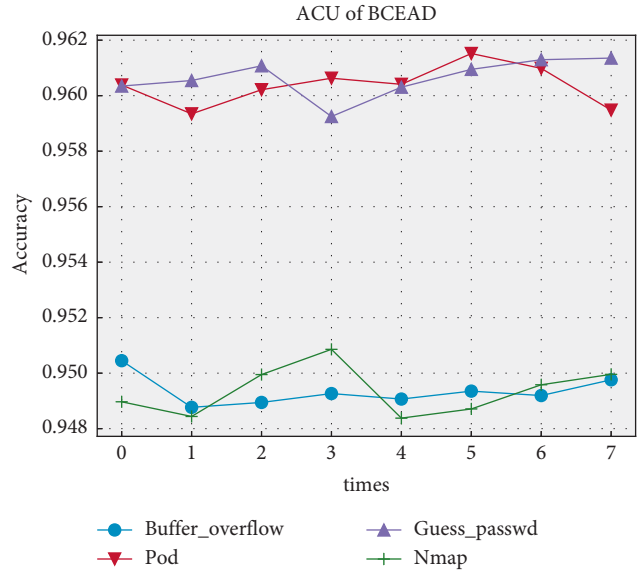

Figure 5: BCEAD's detection accuracy of the data after PCA.

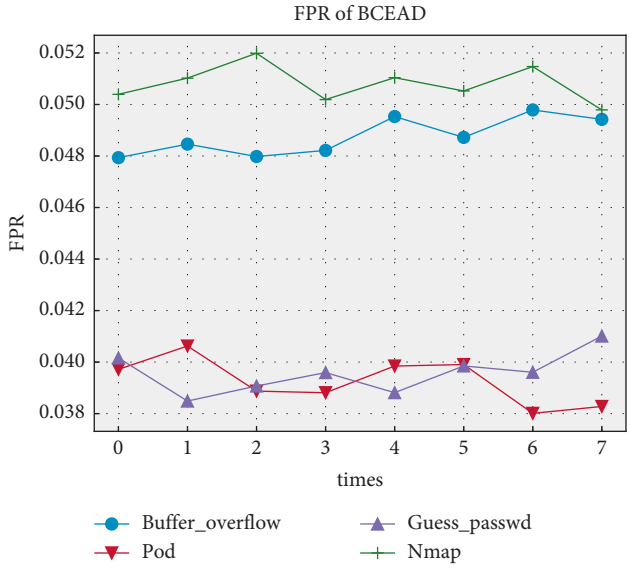

Figure 6: BCEAD's FPR of the data after PCA.

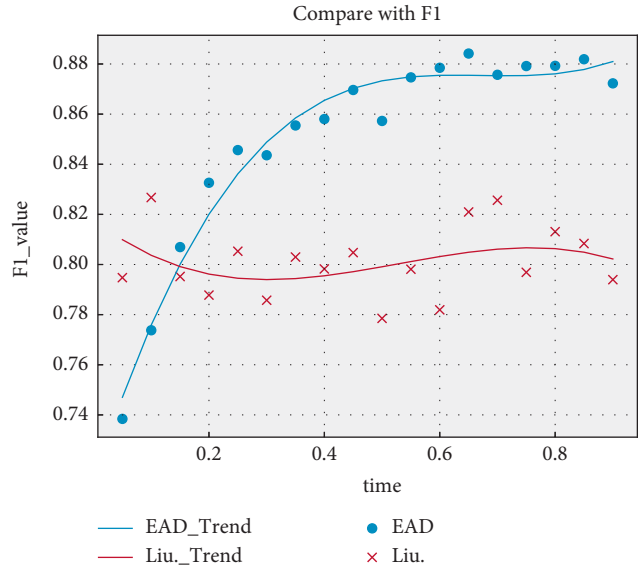

Figure 7: Performance comparison between BCEAD and similar scheme.

fixed range. However, BCEAD will dynamically select the training set to update the detection model in real-time. Therefore, the BECAD will optimize performance in long-term detection according to environmental changes.

Figure 8 shows the cost comparison between the scheme proposed by Liu et al. and BCEAD during the detection. The experiment specifically compared the communication overhead and storage overhead. As can be seen from the figure, because Liu et al.'s scheme stores detection data by the blockchain, the overhead will continue to grow during the long-term detection process. However, BCEAD uses the blockchain to store the detection model, so the cost in the early detection stage will be slightly higher, but the increase in cost is low. After a period of detection, the cost of BCEAD is significantly lower than similar blockchain-based detection schemes.

