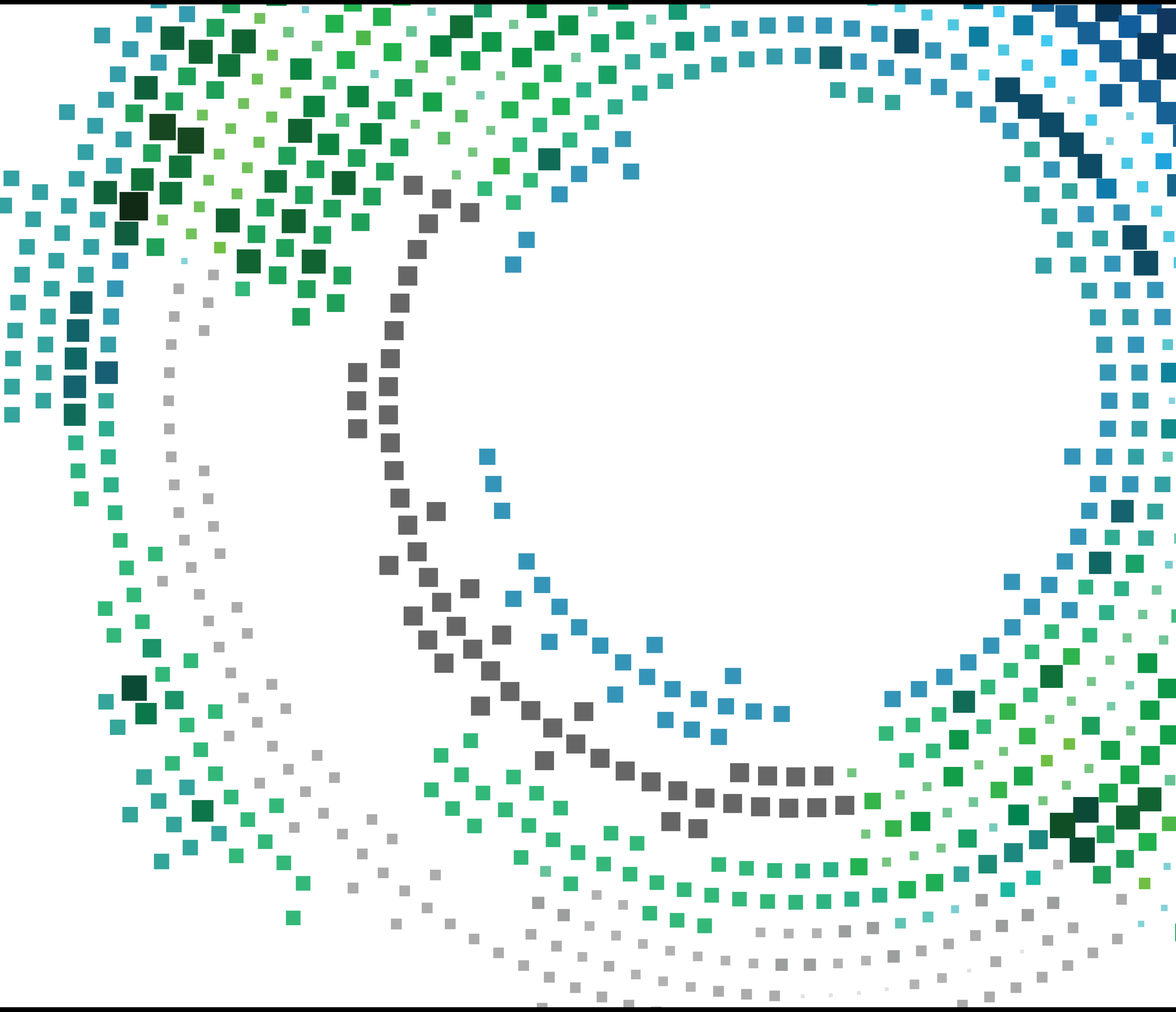
Mobile Information Systems

Distributed Secure Computing for Smart Mobile IoT Networks 2021

Lead Guest Editor: Vishal Sharma
Guest Editors: Daniel G. Reina and Zengpeng Li



Distributed Secure Computing for Smart Mobile IoT Networks 2021

Mobile Information Systems

Distributed Secure Computing for Smart Mobile IoT Networks 2021

Lead Guest Editor: Vishal Sharma
Guest Editors: Daniel G. Reina and Zengpeng Li

Chief Editor

Academic Editors

Contents

Hindawi
Mobile Information Systems
Volume 2021, Article ID 8091363, 13 pages
https://doi.org/10.1155/2021/8091363

Research Article

Feature Entropy Estimation (FEE) for Malicious IoT Traffic and Detection Using Machine Learning

Tarun Dhar Diwan,[1] Siddartha Choubey,[2] H. S. Hota,[3] S. B Goyal,[4] Sajjad Shaukat Jamal,[5] Piyush Kumar Shukla,[6] and Basant Tiwari[7]

[1]Chhattisgarh Swami Vivekananda Technical University, Bhilai, Chhattisgarh, India
[2]Shri Shankaracharya Technical Campus, Bhilai, Chhattisgarh, India
[3]Atal Bihari Vajpayee University, Bilaspur, Chhattisgarh, India
[4]City University, Petaling Jaya, Malaysia
[5]Department of Mathematics, College of Science, King Khalid University, Abha, Saudi Arabia
[6]Computer Science & Engineering Department, University Institute of Technology, Rajiv Gandhi Proudyogiki Vishwavidyalaya, (Technological University of Madhya Pradesh), Bhopal 462023, India
[7]Hawassa University, Awasa, Ethiopia

Correspondence should be addressed to Basant Tiwari; basanttiw@hu.edu.et

Received 14 September 2021; Revised 15 November 2021; Accepted 30 November 2021; Published 28 December 2021

Academic Editor: Vishal Sharma

Identification of anomaly and malicious traffic in the Internet of things (IoT) network is essential for IoT security. Tracking and blocking unwanted traffic flows in the IoT network is required to design a framework for the identification of attacks more accurately, quickly, and with less complexity. Many machine learning (ML) algorithms proved their efficiency to detect intrusion in IoT networks. But this ML algorithm suffers many misclassification problems due to inappropriate and irrelevant feature size. In this paper, an in-depth study is presented to address such issues. We have presented lightweight low-cost feature selection IoT intrusion detection techniques with low complexity and high accuracy due to their low computational time. A novel feature selection technique was proposed with the integration of rank-based chi-square, Pearson correlation, and score correlation to extract relevant features out of all available features from the dataset. Then, feature entropy estimation was applied to validate the relationship among all extracted features to identify malicious traffic in IoT networks. Finally, an extreme gradient ensemble boosting approach was used to classify the features in relevant attack types. The simulation is performed on three datasets, i.e., NSL-KDD, USNW-NB15, and CCIDS2017, and results are presented on different test sets. It was observed that on the NSL-KDD dataset, accuracy was approx. 97.48%. Similarly, the accuracy of USNW-NB15 and CCIDS2017 was approx. 99.96% and 99.93%, respectively. Along with that, state-of-the-art comparison is also presented with existing techniques.

1. Introduction

The IoT is the new era of technology in the digital world. IoT is empowering physical objects in processing data seamlessly [1]. It makes the physical objects interactive and makes them responsive without any human intervention. According to a recent Gartner report, there will be around 8.4 billion connected physical things worldwide in 2020 and it is expected that this number will increase to 20.4 billion by the year 2022 [2]. These applications are highly promising and serve the best of them. This number boosts the scholars to work on IoT in terms of its potential, performance, efficiency, challenges, threats, and security as well. Therefore, it is optative to have high security, privacy, authentication, and recovery from attacks. Spoofing, eavesdropping, DoS, and DDoS are some attacks on IoT applications [1], and to safeguard these applications, we need methods that can prevent these types of attacks [3]. Fog is also a new emerging technology that lets the user to virtual store and process the data between the cloud and devices, and fog can play a vital role in IoT security. Fog nodes have the potential to produce an alarm or warning to IoT systems if they encounter any suspicious data or requests [1]. Some researchers applied edge computing which is also one of the fastest-growing technologies that can be embedded with other technologies to improve its security, potential, performance, mobility, and data management. So, it can also be applied to IoT applications as well. Edge computing techniques provide a shorter response time [4]. This helps in better latency and system performance especially with the data generated by IoT applications. It can serve in the prevention of eavesdropping and data breaches in IoT applications. Moreover, IoT works on three-layer architecture and it has perception, network, and application as three layers. To achieve maximum security and privacy in IoT systems, it is essential to have security at various layers. Many architectures have been proposed for the security of IoT on various layers using machine learning (ML) or deep learning (DL). Scholars have studied various issues, challenges, and threats in IoT. Also, the existing security systems are not enough to handle all the aspects of security. So, more advanced and enhanced security systems are required; otherwise, IoT may lose its high potential and high demand. This advanced and improved system can be deployed with the help of the latest technologies that can be replaced with classical algorithms in IoT security. Machine learning (ML), deep learning (DL), and artificial intelligence (AI) serve methods that will improve the performance and efficiency of algorithms. Efficient intrusion detection systems can help reduce malicious IoT traffic. The incoming IoT packet streams are monitored continuously by intrusion detection systems [5]. There are two main threat detection approaches: signature-based and anomaly-based. A pattern is designed from previously recognized attacks by a signature-based approach. Therefore, an IDS based on signatures equates the signatures with the events seen and reports a hazard when matched. There are several problems relating to IDSs based on signatures, and the following are summarized:

(i) The very first problem is that the only known attacks with well-investigated features can be found, while zero-day (i.e., unknown) attacks cannot be detected. Unfortunately, attackers continue to develop their strategies to bypass conventional security mechanisms in various attack behaviors [6].

(ii) The second problem is that even though the numbers of newly discovered attacks grow, so does the number of signatures, leading to further similarities between stored patterns and new occurrences. This raises the detection systems more complex, which has a direct impact on the system's response time, making this a critical problem for real-time intrusion detection systems. Therefore, under certain conditions, these IDSs' system performance degrades due to limited source availability [7].

Anomaly-based detection techniques can address all of the above limitations. A system based on anomalies observes a sequence of incoming packets and builds the normal behavior model of the system. The learned model then identifies abnormalities relying on an index of similarities between normal and abnormal packets. In this approach, the major challenge is that to build a model with unique normal system behavior, a reference with different underlying behaviors, generated by individual data sources. Admittedly, various types of data references can produce an increased false-positive rate by reducing the resemblance between malicious and normal learned behaviors [8]. After reading several research works in this field, the authors mainly raised concerns about the validity of signature-based IDSs in scenarios where attacks could not be found [9]. This paper gave this assumption a shape. Indeed, an IDS-based signature could not detect unknown attacks because its vocabulary of attacks could not contain those definitions. Worse still, an IDS is deployed over the distance, if not on end devices or low-cost IoT gateways. As required in a signature-based approach, the regular update of attack definitions is more difficult.

The major issue that arises while implementation of intrusion detection model is that it has to handle large amount of data. The large, irrelevant, and redundant data may cause negative impact on performance of machine learning. Therefore, building machine learning algorithms, feature selection plays a major issue. The accuracy and time complexity of the model is affected due to the presence of irrelevant features. In this paper, intrusion detection model is presented in the presence of feature selection methods. In this paper, a wrapper feature selection algorithm for IDS is proposed to handle large number and high-dimensional dataset. Redundant or irrelevant features are filtered that significantly improve the training time and accuracy of the machine learning. Filter, wrapper, and embedded methods are three types of feature selection. In filter method, each feature is assigned with a weight and feature subsets are used with machine learning for classification. Filter method has benefits such that it requires fewer computing resources and time but the main issue with this type of feature selection is that it lacks compatibility with classification process and thus results in low accuracy. Another feature selection method is wrapper method that considered the classification performance while selecting feature subset. This method results in high accuracy but takes more computational time. The embedded feature selection is another method that shows performance in between filter method and wrapper method. In IDS, data need high accuracy as training time was not of much concern. Therefore, in this paper, wrapper feature selection method is adopted. Pearson correlation, f_score correlation, and rank-based chi-square feature selection techniques are combined together to design a hybrid wrapper-type feature selection method that selects optimal or relevant features out of number of feature sets. Correlation feature selection method finds the association among features. But one of the drawbacks is that if it takes entire population (i.e., entire data), it does not result in good performance. Therefore, in this paper, hybridization of feature selection techniques is performed on different sets of data. This results in accurate association among features and gives high accuracy.

The key contributions of this paper are as follows:

(i) In this paper, state-of-the-art about intrusion detection frameworks for detecting malicious IoT traffic and their challenges are also presented along with

(ii) This paper also presented a model based on feature selection techniques intending to design a lightweight algorithm for IoT traffic

(iii) The proposed framework is effective in the IoT scenario as the methodology can handle big data along with the best features

(iv) Finally, this paper gives a comparative state of the art with existing techniques

The remaining section of this paper is illustrated to be as follows: Section 2 introduced the background knowledge of intrusion detection techniques for IoT security, and their challenges are also discussed. In Section 3, paper gives an overview of the proposed methodology and training algorithms. In Section 4, result analyses of the proposed model on the different datasets are presented. This section also gives the comparative state of the art with existing techniques. Finally, in Section 5, the conclusion and future research scope are discussed.

2. Related Works

The involvement of IoT devices, in our daily lives, is increasing, and the critical issue of security of the collected data by these devices is also rapidly increasing. Thus, in [10], a three-layer intrusion detection system is introduced. The perspective of the system is to determine the domain of the IoT network based on cyberattacks [11]. Overall IDS architecture consists of three layers; first layer consists of a tool that scans the network and recognizes the linked IoT devices based on the Mac addresses and categorizes them based on their behavior. The second layer identifies the genuine or spiteful packets from the connected IoT devices. If any spiteful or malicious packet is found in the second layer, then the third layer will determine which kind of attack it is. Some commercial IoT devices are connected in the home, and to collect the created traffic from these devices, tcpdump was programmed to run on an access point. Finally, in the Syslog server, this gathered to traffic in the form of PCAP logs is transferred and then stored. As the network process of collecting data from the testbed is started, a time span for both useful and harmful data was decided. The testbed of IoT devices was well arranged and implemented so that overall inbound and outbound traffic which was processing on access point was recorded by using the tool tcpdump. To increase the complexity of the network, four automated multilevel spiteful attacks, some scenarios were established on the network. When an individual attack takes place, scripts were developed to generate logs. This is important for the data labeling task which will be further done to supervise the machine learning. The next process is the feature selection in which the development of machine learning is based on an intrusion detection system and IoT. The limitation of this system is that this is not a real-time implemented system.

In [12], the intrusion detection system was designed by using the fog computing method to implement it in the network which is spread. The introduced system consists of two sections; i.e., the first one is attack observation at fog nodes which uses the OS-ELM algorithm which detects that the packet which is coming through IoT traffic is genuine or just to create attacks. The second section of the proposed system is summating at cloud service which provides the global view to examine and observe the ongoing security condition of IoT applications. This is used to forecast the upcoming action of the attacker. The results of this experiment are estimated based on accuracy and response time, the given system achieves 97.36% of accuracy, and response time is evaluated as this system determines attacks 25% faster as compared to other algorithms. This system is to be protected from proactive attacks, which is the limitation of this system.

In [13], the author introduced a confidential preserving distributed intrusion detection system structure based on progressive learning. This model is used to recognize denial of service attack because many researchers remain unsuccessful in determining real-time traffic dataset and thus many attackers insert spiteful traffic patterns to corrupt the training structure. The proposed structure consists of three networks; they are generative network which gathers all the incoming traffic from all IoT devices, and then, unique feature from all the devices was extracted with the help of autoencoders. The second network is the bridge network, and all the collected data of useful features which is extracted from the first network (generative network) are sent to bridge network. The gathered data are analyzed and then compared with the available data in the third network which is classifier network. In this network, only important data are sent by the bridge network so that the model will not behave as time-consuming. To save time in execution, CNN (convolutional neural network) model is used to minimize false alarms and inessential service visits. Simply, it can be defined as the overall process can be divided into three phases; the first one is the preprocessing phase; the second phase is the comparison phase; the third phase is the classification phase based on separate coding extraction of feature and fusion, incremental maintaining module, and finally classification process. This structure has given classification accuracy with minimum space and low computational cost, but the limitation of this structure is that it is not able to identify new attacks in multiple attack scenarios.

In [14], the author proposed a model of intrusion detection honeypot based on SoLA to identify malware attacks.

Honeypot is nothing but an unreal environment that collects the data of attacks and attackers to only trap them but not to prohibit them. The introduced system of intrusion detection honeypot's architecture consists of a low interaction honeypot server and IDS network. Both of them collect all the information from the incoming traffic and investigate the collected data. Complex event processing (CEP) engine connects different events of direct attacks from host and network, honeypot agent, SDN controller, etc. Depending on the CEP outcome, the spiteful process is determined and destroyed. Honeypot agent is structured by applying social leopard algorithm (SoLA). The introduced system of intrusion detection honeypot uses the complex event processing technique to interrelate between features of the host, network, and several events. The ransomware encryption process here takes place as read, encrypt, write, and delete. When there is progression or movement from one state to another one in the fake folder, it investigates whether the activity is done by the user or it is the activity of the attacker. Once the file is read and encrypted, it denotes it as doubtful activity and examines the particular variable and then the outcome of this process is sent to CEP engine and firewall; here, engine interrelates the values from the honey folder, audit watch, and SDN application and creates an alert which is based on high accuracy, and thus, it determines the malware attacks with minimized loss and high accuracy. The software-defined networking (SDN) applications upgrade network security by applying simple commands. It does not work on healthcare implants, so further this model can be amplified for Internet-connected toys to identify the malware attacks.

In [15], to identify the harmful data that are inserted within the IoT network, a light-weighted intrusion detection system is used. The attack identification is done by using machine learning which relies on a support vector machine (SVM). The architecture of the introduced system comprises two stages: the first one is the training stage and the second one is the evaluation stage. The overall system is conducted by varying the traffic intensity. Training databases carrying labeled samples are acquired in the training stage, these databases are then used to acquire their features, and a pool of features is generated which can be called a feature pool. This pool accompanied by a vector label is sent to train the classifier, and this trained classifier then categorizes the samples as labeled samples and unlabeled samples. To calculate the performance of the classifier, alike features used in the training stage are extricated from data samples. Several experiments were done in this work with different traffic intensities, and it is proved that packet arrival rate feature and support vector machine-based classifier are sufficient to detect the intrusion in the network as compared to other classifiers like NN, k-NN, and DT. The outcomes of using this model are the intrusion detection system which relies on a support vector machine (SVM) to achieve adequate detection of attacks.

In [16], the author introduced the intrusion detection system which relies on brilliant deviation, and the researcher named it Passban; this system is used to secure the Internet of things that are linked to this system. Passban is comprised of packet flow discovery, feature extraction, training and loading of the model, action manager, and web management interface. In this work, author focuses on the performance of network interchange having the objective of identifying the patterns which slightly vary from them, and names them as anomalies. These different patterns are the pattern of attacks which takes place in the network. One-class classification, a type of learning strategy, is accessible for the present unambiguous condition. There are many algorithms obtained that rely on basically two techniques; the first one is profiling, and the second one is isolation. Feature extraction is another step on which machine learning is then applied; several features are extracted from the data. Trained data are saved in the local memory of the edge device and the prediction; phase-predicted anomalies are also detected. Finally, all the anomalies are then forwarded to the action manager. In this, Passban is worked on two scenarios and then results are declared. In the first scenario, LOF and i-forest are capable of detecting all the attacks with adequate accuracy, while in the second scenario arrangements are not requisite, which means Passban is connected individually to the network which is to be identified for the attack; it can scan overall traffic of the device linked with the network. This technique is useful in threat determining with the accurate performance, and it can be applied on inexpensive devices also.

In [5], the author discussed multiple works that rely on IoT devices, security procedures, and machine learning procedures. The main aim of the research is to create a junction in between the above given three areas. The first junction is between IoT and security procedure, IoT architecture consists of three layers: perception layer, network layer, and application layer, and every layer has different attacks on it. Many machine learning techniques for intrusion detection were discovered. This survey bestows complete analysis of network intrusion detection for IoT security established in various aspects of learning techniques. Here, IoT attacks are categorized based on challenges they are spoofing, routing attacks like sinkhole attack, selective forwarding attack, black hole attack, wormhole attack, replay attack, tampering attack, repudiation attack, and man-in-the-middle attack; these are technical terms of attacks. Other kinds of attacks are based on design challenges, and some of them are interoperability and diverseness, security and privacy, etc., based on mechanism filter packets, adopt encryption, employ robust password authentication schemes, and audit and log activities. Various learning techniques are described by the different researchers; some of them are machine learning and deep learning, and based on these learning strategies, different algorithms are generated like decision tree, artificial neural network, Naïve Bayes, optimum path forest algorithm, logistic recursion, support vector machine, etc. This research work is helpful from an academic point of view and also industrial research.

In [17], author proposed a 2-stage AI IDS in SD-IoT network that has flow classification and feature extraction as its two stages. The architecture has self-learning ability. To get the best features been extracted, improved bat algorithm is used and for its optimal performance swarm division and binary differential mutation are applied with it. Then, an improved random forest is applied for network flow classification, and to improve the classification, weighing mechanism is also used. Experimental results are performed on a subset of the KDD Cup 1999 dataset after downsampling. The experimental results validate the better accuracy and lower overhead of an architecture which is better than the previous solution, and as far as it is concerned with future work, it can be extended to be applied in the real network for classification of traffic.

In [18], author proposed a novel SDRK-ML algorithm, that is, supervised DNN, and further extended it to an unsupervised clustering technique. The algorithm is placed between IoT and cloud layers to make it work better. Fog nodes are set as a gateway and perform data acquisition, and later, feature extraction is done which has been inputted to trained SDRK in which deep feedforward NN and K-means work on its core but in this paper, due to the unsuitability of K-means, a variation of K-means named as "RRS-K-means" is used to overcome the issue of NP-completeness. In a testbed, it is mentioned that a programmable feature of fog is installed that mitigates the attack for evaluation. The experimental results are performed on the benchmark NSL-KDD dataset. The limitation of this paper was that the fog nodes themselves can also be a point of attack to hack, so identification of same and reducing the retraining time are some suggested study works.

In [19], the author developed a CNN-based architecture that extracts the properties of the link load to detect roadside unit intrusion. This deep architecture consists of six covert layers, three of which are convolutional layers and three of which are pooling layers that implement average pooling with factor two to obtain abnormal fluctuations. As activation function, the sigmoid function is taken. The spatial characteristics of link charges are indicated as a matrix, and a loss function based on the standard L1 is presented to train the model's backpropagation algorithm. The first assessment parameter of sensitivity, calculated using different weights and biases, is precision for the evaluation of performance. Low-orbit ion cannon is installed for the experimental result, and the result is driven from four attacks, TCP, UDP, SYN, and HTTP. To implement DDoS, LOIC is installed.

The author of [20] proposed a CorrAUC approach for the effective selection of traffic through the use of algorithms to improve traffic detection in the IoT network. The method proposed works in four steps. In the first step, a function selection metric called CorrAUC extracts the characteristics. In the second step, a wrapper technique is used to develop and design an algorithm based on the same metric. In the third step, it combines the ROC curve correlation attribute (CAE) and ROC curve area (AUC) to select the effective function to detect the bot-IoT. In the last step, integrated TOPSIS and Shannon entropy will be used for the validation of selected features on a bijective soft set. The Pearson correlation coefficient is utilized between M and N attributes in the feature selection matrix, but a feature becomes effective if the relationship between feature and class is not strongly correlated so that the correlation is calculated for greater precision. A newly developed dataset called bot-IoT is used for experimental evaluation. The most significant works reviewed in this section are summarized in Table 1 with their limitations.

3. Methodology

The proposed framework for intrusion detection in IoT is illustrated in Figure 1. This framework gives a general overview of the framework that is composed of basically three layers: data layer, communication or network layer, and the application layer. The data layer is composed of smart sensor data or IoT nodes. The collected data, either it is from sensors or any data user, are collectively communicated to the next layer, i.e., communication or network layer. This layer is composed of gateway or switching devices that are responsible for analyzing the collected network data. The abnormal packets are analyzed using the proposed approach and report to an administrator, whereas the normal data packets are transmitted to the next layer for their storage and analysis purposes. We explained the technique proposed in this section step by step with details. Our proposed method includes three steps for effective selection in the IoT network: data collection and extraction of features, optimal selection of features, and classification (as shown in Figure 2). The IoT packets are captured and transmitted for preprocessing and feature extraction at the data collection stage. Second, the proposed feature selection method is used which selects functions that contain sufficient information and then selects the feature to accurately filter it and select effective features for the selected ML algorithm based on these optimal features. The proposed algorithm is an assessment of correlations with an entropy feature estimate to solve the problem by a specific machine learning (ML) algorithm [21] of effective intrusion detection selection. The entropy estimate, which provides more detailed information on whether or not selected features are similar, is a mathematical method used for homogeneity measurement. In terms of the effective function selection for IoT attack detections in the IoT network environment, this technology produces very effective results. In addition, our method selects features that carries sufficient information for identifying IoT attacks [22] on the IoT network. To understand clearly, the methods for effective feature selection in the IoT network are discussed below, taking into consideration the detection of attacks by IoT.

3.1. Data Collection and Feature Extraction.

The incoming traffic flow is captured, and further necessary features are extracted out of normalized incoming data packets [23]. These extracted features help out to find the type of attack and its identification. But before the feature selection process, it is required to preprocess the extracted data or features because data preprocessing plays a vital role in network traffic as the volume of data handled is huge. The algorithm for this stage is illustrated in Algorithm 1.

In preprocessing process, reduction of redundant data and normalization is an important step. This results in balanced data formation within a specified range. The

Table 1: Overview of reviewed IDS for IoT security.

Ref	Techniques	Attack types	Dataset	Drawbacks
[10]	Machine learning	DoS, man-in-the-middle attack, spoofing, reply attack, etc.	—	Does not support real-time detection
[12]	Online sequential extreme learning machine	DoS, R2L, probe, U2R	NSL-KDD	It cannot analyze all kinds of attacks evolving in the highly dynamic IoT environment
[13]	Autoencoders	DoS, R2L, probe, U2R	NSL-KDD, KDD99, real time	Not suitable for multiclass attack scenarios
[14]	Social leopard algorithm	Ransomware attacks	UNSW-NB15	Only applicable on ransomware attacks
[15]	Support vector machine	DoS attacks	CICIDS2017	Not suitable for changing traffic flow.
[16]	Machine learning	Port scanning, HTTP and SSH brute force, and SYN flood attacks	Real IoT testbed	Operable on limited data rate of incoming packets
[17]	Random forest	DoS, R2L, probe, U2R	KDD99	Does not support real-time detection
[18]	Deep feedforward neural network	DoS, R2L, probe, U2R	NSL-KDD	Does not support real-time detection
[19]	Convolutional neural network	Flooding DDoS attack	Real IoT testbed	Training error shows steep convergence curve
[20]	Machine learning	Botnets attacks	Bot-IoT dataset	Not suitable for multiclass attack scenarios

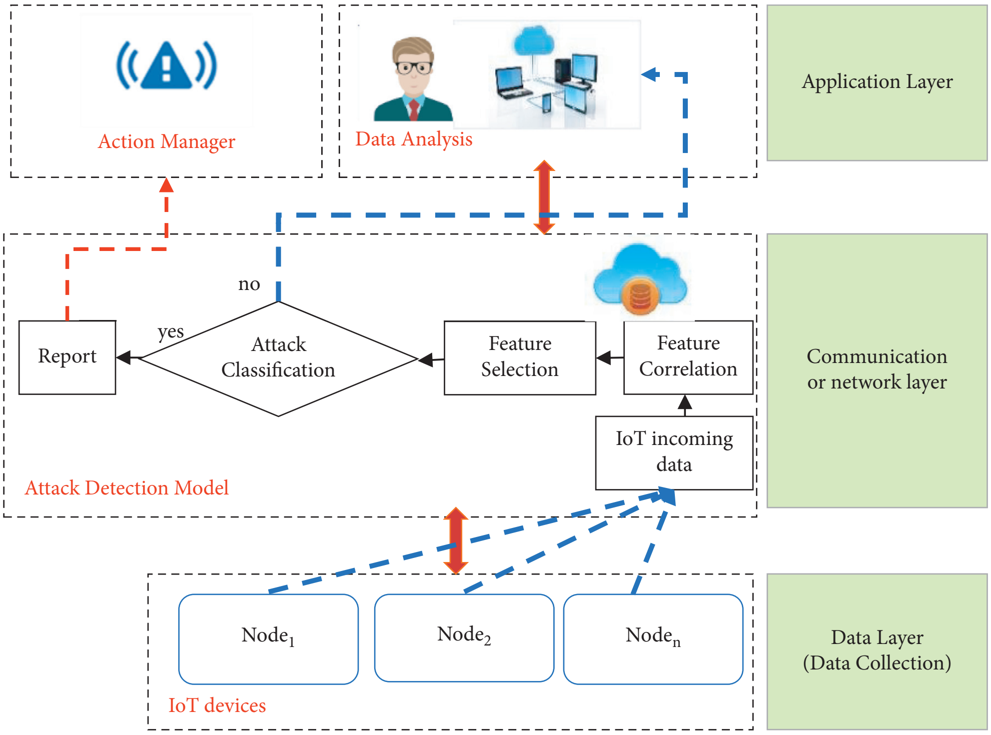


Figure 1: Framework for IoT security.

z-score technique is used to normalize the incoming data packets, as illustrated in the following equation:

$$z_i = \frac{x_i - \text{mean}(x_i)}{\text{std}(x_i)}, \quad (1)$$

where $x_i = i_{th}$ feature set, $\text{mean}(x_i) =$ mean of i_{th} feature set, and $\text{std}(x_i) =$ standard deviation of i_{th} feature set.

3.2. Feature Selection. For effective feature selection to solve the problem associated with malicious attack detection for IoT networks, three feature selection methodologies are hybridized with entropy estimation. This will result in the selection of the best and optimal feature selection. Correlation among features that selects effective features reduces volume of data as well as calculations also. Further feature entropy estimation is performed to overcome from class imbalance problem by eliminating irrelevant features.

A rank-based chi-square feature selection algorithm [24] is used to evaluate the dependency level of feature sets (f_i) on the class label (c_1). Mathematically, it is represented as in the following equation:

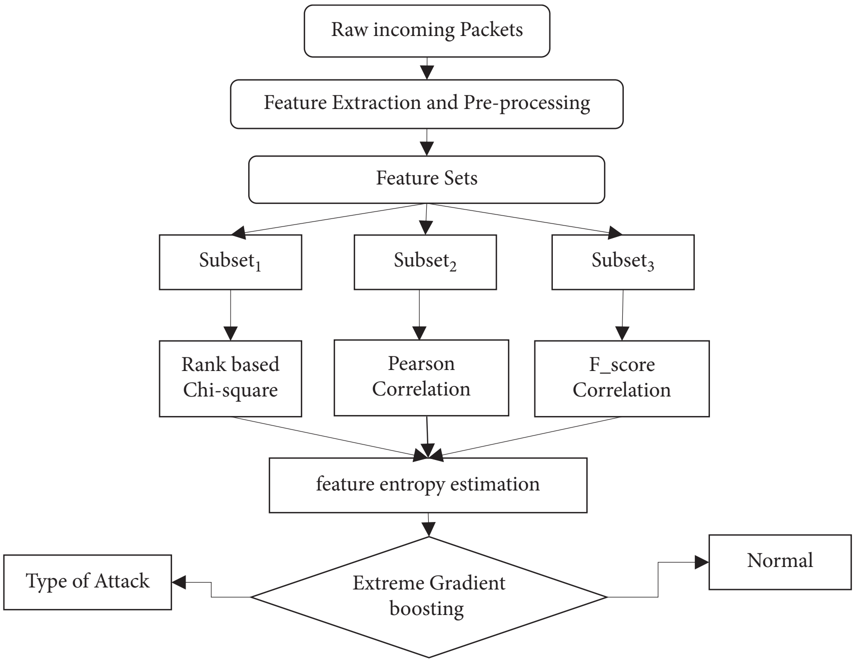


Figure 2: Proposed flowchart.

$$\text{Chi}^2_{(f_i, c_l)} = \frac{(A_i - C_i)^2}{C_i}, \tag{2}$$

where A_i = number of observations in a class c_l and C_i = number of expected observations in a class c_l.

In addition, the technique of Pearson moment correlation was adopted. This technique was used to study the relationship between independent and target class characteristics more thoroughly. P_{corr}, a range of values from +1 to −1, is available for the Pearson correlation coefficient. A value of 0 indicates that the two variables are unrelated. A value above 0 is a positive association, and a value below 0 is a negative one. This technique is used to identify relationships between different characteristics or attributes. Mathematically, it is represented as follows:

$$P_{\text{corr}} = \frac{\sum\left(A_i - \widehat{A}_i\right)\left(B_i - \widehat{B}_i\right)}{\sqrt{\sum\left(A_i - \widehat{A}_i\right)^2 \sum\left(B_i - \widehat{B}_i\right)^2}}, \tag{3}$$

where A_i and B_i = feature sets and $\widehat{A}_i$ and $\widehat{B}_i$ = mean of feature sets A_i and B_i, respectively.

Lastly, the F-score correlation feature selection method was adopted. F-Score correlation is an algorithm that is used to determine the direct or indirect relation among data values. If this F-score value is smaller among feature sets, then those features are not related to each other, whereas if the value is higher, then that feature is highly related and can be added to the feature subset. Mathematically, it is represented as in the following equation:

$$F_{\text{score}} = \frac{\sum_{j=1}^{J}\left(\overline{f}_i^j - \overline{f}_i\right)^2}{\sum_{j=1}^{J} 1/N_j - 1 \sum_{n=1}^{n_j}\left(f_{n,i}^j - f_i^j\right)^2}, \tag{4}$$

where $\overline{f}_i$ = mean of f_i feature set, $\overline{f}_i^j$ = mean of i_{th} attribute of the f_j feature set, $f_{n,i}^j$ = i_{th} attribute of the n_{th} instance in f_j feature set, and N_j = number of attributes in j_{th} feature set.

After finding correlations from three different algorithms, different feature sets are identified. These feature sets are further input into the feature entropy estimation algorithm which results in optimal feature selection that contributes to finding the class of input data packets. This is considered to be an ensembling of feature selection that is preferred to give a precise result. This is illustrated in Algorithm 2.

Feature entropy estimation (FEE) was used to find the best-related feature extraction; information gain formula is used for feature selection. In the context of the target variable, it evaluates the gain for each variable. The calculation is referred to as the mutual information between the two random variables in this slightly different application. The best characteristic is determined by the entropy calculation. Entropy is an uncertainty measure that can be used to deduce the distribution of characteristics in a concise way. It is mathematically evaluated as in the following equation:

$$\text{FEE} = E_f - \sum_{n=1}^{N} \frac{f_i}{f}\left(E_{f_n}\right), \tag{5}$$

where E_f = entropy of feature sets, f_i = i_{th} feature set, and E_{f_n} = entropy of n_{th} subset of feature sets.

Therefore, for attack detection, these feature selection techniques are applied to generate important, relevant features and remove unnecessary features to reduce computational complexity that will result in a reduction of execution time for malicious IoT traffic.

3.3. Extreme Gradient Boosting Classification. Gradient boosting is a kind of collective machine learning algorithm that can be used to solve categorization [25] and regression modeling issues. Decision tree models are used to create

```
Begin
Input: Incoming IoT traffic ($IoT_{traffic}$)
Output: Extracted Features ($F_v$)
    $i$ = input packets, $g_n$ = gateway nodes,
    $i_{norm}$ = normalized input packet
For each $i \epsilon IoT_{traffic}$
    $g_n \longrightarrow$ Gateway ($i$)
$z\_score$ ($i$)$\longrightarrow$ $i_{norm}$
Extract ($i$)$\longrightarrow F_v$
Return ($F_v$)
End
```

Algorithm 1: Data collection and feature extraction.

```
Begin
Input: $F_v$
Output: $F_{v_{optimal}}$
$R \leftarrow$ nrows ($F_v$)
$C \leftarrow$ ncols ($F_v$)
$F_{v1} \leftarrow Chi^2$
$F_{v2} \leftarrow P_{corr}$
$F_{v3} \leftarrow F_{score}$
for each $i$ in $C$ do
    $F_{v_{optimal}}$ FEE ($F_{v1}$, $F_{v2}$, $F_{v3}$)
end for
return $F_{v_{optimal}}$
```

Algorithm 2: Feature selection.

ensembles. To correct the forecasting misclassification caused by past models, trees are introduced to the array at the same time and matched. Gradient boosting gets its name from the fact that the loss gradient is reduced when the model is fitted, almost like a neural network. Simulations are fitted using a gradient-based approach and any configurable differentiable loss function. Because the GBDT algorithm is prone to overfitting, the XGBoost technique incorporates normalization factors into the original GBDT algorithm. XGBoost has been extensively enhanced in contrast to previous algorithms, as evidenced by the greatly enhanced training time and accuracy. Let us consider input as x_i, output as o_i, and $\widehat{o}_i$ as the observed and predicted label, respectively. Mathematically [26], the learning model is represented as

$$\begin{aligned}\widehat{o}_i^{(1)} &= \widehat{o}_i^{(0)} + f_1(x_i),\\ \widehat{o}_i^{(2)} &= \widehat{o}_i^{(1)} + f_2(x_i),\\ \widehat{o}_i^{(t)} &= \widehat{o}_i^{(t-1)} + f_t(x_i),\end{aligned} \tag{6}$$

where $f_t(x_i)$ = the weak learning function.

The loss function, while training, is mathematically represented as in the following equation:

$$\text{loss}_t = \sum_{i=1}^{n} \text{loss}(o_i, \widehat{o}_i) + \sum_{t=1}^{T} \text{loss}(f_t(x_i)), \tag{7}$$

where loss_t = training loss function, $\text{loss}(o_i, \widehat{o}_i)$ = empirical loss between observed and predicted labels, And $\text{loss}(f_t(x_i))$ = loss of boosted learner.

The entire training process is illustrated in Algorithm 3.

4. Results and Discussion

In this section, first, we have illustrated the dataset used and the environment of the experiment. Then, the metrics used are discussed for the measurement of performance of proposed models, and later, results are discussed.

4.1. Experimental Setup. We selected three datasets for the performance evaluation of the proposed approach, namely, the NSL-KDD dataset [27], the UNSW NB15 dataset [28], and the CICIDS2017 dataset [29]. The NSL-KDD dataset was generated from the KDD Cup'99 dataset to eliminate identical datasets and alleviate the problems involved with the KDD Cup'99 dataset. There are 125,973 data records in the NSL-KDD train database and 22,544 data files in the test data file. The size of the NSL-KDD record is large enough that the full record can be used without the use of a representative sample. The given dataset is comprised of 41 characteristics and 22 training intrusion attacks. Here, the connection has 21 characteristics, and the type of connection all together in the same host has 19 characteristics. The IXIA Perfect Scenario program in the Australian Center for Cyber Security (ACCS), Cyber Range Lab, established a combination of spatial and temporal activities from the unprocessed network packets of the UNSW-NB 15 dataset [28]. 100 GB of unprocessed data traffic is collected using the

```
Begin
Input: $v(x, y)_i$, $N$ = number of iterations
Initialize: $f_t$, $t = 1, 2, \ldots .T$
$F_v$ = Algorithm 1 (IoT$_{\text{traffic}}$)
$F_{v_{\text{optimal}}}$ = Algorithm 2 ($F_v$)
for $t = 1$ to $T$ do
    compute gradients ($F_{v_{\text{optimal}}}$)
    $\widehat{o}_i^{(t)} = \widehat{o}_i^{(t-1)} + f_t(x_i)$
    loss$_t \leftarrow \sum_{i=1}^{n} \text{loss}(o_i, \widehat{o}_i) + \sum_{t=1}^{T} \text{loss}(f_t(x_i))$
    if loss$_t$ = = min
        return $\widehat{o}_i^{(t)}$
    end if
end for
```

Algorithm 3: Training process.

tcpdump utility (e.g., pcap files). There are nine different security threats in this dataset. And the last dataset is CICIDS2017 [29] which includes even more fresh data packets, both with and without assault, that is remarkably similar to real-world communication networks. This database comprises current real-world network-like information that has been gathered over five days and included a variety of malware as well as normal data. This work is employed on a 64-bit Intel Core-5 CPU with 8 GB RAM in Windows 10 environment. Machine learning algorithm is implemented in MATLAB 2020a.

4.2. Performance Parameters. The proposed work was evaluated on the basis of the following parameters:

$$\begin{aligned}
\text{Accuracy} &= \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})} * 100,\\
\text{Precision} &= \frac{\text{TP}}{(\text{TP} + \text{FP})} * 100,\\
\text{Recall} &= \frac{\text{TP}}{(\text{TP} + \text{FN})} * 100,\\
\text{F\_Measure} &= \frac{2 * \text{Precision} * \text{Recall}}{(\text{Precision} + \text{Recall})},
\end{aligned} \tag{8}$$

where TP stands for true positive that means if actual and predicted data samples are an anomaly in nature, then TP is evaluated, TN stands for true negative that means if actual and predicted data samples are not an anomaly in nature, then TN is evaluated, FP stands for false positive that means if actual and predicted data samples are normal and anomaly in nature, respectively, then FP is evaluated, and FN stands for false negative that means if actual and predicted data samples are an anomaly and normal, respectively, then FN is evaluated.

4.3. Result Analysis. Table 2 shows the performance evaluation of the proposed intrusion detection system on the NSL-KDD dataset with 5-fold validation. Table 2 represents performance parameters of accuracy, precision, recall, and F_Measure. Similarly, Table 3 represents the performance evaluation on the CCIDS2017 dataset. And Table 4 represents the performance evaluation on the UNSW_NB15 dataset. In this analysis, the random samples from the test dataset are selected and evaluated. In this work, 5-fold validation is performed. The dataset is divided into 5 parts randomly, one part is selected testing, and other parts are used for training. Similarly, Table 5 represents the time complexity of the proposed algorithm. From Table 5, it was observed that the average time complexity on the NSL-KDD dataset was approx. 38 sec, whereas for USNW_NB15 and CCIDS 2017 the time complexity was approx. 2 sec and 3 sec, respectively. The proposed methodology results in a lightweight low-cost feature selection method for IoT devices. This is due to its low computational time complexity that is illustrated in Figure 3. This figure justifies the time taken for the selected number of features from incoming IoT network traffic.

4.4. State-of-the-Art Comparison. The IoT or edge nodes are vulnerable to network attacks, and network connectivity enables malware injection from the Internet. In most of the attack detection learning models, vanishing gradient problem occurs and faces overfitting issues during the latter stages of training. Nowadays, it has become one of the most promising research areas for researchers as daily new attacks are introduced in the network. This section is dedicated to exploring the work of other researchers in the field of intrusion detection. A comparative state of the art with other existing works is illustrated in Table 6.

5. Conclusion

Attack detection in IoT is quite an essential task to keep track of the security of IoT traffic. In the past few years, many researchers have implemented machine learning (ML) techniques to track and block malicious IoT traffic. But in the presence of inappropriate features, these ML models lead to misclassification issues along with time complexity during the learning process. This noteworthy issue needed to be resolved by designing a framework for optimal and accurate feature selection from malicious IoT traffic. For this purpose, a new framework

Table 2: Evaluation of proposed methodology on NSL-KDD dataset.

Testsets		Accuracy	Precision	Recall	*F*_measure
Testset_1	DoS	94.54444	91.10397	94.22564	92.63851
	Probe	95.93165	84.76717	68.65544	75.86532
	R2L	99.50386	95.67901	38.94472	55.35714
	U2R	99.96626	99.97023	99.99603	99.98313
Testset_2	DoS	94.57939	91.14358	94.29584	92.69292
	Probe	95.12826	90.0986	53.4878	67.1258
	R2L	99.54866	94.96855	44.15205	60.27944
	U2R	99.96825	99.97505	99.99319	99.98412
Testset_3	DoS	94.64966	91.31289	94.31761	92.79093
	Probe	95.21327	90.21739	54.37197	67.85143
	R2L	99.51577	96.18321	41.44737	57.93103
	U2R	99.97354	99.97883	99.99471	99.98677
Testset_4	DoS	94.62738	91.27438	94.31897	92.7717
	Probe	94.47814	90.65934	45.15908	60.28774
	R2L	99.50783	94.17476	39.43089	55.58739
	U2R	99.98095	99.98412	99.99682	99.99047
Testset_5	DoS	94.57432	91.14699	94.23958	92.66749
	Probe	95.92379	84.40888	69.27966	76.0996
	R2L	99.51578	96.20253	38.97436	55.47445
	U2R	99.98015	99.98412	99.99603	99.99008

Table 3: Evaluation of proposed methodology on UNSW_NB15 dataset.

Testsets	Accuracy	Precision	Recall	*F*_measure
Testset_1	99.95465	100	99.90926	99.95461
Testset_2	99.96599	100	99.93179	99.96588
Testset_3	99.96112	100	99.9223	99.96113
Testset_4	99.94556	100	99.8913	99.94562
Testset_5	99.93197	100	99.8645	99.9322

Table 4: Evaluation of proposed methodology on CICIDS2017 dataset.

Testsets	Accuracy	Precision	Recall	*F*_measure
Testset_1	99.93197	100	99.8645	99.9322
Testset_2	99.9114	99.97733	99.86413	99.9207
Testset_3	99.9114	99.97371	99.8687	99.92118
Testset_4	99.89367	99.96824	99.84142	99.90479
Testset_5	99.93355	99.96047	99.92098	99.94072

Table 5: Time evaluation of proposed methodology.

Time (in sec)	NSL-KDD	UNSW_NB15	CICIDS2017
Testset_1	34.64	1.21	2.99
Testset_2	38.78	1.25	3.06
Testset_3	39.18	1.18	3.09
Testset_4	40.79	1.12	3.99
Testset_5	38.51	1.22	3.88
Average	38.38	1.196	3.402

model is proposed. Firstly, the proposed feature selection approach is developed by combining rank-based chi-square, Pearson correlation, and f_score correlation to extract relevant features out of all available features from the dataset. These algorithms are a type of wrapper technique that filters out the features more accurately and effectively for classification. Then, feature entropy estimation was applied to validate the relationship among all extracted features to identify malicious traffic in IoT networks. The experimental simulation was performed by using three datasets, NSL-KDD, UNSW-NB15, and CCIDS2017, and compared with some existing works. It was observed that on the NSL-KDD dataset, accuracy was approx. 97.48%. Similarly, the accuracy of USNW-NB15 and CCIDS2017 was approx. 99.96% and 99.93%, respectively.

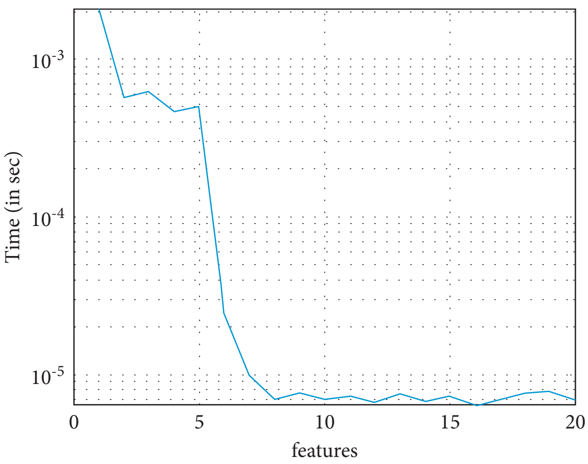


Figure 3: Feature selection time complexity analysis.

Table 6: Comparative performance evaluation.

	OS-ELM [12]	ELM [30]	Ours	XGBoost [30]	CART [31]	ANN [32]	Ours	SVM [15]	Ours
	NSL-KDD			UNSW-NB15				CCIDS2017	
Accuracy	96.54	94.45	97.48	88	88	84	99.96	98.03	99.93
Training time (in sec)	—	~600	~50	—	~6	—	~2	~16	~4

The following conclusions can be derived from the implementation of the proposed algorithm:

(i) The proposed framework can enforce security and trustworthiness on the Internet of things (IoT)

(ii) The feature selection techniques remove the drawback of a local minimum, and they converge faster

(iii) By selecting optimal features, training time is reduced

(iv) Highly related features are needed for improvement in performance level. Unnecessary features will cause calculation complexity

(v) Faster execution with reduced features results in faster alert of intrusion, and prevention measures can be applied accordingly

In future work, this work would be extended for other datasets also and more real-time attack detection would be explored. This would create fine-grained usage limitations to ensure privacy characteristics across big datasets even while enabling classification algorithms and analyses to operate on top of them. Internet of things (IoT) application frameworks would develop the necessary technical capabilities to impose sufficient security controls as even more data are collected, transferred, and analyzed over a common infrastructure.

Notations

$\text{IoT}_{\text{traffic}}$:	Incoming IoT traffic
F_v:	Extracted features
g_n:	Gateway nodes
i_{norm}:	Normalized input packets
x_i:	i_{th} feature set
Mean (x_i):	Mean of i_{th} feature set
std (x_i):	The standard deviation of i_{th} feature set
z_i:	Z-score of feature sets
f:	Feature sets
c_l:	Class label
$\text{Chi}^2_{(f_i, c_l)}$:	Rank-based chi-square feature selection
A_i:	Number of observations in class c_l
C_i:	Number of expected observations in class c_l
P_{corr}:	Pearson moment correlation
A_i, B_i:	Any feature sets
$\widehat{A}_i$, $\widehat{B}_i$:	Mean of feature sets
$\underline{F}_{\text{score}}$:	F-score correlation
$\underline{f_i}$:	Mean of f_i feature set
$\underline{f}_{i,}^{j}$:	Mean of the i^{th} attribute of the f_j feature set
$f_{n,i}^{j}$:	The i^{th} attribute of the n^{th} instance in f_j feature set
N_j:	Number of attributes in j_{th} feature set
E_f:	The entropy of feature sets
f_i:	i_{th} feature set
E_{f_n}:	The entropy of n_{th} subset of feature sets
x_i:	Input data
o_i:	Observed class label
$\widehat{o}_i$:	Predicted class label
$f_t(x_i)$:	Weak learning function
loss_t:	Training loss function
$\text{loss}(o_i, \widehat{o}_i)$:	

Empirical loss function between observed and predicted labels

$\text{loss}(f_t(x_i))$: The loss function of the boosted learner

F_v: Feature vector

$F_{v_{\text{optimal}}}$: Optimal feature vector.

Data Availability

The data that support the findings of this study are available on request from the corresponding author.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This work was supported by the King Khalid University Researchers Supporting Project Number (R. G. P. 1/77/42), King Khalid University, Saudi Arabia.

References

[1] V. Hassija, V. Chamola, V. Saxena, D. Jain, P. Goyal, and B. Sikdar, “A survey on IoT security: application areas, security threats, and solution architectures,” *IEEE Access*, vol. 7, 2019.

[2] R. Kandaswamy and D. Furlonger, “Blockchain-based transformation: a gartner trend insight report,” 2021, https://www.gartner.com/en/doc/3869696-blockchain-based-transformation-a-gartner-trend-insight-report.

[3] L. Xiao, Y. Li, G. Han, G. Liu, and W. Zhuang, “PHY-layer spoofing detection with reinforcement learning in wireless networks,” *IEEE Transactions on Vehicular Technology*, vol. 65, no. 12, pp. 10037–10047, 2016.

[4] R. Ullah, S. H. Ahmed, and B.-S. Kim, “Information-centric networking with edge computing for IoT: research challenges and future directions,” *IEEE Access*, vol. 6, pp. 73465–73488, 2018.

[5] N. Chaabouni, M. Mosbah, A. Zemmari, C. Sauvignac, and P. Faruki, “Network intrusion detection for IoT security based on learning techniques,” *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2671–2701, 2019.

[6] E. Benkhelifa, T. Welsh, and W. Hamouda, “A critical review of practices and challenges in intrusion detection systems for IoT: toward universal and resilient systems,” *IEEE Communications Surveys & Tutorials*, vol. 20, 2018.

[7] E. Albin and N. C. Rowe, “A realistic experimental comparison of the Suricata and Snort intrusion-detection systems,” in *Proceedings of the 26th IEEE International Conference on Advanced Information Networking and Applications Workshops, WAINA*, pp. 122–127, Fukuoka, Japan, March 2012.

[8] M. Al-Hawawreh, N. Moustafa, and E. Sitnikova, “Identification of malicious activities in industrial internet of things based on deep learning models,” *Journal of Information Security and Applications*, vol. 41, pp. 1–11, 2018.

[9] H. H. Al-Maksousy, M. C. Weigle, and C. Wang, “NIDS: neural network based intrusion detection system,” in *Proceedings of the 2018 IEEE International Symposium on Technologies for Homeland Security, HST 2018*, October 2018.

[10] E. Anthi, L. Williams, M. Slowinska, G. Theodorakopoulos, and P. Burnap, “A supervised intrusion detection system for smart home IoT devices,” *IEEE Internet Things Journal*, vol. 6, pp. 9042–9053, 2019.

[11] L. Bilge and T. Dumitras, “Before we knew it: an empirical study of zero-day attacks in the real world,” in *Proceedings of the ACM Conference on Computer and Communications Security*, pp. 833–844, Raleigh, NC, USA, October 2012.

[12] S. Prabavathy, K. Sundarakantham, and S. M. Shalinie, “Design of cognitive fog computing for intrusion detection in Internet of Things,” *Journal of Communications and Networks*, vol. 20, no. 3, pp. 291–298, 2018.

[13] A. Tabassum, A. Erbad, A. Mohamed, and M. Guizani, “Privacy-preserving distributed IDS using incremental learning for IoT health systems,” *IEEE Access*, vol. 9, pp. 14271–14283, 2021.

[14] S. Sibi Chakkaravarthy, D. Sangeetha, M. V. Cruz, V. Vaidehi, and B. Raman, “Design of intrusion detection honeypot using social leopard algorithm to detect IoT ransomware attacks,” *IEEE Access*, vol. 8, pp. 169944–169956, 2020.

[15] S. U. Jan, S. Ahmed, V. Shakhov, and I. Koo, “Toward a lightweight intrusion detection system for the internet of things,” *IEEE Access*, vol. 7, pp. 42450–42471, 2019.