4.3. Security. Yang et al. proposed the security entropy to evaluate the benefits of system offense and defense [26], which can quantify the security of the system. We can suppose that the insecure factors caused by the attacker are $q_1, q_2, \ldots, q_n$, and the insecure entropy of the detection system at time t is $Q(t, q_1, q_2, \ldots, q_n)$, or simply $Q(t)$. If $Q(t)$ grows with time, that is, differential $dQ(t)/dt > 0$, then the system will become more and more unsafe. If $Q(t)$ decreases with time (differential $dQ(t)/dt < 0$), then the system will become more and more secure.

In BCEAD, the factors that affect the security of the system are generally classified into two: (1) the attacker publishes bad blocks, depletes the global model, and increases insecure entropy. (2) Ordinary nodes publish normal

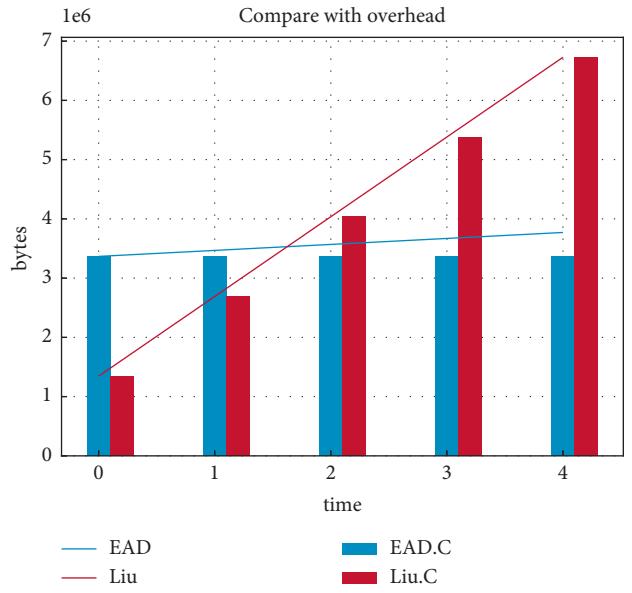

Figure 8: Overhead comparison between BCEAD and similar scheme.

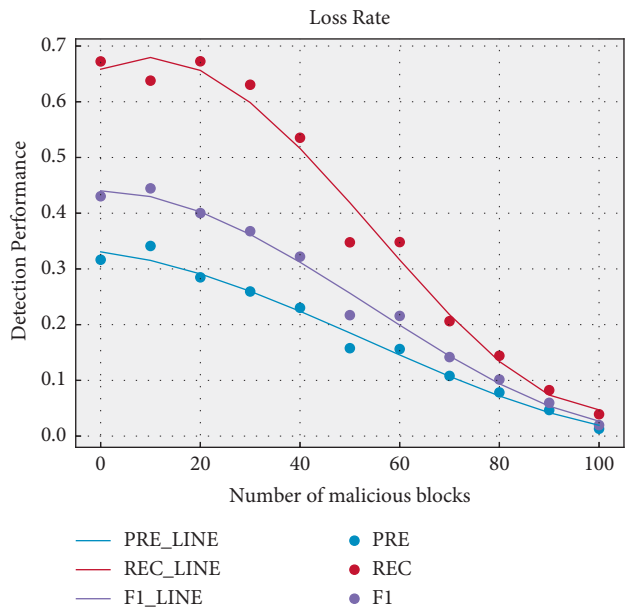

Figure 9: Changes in detection performance with the release of malicious blocks.

blocks, gaining the global model, and not increasing the insecure entropy. Then, the changes in the security entropy of the two factors are as follows:

$$\frac{dQ_1}{dt} = a_1 Q_1 \text{and} \frac{dQ_2}{dt} = a_2 Q_2. \tag{10}$$

Solving this system of equations,

$$Q_1 = c_1 e^{a_1 t} \text{and} Q_2 = c_2 e^{a_2 t}. \tag{11}$$

The time t is

$$t = \frac{\ln Q_1 - \ln c_1}{a_1} = \frac{\ln Q_2 - \ln c_2}{a_2}. \tag{12}$$

Set $a = a_1/a_2$, set $b = c_1/c_2^a$; then,

$$Q_1 = b(Q_2)^a. \tag{13}$$

In addition, it can be obtained that

$$\left\{\frac{dQ_1}{dt}\frac{1}{Q_1}\right\} : \left\{\frac{dQ_2}{dt}\frac{1}{Q_2}\right\} = a. \tag{14}$$