[16] M. Eskandari, Z. H. Janjua, M. Vecchio, F. Antonelli, and I. D. S. Passban, “Passban IDS: an intelligent anomaly-based intrusion detection system for IoT edge devices,” *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 6882–6897, 2020.

[17] J. Li, Z. Zhao, R. Li, and H. Zhang, “AI-based two-stage intrusion detection for software defined IoT networks,” *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2093–2102, 2019.

[18] N. Ravi and S. M. Shalinie, “Semisupervised-learning-based security to detect and mitigate intrusions in IoT network,” *IEEE Internet of Things Journal*, vol. 7, no. 11, pp. 11041–11052, 2020.

[19] L. Nie, Z. Ning, X. Wang, X. Hu, J. Cheng, and Y. Li, “Data-driven intrusion detection for intelligent internet of vehicles: a deep convolutional neural network-based method,” *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 4, pp. 2219–2230, 2020.

[20] M. Shafiq, Z. Tian, A. K. Bashir, X. Du, and M. Guizani, “CorrAUC: a malicious bot-IoT traffic detection method in IoT network using machine-learning techniques,” *IEEE Internet of Things Journal*, vol. 8, no. 5, pp. 3242–3254, 2021.

[21] R. K. Gupta, A. R. Gupta, N. Pathik et al., “Novel deep neural network technique for detecting environmental effect of COVID-19,” *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, pp. 1–19, 2021.

[22] M. Gupta, K. K. Gupta, and P. K. Shukla, “Session key based fast, secure and lightweight image encryption algorithm,” *Multimedia Tools and Applications*, vol. 80, no. 7, pp. 10391–10416, 2021.

[23] L. Sun, R. K. Gupta, and A. Sharma, “Review and potential for artificial intelligence in healthcare,” *International Journal of System Assurance Engineering and Management*, 2021.

[24] G. Forman, “An extensive empirical study of feature selection metrics for text classification,” *Journal of Machine Learning Research*, vol. 3, pp. 1289–1305, 2003.

[25] R. Bhatt, P. Maheshwary, P. Shukla, P. Shukla, M. Shrivastava, and S. Changlani, “Implementation of fruit fly optimization algorithm (FFOA) to escalate the attacking efficiency of node capture attack in wireless sensor networks (WSN),” *Computer Communications*, vol. 149, pp. 134–145, 2020.

[26] R. Gupta, P. K. Shukla, and P. Kumar Shukla, “Performance analysis of anti-phishing tools and study of classification data mining algorithms for a novel anti-phishing system,”

International Journal of Computer Network and Information Security, vol. 7, no. 12, pp. 70–77, 2015.

[27] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," in *Proceedings of the 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications, CISDA*, Ottawa, Canada, July 2009.

[28] H. M. Anwer, M. Farouk, and A. Abdel-Hamid, "A framework for efficient network anomaly intrusion detection with features selection," in *Proceedings of the 2018 9th International Conference on Information and Communication Systems, ICICS 2018*, pp. 157–162, Institute of Electrical and Electronics Engineers Inc., Irbid, Jordan, April 2018.

[29] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," in *Proceedings of the 4th International Conference on Information Systems Security and Privacy*, Funchal, Madeira, Portugal, 2018.

[30] R. Kozik, M. Choraś, M. Ficco, and F. Palmieri, "A scalable distributed machine learning approach for attack detection in edge computing environments," *Journal of Parallel and Distributed Computing*, vol. 119, pp. 18–26, 2018.

[31] A. Alsaedi, N. Moustafa, Z. Tari, A. Mahmood, and A. Anwar, "TON_IoT telemetry dataset: a new generation dataset of IoT and IIoT for data-driven intrusion detection systems," *IEEE Access*, vol. 8, pp. 165130–165150, 2020.

[32] S. Hanif, T. Ilyas, and M. Zeeshan, "Intrusion detection in IoT using artificial neural networks on UNSW-15 dataset," in *Proceedings of the HONET-ICT 2019-IEEE 16th International Conference on Smart Cities: Improving Quality of Life Using ICT, IoT and AI*, pp. 152–156, Institute of Electrical and Electronics Engineers Inc., Charlotte, NC, USA, October 2019.

Hindawi
Mobile Information Systems
Volume 2021, Article ID 2427434, 10 pages
https://doi.org/10.1155/2021/2427434

Research Article

Cost-Effective Proxy Signcryption Scheme for Internet of Things

Insaf Ullah,[1] **Ali Alkhalifah**,[2] **Muhammad Asghar Khan**,[1] **and Samih M. Mostafa**[3]

[1] *HIET, Hamdard University Karachi, Islamabad Campus, Islamabad 44000, Pakistan*
[2] *Department of Information Technology, College of Computer, Qassim University, Buraydah 51452, Saudi Arabia*
[3] *Faculty of Computers and Information, South Valley University, Qena 83523, Egypt*

Correspondence should be addressed to Muhammad Asghar Khan; khayyam2302@gmail.com

Received 8 July 2021; Revised 31 August 2021; Accepted 15 October 2021; Published 30 November 2021

Academic Editor: Zengpeng Li

The Internet of things (IoT) has emerged into a revolutionary technology that enables a wide range of features and applications given the proliferation of sensors and actuators embedded in everyday objects, as well as the ubiquitous availability of high-speed Internet. When nearly everything is connected to the Internet, security and privacy concerns will become more significant. Furthermore, owing to the resource-constrained nature of IoT devices, they are unable to perform standard cryptographic computations. As a result, there is a critical need for efficient and secure lightweight cryptographic scheme that can meet the demands of resource-constrained IoT devices. In this study, we propose a lightweight proxy in which a person/party can delegate its signing authority to a proxy agent. Existing proxy signcryption security approaches are computationally costly and rely on RSA, bilinear pairing, and elliptic curves cryptography (ECC). The hyperelliptic curve cryptosystem (HECC), on the other hand, employs a smaller key size while maintaining the same level of security. When assessed using the random oracle model (ROM), the proposed scheme provides resilience against indistinguishable under adaptive chosen ciphertext attacks (IND-CCA) and unforgeable under adaptive chosen message attacks (UU-ACMA). To demonstrate the viability of the proposed scheme, security analyses and comparisons with existing schemes are performed. The findings show that the proposed scheme provides high security while reducing computational and communication costs.

1. Introduction

Modern enterprises and business organizations require the delegation of signing rights due to a lack of processing capability or the temporal absence of an agent. Similarly, it attracted e-commerce applications like signing the business contract and online proxy auction. To provide the delegation of rights, Mambo et al. [1, 2] were the first who contributed a new method called a proxy signature. This approach includes three participants: original signer, proxy signer, and a verifier/receiver. The original signer can delegate its signing rights to the delegated agent/proxy signer. Later, the delegated agent uses this sign on the behalf of its delegator and delivers it to the respective verifier/receiver. Unfortunately, the schemes in [1, 2] do not provide any solution to prevent it from misuse. Another attempt in enhancing proxy signature was made by Kim et al. [3]. They claim that the partial delegation with a warrant is more impactful and secure than full delegation in terms of computations and more processing speed. But it gives unlimited delegation resulting misuse of delegation. Another scheme proposed in [4] gives the concept of nonrepudiation by devising the threshold proxy signature scheme (TPSS). The scheme successfully preserves the nonrepudiation between the original sender and proxy groups without involving the trusted third party.

Though, the proxy signature will fail when communication includes some commercial secrets. Thus, to resolve this problem, Gamage et al. [5] designed a proxy signcryption approach by combining proxy signature and the encryption in a single logical step. The proficiency and security strength of the given approach relies upon the discrete logarithm problem which causes making it more costly in terms of both computation and communication. In addition, the proposed approach does not provide some security services like forward secrecy and public verifiability. Zhang [6] contributed publicly verifiable and forward secure proxy signcryption scheme. His proposed scheme is inefficient as it needs a secure channel between the sender and proxy. In addition, the proposed scheme creates more computational cost and requires more bandwidth for communication. Li and Chen [7] used pairing phenomena in the identity-based proxy signcryption (IDBPYS) scheme that necessarily requires a safe medium for transferring the secret key to the user. Wang et al. [8] proposed an IDBPYS scheme that satisfies the security parameters like forward secrecy and public verifiability. But their proposed approach faces the key escrow problem. Duan et al. [9] presented a secure delegation-by-warrant IDBPYS scheme which is secure under the random oracle model (ROM). In this approach, efficiency and hardiness of security are based on bilinear pairing. It requires more communication bandwidth and creates high computation cost. Elkamshouchy et al. [10] improved the proxy signcryption techniques and proposed a new publicly verifiable proxy signcryption scheme based on the discrete logarithm problem (DLP). The authors claim that the given approach achieves the security properties of confidentiality and authenticity through an unsecured channel. Since, it depends upon DLP, which consumes more computing power. Furthermore, the proxy signcryption idea was furnished by Elkamshouchy et al. [11]. They attempted to improve the security of this scheme, but the scheme is affected by high computing power and extra bandwidth due to utilizing hard problems, i.e., integer factorization problem (IF), DLP, Diffie–Hellman problem (DHP), and DSA problem. So, IF, DHP, and DLP require more machine cycles and more computational power. Elkamchouchi and Abouslseoud [12] successfully enabled the partial delegation rights in their scheme by utilizing bilinear pairings on EC. However, in the given approach, the proxy signcrypter utilizes the signcrypting right incorrectly in light of the fact that in partial delegation, there is no limitation on proxy signcrypter. Lin et al. [13] designed a new provable secure proxy signcryption approach utilizing bilinear pairing. Unluckily, their proposed approach does not ensure the security requirement of warrant unforgeability. For further improving, Elkamchouchi et al. [14] proposed the notion of warrants-based proxy signcryption which is good for low resource devices. The security hardness and efficiency of this scheme are completely based on the elliptic curve cryptography that leads to more power consumption of the machine. Yanfeng et al. [15] presented a secure certificateless proxy identity-based signcryption scheme. They proposed elliptic curve discrete logarithm problem (ECDLP) for the efficiency and security in their scheme. But the scheme needs a secured channel for the partial private key distribution to the users. Elkamchouchi et al. [16] introduced two proxy signcryption schemes: one relies on DLP and other on ECDLP, respectively. They claim that this approach has less computational and communication costs. The scheme is still affected by more machine power consumption and extra communication bandwidth. Furthermore, the proposed scheme was not provable secured. Lo and Tsai [17] coined a provable secure proxy signcryption scheme depending on the bilinear pairings. They demonstrate better performance and secrecy in terms of in-distinguishability and unforgeability. Furthermore, they proved the security requirements of the given approach under the ROM. Then, for improving security services, Ming and Wang [18] proposed a provable secured proxy signcryption on the standard model. Because of heavy computations due to bilinear pairing, the proposed approach can still be affected by more machine control usage and extra communication of information transmission. Insafullah et al. [19] presented a lightweight proxy signcryption approach based on HECC. They claim that their newly designed scheme ensures all the security services with low computational and communication costs. Unfortunately, the scheme is affected by using more major operations over the hyperelliptic curve. Abdelfatah [20] coined a novel proxy signcryption approach and its EC variant. Hui and Lunzhia [21] coined a new proxy signcryption with EC. Waheed et al. [22] coined a new proxy signcryption with EC. Hundera et al. [23] coined a novel proxy signcryption approach with bilinear pairing for cloud data sharing. However, the designed approaches in [20–23] have been affected by more computational cost and extra communication bandwidth due to EC and bilinear pairing.

1.1. Motivations and Contributions.

Keeping in view all the above proxy signcryption approaches, we identified that there is still a need for improvement in computational cost and bandwidth utilization. Though the abovementioned techniques are based on some prominent security techniques, i.e., RSA, bilinear pairing, and EC, HECC provides an equal level of security with 80 bits key size as compared to the elliptic curve with 160 bits key size and RSA and bilinear pairing with 1024 bits key size, respectively. Therefore, in order to decrease computational costs and channel bandwidth consumption, we design a cost-effective proxy signcryption scheme based on HECC that perform three roles of proxy delegator/original signcrypter, proxy signcrypter, and proxy unsigncrypter. The following are the main contributions of this study:

(i) We make a new proxy signcryption approach with the help of the hyperelliptic curve cryptosystem

(ii) We prove that the proposed approach is resilient against indistinguishable under adaptive chosen ciphertext attacks (IND-CCA) and unforgeable under adaptive chosen message attacks (UU-ACMA), when it is tested through the random oracle model (ROM).

(iii) Our approach reduces the computational cost and communication costs as compared to its counterpart schemes

1.2. Organization of the Study.

The organization of the study is as follows. Section 2 defines the basic preliminaries and threat model. The proposed model and the algorithm are defined in Section 3. Section 4 contains the security analysis of the proposed approach. Furthermore, in Section 5, we describe the computation and communication overheads analysis. Section 6 discusses the communication overhead, and Section 7 presents the conclusion.

2. Preliminaries

This section includes some formal definitions of the hyperelliptic curve discrete logarithm problem and hyperelliptic curve Diffie–Hellman problem; furthermore, the explanation of the threat model is provided.

Definition 1. Suppose a devisor $\mathscr{D}$ of order n and an instance $\xi = \delta \cdot \mathscr{D}$ is given, so, to extract δ from ξ is said to be hyperelliptic curve discrete logarithm problem (HDL).

Definition 2. Suppose a devisor $\mathscr{D}$ of order n and an instance $\xi = \ell \cdot \delta$ is given, so, to extract δ and ℓ from ξ is said to be hyperelliptic curve Diffie–Hellman problem (HDDH).

2.1. Threat Model. Here, we are trying to explain the threats against our proposed scheme regarding the security requirements of indistinguishable under adaptive chosen ciphertext attacks (IND-CCA) and unforgeable under adaptive chosen message attacks (UU-ACMA) by adversary $\mathscr{A}$. The following Definitions 3 and 4 can be better explaining the threats against our newly proposed scheme.

Definition 3. The newly proposed scheme can be IND-CCA secure, if $\mathscr{A}$ with the help of challenger $\mathscr{A}$ cannot win with nonnegligible benefit in the following game.

Setup: $\mathscr{A}$ executes the setup part to make the global parameter param and sends it to $\mathscr{A}$.

2.1.1. Phase 1

Hash queries: $\mathscr{A}$ submits these queries and Φ can check the value for the ask queries if the value is found in the list; then, it gives the value to $\mathscr{A}$; otherwise, Φ selects a random value for each ask query and sends them to $\mathscr{A}$.

Private key generation query: $\mathscr{A}$ can submit queries for private key of signer and Φ executes the key generation algorithm to produce the required private key and dispatch it to $\mathscr{A}$.

Proxy delegation query: when this query is submitted by $\mathscr{A}$, Φ responds as valid delegation for ask query to $\mathscr{A}$.

Proxy signcryption query: when this query is submitted with message and private key of proxy and delegation by $\mathscr{A}$, Φ responds as valid proxy signcryption tuple for asking query to $\mathscr{A}$.

Proxy unsigncryption query: when this query is submitted with proxy signcryption tuple by $\mathscr{A}$, Φ responds as valid plaintext which is generated through proxy unsigncryption for asking query to $\mathscr{A}$.

Challenge: two equal lengths plaintext $\mathfrak{m}_a$ and $\mathfrak{m}_b$ will send by $\mathscr{A}$, and Φ uniformly chooses a bit b $\in$ {0, 1} and computes a ununderstandable text ψ^* on $\mathscr{M}$b.

2.1.2. Phase 2. In this phase, $\mathscr{A}$ should make same queries as phase 1 with the following constraints:

(i) $\mathscr{A}$ will not send a request for any user private key

(ii) $\mathscr{A}$ never asks for proxy unsigncryption for ciphertext ψ^*

(iii) At the end of this phase, $\mathscr{A}$ generates a bit b^* and succeeds this game if $b^* = b$.

Definition 4. The newly proposed scheme can be UU-ACMA secure, if $\mathscr{A}$ with the help of challenger Φ cannot win with nonnegligible benefit in the following game.

Setup: same as above IND-CCA game.

Query: same as above IND-CCA game.

Forgery: finally, $\mathscr{A}$ outputs a proxy signcryption tuple and succeeds in this game if the following events happen successful.

(i) The generated proxy signcryption text is valid

(ii) The private key of proxy signcrypter never been asked

(iii) The proxy signcryption text is not generated using proxy signcryption query

3. Proposed Model

We present here our cost-effective proxy signcryption scheme for low constraint environment. Our proposed scheme is comprised of four phases such as public key verification, proxy delegation, proxy signcryption, and proxy unsigncryption, respectively. The block diagram of our cost-effective proxy signcryption scheme is shown in Figure 1 and the symbols used in algorithm in Table 1. Four types of roles used in our scheme are public key verification, proxy delegator/original signcrypter, proxy signcrypter, and proxy unsigncrypter. First of all, each user verifies the requested user public key from certificate authority (CA). A proxy delegator first sends a warrant message with the signature to delegate the signcryption privileges to proxy signcrypter. Later, proxy signcrypter verifies the received message and computes the signcryption on behalf of the proxy delegator and then delivers it to the proxy unsigncrypter. After receiving a proxy signcryption tuple, proxy unsigncrypter verifies the authentication and performs the steps of unsigncryption.

3.1. Setup. In this section, the certificate authority (CA) pick HEC with 80 bits parameter size, make a system parameter set as $\ell = \{\text{HEC}, \mathscr{D}, \text{h}_0, \text{h}_1, \text{h}_2, \text{h}_3, \zeta\}$, where ζ is the public key CA and made as by selecting π at random, and then compute $\zeta = \pi \cdot \mathscr{D}$. Finally, CA makes sure the availability of ℓ in a network publicly.

3.2. Key Generation. In this subsection, the participants $(i = \mathscr{U}o, \mathscr{PS}, \mathscr{PU})$ first compute their public and private keys in the following way. The participants (i) randomly selects a number $a_i \in \{1, 2, \ldots, q-1\}$ and calculates $\text{f}_i = a_i \cdot D$. So, a_i and f_i represent the participants (i) private and public keys.

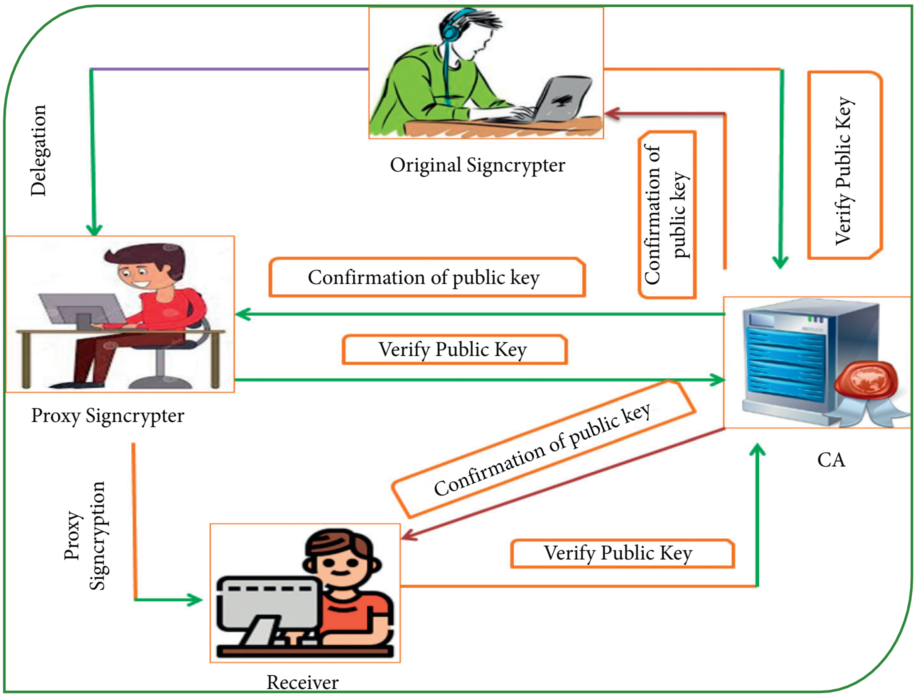


Figure 1: Framework model of the proposed proxy signcrypton scheme.

Table 1: Symbols used in the algorithm.

Notations of algorithm	Descriptions
$\mathscr{D}$	Divisors of the hyperelliptic curve
$\mathscr{U}o, \mathscr{P}\mathscr{S}, \mathscr{P}\mathscr{U}$	Represents the role of delegator, proxy signcrypter, and unsigncrypter
$a_{\mathscr{U}}$, $a_{\mathscr{P}}$, a_r	Private keys of delegator, proxy signcrypter, and unsigncrypter
$\mathrm{f}_{\mathscr{U}}$, $\mathrm{f}_{\mathscr{P}}$, f_r	Public keys of delegator, proxy signcrypter, and unsigncrypter
mw, m	Warrant message and message (plain text)
N_a, N_p	Nonce for delegator and proxy signcrypter
$E_{\mathscr{K}}$, $D_{\mathscr{K}}$	Encryption and decryption
$\mathrm{h}_0, \mathrm{h}_1, \mathrm{h}_2, \mathrm{h}_3$	Hash functions
$\mathscr{K}, \mathscr{K}_{sr}$	Preshared, computed shared key among proxy signcrypter and unsigncrypter
$\mathscr{X}_{\mathscr{K}}, \mathscr{Y}_{\mathscr{K}}$	Secret and public key for proxy signature generation and verifications

3.3. Proxy Delegation. In this subsection, the original signcrypter/proxy delegator $\mathscr{U}_o$ gives the right of the sign to proxy signcrypter $\mathscr{P}\mathscr{S}$.

(i) The original signcrypter selects $\ell \in \{1, 2, \ldots, q-1\}$

(ii) Compute $\mathscr{W} = \mathscr{Z} \cdot \mathscr{D}$

(iii) Compute $\mathscr{T} = \mathrm{h}_0(\mathrm{mw}, \mathscr{W})$ and also compute $\mathscr{V} = (\mathscr{Z} - a_{\mathscr{U}} \cdot \mathscr{T}) \bmod q$

(iv) Sends $\delta = (\mathscr{V}, \mathscr{J}, \mathrm{mw})$ to proxy signcrypter PS

After receiving $\delta = (\mathscr{V}, \mathscr{J}, \mathrm{mw})$ for validation, $\mathscr{P}\mathscr{S}$ performs the following equations:

$$\begin{aligned}
\mathscr{W} &= \mathscr{V} \cdot \mathscr{D} + \mathrm{h1}(\mathrm{mw}, N_a, \mathscr{J}) \cdot \mathrm{f}_{\mathscr{U}} \\
&= \mathscr{V} \cdot \mathscr{D} + \mathrm{h1}(\mathrm{mw}, N_a, \mathscr{J}).\mathrm{f}_{\mathscr{U}} = (\mathscr{Z} - a_{\mathscr{U}}.\,\mathrm{h1}(\mathrm{mw}, N_a, \mathscr{J})).\mathscr{D} + \mathrm{h1}(\mathrm{mw}, N_a, \mathscr{J}) \cdot a_{\mathscr{U}} \cdot \mathscr{D} \\
&= \mathscr{D} \cdot (\mathscr{Z} - a_{\mathscr{U}}.\,\mathrm{h1}(\mathrm{mw}, N_a, \mathscr{J}) + a_{\mathscr{U}}.\,\mathrm{h1}(\mathrm{mw}, N_a, \mathscr{J})) \\
&= \mathscr{D} \cdot (\mathscr{Z}) = \mathscr{Z} \cdot \mathscr{D} = \mathscr{W}.
\end{aligned} \tag{1}$$

After validation, the proxy signcrypter $\mathscr{P}\mathscr{S}$ generates the secret key $\mathscr{X}_{\mathscr{K}} = (\mathscr{V} + a_{\mathscr{P}}) \bmod q$ and then calculates and publishes the public key $\mathscr{Y}_{\mathscr{K}} = \mathscr{X}_{\mathscr{K}} \cdot \mathscr{D}$.

3.4. Proxy Signcryption. In this subsection, proxy signcrypter $\mathscr{P}\mathscr{S}$ performs the following steps to generate signcryption on the message (m).

(i) First choose a random number $\mathrm{j} \in \{1, 2, \ldots, q-1\}$

(ii) Compute $\mu = \mathrm{j} \cdot \mathscr{D}$, where $\mathscr{D}$ is the divisor over the hyper elliptic curve

(iii) Compute $\mathscr{K} = \mathrm{h}_1(\mathscr{K}_{\mathrm{sr}} + \mu)$, where $\mathscr{K}_{\mathrm{sr}}$ is the shared secret key between proxy and recipient

(iv) Compute the ciphertext $\mathscr{C} = E_{\mathscr{K}}(\mathrm{m})$, where m is the plain text

(v) Compute the hash function $\Omega = \mathrm{h}_2(\mathscr{C}, \mu)$

(vi) Compute the signature $\mathscr{S} = ((\mathrm{j}/\mathscr{X}_{\mathscr{K}}) - \Omega) \mathrm{mod} q$, where $\mathscr{X}_{\mathscr{K}}$ is the proxy signcrypter secret key

(vii) Then, send $\psi = (\mathscr{C}, \mathscr{S}, \Omega)$ to the proxy unsigncrypter

3.5. Proxy Verification and Unsigncryption. In this subsection, receiving the tuple $\psi = (\mathscr{C}, \mathscr{S}, \Omega)$ proxy unsigncrypter carry out the subsequent steps for verification and decryption of the proxy signcrypted text.

(i) First recover $\mu = \mathscr{Y}_{\mathscr{K}} \cdot (\mathscr{S} + \Omega)$ and $\mu = \mathscr{X}_{\mathscr{K}} \cdot \mathscr{D} \cdot ((\mathrm{j}/\mathscr{X}_{\mathscr{K}}) - \Omega + \Omega) = \mathrm{j} \cdot \mathscr{D}$

(ii) After this, verify the signature $\Omega^* = \mathrm{h}_2(\mathscr{C}, \omega)$ and accept if $\Omega = \Omega^*$

(iii) Compute $\mathscr{K} = \mathrm{h}_1(\mathscr{K}_{\mathrm{sr}} + \mu)$ and decrypt $(\mathrm{m}) = D_{\mathscr{K}}(\mathscr{C})$

4. Security Analysis

Our scheme meets the security requirements of indistinguishable under adaptive chosen ciphertext attacks (IND-CCA) and unforgeable under adaptive chosen message attacks (UU-ACMA) by adversary $\mathscr{A}$. The following Theorems 1 and 2 can be better explaining the threats against our newly proposed scheme.

Theorem 1. *The newly proposed scheme can be IND-CCA secure, if $\mathscr{A}$ with the help of challenger Φ cannot win with nonnegligible benefit in the following steps.*

Proof. The instance of the hyperelliptic curve $(\mathscr{Q}, \mathscr{D}, \mathscr{V} \cdot \mathscr{D})$ is given to Φ and the task of Φ to compute $\mathscr{Q} = \mathscr{O} \cdot \mathscr{V} \cdot \mathscr{D}$.

Setup: Φ executes the setup part for to make the global parameter param and sends it to $\mathscr{A}$. □

4.1. Phase 1

Hash (h_0) queries: if $\mathscr{A}$ submits $(\mathrm{mw}, \mathscr{J})$ query and Φ can check the value for a query if the value found in the list is (LH0), then it gives the values $(\mathscr{T}_i)$ to $\mathscr{A}$; otherwise, Φ selects $\mathscr{T}_i$ randomly and send them to $\mathscr{A}$.

Hash (h_1) queries: if $\mathscr{A}$ submits a query and Φ can check the value for a query if the value found in the list is (LH1), then it gives the values $(\mathscr{K}_i)$ to $\mathscr{A}$; otherwise, Φ selects $\mathscr{K}_i$ randomly and send them to $\mathscr{A}$.

Hash (h_2) queries: if $\mathscr{A}$ submits a query and Φ can check the value for a query if the value found in the list is (LH2), then it gives the values (Ω_i) to $\mathscr{A}$; otherwise, Φ selects Ω_i randomly and send them to $\mathscr{A}$.

Private key generation query: if $\mathscr{A}$ submits query for private key and public key of signer and Φ randomly select $a_i \in \{1, 2, \ldots, q-1\}$, calculate $\mathrm{f}_i = a_i \cdot D$, and dispatch (a_i, f_i) to $\mathscr{A}$.

Proxy delegation query: when this query is submitted by $\mathscr{A}$, Φ responds as valid delegation δ to $\mathscr{A}$ in the following way.

(i) Φ randomly selects ℓ and $\mathscr{T}$ form $\{1, 2, \ldots, q-1\}$ and compute $\mathscr{W} = \mathscr{Z} \cdot \mathscr{D}$

(ii) Compute $\mathscr{V} = (\mathscr{Z} - a_{\mathscr{U}} \cdot \mathscr{T})$, set $\delta = (\mathscr{V}, \mathscr{J}, \mathrm{mw})$, and respond δ to $\mathscr{A}$ as a delegation

Proxy signcryption query: when this query is submitted with message (m) and private key of proxy (a_p) and delegation (δ) by $\mathscr{A}$, Φ responds as valid proxy signcryption ψ to $\mathscr{A}$ in the following way.

(i) Φ chooses random numbers $\mathrm{j}, \mathscr{K}, \Omega, \mathscr{X}_{\mathscr{K}} \in \{1, 2, \ldots, q-1\}$

(ii) Computes the ciphertext $\mathscr{C} = E_{\mathscr{K}}(\mathrm{m})$

(iii) Computes the signature $\mathscr{S} = ((\mathrm{j}/\mathscr{X}_{\mathscr{K}}) - \Omega)$

(iv) Set $\psi = (\mathscr{C}, \mathscr{S}, \Omega)$ and respond ψ to $\mathscr{A}$ as a proxy signcryption

Proxy unsigncryption query: when this query is submitted to ψ by $\mathscr{A}$, if this query is not for target participant, Φ responds as valid plaintext which is generated through proxy unsigncryption to $\mathscr{A}$. Otherwise, Φ outputs ψ as an invalid proxy signcryption tuple.

Challenge: two equal lengths plaintext $\mathfrak{m}_a$ and $\mathfrak{m}_b$ will send by $\mathscr{A}$, Φ uniformly chooses a bit $\mathrm{b} \in \{0, 1\}$, and computes an un-understandable text ψ^* on $\mathscr{M}$b as follows.

(i) Φ choose random numbers $\mathscr{X}_{\mathscr{K}}, \mu, \mathscr{K}_{\mathrm{sr}} \in \{1, 2, \ldots, q-1\}$

(ii) Compute $\mathscr{K} = (\mu + \mathscr{K}_{\mathrm{sr}})$, $\mathscr{C}^* = E_{\mathscr{K}}(\mathrm{m})$, and $\Omega^* = \mathrm{h}_3(\mathscr{C}^*, \mu)$

(iii) Compute the signature $\mathscr{S}^* = ((\mathrm{j}/\mathscr{X}_{\mathscr{K}}) - \Omega)$, set $\psi^* = (\mathscr{C}^*, \mathscr{S}^*, \Omega^*)$, and respond ψ^* to $\mathscr{A}$ as a proxy signcryption on $\mathscr{M}$b to $\mathscr{A}$.

4.2. Phase 2. Just like phase 1, $\mathscr{A}$ can submit the identical queries, but it does not make a query for receiver private key and a massage corresponding to the ψ^*.

After that, $\mathscr{A}$ results $\mathrm{b}^* \in \{0, 1\}$, and if $\mathrm{b}^* = \mathrm{b}$, then Φ results 1. Otherwise, Φ results 0. If $\mathscr{Q} = \mathscr{O} \cdot \mathscr{V} \cdot \mathscr{D}$, ψ^* is valid signcrypted text, and for this reason can extricate b by utilized advantage π. Accordingly, $\Pr[\Phi \longrightarrow 1 | \mathscr{Q} = \mathscr{O} \cdot \mathscr{V} \cdot \mathscr{D}] = \Pr[\mathrm{b}/= \mathrm{b} | \mathscr{Q} = \mathscr{O} \cdot \mathscr{V} \cdot \mathscr{D}] = 1/2 + \pi$.

If $\mathscr{Q} \neq \mathscr{O} \cdot \mathscr{V} \cdot \mathscr{D}$, $\mathscr{A}$ cannot extricate b without advantages. Accordingly, $\Pr[\Phi \longrightarrow 1 | \mathscr{Q} \neq \mathscr{O} \cdot \mathscr{V} \cdot \mathscr{D}] = \Pr[\mathrm{b}^* = \mathrm{b} | \mathscr{Q} \neq \mathscr{O} \cdot \mathscr{V} \cdot \mathscr{D}] = 1/2$.

Probability analysis: suppose the queries $(\mathrm{h}_0, \mathrm{h}_1, \mathrm{h}_2)$, qpk, $qp\ d$, and $qpsn$ represent hash queries, private key queries, proxy delegation queries, and proxy signcryption queries, separately.

Thus, we signify some measures ($\mathscr{MER}$) as follows:

(i) $\mathscr{MER}1$: Φ output is positive in private key queries, and the probability is $1 - qpk/2^k$.
(ii) $\mathscr{MER}2$: Φ output is positive in proxy unsigncryption queries, and the probability is $1 - 1/2^k$.
(iii) $\mathscr{MER}3$: Φ output is positive in challenge part, and the probability is $1/qpk - 2^k$.

So, the total probability will be as follows:

$$\begin{aligned}\Pr[\Phi \longrightarrow \pi] &= \Pr[\mathscr{MER}1\wedge\mathscr{MER}2\wedge\mathscr{MER}3]\\ &= \Pr\left[1-\frac{qpk}{2^k}\wedge 1-\frac{1}{2^k}\wedge\frac{1}{qpk}-2^k\right]\\ &=\left(1-\frac{qpk}{2^k}\right)\left(1-\frac{1}{2^k}\right)\left(\frac{1}{qpk}-2^k\right)\cdot\pi.\end{aligned} \tag{2}$$

Theorem 2. *The newly proposed scheme can be UU-ACMA secure, if $\mathscr{A}$ with the help of challenger Φ cannot win with nonnegligible benefit in the following steps.*

Proof. The instance of the hyperelliptic curve $(\mathscr{Q}, \mathscr{V}\cdot\mathscr{D})$ is given to Φ and the task of Φ to compute $\mathscr{Q} = \mathscr{V}\cdot\mathscr{D} = \mathscr{Y}_{\mathscr{K}}$.

Setup: Φ execute the setup part for to make the global parameter param ℓ and sends it to $\mathscr{A}$.

Phase 1

Queries: same like Theorem 1.

Forgery: according to forking lemma [24], Φ can get two valid proxy signcryption text that are $\psi = (\mathscr{C}, \mathscr{S}, \Omega)$ and $\psi^* = (\mathscr{C}, \mathscr{S}, \Omega^*)$. Then, for the verification, we get two equations that are $\mu = \mathscr{Y}_{\mathscr{K}}\cdot(\mathscr{S}+\Omega)$ and $\mu^* = \mathscr{Y}_{\mathscr{K}}\cdot(\mathscr{S}+\Omega^*)$. So, after subtraction, we can get the following results.

$$\begin{aligned}\mu-\mu^* &= \mathscr{Y}_{\mathscr{K}}\cdot(\mathscr{S}+\Omega)-(\mathscr{Y}_{\mathscr{K}}\cdot(\mathscr{S}+\Omega^*))\\ &= \mathscr{Y}_{\mathscr{K}}\cdot\mathscr{S}+\mathscr{Y}_{\mathscr{K}}\cdot\Omega-\mathscr{Y}_{\mathscr{K}}\cdot\mathscr{S}+\mathscr{Y}_{\mathscr{K}}\cdot\Omega^*\\ &= \mathscr{Y}_{\mathscr{K}}\cdot\Omega-\mathscr{Y}_{\mathscr{K}}\cdot\Omega^* = \mu-\mu^* = \mathrm{j}\cdot\mathscr{D}-\mathrm{j}^*\cdot\mathscr{D}\\ &= \mathscr{V}\cdot\mathscr{D}\cdot\Omega-\mathscr{V}\cdot\mathscr{D}\cdot\Omega^*\\ &= (\mathrm{j}-\mathrm{j}^*)\cdot\mathscr{D} = \mathscr{V}\cdot\mathscr{D}\cdot(\Omega-\Omega^*) = (\mathrm{j}-\mathrm{j}^*)\\ &= \mathscr{V}((\Omega-\Omega^*).\end{aligned} \tag{3}$$

$\mathscr{V} = (\mathrm{j}-\mathrm{j}^*)/(\Omega-\Omega^*)$; hence, this is the solution for solving the hyperelliptic curve discrete logarithm problem.

Probability analysis: suppose the queries $(\mathrm{h}_0, \mathrm{h}_1, \mathrm{h}_2)$, qpk, $qp\ d$, and $qpsn$ represent the hash queries, private key queries, proxy delegation queries, and proxy signcryption queries, separately.

Thus, we signify some measures ($\mathscr{MER}$) as follows:

(i) $\mathscr{MER}1$: Φ output is positive in private key queries, and the probability is $1 - qpk/2^k$.
(ii) $\mathscr{MER}2$: Φ output is positive in proxy unsigncryption queries, and the probability is $1 - 1/2^k$.
(iii) $\mathscr{MER}3$: Φ output is positive in challenge part, and the probability is $1/qpk - 2^k$.

So, the total probability will be as follows:

$$\begin{aligned}\Pr[\Phi \longrightarrow \pi] &= \Pr[\mathscr{MER}1\wedge\mathscr{MER}2\wedge\mathscr{MER}3]\\ &= \Pr\left[1-\frac{qpk}{2^k}\wedge 1-\frac{1}{2^k}\wedge\frac{1}{qpk}-2^k\right]\\ &=\left(1-\frac{qpk}{2^k}\right)\left(1-\frac{1}{2^k}\right)\left(\frac{1}{qpk}-2^k\right)\cdot\pi.\end{aligned} \tag{4}$$

□

5. Computational Cost

The comparisons of the proposed and existing proxy signcryption schemes in terms of major operations are offered in table. In Table 2, computational cost in ms is provided. The symbols $\mathscr{EML}$, $\mathscr{P}r$, and $\mathscr{HEML}$ represent the exponential computations, elliptic curve multiplications, pairing operations, and hyperelliptic curve devisor multiplication, respectively. The other operations such as addition, subtraction, hash, and division are ignored because they require fewer computations time.

To show more clearly the comparisons between the proposed scheme and existing schemes, it has been observed from [25], by using the Multiprecision Integer and Rational Arithmetic C Library (MIRACL) and test the run time of the basic cryptographic operations. The running time for basic cryptographic operations is given in Table 3 (tested it 100 of times), an experiment donned through:

(i) Raspberry PI 3 B + Rev 1.3
(ii) OS: Ubuntu 20.04 LTS, 64-bit
(iii) with CPU: 64-bit, processor: 1.4 GHz Quad-Core
(iv) With 1 GB of memory

Also, we assume the half-time elliptic curve for the hyperelliptic curve because it is the generalized form of elliptic curve [26–30]. So, Table 3 provides details about the average time. Table 4 and Figure 2 show that our proposed scheme is computationally efficient from existing schemes. In Table 2, we provide the computational cost comparisons in milliseconds.

6. Communication Overhead

To design a cryptographic protocol for wireless communication, media is an important element because wireless protocols need lower communication overhead. Selecting a larger size of parameters can greatly affect the efficiency. In this section, we compare our newly designed scheme with

Table 2: Computational cost comparisons.

Schemes	Proxy delegation	Proxy signcryption	Proxy verification and unsigncryption	Total
Insafullah et al. [19]	3.42	1.14	3.42	7.98
Abdelfatah [20]	6.84	2.28	2.28	11.40
Guo and Deng [21]	9.12	11.40	11.40	31.92
Waheed et al. [22]	2.28	6.84	11.40	20.52
Hundera et al. [23]	64.16	0	128.32	192.48
Proposed	3.42	1.14	1.14	5.70

Table 3: Running time in milliseconds.

Primitive	Average time (in milliseconds)
$\mathscr{EML}$	2.28
$\mathscr{Pr}$	32.08
$\mathscr{HEML}$	1.14

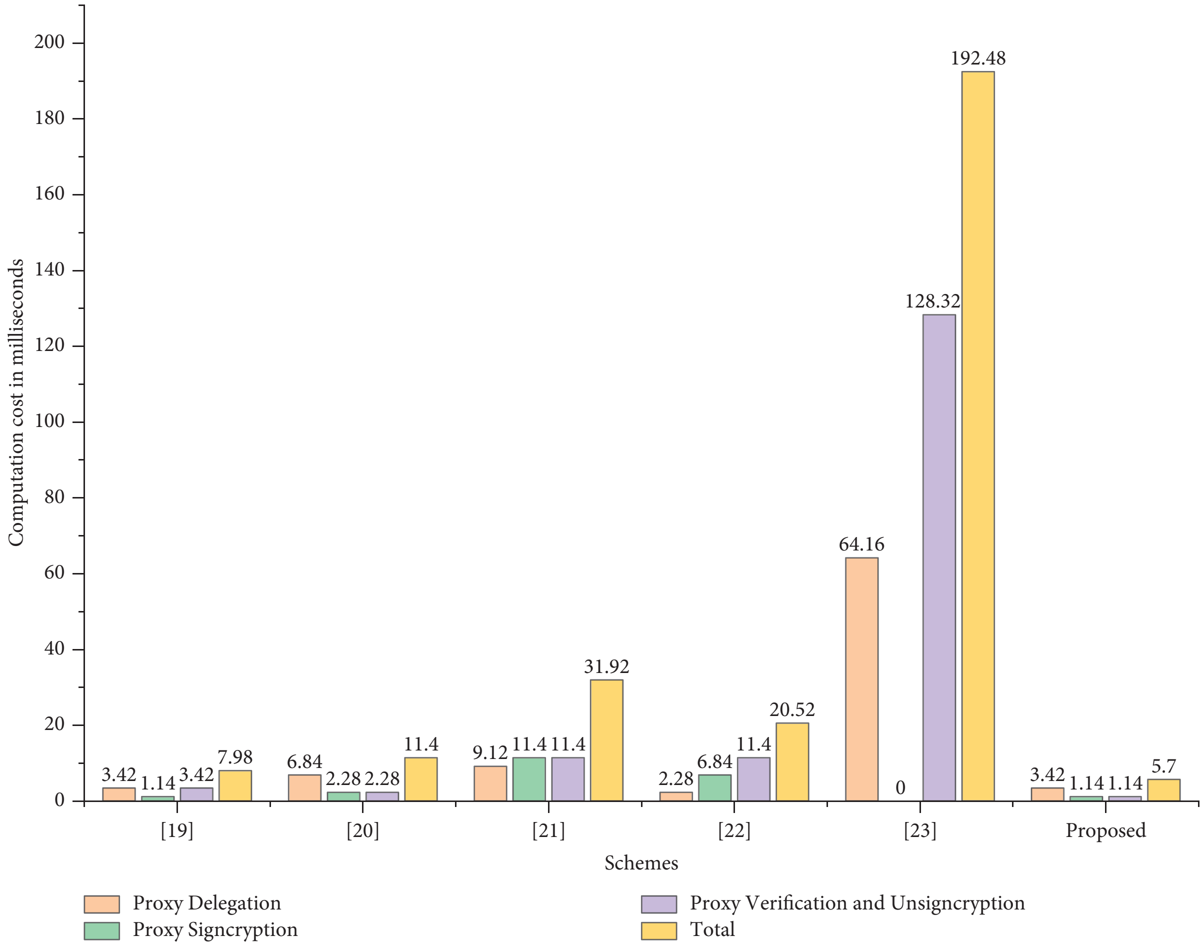


Figure 2: Computational cost comparisons.

Table 4: Major operation comparisons.

Schemes	Proxy delegation	Proxy signcryption	Proxy verification and unsigncryption	Total
Insafullah et al. [19]	$3\,\mathscr{HEML}$	$1\,\mathscr{HEML}$	$3\,\mathscr{HEML}$	$7\,\mathscr{HEML}$
Abdelfatah [20]	$3\,\mathscr{EML}$	$1\,\mathscr{EML}$	$1\,\mathscr{EML}$	$5\,\mathscr{EML}$
Guo and Deng [21]	$4\,\mathscr{EML}$	$5\mathscr{EML}$	$5\mathscr{EML}$	$14\,\mathscr{EML}$
Waheed et al. [22]	$1\,\mathscr{EML}$	$3\,\mathscr{EML}$	$5\mathscr{EML}$	$9\mathscr{EML}$
Hundera et al. [23]	$2\mathscr{Pr}$	–	$4\mathscr{Pr}$	$6\mathscr{Pr}$
Proposed	$3\,\mathscr{HEML}$	$1\,\mathscr{HEML}$	$1\,\mathscr{HEML}$	$5\,\mathscr{HEML}$

Table 5: Communication overhead comparisons in terms of extra parameters.

Schemes	Proxy delegation	Proxy signcryption	Total
Insafullah et al. [19]	$2\|q\| + \|mw\|$	$3\|q\| + \|mw\| + \|C\| + \|H\|$	$5\|q\| + \|mw\| + \|C\| + \|H\|$
Abdelfatah [20]	$2\|p\| + \|mw\|$	$1\|p\| + \|C\| + \|H\|$	$3\|p\| + \|mw\| + \|C\| + \|H\|$
Guo and Deng [21]	$2\|p\| + \|mw\|$	$5\|p\| + \|C\| + \|mw\|$	$7\|p\| + \|C\| + 2\|mw\|$
Waheed et al. [22]	$\|mw\| + \|H\|$	$3\|p\| + \|C\| + \|H\|$	$3\|p\| + \|mw\| + \|C\| + \|H\|$
Hundera et al. [23]	$2\|\mathscr{G}\| + \|mw\|$	$1\|\mathscr{G}\| + \|C\| + \|H\| + \|mw\|$	$3\|\mathscr{G}\| + \|C\| + \|H\| + 2\|mw\|$
Proposed	$2\|q\| + \|mw\|$	$1\|q\| + \|C\| + \|H\|$	$3\|q\| + \|mw\| + \|C\| + \|H\|$

Table 6: Communication overhead comparisons for 1 kb ciphertext and warrant size.

Schemes	Proxy delegation	Proxy signcryption	Total
Insafullah et al. [19]	1184	2800	3984
Abdelfatah [20]	1344	1696	3040
Guo and Deng [21]	1344	2848	4192
Waheed et al. [22]	1536	2016	3552
Hundera et al. [23]	2048	3072	5120
Proposed	1184	1616	2800

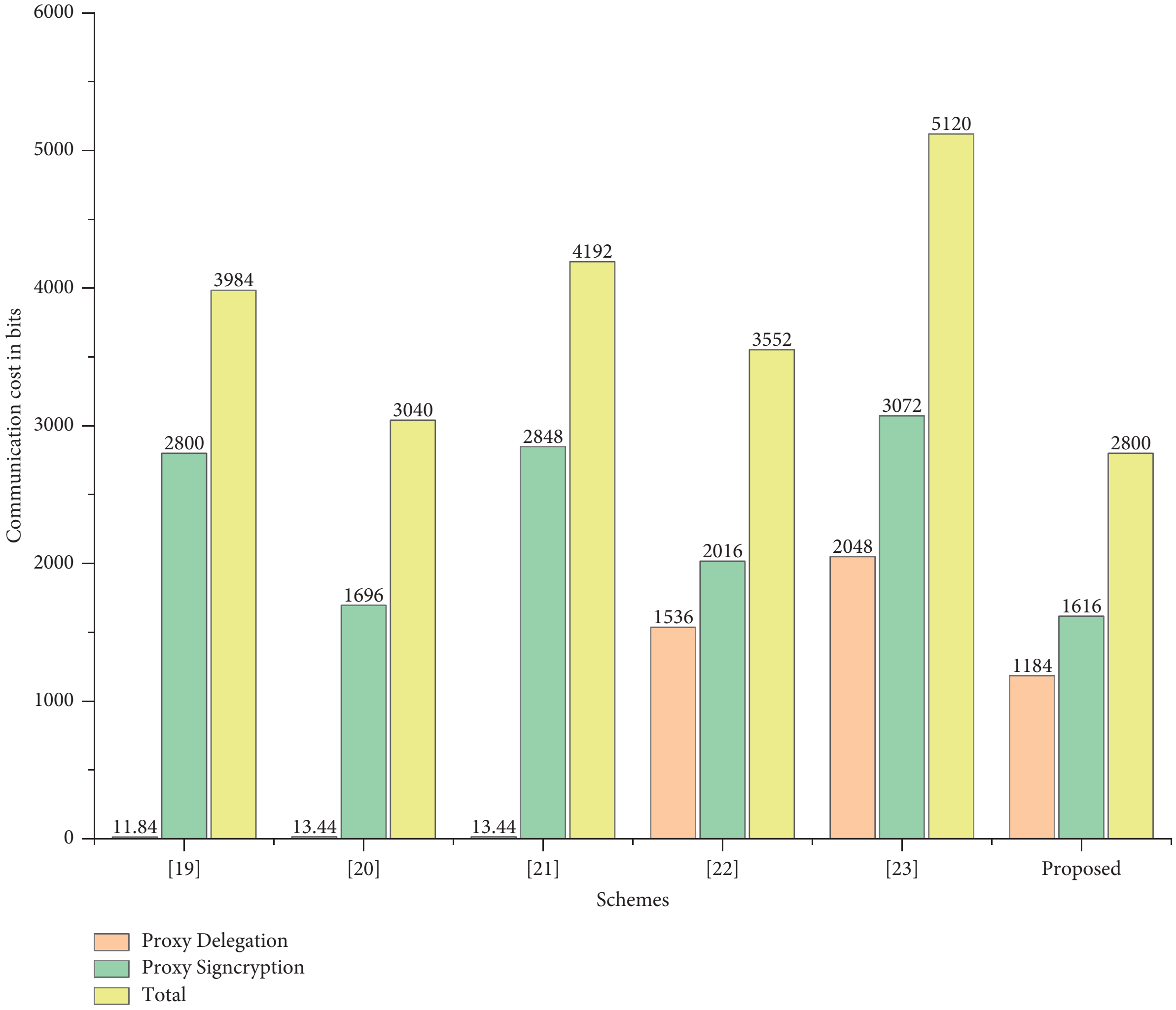


Figure 3: Communication cost comparisons.

previous schemes in terms of communication overhead. For generalization, we suppose that

(i) $|p|$ is a prime number $\geq 2^{160}$

(ii) $|q|$ is a prime number $\geq 2^{80}$

(iii) $|\mathscr{G}|$ where $\mathscr{G}$ be a group $\geq 2^{512}$

(iv) $|H|$ is a hash with 512 bits

Table 5 represents the communication cost of the designed and previous schemes; furthermore, Table 6 and Figure 3 show that when we consider 1 kb message or warrant, then our scheme is best from existing schemes.

7. Conclusion

In this article, we proposed a cost-effective proxy signcryption scheme for IoT devices. The proposed approach ensures the security properties such as unforgeability and confidentiality when it is tested through the ROM. The proposed scheme is lightweight due to the usage of HECC, which provides the same level of security with a lower-key size. A detailed security as well as performance analysis is conducted with the relevant existing schemes. The results demonstrate that the proposed scheme improves the overall computational cost and communication overhead, these being 5.7 ms and 2800 bits, respectively, which authenticates the superiority of our scheme from the existing schemes. Finally, we concluded that the proposed scheme could be of prime importance for the Internet of things devices.

In the future, we are intended to implement the same scheme on multimessage multireceiver environment using genus 3 of HECC.

Data Availability

The data generated or analyzed during this study are included within this article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] M. Mambo, K. Usuda, and E. Okamoto, "Proxy signature: delegation of the power to sign messages," *IEICE Transactions on Fundamentals*, vol. E79-A, no. 9, pp. 1338–1353, 1996.

[2] M. Mambo, K. Usuda, and E. Okamoto, "Proxy signatures for delegating signing operation," in *Proceedings of the 3rd ACM Conference on Computer and Communications Security*, pp. 48–57, New Delhi, India, March 1996.

[3] S. Kim, S. Park, D. Won, S. Park, and D. Won, "Proxy signatures, revisited," *Information and Communications Security*, vol. 1334, pp. 223–232, 1997.

[4] K. Zhang, "Threshold proxy signature schemes," in *Proceedings of the ISW'97, Information Security Workshop*, pp. 191–197, Ishikawa Japan, September 1997.

[5] C. Gamage, J. Leiwo, and Y. Zheng, "An efficient scheme for secure message transmission using proxy signcryption," Technical report 9801, Monash University, Melbourne, Australia, 1998.

[6] Z. A. Zhang, "New publicly verifiable proxy signcryption scheme," *Progress on Cryptography*, 2004.

[7] X. Li and K. Chen, "Identity based proxy-signcryption scheme from pairings," in *Proceedings of the 2004 IEEE International Conference on Services Computing (SCC'04)*, pp. 494–497, Washington, DC, USA, September 2004.

[8] M. Wang, H. Li, and Z. Liu, "Efficient identity based proxy-signcryption schemes with forward security and public verifiability," in *Proceedings of the 3rd International Conference on Networking and Mobile Computing (ICCNMC)*, pp. 982–991, Zhangjiajie, China, August 2005.

[9] S. Duan, Z. Cao, and Y. Zhou, "Secure delegation-bywarrant ID-based proxy signcryption scheme," in *Proceedings of the Computational Intelligence and Security Conference (CIS '05)*, pp. 445–450, Springer, Xian, China, December 2005.

[10] D. H. Elkamshoushy, A. K. AbouAlsou, and M. Madkour, "New proxy signcryption scheme with DSA verifier," in *Proceedings of the Twenty Third National Radio Science Conference (NRSC'2006)*, pp. 1–8, Al Minufiyah, Egypt, March 2006.

[11] H. Elkamshouchy, M. Nasr, and R. Ismail, "A new efficient strong proxy signcryption scheme based on a combination of hard problems," in *Proceedings of the International Conference on Systems, Man, and Cybernetics San Antonio (ICSMC'09)*, pp. 5123–5127, San Antonia, TX, USA, October 2009.

[12] H. Elkamchouchi and Y. A. Abouslseoud, "New proxy identity-based signcryption scheme for partial delegation of signing rights," *IACR Cryptology Eprint Archive*, vol. 41, 2008.

[13] H.-Y. Lin, T.-S. Wu, S.-K. Huang, and Y.-S. Yeh, "Efficient proxy signcryption scheme with provable CCA and CMA security," *Computers & Mathematics with Applications*, vol. 60, no. 7, pp. 1850–1858, 2010.