It can be seen from Equations (13) and (14) that the attacker's loss rate and the system's loss rate of the detection model is a power function with each other, and the ratio of the two to the model's loss change rate is a constant. In BCEAD, WSN's device authentication has prevented external attacks, and internal attackers can only undermine system security through the loss detection model. Because the blockchain guarantees the format of the network communication content, the attacker's loss model can only be done by publishing unhelpful isolated tree blocks. Therefore, restricting this single evil means can ensure the safety of the detection system.

The experimental results of malicious attacks on the loss of the system are shown in Figure 9. When malicious blocks continue to be released, the performance indicators of the detection system continue to decrease. It can be seen from the figure that after the release of 30 malicious blocks, the detection performance has dropped significantly. Since the BCEAD verification block is randomly selected, the probability of continuous bad blocks is low, and the submission of normal blocks has a gain effect on the detection model, so the solution itself has a certain attack resistance. From the above analysis, it can be seen that the ratio of the attacker's loss to the model and the program's gain to the model is constant, so this confrontation can find a balanced threshold. By disabling the node's continuous release of blocks, the security resistance of the detection system can be sufficient to resist malicious internal attacks.

5. Conclusion

This paper studies the security of wireless sensor networks, applies distributed anomaly detection to WSNs by the blockchain technology, and proposes the BCEAD scheme. The scheme divides the WSN into multiple layers. The sink layer performs anomaly detection for each network domain in the sensor layer. The detection model (isolation forest) is stored in the blockchain (tangle). Besides, this paper compares and analyzes the detection performance of BCEAD through experiments and proves its superiority. However, the actual deployment of the blockchain may affect the performance of the detection system, and the communication and storage overhead of the blockchain technology is also difficult to be balanced. Therefore, we will do some practice for the blockchain in the future. Nevertheless, the experiment proved that the scheme is compatible with some existing security mechanisms [27] for detection, which is enough to guarantee the application potential of the scheme.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Natural Science Foundation under Grant no. 61962009, Major Scientific and Technological Special Project of Guizhou Province under Grant no. 20183001, Science and Technology Support Plan of Guizhou Province ([2020]2Y011), and Foundation of Guangxi Key Laboratory of Cryptography and Information Security (GCIS202118).

References

[1] L. Qiang, H. Xiaohong, L. Supeng, L. Longjiang, and M. Yuming, "Deployment strategy of wireless sensor networks for internet of things," *China Commun*, vol. 8, pp. 111–120, 2011.

[2] N. Marriwala and P. Rathee, "An approach to increase the wireless sensor network lifetime, Proc. 2012 World Congr," *Inf. Commun. Technol. WICT*, vol. 2012, pp. 495–499, 2012.

[3] C. Mahmoud and S. Aouag, "Security for internet of things: a state of the art on existing protocols and open research issues," *ACM Int. Conf. Proceeding Ser.*vol. 17, pp. 1294–1312, 2019.

[4] Y. Chen, J. Sun, Y. Yang, T. Li, X. Niu, and H. Zhou, "PSSPR: a source location privacy protection scheme based on sector phantom routing in WSNs," 2021, https://arxiv.org/abs/2109.13774.

[5] A. M. FSabahi, "Intrusion detection: a survey," in *Proceedings of the 3rd Int. Conf. Syst. Networks Commun. ICSNC 2008 - Incl. I-CENTRIC 2008 Int. Conf. Adv. Human-Oriented Pers. Mech. Technol. Serv.*, pp. 23–26, Sliema, Malta, October 2008.

[6] J. J. Davis and A. J. Clark, "Data preprocessing for anomaly based network intrusion detection: a review," *Computers & Security*, vol. 30, no. 6-7, pp. 353–375, 2011.

[7] M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, "Network anomaly detection: methods, systems and tools," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, pp. 303–336, 2014.