[14] H. M. Elkamchouchi, Y. Abouelseoud, and W. S. Shouaib, "A new proxy signcryption scheme using warrants," *International Journal of Intelligent Engineering Informatics*, vol. 1, no. 3/4, pp. 309–327, 2011.

[15] Q. Yanfeng, T. Chunming, L. Yu, X. Maozhi, and G. Baoan, "Certificateless proxy identity-based signcryption scheme without bilinear pairings," *China Communications*, vol. 10, no. 11, pp. 37–41, 2013.

[16] H. Elkamchouchi, E. Abu Elkhair, and Y. Abouelseoud, "An efficient proxy signcryption scheme based on the discrete logarithm problem," *International Journal of Information Technology, Modeling and Computing*, vol. 1, no. 2, pp. 7–19, 2013.

[17] N.-W. Lo and J.-L. Tsai, "A provably secure proxy signcryption scheme using bilinear pairings," *Journal of Applied Mathematics*, vol. 2014, Article ID 454393, 10 pages, 2014.

[18] Y. Ming and Y. Wang, "Proxy signcryption scheme in the standard model," *Security and Communication Networks*, vol. 8, no. 8, pp. 1431–1446, 2015.

[19] A. Insafullah, I. Haq, A. Amin, A. I. Umar, and H. Khattak, "Proxy signcrypion scheme based on hyper elliptic curves," *International Journal of Computer*, vol. 20, no. 1, pp. 157–166, 2016.

[20] R. I. A. Abdelfatah, "Novel proxy signcryption scheme and its elliptic curve variant," *International Journal of Computer Applications*, vol. 165, no. 2, pp. 36–43, 2017.

[21] H. Guo and L. Deng, "An identity based proxy signcryption scheme without pairings," *International Journal of Network Security*, vol. 22, no. 4, pp. 561–568, 2020.

[22] A. Waheed, A. I. Umar, M. Zareei et al., "Cryptanalysis and improvement of a proxy signcryption scheme in the standard computational model," *IEEE Access*, vol. 8, pp. 131188–131201, 2020.

[23] N. W. Hundera, Q. Mei, H. Xiong, and D. M. Geressu, "A secure and efficient identity-based proxy signcryption in cloud data sharing," *KSII Transactions on Internet and Information Systems*, vol. 14, no. 1, pp. 455–472, 2020.

[24] M. Bellare and G. Neven, "Multi-signatures in the plain public key model and a general forking lemma," in *Proceedings of the 13th ACM Conference on Computer and Communications Security (CCS'06)*, pp. 390–399, October 2006.

[25] B. Bera, S. Saha, A. K. Das, N. Kumar, P. Lorenz, and M. Alazab, "Blockchain-envisioned secure data delivery and collection scheme for 5G-based IoT-enabled internet of drones environment," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 8, pp. 9097–9111, 2020.

[26] M. A. Khan, I. Ullah, S. Nisar et al., "Multiaccess edge computing empowered flying ad hoc networks with secure deployment using identity-based generalized signcryption," *Mobile Information Systems*, vol. 2020, Article ID 8861947, 15 pages, 2020.

[27] M. A. Khan, "Efficient and provably secure certificateless key-encapsulated signcryption scheme for flying ad-hoc network," *IEEE Access*, vol. 8, pp. 36807–36828, 2020.

[28] M. A. Khan, "An efficient and secure certificate-based access control and key agreement scheme for flying ad hoc networks," *IEEE Transactions on Vehicular Technology*, vol. 99, 2021.

[29] M. A. Khan, I. Ullah, S. Nisar et al., "Multiaccess edge computing empowered flying ad hoc networks with secure deployment using identity-based generalized signcryption," *Mobile Information Systems*, vol. 2020, Article ID 8861947, 15 pages, 2020.

[30] M. A. Khan, "Securing internet of drones with identity-based proxy signcryption," *IEEE Access*, vol. 9, pp. 89133–89142, 2021.

Hindawi
Mobile Information Systems
Volume 2021, Article ID 5972486, 12 pages
https://doi.org/10.1155/2021/5972486

Research Article

Towards Region Queries with Strong Location Privacy in Mobile Network

Songtao Yang[1] **and Qingfeng Jiang**[2]

[1]*College of Computer Science and Information Engineering, Bengbu University, Bengbu 233030, China*
[2]*College of Computer Science and Engineering, Changshu Institute of Technology, Changshu 225500, China*

Correspondence should be addressed to Songtao Yang; yst@bbc.edu.cn

Received 26 April 2021; Revised 26 August 2021; Accepted 28 October 2021; Published 18 November 2021

Academic Editor: Daniel G. Reina

With the interaction of geographic data and social data, the inference attack has been mounting up, calling for new technologies for privacy protection. Although there are many tangible contributions of spatial-temporal cloaking technologies, traditional technologies are not enough to resist privacy intrusion. Malicious attackers still steal user-sensitive information by analyzing the relationship between location and query semantics. Reacting to many interesting issues, oblivious transfer (OT) protocols are introduced to guarantee location privacy. To our knowledge, OT is a cryptographic primitive between two parties and can be used as a building block for any arbitrary multiparty computation protocol. Armed with previous privacy-preserving technologies, for example, OT, in this work, we first develop a novel region queries framework that can provide robust privacy for location-dependent queries. We then design an OT-assist privacy-aware protocol (or OTPA) for location-based service with rigorous security analysis. In short, the common query of the client in our solution can be divided into two parts, the region query R_q and the content query C_q, to achieve location k-anonymity, location m-diversity, and query r-diversity, which ensure the privacy of two parties (i.e., client and server). Lastly, we instantiate our OTPA protocol, and experiments show that the proposed OTPA protocol is reasonable and effective.

1. Introduction

Location-based services (LBS) are one of the successful mobile applications in our daily life. Armed with the help of LBS, it will be easy for you whether you want to let others track your movements, find or get to somewhere, or simply know your current location and what is around you. Obliviously, LBS along with its corresponding applications greatly improve the public living style in terms of richness and diversity. However, problems of location privacy disclosure have not raised the concern of the public. For instance, some malicious attackers will track the trace of the location using the LBS. Attackers can monitor and identify goal-oriented people, but the goal-oriented people could not be aware of being tracked [1, 2]. In this case, researchers were beginning to engage in how to conceal location and identity of users.

In reality, the user can submit the points of interest (POIs) queries (e.g., "find the nearest mall") to the LBS provider like Google Map. To conceal location and identity, users could mask the query via an anonymity tool, such as k-anonymity and obfuscation. But the attacker can deduce the user's identity from the content of query, background knowledge, and the observation information if we just adopt a simple pseudonym to cloak the location and the identity [3]. To overcome these limitations, these proposed research schemes can be divided into three major types: (*a*) location k-anonymity, (*b*) location obfuscation, and (*c*) private information retrieval (PIR). However, these existing techniques cannot efficiently address the following two major problems in more detail: (*a*) most of the existing proposals assume that all anonymous participants are completely reliable. In contrast, the participants stay at the same level of security. Apparently, this assumption is unrealistic and inconsistent. It is often questionable with the actual scenario. Collaborator may be disclosing the accurate location information or the accurate queries information, either directly or indirectly. (*b*) In fact, intermediary servers or query issuers obtain a large amount of redundant data during perquery. However, these data are employed in charging customers according to actual use, whether directly or indirectly, and they are valuable assets of the LBS server.

Consider an application scenario shown in Figure 1. Alice wants to get a discount list of this mall located in a certain area or obtain what movie will be released in the nearby cinema. Although there are a lot of POIs around her, she is only concerned with a certain category of these POIs information. For example, she issues a service request, "find the discount price of the mall which is away from my current location about 5 km", to trade with the LBS provider. According to LBS service mode, the NN of Alice is P6, where the set P1, P3, P4, P6, P8 represents some malls.

To our knowledge, previous schemes are not conducive to the embodiment of the commercial value of the LBS information. Hence, we will ask the following question: is it possible to address the two above-mentioned problems using OT-assist privacy-aware protocol? To answer this question, in this paper, we first develop a novel region query framework that supports the private location-dependent query. We then design an OT-assist privacy-aware protocol (or OTPA) for location-based service with rigorous security analysis.

The contributions can be summarized as follows:

(1) A novel region queries framework: We first developed a novel region queries framework that supports private location-dependent queries. Our framework achieves noncooperative privacy preserving via cryptographic techniques, and it does not require a trusted third party. We proposed a new fair exchanging pattern with semitrusted three parties, which includes an intersection with three subjects: users, location cloaking server, and LBS server. Assume that all the participants are semihonest in this architecture; they will honestly follow the protocol but they are curious to find out as much as information from the data that it receives and stores.

(2) An OT-assist privacy-aware protocol: We designed a privacy-aware query protocol, which guarantees the untraceability of user trajectory and unlinkability of the content. A common query is split into a region query and a content query in our solution. Further, we analyzed the user's privacy through theory analysis and demonstrated the effectiveness by experiments.

1.1. Roadmap. The rest of this paper is organized as follows. We reviewed the related work in Section 2. In Section 3, we presented some definitions and gave some terms. In Section 4, we introduced the region queries and designed a system model and expression. The proposed privacy-aware region queries and OTPA protocol are presented in Section 5, followed by the security analysis in Section 6 and the experiment evaluation in Section 7. Finally, we concluded the paper in Section 8.

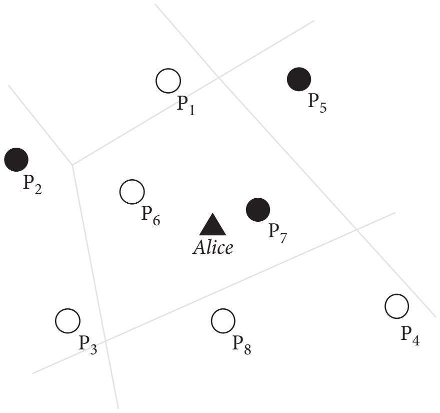


Figure 1: Application scenario of LBS.

2. Related Works

In numerous studies, the location k-anonymity [4–6] is always the predominant approach. The essence of location k-anonymity is that the probability of identifying the query user cannot exceed $1/k$, which is mainly focused on query privacy. Instead of sending the query to the LBS server, the user interacts with the anonymizer, which cuts off the association between user's identities and query contents to prevent the attacker from analyzing the user's sensitive information. However, k is not a representative of the actual location privacy of mobile users. In fact, these cloaking techniques based on the location k-anonymity metric could even be counterproductive and give the illusion of a higher location privacy level. Shokri et al. argued that the k-anonymity scheme is insufficient for protecting location privacy [7]. For example, if k users within an anonymous spatial region (ASR) are located in the same semantic location, the ASR guarantees the requester's query privacy but discloses their location privacy. On the other hand, if k users have similar query content and distribute in different locations, the ASR guarantees users location privacy and exposes the user's query privacy. Therefore, some researchers develop further studies for location diversity and context-aware and location semantics [8]. A complementary technique to the location k-anonymity is the location obfuscation technique. These location obfuscation techniques are achieved by deliberately reducing the resolution of the user's location to protect user privacy, namely, using a cloaking region instead of the user's actual location. To release ambiguous location is often used as a simple and effective technique [9–11]. Space Twist framework avoids the high computational cost and communication cost caused by the ASR. However, a lower resolution of location may cause coarse-grained service provided by the LBS server. The size of cloaking region is proportional to degree of privacy and is inversely proportional to the quality of service. Therefore, the adversary can deduce the approximate location of the user according to the context of background environment, which means leading weak privacy [12]. Collusion of LBS leads to complexity of privacy preserving in real-world applications. The correlation between geographic data and social data leads to losing effectiveness for spatial-temporal anonymity technology. As a cryptography-based oblivious transfer method, private information retrieval (PIR) was also adopted to secure the location privacy [13, 14]. OT and PIR are similar: cryptographic protection against information disclosure. The methods which employed PIR protocol or OT protocol provide provable privacy guarantees against correlation attacks and eliminate the requirement for any trusted third party. Computational PIR-based approach utilizes a PIR protocol to implement a simple query pattern, which retrieves a specific database block from the LBS server without discovering which block is retrieved. However, it leads to a prohibitive computational cost and communication cost even for a small POIs databases. Therefore, secure hardware-aided PIR proven efficient is currently considered as a practical mechanism for PIR. Some cryptographic technologies (such as attribute based encryption [15–17] and data integrity checking [18–20]) have potential application in location privacy, which not only guarantees secure data share but also ensures remote data integrity [21].

3. Preliminaries

In this section, we present some definitions for follow-up work, including the framework, privacy-aware protocol, and privacy-aware queries in LBS.

Definition 1 (P). A point of interest (POIs) is a landmark or specific location that someone may feel useful or be interested in, such as a hotel, hospital, and school. It can be formalized into a triple set: $P = \langle l, c, i \rangle$. Here, we denote the $i -$ th POIs as P_i, where, $P_i[l]$ is location coordinates of a P_i, $P_i[c]$ is category of a P_i, $P_i[i]$ is service content of a P_i.

Definition 2 (R_q). Region query can be formalized as follows: $R_q = \langle R, k, m \rangle$. Here, R represents a geographic region illustrated by the query submitter. The k is the user-desired degree of anonymity. m is the user-desired number of different semantic locations within R.

Definition 3 (C_q). Content query can be formalized as follows: $C_q = \langle R', C' \rangle$. Here, R' represents the minimum area meeting users' privacy. C' is a subset of C. It is selected by the user. C is a comprehensive POIs taxonomy set.

Definition 4 ($H(l_i)$). Given an anonymous spatial region, a set of m location points $L = \{l_1, \ldots, l_m\}$. For any location in an anonymous spatial region, the Location Entropy is denoted as $H(l_i) = -\sum_{i=1}^{m} p_i^{(l_i)} \log_2^{p_i^{(l_i)}}$. Here, $p_i^{(l_i)}$ is the probability of user u_i locating in l_i.

Definition 5 ($H(q_i)$). Given an anonymous set, a set of n users $U = \{u_1, \ldots, u_n\}$. For any use in an anonymous set, the Query Entropy is denoted as $H(q_i) = -\sum_{i=1}^{m} p_i^{(q_i)} \log_2^{p_i^{(q_i)}}$. Here, $p_i^{(q_i)}$ is the probability of user u_i issuing query q_i.

Definition 6 (POIIR). The POIIR is the abbreviation of "POIs influence range." Let $P = \{p_1, \ldots, p_n\}$ indicate a set of POIs that possess identical datatype in the LBS database. Thus,

$$\mathrm{POIIR}(p_i) = \{p \mid \mathrm{dist}(p, p_i) \le \mathrm{dist}(p, p_j), i \ne j\}, \tag{1}$$

where p is an arbitrary point in the service range.

For ease of description, we define some terminology about location privacy. The definition of notations in our work is shown in Table 1.

4. Region Queries Framework

4.1. Region Queries. We map the experimental area onto a grid G composed of cells. Each cell corresponds to a Hilbert value, covers an $\alpha \times \alpha$ square area, where α indicates the parameter that defines the cell size of the grid G. Users regularly upload location information to a location cloaking server. The current cell of a user contains the current position of the moving object.

In our solutions, the objective for R_q is to find some Minimum Cloak Regions (MCRs). All these MCRs meet the requirement of user privacy. Similar to the Hilbert Cloak, given a query from the mobile client (MC) with anonymity requirement k, Location Cloaking Server (LCS) ranks the Hilbert Values and splits them into k-buckets. The LCS calculates the start and end positions defining the k-bucket that includes requester and constructs k-ASR using all users in the same bucket. The difference is that our solutions meet the requirements of the location m-diversity, while building k-ASR for each user.

For example, as shown in Figure 2, suppose u_2 issues the query "$R_q = \langle (0,0), (7,7), 4, 3 \rangle$." We can easily calculate that one of k-ASR is $\{(0,0), (3,3)\}$. Moreover, k-ASR offered by LCS is not unique, which may be $\{(0,5), (5,7)\}$. What it is designed to do is to be against inference attacks of LCS. The MC chooses a correct k-ASR that contains its real coordinates as the basis for the C_q.

4.2. System Model and Expression. In a LBS system, a large number of mobile users move within a two-dimensional square unit space. Users can issue location-dependent queries, answered by LBS providers. We adopt the three-tier centralized architecture consisting of three key parts: mobile user, location cloaking server, and LBS server.

Mobile clients (MC) are equipped with a positioning device, for example, GPS or sensor-based local positioning systems, to determine its current location information l. All of the users who held MC in our model enjoy location-dependent service by the LBS server. This device is trusted, and no malicious software component running on the mobile device has access to the location sensor. That can be assured by using a trusted computational approach.

LBS servers (LBSS) are the service providers of the LBS system. These LBSS are nontrusted since an attacker is aware of all the information that users provided to the LBS server and compromise user privacy. In addition, we assume that the attacker has statistical background information about the

Table 1: Notations.

Symbol	Description
l	User location
R	Query region presented by user
R'	A minimum cloaking region selected by MC
k	The degree of k-anonymity
m	Location m-diversity
r	Query r-diversity
C	The set of POIs category
C'	A subset of C
S_n	Candidate set
P_r	Coordinate set of POIs picked by user
M_i	Service information of POIs
K_i	Encryption key

users, although in practice, it is difficult to model the exact knowledge.

Location cloaking servers (LCS) are also the semitrusted party placed between MC and LBSS. All registered mobile users periodically update their location information to the LCS. These LCS construct MCRs, which meets users' requirements of location k-anonymity and location m-diversity.

Users establish a secure connection (e.g., an SSL) with LCS, hiding the query issuer's identity and IP address. As a hypothesis for our model, we further consider that the anonymity algorithm used by LCS is public. We support that the distribution of the population in the geographical space is uneven to conform with laws of nature.

The general procedure of continuous region query processing and specification processing is shown in Figure 3.

(1) A user sends a query R_q that contains the user's privacy requirement to LCS

(2) LCS executes MCRs Finding Algorithm to form MCRs and initiates C_q to LBSS

(3) LBSS retrieve the spatial database and interact with LCS

(4) LCS minifies the candidate set before sending the results to the user

4.3. MCRs Finding Algorithm. The LCS executes MCRs Finding Algorithm to calculate MCRs. We use the notation $G = \{g_1, \ldots g_n\}$ to denote Hilbert curve space; $g_i\,(1 \le i \le n)$ is some of the cells. P represents the criterion of judgment and $P(g_i) = \text{TRUE}$ means that g_i only consists of a cell. The MCRs Finding Algorithm consists of three phases.

Firstly, as shown in Figure 4, region segmentation starts from a set of seed points. An alternative is to start with a single region ($R_q[R]$) and subdivide the regions that do not satisfy a condition of $P(g_i)$. In other words, split into four disjoint quadrants any region $P(g_i)$ for which $P(g_i) = \text{FALSE}$. Secondly, region merging is the opposite of region splitting. It starts with small regions and merges the regions that have similar characteristics. The aim of merging any adjacent region g_j and g_k is to find MCRs. Thirdly, we adopt an R-tree to index G. The process of constructing a G is iterative. The processing is repeated until all of MCRs satisfying privacy requirements ($R_q[k]$ and $R_q[m]$) are found. The LCS randomly selects some of MCRs and sends them to the MC.

5. Privacy-Aware Region Queries

5.1. Motivation. Our framework focuses on continuous region query that is distinct from previous studies of single-point top k-nearest-neighbor query. Consider an application scenario shown in Figure 5; the same icons represent that these POIs belong to the same classification; and A, B, and C represent different mobile users, respectively.

The common region queries are classified into three categories: (1) A uses its location as the center of region queries; (2) B uses one certain POI as the center of region queries, but B is not in the particular area; and (3) C uses one certain POI as the center of region queries, but C is in the particular area. Suppose that a user named Alice is moving in a bidimensional road network.

The above description faces two problems. Firstly, users desire to experience both high-quality service and not to expose location and identity. Therefore, users are more concerned about privacy issues. Secondly, the LBSS do not want to publish more information about POIs, which means the LBSS also express concern about the quality of service issues and business profits. From the privacy perspective, both LBSS and MC are attackers. In addition, the IP address issue is orthogonal to our problem. It can be achieved through a widely available anonymous web browsing service.

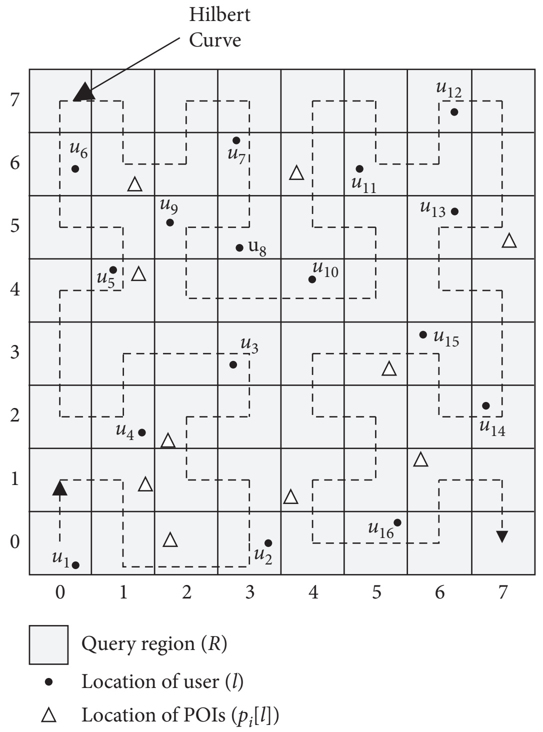
5.2. OT-Assist Privacy-Aware Protocol. Oblivious transfer protocol normally runs as a building block for more complex secure protocols or as a stand-alone protocol for privacy-preserving in LBS. Efficient 1-out-of-r oblivious transfer schemes (OT_r^1) rely on the hardness of the decisional Diffie–Hellman problem to achieve unconditional security. Assume an order-q group G_q with a short description, where q is a large prime number. Let g and h be two generators of G_q. Parameters g, h, q, G_q are publicly accessed by every entity in our protocol, where senders and receivers refer to MC and LBSS, respectively. LBSS have r keys $K_1, \ldots K_r$. The MC knows one of the key $K_a\,(1 \le a \le r)$ is his/her own choice and does not want LBSS to have that data. Meanwhile, the LBSS only provide K_a for the MC but do not want MC to get more information. The implementation process of OT-assist privacy-aware protocol is shown as follows:

(1) MC chooses $a\,(1 \le a \le r)$, generates a random number d, calculates $y = g^d h^a \mod q$, and sends y to LBSS.

(2) LBSS calculate two tuples of sequence D and send D to MC. Here, $D = \{(s_1, t_1), \ldots, (s_r, t_r)\}$, $s_i = g^{k_i} \text{mod} q$, $t_i = K_i\,(y/h^i)^{k_i} \text{mod} q$, $k_i \in Z_q\,(1 \le i \le r)$.

(3) LBSS send D to MC.

(4) MC calculates $K_a = (t_a/(s_a)^d) \text{mod} q$.



User	Hilbert value	Grid no.
u_1	1	0,0
u_2	2	0,3
u_3	3	3,3
u_4	4	2,1
u_5	5	4,1
u_6	6	6,0
u_7	7	6,3
u_8	8	5,3
u_9	9	5,2
u_{10}	10	4,4
u_{11}	11	6,5
u_{12}	12	7,6
u_{13}	13	5,6
u_{14}	14	2,7
u_{15}	15	3,6
u_{16}	16	0,5

Figure 2: Hilbert curve map for a 2D space with 8 × 8 space partition.

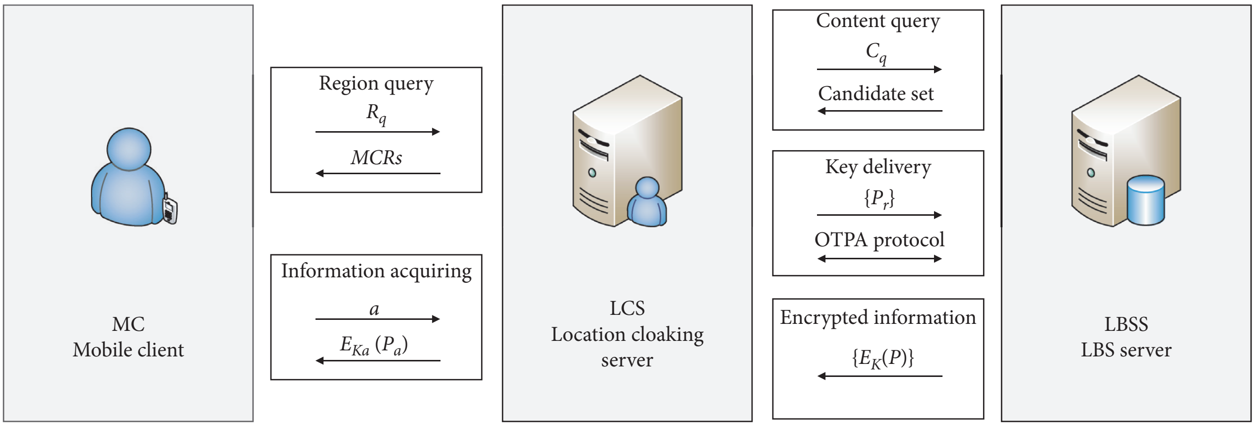


Figure 3: The system model that contains three entities: MC, LCS, and LBSS.

The purpose of the OTPA protocol is to obtain one and only one key from LBSS. This scheme meets the following privacy requirements. For any a, there is d that satisfies $y = g^d h^a \bmod q$. Therefore, LBSS cannot get any information related to a, even if it has unlimited computing power. When MC and LBSS gradually follow the protocol, although MC receives LBSS's secrets $K_1, \ldots K_r$ and cannot get two secrets, there is no way of getting information other than K_a.

5.3. Bidirectional Security Processing. Assume that the LBSS have r POIs information $P = (p_1, \ldots, p_r)$ and randomly generate the r key $K = (k_1, \ldots, k_r)$. Query senders desire p_a, but they do not wish the LBSS to know what they will get. Moreover, the LBSS also employ k_a to prevent users from accessing unauthorized content. We define this query process as bidirectional security processing. We implement our solutions with secure multiparty computation theories. It is reasonable to make an assumption about which the LCS does not collude with the LBSS since the LCS stores query examples of the MC. Otherwise, it will completely subvert any method for location privacy preserving if the LCS is allowed to collude with LBSS. We consider that all the participants in a query session are semihonest. The MC and the LCS try to obtain more data than authorized. The LBSS tries to associate a user with a location or some POIs. More details of bidirectional security processing are depicted as follows, as shown in Figure 6:

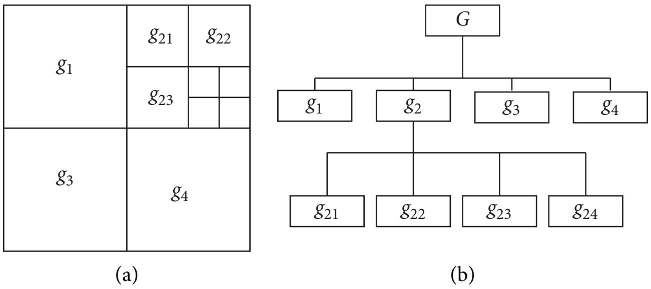


FIGURE 4: An example of MCRs Finding Algorithm. (a) Splitted region. (b) R-tree.

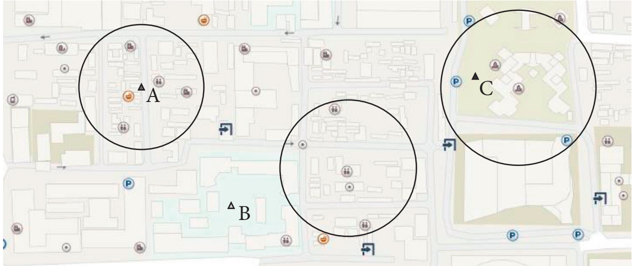


FIGURE 5: An example of a realistic road network environment.

(1) The MC submits a region query R_q to the LCS to find some MCRs.

(2) The LCS responds to the request of the user according to the privacy requirements of the user, executes MCRs Finding Algorithm, and sends some of MCRs to the MC.

(3) The MC randomly selects MCRs as $C_q[R']$ and submits a content query C_q to the LBSS for obtaining POIs candidate set S_n. $C_q[C']$ contains actual POIs category (c_i) about this query.

(4) The LBSS calculate all candidate POIs of $C_q[R']$' and send candidate set S_n to MC. S_n is formulated as the following form:

$$S_n = \left\{p_i[l] | p_i[l] \subset C_q[R'] \text{ and } p_i[c] \in C_q[C']\right\}. \quad (2)$$

Further, we can also express S_n as $S_n = \left\{S^{(1)}, \ldots, S^{(r)}\right\}$, where $S^{(j)} (1 \le j \le r)$ is a location set retrieved by c_j.

(5) MC calculates the obstacle distance between its current coordinate and each element of $S^{(i)}$ and adds the nearest point to the set P_r. MC randomly also extracts an element from $S^{(j)} (1 \le j \le r, i \ne j)$ and adds it to the set P_r. The MC disrupts the order of P_r and sends it to the LBSS.

(6) The LBSS retrieve the spatial database and find all of POIs information in terms of P_r. It is referred to as M.

(7) The MC and the LBSS perform OTPA protocol.

(8) The LBSS can encrypt $M = \left\{E_{K_1}(p_1), \ldots, E_{K_r}(p_r)\right\}$ to prevent LCS from reading it and send it back to the LCS.

(9) The MC retrieves a particular record for $E_{K_a}(p_a)$, which is precisely what the user needs.

6. Security Analysis

Data security and user's privacy have the absolute critical priority for a LBS system. There is much more risk of sensitive data being stolen or leaked because LBSS gather mass data from social media users. In this section, firstly, we explain the privacy threats caused by location and measurement of the privacy leakage. Moreover, we compare the proposed solution with existing works in terms of location k-anonymity, location m-diversity, and query r-diversity.

6.1. Attack Expression and Privacy Metric. Location privacy is the nature of an individual to control access to their current and past location information. Figure 7 shows the importance of location. There are four key factors affecting personal privacy in LBS system: identity, location, time stamp, and candidate POIs. As long as it is not associated with the particular user's identity, query context does not lead to privacy disclosure. However, the user's trajectory is the key link in query context and user's identity. For example, continuous location samples have been tracked by attacks and then used to infer a user's identity. The relationship feature between trajectory and POIs can also be used to define a user's behavior. The combination of identity and behavior exposed the sensitive data of the user.

All research related to location privacy stems from the assumption that untrusted LBS providers are the most critical threat to privacy. The LBS attacks involve two aspects: location tracing and user identification. Meanwhile, the prior knowledge of the attacker is unable to measure, and

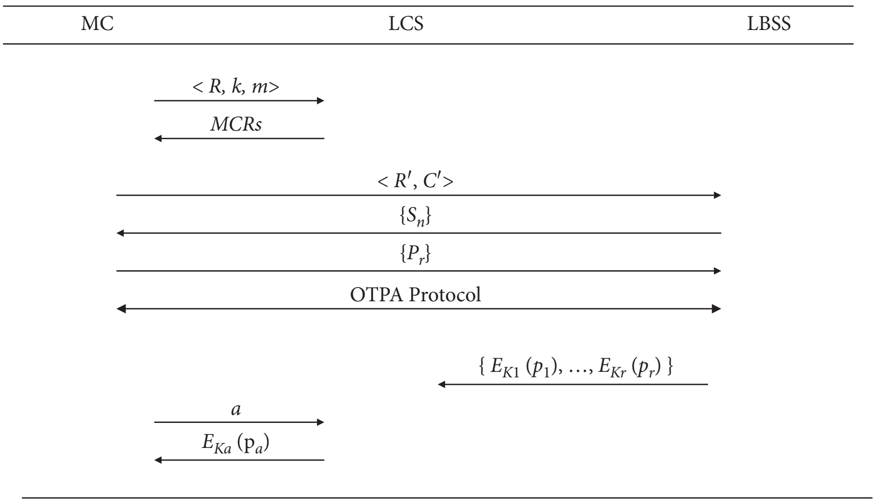


Figure 6: The interactions between MC, LCS, and LBSS.

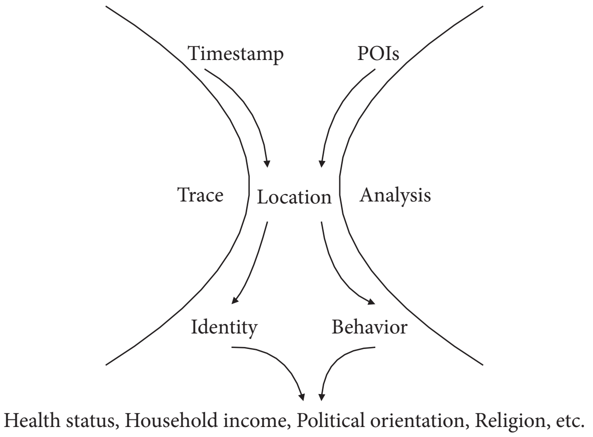

Timestamp
POIs
Trace
Location
Analysis
Identity
Behavior
Health status, Household income, Political orientation, Religion, etc.

Figure 7: Location is a key factor in the LBS system.

the invade mode taken by the attacker is unpredictable. As the diversity of profiles, such as user profile or user velocity, are not the same, the spatial cloaking faces continuous multiquery attacks, inference attacks, and correlation attacks.

Theorem 1. *The combination of identity, location, timestamp, and candidate POIs poses a serious threat to user privacy, and location plays a significant role in the LBS system.*

The concept of entropy was rooted in Shannon entropy. It gives an accurate metric of the uncertainty that an attacker infers for the user's information. Shannon entropy also can be used to evaluate location privacy or query privacy. Before a user submits a query, the uncertainty over location obtained by LBSS has been called Priori Location Entropy. However, we can improve the degree of privacy using some techniques such as anonymity, fuzzy, and obfuscation. The uncertainty over location obtained by LBSS has been called Posterior Location Entropy after applying these techniques. The inherent feature of Location Entropy is mainly embodied in the following aspects. Firstly, when the LBSS have real-time location information of users, the Priori Location Entropy $H(l_i) = 0$. *Secondly, when the LBSS do not have any background knowledge, the maximum Priori Location Entropy* $H(l_i) = \log_2^m$. *Thirdly, when the LBSS have some background knowledge which is achieved through statistical analysis, the Priori Location Entropy* $0 < H(l_i) < \log_2^m$.

We can easily recognize that higher entropy is associated with three things: location k-anonymity degree, location m-diversity, and query r-diversity. In our solutions, the probabilities of location anonymity, location diversity, and query diversity are $1/k$, $1/m$, *and* $1/r$, *respectively. Users can freely control their privacy requirements because all of these parameters are determined by themselves. Therefore, our approaches achieved the purpose of hiding user privacy. Obviously, it is inevitable that each query provides some new knowledge for LBSS, which is more conducive to inferring the user's sensitive data. However, our solutions improved the complexity of the invasion of privacy, although they do not overcome the inherent limitations of spatial and temporal cloaking methods. We will be establishing a privacy measure model in subsequent studies.*

6.2. Comparison of OTPA, Spatial Cloaking, and PIR. Because it is required to submit a k-ASR to LBSS in spatial cloaking methods, the user issuing queries must be appearing in the area. The anonymous area has been gradually diminished by attackers according to user profiles, road network restrictions, and moving speed. Thus, the user's trajectory is traceable. The quality of trajectory details relies heavily on the power of an attacker. At the same time, the candidate result set is a vital component for LBS providers to infer the user's sensitive data. There is a direct correlation between queries content and user identity. An attacker can deduce who is most likely to issue the query. In our solutions, the region submitted to LBSS satisfies four properties: location k-anonymity, location m-diversity, query r-diversity, and reciprocal relationship of ASR. Therefore, our solutions can resist the inference attack for spatial cloaking. Firstly, the user submitting queries does not reveal the accurate location to LCS and LBSS since the calculation program of the nearest neighbor runs on the client device. However, LBSS can calculate the minimum inference region, which is the intersection of the R' and all of disclosed POIs influence regions (POIIR). Consequently, larger value of R' means higher location privacy for the user. The POIIR of disclosed POIs is discrete and random, which makes it difficult to trace the sequence of trajectories. Moreover, the user submitting queries confuses the query content with a plurality of POIs that are selected by themselves. Therefore, the probability of LBSS inference user query content is $1/r$. Consequently, LBSS cannot associate the user with the identity by specific POIs.

Theorem 2. *Assume that all of these attributes of location k-anonymity, location l-diversity, query r-diversity, and reciprocal relationship of ASR can guarantee privacy, which makes our solution have the untraceability and the unlinkability.*

OTPA is parallel to PIR. Both of them are based on encryption techniques to protect user privacy. Computational PIR relies on the quadratic residuosity problem. However, it cannot avoid a linear scan of the entire database for processing each query. The communication complexity of each query is roughly $\sqrt{n}$. The symbol n represents the size of the database. Therefore, the PIR techniques require extreme computational efficiency, where the usage of resources, such as run-time, storage, or data samples, is sublinear in the size of the candidate module. In contrast, OTPA does not have such requirements. Our solutions are superior to PIR techniques because the typical PIR framework does not limit the number of POIs obtained by the user. Thus, it does not provide an effective way to protect the valuable resources of LBS server.

Theorem 3. *Assuming that the OTPA scheme is unconditionally secure, our solution achieves server-oriented security. It can be hard to maliciously get precious data of the LBS server.*

7. Experiment Results and Discussion

We implement a prototype system by extending an existing work of C# program that supports OT protocol. The database is one of the widest and most interesting public data sets to analyze user trajectory which is generated by Brinkhoff's network-based generator of moving objects. We conduct the experiments on a machine with Intel(R) Core(TM) i7-10510U CPU and 40 GB memory and some smartphones with Android 10 OS as the client. Our experimental default parameters are summarized in Table 2. We simulate 1000 users sending queries randomly to the LBS provider through a wireless network. Default values for these parameters constrain the scope of the following experiments; see Table 1 for specific meanings.

In the following experiments, we mainly focus on the communication cost and the computational cost, which is the dominating factor for the proposed solutions. In OT protocol, the cost of computation is often criticized with the comparison of communication cost. OT protocol is characteristically implemented using modular exponentiations, which are involved in the intensive computing. Therefore, researchers are more concerned about the effectiveness and availability of these algorithms in cryptographic applications.

Table 2: Parameters and default values.

Parameters	Default values
R	1 km^2
k	50
m	20
r	10
C	50
M_i	10 kbit
K_i	64 bit

The first experiment aims at studying the time consumption with different numbers of candidate POIs. The efficiency of our approaches depends on parameters R and R'. Without loss of generality, we assume that the number of candidate POIs is directly proportional to the size of R'. The time consumption in two query phases is shown in Figure 8. The result shows that the CPU time of content query is large since the number of modular exponentiation is proportional to the number of candidate POIs.

As shown in Figures 9–11, the CPU time is influenced by these parameters (R, R', k, m, and r) in the region query and content query. We can find that more stringent privacy requirements take longer time.

Figure 12 shows the result of the comparison with the typical method Casper and PIR. Experimental results indicate that the average processing time of the above three methods is linear to the number of candidate POIs. From computation efficiency, modular exponentiation is the most expensive. Therefore, Casper performs better than the other two methods in the average computation time.

The second experiment focuses on studying the communication cost in the two-query phase. Figure 13 shows that the communication cost in the region query is lower since the main communications are composed of some coordinates of POIs transferred from server to clients. The communication cost of the content query will just keep growing. However, its upper limit is around 550 kb since the category of POIs is no more than 50. The R and α affected communication cost in region query stage, and R and α are larger, which makes the traffic greater. R' and C' affected the communication cost in the content query stage. The larger the R' and C', the greater the traffic loads. At the same time, k and m have decided the area of R', and r have limited the dimensions of C'. Therefore, the higher the user's privacy requirement, the greater the traffic loads.

Finally, we observe the number of POIs that users obtain from each query since users are often charged by the LBS provider according to the number of retrieved POIs. We conduct experiments to compare with other techniques. Figure 14 shows that the number of candidate POIs is linear to the number of users. The difference is due to the diversity of the querying methods. These results indicate that, in order to maintain an appropriate number of disclosed POIs,

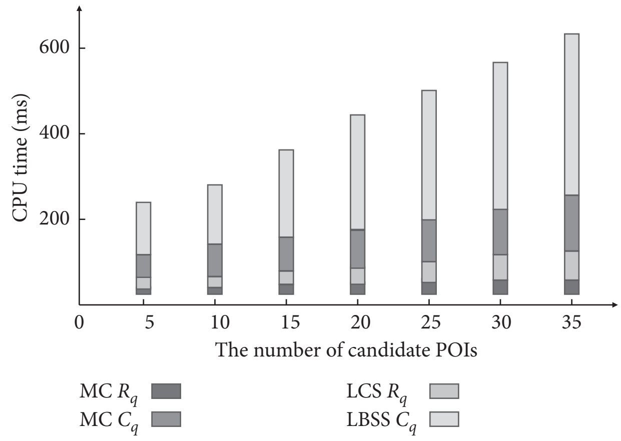


Figure 8: The time consumption in two query phases.

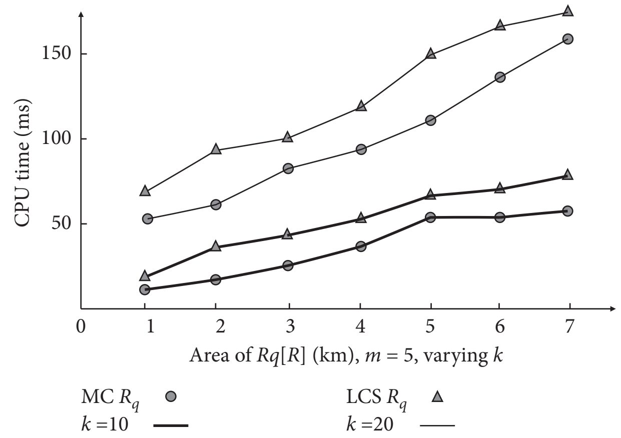


Figure 9: The R and k influence on CPU time.

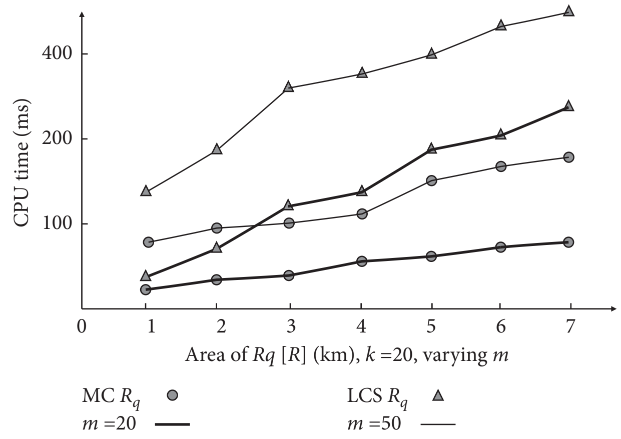


Figure 10: The R and m influence on CPU time.

cloaking-based methods have to collect a large number of users. These result in a high cost of location updates and pose privacy concerns since all users must be trustworthy. The number of disclosed POIs is constant for PIR methods because no other users are required to construct a cloaking set. The number of candidate POIs gradually decreases from 50 to 1 as the user number increases in our solutions. However, only one candidate POI is exposed to the user submitting query. Therefore, we provide security guarantees for the resources of the LBS server.

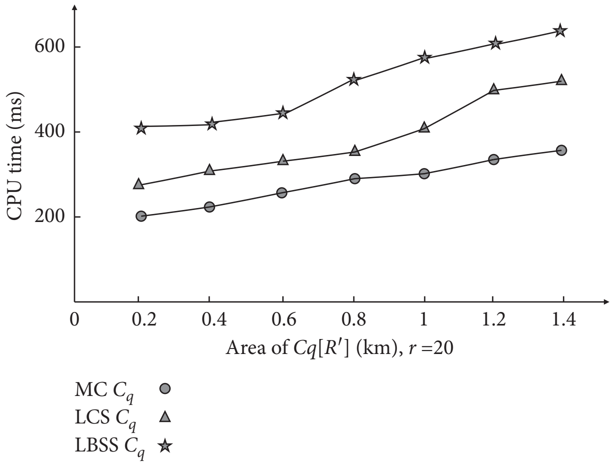


Figure 11: The R' and r influence on CPU time.

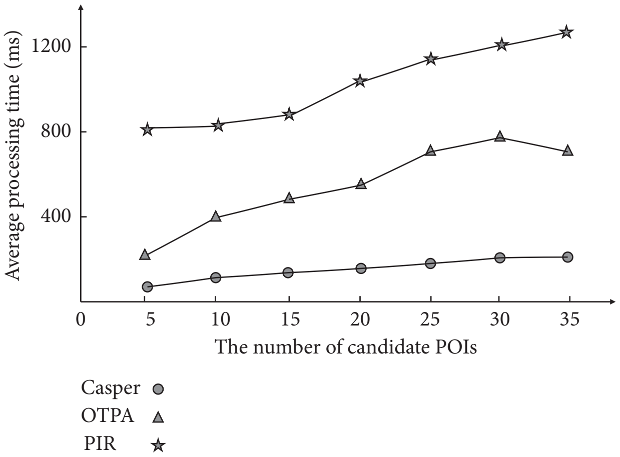


Figure 12: The comparison with the Casper and PIR.

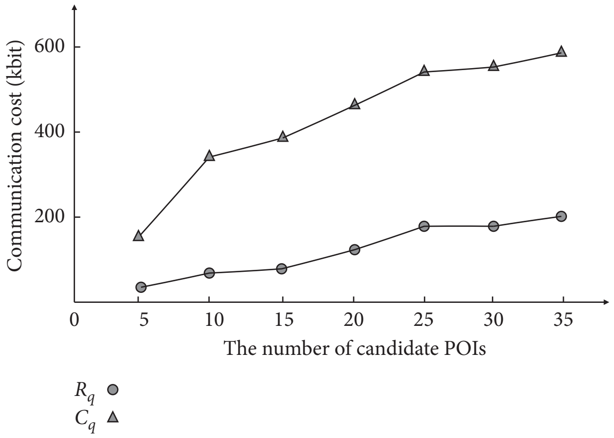


Figure 13: The communication cost in the region query and content query.

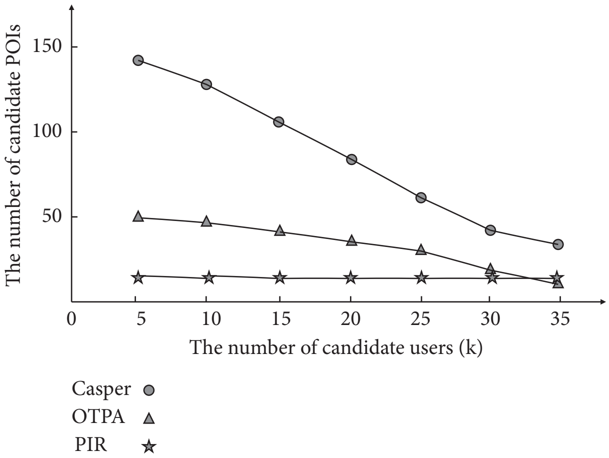


Figure 14: The relationship between candidate POIs and the number of users.

8. Conclusion

Our awareness of privacy has been heightened lately because some platforms abuse our personal data gathered by LBSS or LCS. Two prominent issues need to be further explored in the field of LBS privacy. Many studies assumed that the parties involved in anonymity are entirely trustworthy. In reality, participants could reveal the other location information because of the inconsistency of privacy degree of anonymous. In addition, the strategy that the LBSS confuse attackers with a plurality of redundant POIs information is not conducive to the operation of the LBS market and hinders the development of LBS. We developed a region queries framework and designed a privacy-aware query protocol-based oblivious transfer protocol, mainly to solve the aforementioned problems. Our solution has met the requirement of untraceability and unlinkability under the premise of preserving personal privacy. Therefore, it is certified that authenticated users can only obtain service information what they need, but malicious users cannot steal LBS server resources. Simulation results show a mutual influence and interactive relationship between the query processing time, the communication cost, the privacy degree, and the candidate POIs. Although it is inevitable that strict privacy requirements must confront a sacrifice of service quality, we will enhance our understanding of LBS to strengthen future work from reducing operating costs to improving efficiency and reinforcing privacy.

Data Availability

The location data used to support the findings of this study may be released upon application to the Microsoft GeoLife GPS Trajectories, who can be contacted at http://research.microsoft.com/en-us/downloads/b16d359d-d164-469e-9fd4-daa38f2b2e13/default.aspx.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was partially supported by grants from the Scientific Research Fund for Advanced Talents of Bengbu University (no. BBXY2021KYQD01); the Scientific Research Projects of Universities in Anhui Province of China; the Humanity and Social Science Youth Foundation of Ministry of Education of China (no. 18YJCZH068); and the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (no. 18KJB520002).

References

[1] A. Das, M. Degeling, D. Smullen, and N. Sadeh, "Personalized privacy assistants for the Internet of Things: providing users with notice and choice," *IEEE Pervasive Computing*, vol. 17, no. 3, pp. 35–46, 2018.

[2] K. H. Mohammadani, K. A. Memon, I. Memon, N. N. Hussaini, and H. Fazal, "Preamble time-division multiple access fixed slot assignment protocol for secure mobile ad hoc networks," *International Journal of Distributed Sensor Networks*, vol. 16, no. 5, Article ID 155014772092162, 2020.

[3] H. Li, H. Zhu, S. Du, X. Liang, and X. Shen, "Privacy leakage of location sharing in mobile social networks: attacks and defense," *IEEE Transactions on Dependable and Secure Computing*, vol. 15, no. 4, pp. 646–660, 2018.

[4] R.-H. Hwang, Y.-L. Hsueh, and H.-W. Chung, "A novel time-obfuscated algorithm for trajectory privacy protection," *IEEE Transactions on Services Computing*, vol. 7, no. 2, pp. 126–139, 2014.

[5] J. Wang, Z. Cai, Y. Li, D. Yang, J. Li, and H. Gao, "Protecting query privacy with differentially private k-anonymity in location-based services," *Personal and Ubiquitous Computing*, vol. 22, no. 3, pp. 453–469, 2018.

[6] S. Zhang, X. Li, Z. Tan, T. Peng, and G. Wang, "A caching and spatialK-anonymity driven privacy enhancement scheme in continuous location-based services," *Future Generation Computer Systems*, vol. 94, pp. 40–50, 2019.

[7] R. Shokri, C. Troncoso, C. Diaz, J. Freudiger, and J. P. Hubaux, "Unraveling an old cloak: K-anonymity for location privacy,"

in *Proceedings of the 9th Annual ACM Workshop on Privacy in the Electronic Society-WPES'10*, 2010.

[8] X. He, R. Jin, and H. Dai, "Leveraging spatial diversity for privacy-aware location-based services in mobile networks," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 6, pp. 1524–1534, 2018.

[9] R. Schlegel, C.-Y. Chow, Q. Huang, and D. S. Wong, "User-defined privacy grid system for continuous location-based services," *IEEE Transactions on Mobile Computing*, vol. 14, no. 10, pp. 2158–2172, 2015.

[10] P. Perazzo and G. Dini, "A uniformity-based approach to location privacy," *Computer Communications*, vol. 64, no. 64, pp. 21–32, 2015.

[11] A. S. Saxena, D. Bera, and V. Goyal, "Modeling location obfuscation for continuous query," *Journal of Information Security and Applications*, vol. 44, no. 44, pp. 130–143, 2019.

[12] A.-M. Olteanu, K. Huguenin, R. Shokri, M. Humbert, and J.-P. Hubaux, "Quantifying interdependent privacy risks with location data," *IEEE Transactions on Mobile Computing*, vol. 16, no. 3, pp. 829–842, 2017.

[13] R. Paulet, M. G. Kaosar, Y. Xun, and E. Bertino, "Privacy-preserving and content-protecting location based queries," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 5, pp. 1200–1210, 2014.

[14] H. Jannati and B. Bahrak, "An oblivious transfer protocol based on elgamal encryption for preserving location privacy," *Wireless Personal Communications*, vol. 97, no. 2, pp. 3113–3123, 2017.

[15] N. Chen, J. Li, Y. Zhang, and Y. Guo, "Efficient CP-ABE scheme with shared decryption in cloud storage," *IEEE Transactions on Computers*, p. 1, 2020.

[16] J. Li, Y. Wang, Y. Zhang, and J. Han, "Full verifiability for outsourced decryption in attribute based encryption," *IEEE Transactions on Services Computing*, vol. 13, no. 3, pp. 478–487, 2020.

[17] J. Li, Y. Zhang, J. Ning, X. Huang, G. S. Poh, and D. Wang, "Attribute based encryption with privacy protection and accountability for CloudIoT," *IEEE Transactions on Cloud Computing*, p. 1, 2020.

[18] J. Li, H. Yan, and Y. Zhang, "Identity-based privacy preserving remote data integrity checking for cloud storage," *IEEE Systems Journal*, vol. 15, no. 1, pp. 577–585, 2021.

[19] J. Li, H. Yan, and Y. Zhang, "Certificateless public integrity checking of group shared data on cloud storage," *IEEE Transactions on Services Computing*, vol. 14, no. 1, pp. 71–81, 2021.

[20] J. Li, H. Yan, and Y. Zhang, "Efficient identity-based provable multi-copy data possession in multi-cloud storage," *IEEE Transactions on Cloud Computing*, p. 1, 2019.

[21] L. Zhang, H. Xiong, Q. Huang, J. Li, K.-K. R. Choo, and J. Li, "Cryptographic solutions for cloud storage: challenges and research opportunities," *IEEE Transactions on Services Computing*, p. 1, 2020.