[8] F. Huang, G. Xie, and R. Xiao, "Research on ensemble learning," in *Proceedings of the 2009 Int. Conf. Artif. Intell. Comput. Intell. AICI*, pp. 249–252, Shanghai, China, November 2009.

[9] K. Christidis and M. Devetsikiotis, "Blockchains and smart contracts for the internet of things," *IEEE Access*, vol. 4, pp. 2292–2303, 2016.

[10] B. Hu, C. Zhou, Y.-C. Tian, Y. Qin, and X. Junping, "A collaborative intrusion detection approach using blockchain for multimicrogrid systems," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 8, pp. 1720–1730, 2019.

[11] H.-J. Liao, C.-H. Richard Lin, Y.-C. Lin, and K.-Y. Tung, "Intrusion detection system: a comprehensive review," *Journal of Network and Computer Applications*, vol. 36, no. 1, pp. 16–24, 2013.

[12] F. T. Liu, K. M. Ting, and Z. H. Zhou, "Isolation forest," in *Proceedings of the IEEE Int. Conf. Data Mining, ICDM.*, pp. 413–422, Pisa, December 2008.

[13] F. T. Liu and K. M. Ting, "Isolation-based anomaly detection," *ACM Transactions on Knowledge Discovery from Data*, vol. 6, no. 1, pp. 1–44, 2018.

[14] Y. Yuan and F. Y. Wang, "Blockchain: the state of the art and future trends," *Zidonghua Xuebao/Acta Autom. Sin.*vol. 42, pp. 481–494, 2016.

[15] S. Nakamoto, "Bitcoin: a peer-to-peer electronic cash system," *SSRN Electronic Journal*, 2008.

[16] T. Li, Y. Chen, Y. Wang et al., "Rational protocols and attacks in blockchain system," *Security and Communication Networks*, vol. 2020, pp. 1–11, 2020.

[17] W. Li, S. Tug, W. Meng, and Y. Wang, "Designing collaborative blockchained signature-based intrusion detection in IoT environments," *Future Generation Computer Systems*, vol. 96, pp. 481–489, 2019.

[18] W. Li, Y. Wang, J. Li, and M. H. Au, "Toward a blockchain-based framework for challenge-based collaborative intrusion detection," *International Journal of Information Security*, vol. 20, no. 2, pp. 127–139, 2021.

[19] X. Xie, Y. Fang, Z. Jian, Y. Lu, T. Li, and G. Wang, "Blockchain-driven anomaly detection framework on edge intelligence," *CCF Transactions on Networking*, vol. 3, no. 3-4, pp. 171–192, 2020.

[20] X. Liu, F. Jiang, and R. Zhang, "A new social user anomaly behavior detection system based on blockchain and smart contract," *IEEE Int. Conf. Networking, Sens. Control. ICNSC*, vol. 2020, 2020.

[21] I. Makhdoom, M. Abolhasan, H. Abbas, and W. Ni, "Blockchain's adoption in IoT: the challenges, and a way forward," *Journal of Network and Computer Applications*, vol. 125, pp. 251–279, 2019.

[22] T. Li, Z. Wang, G. Yang, Y. Cui, Y. Chen, and X. Yu, "Semi-selfish mining based on hidden Markov decision process," *International Journal of Intelligent Systems*, vol. 36, no. 7, pp. 3596–3612, 2021.

[23] T. Li, Z. Wang, Y. Chen, C. Li, Y. Jia, and Y. Yang, "Is semi-selfish mining available without being detected?" *International Journal of Intelligent Systems*, 2021.

[24] W. F. Silvano, R. Marcelino, and I. Tangle, "Iota Tangle: a cryptocurrency to communicate Internet-of-Things data," *Future Generation Computer Systems*, vol. 112, pp. 307–319, 2020.

[25] B. Jia and Y. Liang, "Anti-D chain: a lightweight DDoS attack detection scheme based on heterogeneous ensemble learning in blockchain," *China Communications*, vol. 17, no. 9, pp. 11–24, 2020.

[26] Y. X. Yang and X. X. Niu, *The General Theory of Information Security*, Publishing House of Electronics Industry, Beijing, China, 2018.

[27] W. Meng, W. Li, and L.-F. Kwok, "EFM: enhancing the performance of signature-based network intrusion detection systems using enhanced filter mechanism," *Computers & Security*, vol. 43, pp. 189–204, 2014.