Hindawi
Mobile Information Systems
Volume 2021, Article ID 1440259, 13 pages
https://doi.org/10.1155/2021/1440259

Research Article

Intrusion Detecting System Based on Temporal Convolutional Network for In-Vehicle CAN Networks

Dongxian Shi[1,2] **Ming Xu**[1] **Ting Wu**[1] **and Liang Kou**[1]

[1]*School of Cyberspace, Hangzhou Dianzi University, Hangzhou 310018, China*
[2]*College of Information Technology, Zhejiang Institute of Economics and Trade, Hangzhou 310018, China*

Correspondence should be addressed to Liang Kou; kouliang@hdu.edu.cn

Received 8 July 2021; Revised 23 August 2021; Accepted 7 September 2021; Published 24 September 2021

Academic Editor: Vishal Sharma

In recent years, deep learning theories, such as Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN), have been applied as effective methods for intrusion detection in the vehicle CAN network. However, the existing RNNs realize detection by establishing independent models for each CAN ID, which are unable to learn the potential characteristics of different IDs well, and have relatively complicated model structure and high calculation time cost. CNNs can achieve rapid detection by learning the characteristics of normal and attack CAN ID sequences and exhibit good performance, but the current methods do not locate abnormal points in the sequence. To solve the above problems, this paper proposes an in-vehicle CAN network intrusion detection model based on Temporal Convolutional Network, which is called Temporal Convolutional Network-Based Intrusion Detection System (TCNIDS). In TCNIDS, the CAN ID is serialized into a natural language sequence and a word vector is constructed for each CAN ID through the word embedding coding method to reduce the data dimension. At the same time, TCNIDS uses the parameterized Relu method to improve the temporal convolutional network, which can better learn the potential features of the normal sequence. The TCNIDS model has a simple structure and realizes the point anomaly detection at the message level by predicting the future sequence of normal CAN data and setting the probability strategy. The experimental results show that the overall detection rate, false alarm rate, and accuracy rate of TCNIDS under fuzzy attack, spoofing attack, and DoS attack are higher than those of the traditional temporal convolutional network intrusion detection model.

1. Introduction

With the development of technologies such as the Internet of Vehicles, unmanned driving, and software-defined cars, modern cars are equipped with more and more advanced sensing devices and intelligent control systems [1], making cars more intelligent and providing people with a more comfortable driving service. However, with the increase of the number of electronic control units (ECU), sensing devices, ports, etc., and the diversity of networking, the attack surface of automobiles has become more and more extensive [2] and many security researchers have demonstrated the vehicles' vulnerability to attacks. For example, Miller et al. used WiFi open ports to invade a car's in-vehicle CAN network [3] by analyzing the CAN communication protocol [4], i.e., sending protocol data to the bus to cause car brake failure and engine stop. Therefore, the in-vehicle network security problem has become the focus of automotive safety, especially the CAN network commonly used in automobiles [5].

Intrusion detection is an effective method to solve the problem of in-vehicle network security, of which the study of CAN data as a sequence is an important research field of current intrusion detection. The normal CAN ID sequence features are extracted through sequence learning, and when a nonexistent sequence appears in the network, the intrusion detection system detects it as an abnormality [6, 7]. Taylor et al. proposed an intrusion detection method based on Long Short-Term Memory (LSTM) [8], which directly inputs the original CAN data packets into the model, and predicted network traffic through a short time sequence of dozens of data packets. This method of learning sequence through recurrent neural network effectively realized intrusion detection, but establishing subsequences and corresponding models for each independent CAN ID will cause the loss of sequence relationships between different IDs and reduce the efficiency of intrusion detection. Song et al. proposed an intrusion detection method based on a deep convolutional neural network [9], which learned normal and attack CAN ID sequence features through the convolutional network and achieved a higher detection rate while using the parallel processing capability of the convolutional network to reduce the time cost. However, it does not locate abnormal points and the abnormal detection of the message level is not realized.

To solve the above problems, an intrusion detection system based on temporal convolution network is proposed in this paper. We choose temporal convolution network because it shows excellent performance and efficiency on different tasks and data sets [10]. In our TCNIDS model, the original CAN data are directly regarded as a sequence, the probability of each CAN ID in the future sequence is predicted by word embedding encoding, and the time convolution model is learned and decoded, so as to realize the anomaly detection at the message level.

Contributions of this paper are the following:

(1) The temporal convolutional network model is applied to the intrusion detection of in-vehicle CAN network for the first time. The model has a simple structure, and effectively realizes the message-level prediction and anomaly detection.

(2) CAN IDs are encoded as words by using the word embedding method, which effectively represents the potential features between IDs and improves the performance of the model. At the same time, word embedding reduces the dimension of data and improves the computational efficiency of the model.

(3) PReLU activation function is used to improve the TCN model, and the performance of this activation function in TCNIDS model is compared and analyzed.

The remainder of this paper is organized as follows. We present the background material about CAN bus and intrusion detecting system in Section 2. The framework of the IDS is proposed and introduced in detail in Section 3. In Section 4, we present our experiment environment, evaluation metrics, and results, and give our conclusions in Section 5.

2. Background

2.1. CAN Bus and Its Features' Analysis. CAN is a field bus with high reliability, strong real-time performance, and low flexibility [4]. It is a standard bus of automobile in-vehicle control system and realizes the communication between in-vehicle electronic control units (ECUs). CAN network is an important part of the entire in-vehicle network. It is a peer-to-peer network, where each ECU node in the CAN network not only receives messages but also sends messages actively. Its main features are as follows:

2.1.1. Realize the Message Exchange between ECUs by Broadcast. Each ECU node in the CAN network sends messages by broadcast, and all ECU nodes in the CAN network receive messages. There are 5 types of messages: data frame, error frame, remote frame, inter-frame space, and overload frame. Figure 1 shows the structure of CAN standard data frame.

2.1.2. Adopt Arbitration Mechanism to Avoid Message Conflict. The CAN network provides an arbitration mechanism to avoid conflicts caused by different ECU nodes sending messages to the CAN network simultaneously. Each ECU carries out line and operation between its own messages to be sent and the ID of other messages, that is, comparing the bits of the arbitration field, if it is the dominant bit 0, it will continue to get the control of the bus; if it is the recessive bit 1, it will lose the arbitration, and turn to be the receiving state from the next bit, until the bus is idle.

2.1.3. Increase ECU and External Interfaces. With the improvement of vehicle intelligence, more and more mechanical parts are replaced by ECU. At present, the number of ECUs in some luxury cars is more than 100 [11], while the increasing demand for network communication and entertainment experience has greatly enriched the external interface of vehicles. For example, Tesla carries out remote software upgrade of ECU through OTA (Over-the-Air) [12], which is a technology to download new software update packages from a remote server through the network to upgrade its own system.

2.1.4. Implement Simple Data Check Code. In order to ensure the real-time performance and functional requirements of the vehicle to the greatest extent, the CAN network only includes a simple data check code when designing the message structure, and does not identify the identity ID of the message sender. Therefore, the protocol lacks security mechanism, such as encryption, access control, and message authentication. At the same time, this broadcast method allows all ECUs to easily obtain message information, which is easy to be sniffed by attackers.

2.2. Intrusion Detecting System. There have been many researches on intrusion detection of in-vehicle CAN networks. Hamada et al. learned the behaviour patterns under normal and attack environments by analyzing the periodicity of CAN messages [11]. Ji et al. believed that although the frequency of the ECU transceiver is fixed, the clock drift [13] would occur because the crystal oscillator was not exactly the same, so the accumulation of clock drift was used as the fingerprinting feature of the ECU [14]. Müter and Asaj applied information entropy to intrusion detection of in-vehicle network through maximum entropy estimation method [15], which can detect abnormal conditions of network traffic. However, these intrusion detection models are targeted at specific attacks, and so their application is limited.

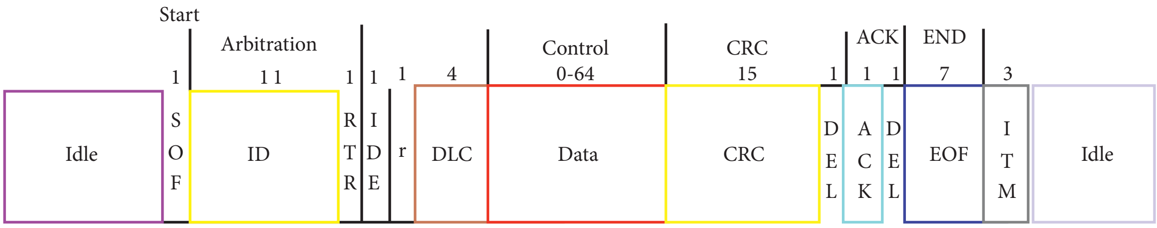


Figure 1: CAN standard data frame.

In view of these limitations, some literature studies [8, 16, 17] build intrusion detection model of in-vehicle CAN network through deep learning theory. We divide deep learning methods into 3 categories: RNN, CNN, and others. RNN, as a deep learning model for time series data processing, has been widely used in many fields. Taylor et al. proposed an intrusion detection method based on LSTM [8] to solve the problems of gradient disappearance, and short memory existed in RNN itself, which directly input original CAN packets into the model and can predict network traffic within a short time scale of dozens of packets. Another advantage of this raw traffic forecasting is that the model does not require domain knowledge. Hanselmann et al. proposed an intrusion detection system CANet based on LSTM and AutoEncoder [18]. The system introduced an independent LSTM model for each CAN ID to learn the time dynamic characteristics of each message-related signal, and then aggregated all IDs and used the AutoEncoder model to learn the interdependence between signals. The AutoEncoder included an Encoder and a Decoder. The Encoder mapped the high-dimensional input data to the low-dimensional embedding space, which could be used for dimensionality reduction. At the same time, the Decoder was used to reconstruct the low-dimensional embedding space of the representation, which could be compared with the original input data for deviation comparison, so as to effectively identify anomalies. In addition, for the first time, it used an Autoencoder to naturally process the data structure of the high-dimensional CAN bus. Wang et al. proposed a distributed anomaly detection system based on the hierarchical time memory (HTM) algorithm [19], which effectively realized the real-time prediction of the original CAN traffic data at the bit level. The method in [8, 18, 19] causes the loss of some information and relationships in the CAN network by establishing a model for an independent CAN ID or ECU [20], and the model becomes more complicated. Kang and Kang proposed a deep neural network (DNN)-based intrusion detection method [21], which used an unsupervised deep belief network (DBN) to pretrain the initialization parameters and test it on the simulation data set generated by the OCTANE platform. Usually, when training a model, it is considered that DNN and LSTM consume more time than CNN. Based on this fact, Song et al. proposed an intrusion detection method based on deep convolutional neural network [9], by simplifying the Inception-ResNet model. The method achieves a higher detection rate and reduces the time cost. However, this method cannot effectively locate the message level detection by detecting whether the sequence has an attack. In addition, some current studies do not use a single method but use the advantages of various methods to mix them. Xiao et al. combined LSTM and CNN to treat CAN network traffic data as a whole from the two dimensions of time and space [20], and proposed a convLSTM model, which can better extract the potential features of normal data flow, so as to predict the deviation attack behaviour of the time series more effectively. However, it needs to be improved in terms of threshold selection and real-time detection performance.

3. Methodology

In this section, first we present the overview of the TCNIDS model for in-vehicle CAN network. Then, we introduce each model component in detail.

3.1. Model Overview. The traffic data in the in-vehicle CAN network appears in the form of sequence. Due to the arbitration mechanism of CAN network and the periodicity of message transmission, there is a dependency on the appearance of the message sequence [22]. On the bus, each ECU in the CAN network follows the CAN protocol to send and receive messages, and there is a certain relationship between the previous message and the next message. Therefore, we convert the intrusion detection of CAN traffic data into sequence prediction for research. We learn to extract normal sequence features, and when a nonexistent sequence appears in the network, the intrusion detection system will identify it as an abnormality and determine which message is inconsistent with the predicted sequence result. At present, in the field of time series forecasting, time convolutional networks have shown excellent performance and efficiency on various data sets and tasks. Therefore, this paper chooses time convolution as the basis of the entire model. Assume that there is an input sequence $X_{t-s:\,t} = \{x_{t-s}, x_{t-s+1}, \ldots, x_{t-1}, x_t\}$ at each time interval t, the objective is that the model can predict the corresponding output sequence $\widehat{Y}_{t-s:\,t} = \{\widehat{y}_{t-s}, \widehat{y}_{t-s+1}, \ldots, \widehat{y}_{t-1}, \widehat{y}_t\}$. Formally, the model is an arbitrary function f: $X_{t-s:\,t} \longrightarrow \widehat{Y}_{t-s:\,t}$:

$$\widehat{Y}_{t-s:\,t} = f\left(X_{t-s:\,t}\right). \tag{1}$$

The goal of the model is to train the function f to minimize the loss function $\text{Loss}\left(Y_{t-s:\,t}, f\left(X_{t-s:\,t}\right)\right)$ between the model output sequence and the real sequence. The loss function of this model training adopts cross entropy, and the specific expression is as follows:

$$\operatorname{Loss}\left(Y_{t-s:\,t}, f\left(X_{t-s:\,t}\right)\right) = -\frac{1}{S}\sum_{1}^{S}\sum_{i=1}^{M} y_i \log_a p_i, \qquad (2)$$

where S denotes the number of messages in the sequence, M denotes the number of CAN message types, y_i denotes the true label of message category i, and p_i is the probability that the model predicts to belong to message category i.

TCN proposed the network structure shown in Figure 2. First, since the output length generated by the network in sequence prediction needs to be consistent with the input length, TCN uses a 1D fully convolutional network (FCN), and each hidden layer uses zero padding for length padding. Second, using future information to predict the past will lead to information leakage [10], so TCN introduces causal convolution [10] to ensure that the output at the current moment comes from the convolution of current and historical information. Third, having a longer historical memory requires a deeper network, but it will increase the number of parameters. Therefore, TCN uses expanded convolution to expand the receptive field of the convolution, thereby reducing the depth of the network as much as possible. Fourth, normalization can solve the problem of gradient vanishing or gradient exploding caused by the increase of network depth to a certain extent, but it will also bring about degradation problems. Therefore, the residual network [23] is introduced to solve this problem in TCN.

This paper, by extending TCN, proposes the intrusion detection model TCNIDS for in-vehicle CAN network. The overview of the model is shown in Figure 3.

The model has two stages: training and detection. The training stage learns the normal CAN data sequence and realizes the prediction of the next sequence by extracting potential sequence features, thereby learning the sequence law of normal behaviour. The detection stage checks all CAN data sequences including attack behaviours. Through observation, there is more than one possibility of the message predicted by the CAN sequence. Therefore, this paper uses the Top g probability strategy to detect anomalies in each message in the prediction sequence. If the predicted real message is in the message set with the top g probability, it is detected as normal, otherwise it is detected as abnormal. The following will introduce each component in the model in detail.

3.2. Data Preprocessing. The data set includes timestamp, CAN ID, DLC, Data, and Label. We only need to extract the two fields of CAN ID and Label to form the original ID sequence. Among them, CAN ID is extracted in the training phase, and CAN ID and Label are extracted in the anomaly detection phase for evaluating the performance of the TCNIDS model proposed in this paper.

3.3. Encoder. Since in One Hot encoding CAN ID, the distance between all IDs is the same, there is a disadvantage that the potential relationship between IDs cannot be extracted during model training; on the other hand, the word embedding coding method maps a word to a point in the semantic space, which makes the semantically similar words relatively close, and it can effectively characterize the relationship between IDs [24]. Therefore, this paper uses the word embedding method to treat each type of CAN ID in the data set as a word, uses a word vector to represent the CAN ID, and learns to extract the potential relationship between IDs, thereby improving the performance of the model. Figure 4 shows the process of CAN ID Embedding:

Step 1. Various types of IDs in the original CAN ID sequence are extracted to construct an embedding matrix. Each type of CAN ID is expressed as a word vector of the same dimension, and the initial vector is assigned a random value.

Step 2. Replace each ID in the original ID sequence according to the embedding matrix of CAN ID, which is represented by the word vector in **Step 1.**

Step 3. The embedded matrix constructed by **Step 1** and the ID sequence represented by **Step 2** are added to the corpus for the input data of model training and testing.

3.4. Temporal Convolutional Network. The TCNIDS model proposed in this paper extends on the general TCN model described in Ref. [10]. The TCN model has two main constraints. The output of the hidden layer in the middle of the model has the same length as the input, and the prediction at time t can only rely on the information before time t. For the first constraint, TCN uses a 1-D fully convolutional network (FCN) to convolve time series data, and uses zero padding to ensure the same length of the front and back network layers. Regarding the second constraint, TCN introduces causal convolution, so that the output at time t can only be convolved with time t and previous information, ensuring that the past cannot be predicted by future information, thereby causing information leakage. As shown in Figure 5, through causal convolution, one-dimensional convolution of past information is realized, and the potential features of CAN data sequence are effectively extracted.

Formally, set the convolution filter $F = (f_1, f_2, \ldots, f_k)$. For any element x_j in sequence $X_{t-k:\,t} = \{x_{t-k}, x_{t-k+1}, \ldots, x_{t-1}, x_t\}$, the causal convolution at x_j is defined as follows:

$$(F \oplus X)\left(x_j\right) = \sum_{i=1}^{K} f_i x_{j-K+i}, \qquad (3)$$

where K denotes the size of the convolution kernel.

3.4.1. Dilated Convolutions. For the prediction of CAN data series, we expect the model to remember more historical information, so that the prediction performance will be more stable. However, with the above causal convolution method, to achieve a larger receptive field, it is necessary to stack many network layers to reach the goal. In order to overcome this problem, the dilated convolution is used to expand the receptive field of the convolution, which greatly reduces the number of intermediate hidden layers, which is also the biggest feature of the dilated convolution. In dilated convolution, filters are applied by skipping a certain number

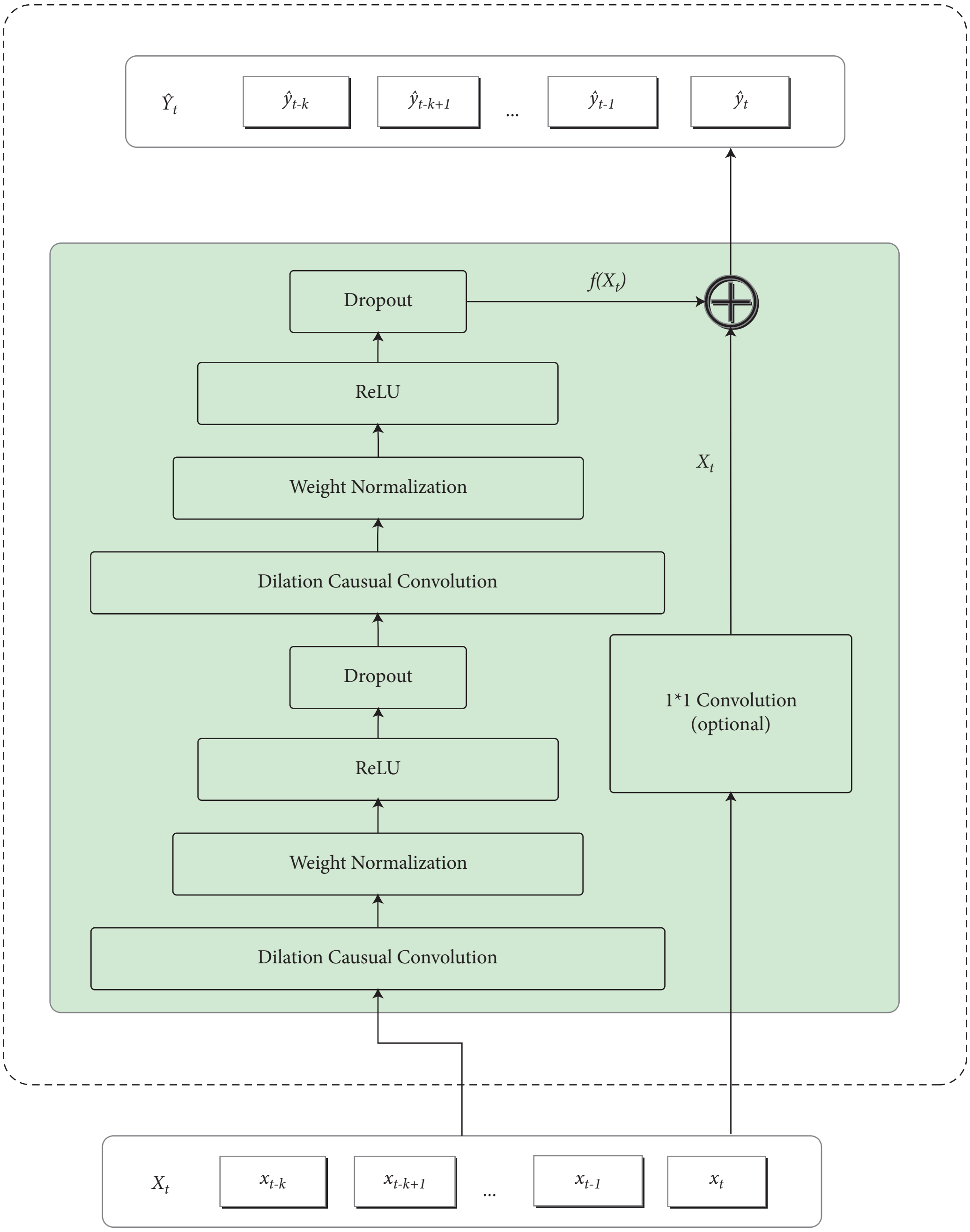


Figure 2: TCN network structure.

of steps according to the expansion factor d to achieve convolution of a larger area. As shown in Figure 6, this growth method of the receptive field is different from pooling operation, as it skips some existing elements. In general, d will increase exponentially as the network depth i increases, so the model can build a long memory.

Formally, set the convolution filter $F = (f_1, f_2, \ldots, f_k)$. For any element x_j in sequence $X_{t-k:\,t} = \{x_{t-k}, x_{t-k+1}, \ldots, x_{t-1}, x_t\}$, the causal convolution at x_j is defined as follows:

$$\left(F\oplus_d X\right)\left(x_j\right) = \sum_{i=1}^{K} f_i x_{j-(K-i)d}. \quad (4)$$

Among them, d is the expansion factor of the dilated convolution, and when $d = 1$, the convolution kernel degenerates into a general convolution operation.

3.4.2. Residual Connections. Since the receptive field of the TCN model depends on the network depth n, filter size k, and expansion factor d, making the TCN deeper and larger is the key to obtain a large enough receptive field [25]. The residual connection can simplify deep network training. The deep network through this structure has been proved to be very effective, which can speed up the training process and avoid the disappearance of gradients. As shown in Figure 2,

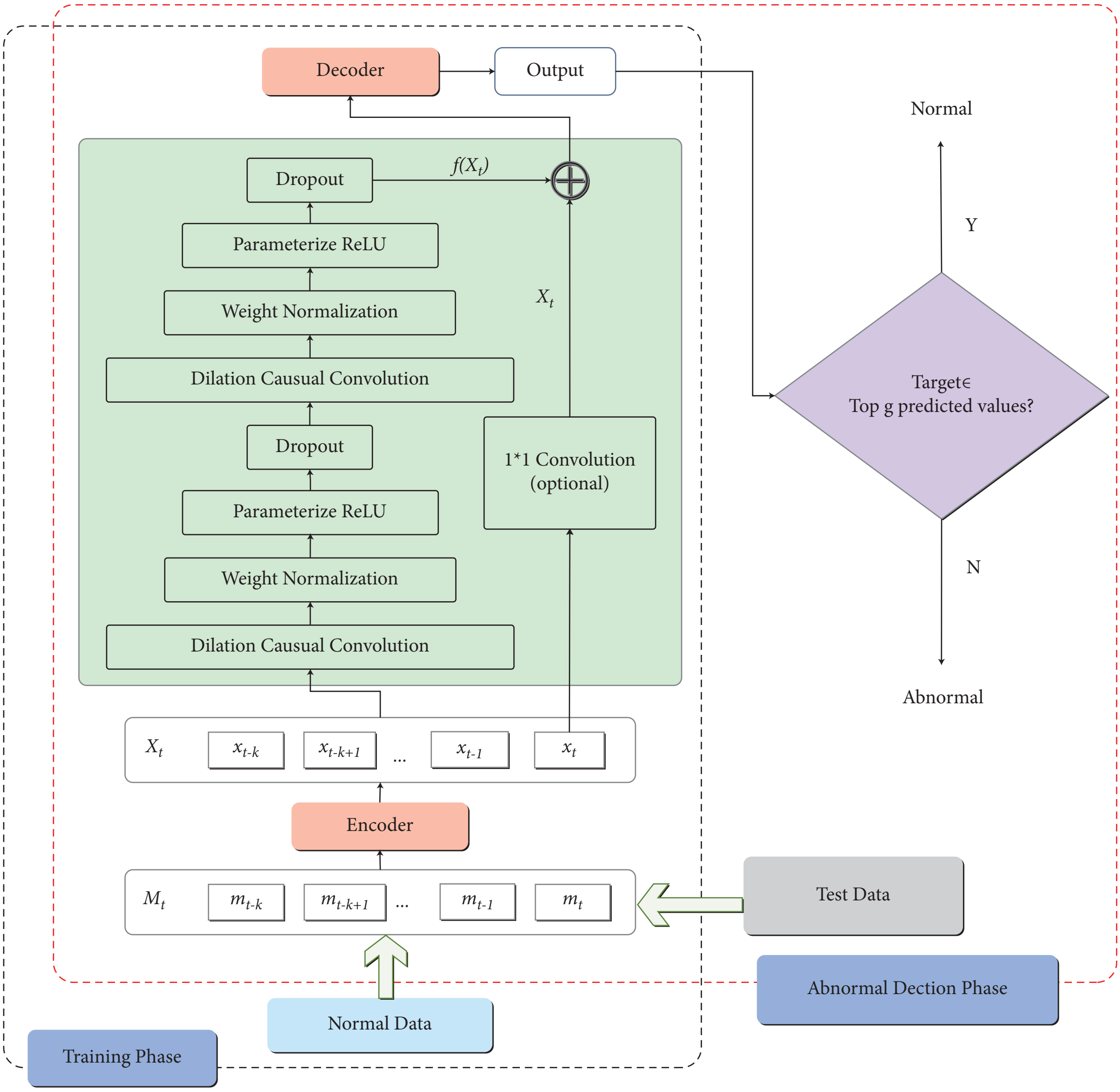
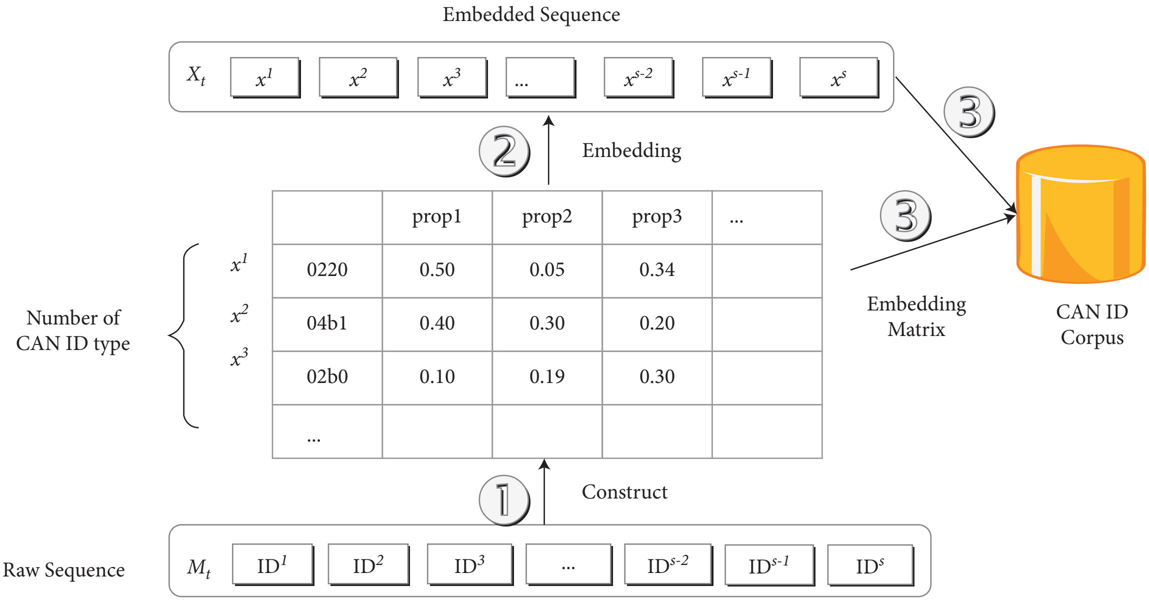


Figure 3: Illustration of the TCNIDS model.

the model is constructed by residual blocks in TCN. Each residual block contains two network layers, and each layer is composed of four parts: causal dilated convolution, normalization, activation function, and regularization. For normalization, we apply weight normalization to the convolution filter. Regularization can effectively prevent the over-fitting phenomenon of the model. In addition, in the standard ResNet [23], the input is directly added to the output of the residual function. While in TCN, the input and output may have different channel dimensions. In order to be able to perform residual operations, we use an additional 1 * 1 convolution to ensure that the output and input of each layer have the same shape.

3.4.3. Activation Function Selection and Improvement. The original TCN model uses a one-dimensional convolutional network to extract features, and uses the Relu activation function to nonlinearly map the features [26]. As shown in Figure 7(a), when $x \geq 0$, the gradient of the Relu activation function is 1, and when $x < 0$, the gradient reduces to 0, so that the network can converge faster. This activation function is widely used in CNN. However, when $x < 0$, the output value of the convolution kernel operation is always 0, which causes many features to be masked, and the network cannot extract effective features. This phenomenon in which the Relu activation function is killed in the negative region is called “Dying” [27].



	prop1	prop2	prop3	...
0220	0.50	0.05	0.34	
04b1	0.40	0.30	0.20	
02b0	0.10	0.19	0.30	
...				

Figure 4: CAN ID encoding process.

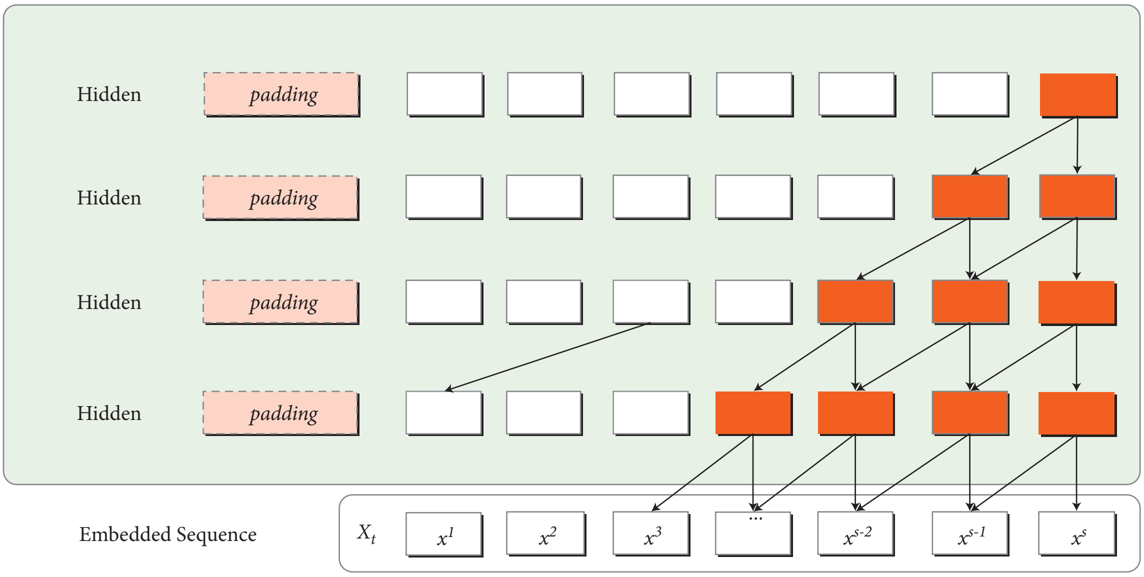


Figure 5: Causal convolutions.

In order to solve the problems of Relu, He et al. proposed a parameterized Relu function method [28], as shown in Figure 7(b). The parameter α is introduced in the parameterized Relu function. When $x < 0$, the gradient of the activation function will automatically change with the learning of the network, so as to obtain the optimal value of the model.

3.5. Decoder. One goal of the model is to predict the probability of which type of CAN ID each message in the sequence belongs to, that is, the target dimension is the number of CAN ID types, but the output dimension obtained through the time convolutional network model is different from this target dimension, so it is necessary to realize the transformation of these two dimensions through decoding. Since the number of CAN IDs in the in-vehicle network is not much, generally within 100, we adopt the simple method of full connection to directly realize the transformation of two dimensions.

4. Experiments

This section firstly introduces the CAN data sets, experimental environment, and evaluation metric, and then illustrates the optimized parameter settings for the training and detecting. Finally, the performance of the model is deeply analyzed through the experimental results.

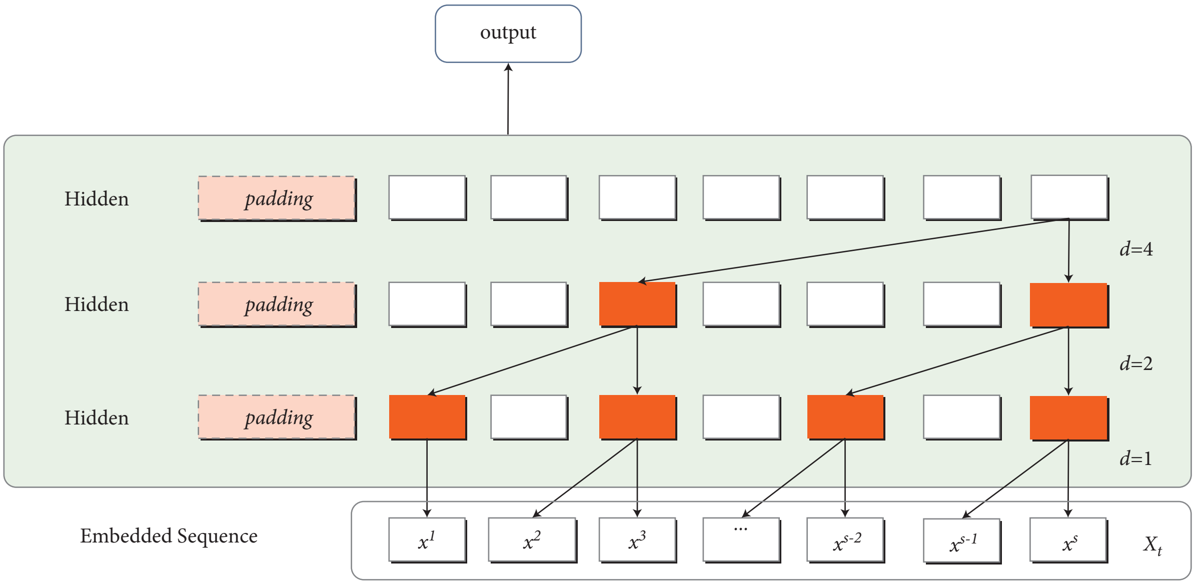


Figure 6: Dilated convolutions.

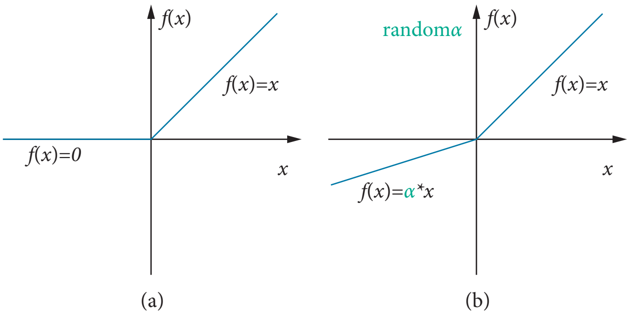


Figure 7: Activation function curve. (a) Relu. (b) Parameterized Relu.

4.1. Data Sets and Experimental Environment. This paper adopts the public CAN data sets provided in [12], which are collected by the Kia Soul test vehicle and contain 17558346 CAN messages. The data sets can be divided into Normal, Fuzzy Attack, Spoof Gear attack, Spoof RPM Attack, and DoS Attack. The information of the CAN data sets is shown in Table 1.

The length of the CAN ID of the data sets used in this paper is 11 bits. Before input to the model, CAN ID of all the data sets above will be extracted to form the corresponding ID sequence data. In this paper, the data sets are not divided according to the fixed time, but are divided according to the sequence length specified in the model parameters, and the method of sliding window is used to extract the next sequence. 80% of the normal data set is selected as the training set, and 20% of the normal data set and the other 4 attack data sets are selected as the test set.

The experimental environment in which the TCNIDS model is tested in this paper is shown in Table 2.

4.2. Evaluation Metric. In order to evaluate the detection performance of the proposed TCNIDS model, we firstly define the confusion matrix for intrusion detection shown in Table 3.

Among them, TP denotes the number of CAN messages that are abnormal and predicted to be abnormal, FN denotes the number of CAN messages that are abnormal but predicted to be normal, FP denotes the number of CAN messages that are normal but predicted to be abnormal, and TN denotes the number of CAN messages that are normal and predicted to be normal. According to this confusion matrix, three indicators are specifically defined to evaluate the ability of real CAN messages to be predicted by TCNIDS as normal or abnormal.

4.2.1. Detecting Rate. Detecting Rate is also known as True Positive Rate (TPR). This paper uses TPR to represent the detection rate, which represents the proportion of abnormal packets predicted to be the total number of abnormal packets. The higher the value, the better the performance. The specific formula is as follows:

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}. \quad (5)$$

Table 1: CAN data set information.

Data set	Total messages	Normal messages	Attack messages
Normal	988871	988871	0
Fuzzy attack	3838860	3347013	491847
Gear spoof	4443142	3845890	597252
RPM spoof	4621702	3966805	654897
DoS attack	3665771	3078250	587521

Table 2: Experiment environment.

Configuration item	Configuration parameter
CPU	Intel Xeon Gold 5118@2.3 GHz * 24
RAM	16.0 GB
GPU	NVIDIA Quadro P4000 GPU
Operating system	Windows Server 2012 R2

Table 3: Confusion matrix of TCNIDS.

Packet	Predicted attack	Predicted normal
True attack	TP	FN
True normal	FP	TN

4.2.2. False Positive. False Positive Rate represents the ratio of the number of normal packets predicted to be abnormal to the total number of normal packets. The lower the value, the better the performance. The specific formula is as follows:

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{TN} + \mathrm{FP}}. \quad (6)$$

4.2.3. Accuracy. Accuracy represents the proportion of the number of correctly predicted packets to the total number of packets. The higher the value, the better the performance. The specific formula is as follows:

$$\text{accuracy} = \frac{\mathrm{TN} + \mathrm{TP}}{\mathrm{TN} + \mathrm{TP} + \mathrm{FP} + \mathrm{FN}}. \quad (7)$$

For the normal data and four types of attack data in the test set, this paper adopts the same sequence length as the training set, and inputs the sequence data to the model for intrusion detection according to the detection process. If the predicted sequence satisfies the ID of the CAN message in the first q message categories with high probability, then the model will update TN or FN according to the real message label, and otherwise update TP or FP.

4.3. Parameter Setting. In this paper, through repeated experiments with different parameter combinations, the optimal parameters of the model are determined, and subsequent experiments all use the optimal parameters for experimental evaluation and comparison. The specific parameter settings are shown in Table 4.

We apply the SGD algorithm to ensure the convergence in the experiment, and for a faster convergence, a learning rate (lr) annealing method is adopted. When the loss is greater than the loss of the previous 5 times, set $\mathrm{lr} = \mathrm{lr}/2$.

Table 4: Parameter setting.

Parameter item	Parameter value
Embedding size	200
Kernel size	5
Layer number	4
Hidden units	150
Convolution dropout	0.45
Embedding dropout	0.25
Initial learning rate	0.50
Gradient clip	0.35
Batch size	16
Sequence length	60
Top g	16

4.4. Result Analysis

4.4.1. Overall Result. The test results of TCNIDS proposed in this paper on the test data sets are shown in Table 5

It can be seen from Table 5 that TCNIDS exceeds 93% on both TPR and Accuracy indicators, and the FPR is not higher than 5%. Especially for normal behaviour, Fuzzy attack, and DoS attack, the TCNIDS model has good detection capabilities. The TPR and Accuracy detected by the model on the normal data set are close to 100%, and the FPR is close to 0. In the case of DoS attacks, TPR is also close to 100%. In addition, the model's TPR and Accuracy indicators for detecting Fuzzy attack are not less than 98%. Since DoS attack and Fuzzy attack themselves are uncommon CAN ID injections, according to the given method of word embedding, their value in the word vector will gradually differ from the normal CAN ID as the model is trained; so, TCNIDS can easily detect these two attacks. At the same time, the model also shows good performance in Gear Spoof Attack and RPM Spoof Attack. TPR and Accuracy also exceed 93%, and FPR is lower than 3%.

4.4.2. Detail Performance. In order to thoroughly analyze the performance of the TCNIDS model in intrusion detection, we collected more experimental results. Figure 8 shows the change of loss during a single epoch of training. It can be seen from Figure 8 that the loss drops and converges rapidly. In this paper, a variety of methods, such as weight normalization, regularization, time convolution network, and residual network, are used to deal with the problem of gradient dispersion and disappearance, which improves the stability of model training and further verifies the effectiveness of the model.

In order to observe the loss of the model on each data set, we trained the model for 50 epochs. Figure 9 shows the changes of loss on each data set. It can be seen from the figure that, in the training stage, with the increase of model training times, loss in the normal data set rapidly declines and converges. In the test stage, the trained model carries out loss calculation for each data set, and it is not difficult to find that the normal data set still maintains the loss similar to that in the training stage, but the loss of each attack data set is at a high level and fluctuates greatly due to the attack behaviour, especially the gear spoofing attack. Therefore, loss can be

Table 5: Detecting results using TCNIDS.

Data set (%)	TPR	FPR	Accuracy
Normal	99.999	0.001	99.999
Fuzzy attack	97.552	0.027	98.345
Gear spoof	93.526	0.039	96.290
RPM spoof	93.078	0.046	94.626
DoS attack	100.000	0.034	98.496

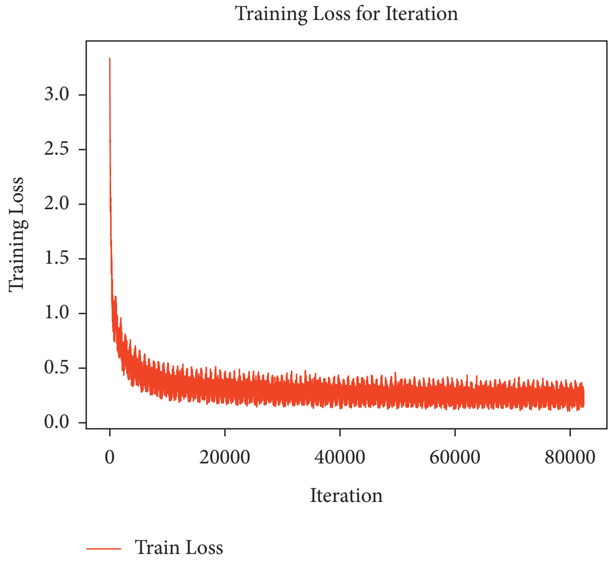


Figure 8: Loss curve for training.

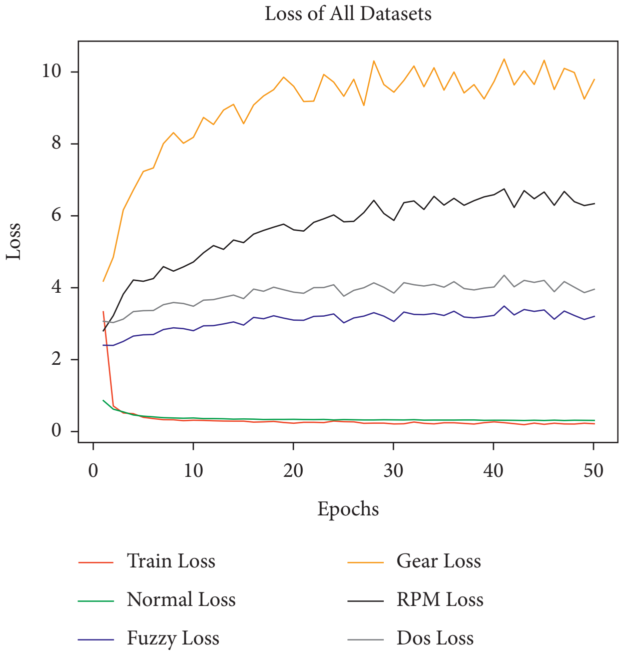


Figure 9: Loss comparison for all data sets.

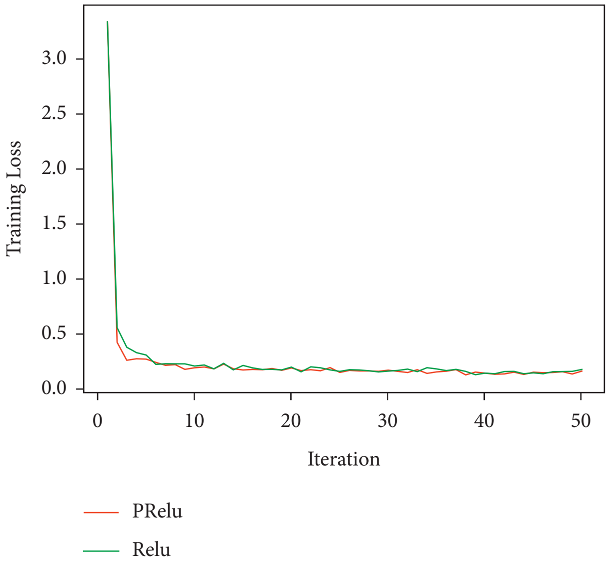


Figure 10: The influence of Relu and parameterized Relu on loss.

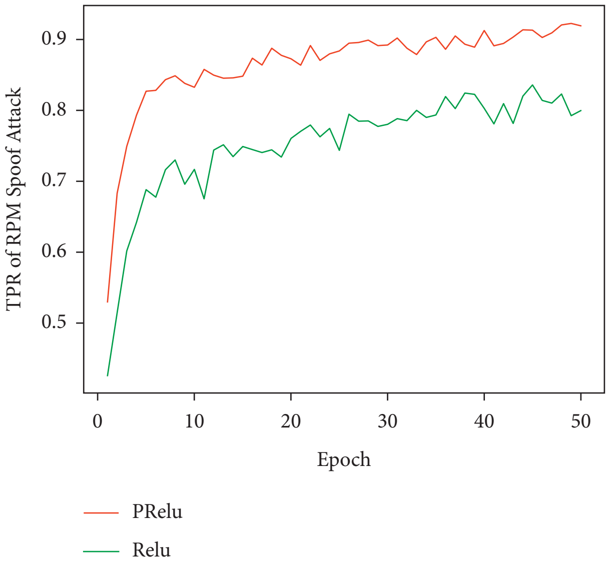


Figure 11: The influence of Relu and parameterized Relu on TPR.

used to distinguish between normal and abnormal behaviour patterns.

It can be seen from Figure 10 that after improving the model by parameterizing the Relu activation function, the convergence rate is faster. It can be seen from Figures 11 and 12 that the TPR and Accuracy have been improved to a certain extent, indicating that the existing features should be preserved as much as possible when nonlinear mapping of the features through the activation function and the full shielding method cannot be adopted when $x < 0$. This also verifies the effectiveness of the parameterized Relu method in Section 4 to improve the model.

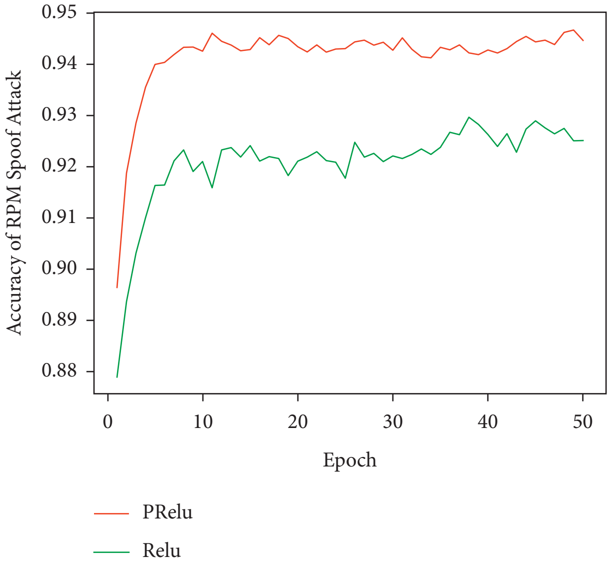


Figure 12: The influence of Relu and parameterized Relu on accuracy.

5. Conclusion

With the increase of the attack surface of modern automobiles, intrusion detection systems have become the most important technology for in-vehicle network security protection. In view of the current problems in the implementation of the in-vehicle CAN network anomaly detection through the deep learning network model, this paper proposes an intrusion detection system based on time convolution network. The structure of the model is simple, and the sequence data are predicted by word embedding encoding, time convolution network and decoding, and the intrusion detection is realized by top g strategy. In the model, the word embedding method encodes CAN ID into words, which effectively characterizes the potential features between IDs, improves the performance of the model, reduces the dimensionality of the data, and improves the computational efficiency of the model. At the same time, the TCNIDS model uses the parameterized Relu activation function to try to retain the characteristics of nonlinear mapping when $x < 0$, and optimize the performance of the model. The experimental results show that the TCNIDS model proposed in this paper has high performance in Fuzzy attack, Spoof attack, and DoS attack, especially Fuzzy attack and DoS attack. At the same time, compared with the ordinary time convolutional network model, the improved model has a certain improvement in detection rate, false alarm rate, and accuracy rate, which also proves the effectiveness of the method. Therefore, the TCNIDS proposed in this paper can strengthen the security of the in-vehicle CAN network. Since the model uses an unsupervised learning method, in the future, we will apply it to more data sets and attack scenarios, and further improve the performance of the model.

Data Availability

The data used to support the findings of the study are available at Korea University and Hacking and Countermeasure Research Lab (https://ocslab.hksecurity.net/Datasets/CAN-intrusion-dataset).

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors thank Korea University and Hacking and Countermeasure Research Lab for publishing the CAN intrusion data sets. This work was partially supported by the "Fundamental Research Funds for the Provincial Universities," Zhejiang Institute of Economics and Trade (Grant Number: 19SBYB06), and the Domestic Visiting Engineer Project of Universities, the Education Department of Zhejiang (Grant Number: FG2020135).

References

[1] W. Wu, R. Li, G. Xie et al., "A survey of intrusion detection for in-vehicle networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 3, pp. 919–933, 2019.

[2] Z. Abdollahi Biron, S. Dey, and P. Pisu, "Real-time detection and estimation of denial of service attack in connected vehicle systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 12, pp. 3893–3902, 2018.

[3] C. Miller and C. Valasek, *Remote Exploitation of an Unaltered Passenger Vehicle*, Black Hat, Washington, DC, USA, 2015.

[4] B. Parikh, "CAN protocol: understanding the controller area network," 2021, https://www.engineersgarage.com/can-protocol-understanding-the-controller-area-network-protocol/.

[5] Q. Luo and J. Liu, "Wireless telematics systems in emerging intelligent and connected vehicles: threats and solutions," *IEEE Wireless Communications*, vol. 25, no. 6, pp. 113–119, 2018.

[6] M. L. Han, B. I. Kwak, and H. K. Kim, "Event-triggered interval-based anomaly detection and attack identification methods for an in-vehicle network," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 2941–2956, 2021.

[7] C. Zhou and R. Paffenroth, "Anomaly detection with robust deep autoencoders," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 665–674, Halifax, Canada, August 2017.

[8] A. Taylor, S. Leblanc, and N. Japkowicz, "Anomaly detection in automobile control network data with long short-term memory networks," in *Proceedings of the 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pp. 130–139, Montreal, Canada, October 2016.

[9] H. M. Song, J. Y. Woo, and H. K. Kim, "In-vehicle network intrusion detection using deep convolutional neural network," *Vehicular Communications*, vol. 21, pp. 1–13, Article ID 100198, 2020.

[10] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," 2018, http://arxiv.org/abs/1803.01271v1.

[11] Y. Hamada, M. Inoue, H. Ueda, Y. Miyashita, and Y. Hata, "Anomaly-based intrusion detection using the density estimation of reception cycle periods for in-vehicle networks," *SAE International Journal of Transportation Cybersecurity and Privacy*, vol. 1, no. 11, pp. 39–56, 2018.

[12] H. Lee, S. H. Jeong, and H. K. Kim, "OTIDS: a novel intrusion detection system for in-vehicle network by using remote frame," in *Proceedings of the 2017 15th Annual Conference On Privacy, Security And Trust (PST)*, pp. 57–5709, Calgary, Canada, August 2017.

[13] H. Ji, Y. Wang, H. Qin, X. Wu, and G. Yu, "Investigating the effects of attack detection for in-vehicle networks based on clock drift of ecus," *IEEE Access*, vol. 6, pp. 49375–49384, 2018.

[14] K. T. Cho and K. G. Shin, "Viden: attacker identification on in-vehicle networks," 2017, https://arxiv.org/abs/1708.08414.

[15] M. Müter and N. Asaj, "Entropy-based anomaly detection for in-vehicle networks," in *Prceedings of the 2011 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1110–1115, Baden-Baden, Germany, June 2011.

[16] E. Seo, H. M. Song, and H. K. Kim, "Gids: gan based intrusion detection system for in-vehicle network," in *Proceedings of the 2018 16th Annual Conference on Privacy, Security and Trust (PST)*, pp. 1–6, Belfast, Ireland, August 2018.

[17] S. Tariq, S. Lee, H. K. Kim, and S. S. Woo, "Detecting in-vehicle CAN message attacks using heuristics and RNNs," in *Proceedings of the International Workshop on Information and Operational Technology Security Systems*, pp. 39–45, Heraklion, Greece, September 2018.

[18] M. Hanselmann, T. Strauss, K. Dormann, and H. Ulmer, "CANet: an unsupervised intrusion detection system for high dimensional CAN bus data," *IEEE Access*, vol. 8, pp. 58194–58205, 2020.

[19] C. Wang, Z. Zhao, L. Gong, L. Zhu, Z. Liu, and X. Cheng, "A distributed anomaly detection system for in-vehicle network using HTM," *IEEE Access*, vol. 6, pp. 9091–9098, 2018.

[20] J. Xiao, H. Wu, and X. Li, "Internet of things meets vehicles: sheltering in-vehicle network through lightweight machine learning," *Symmetry*, vol. 11, no. 11, p. 1388, 2019.

[21] M. J. Kang and J. W. Kang, "A novel intrusion detection method using deep neural network for In-vehicle network security," in *Proceedings of the 2016 IEEE 83rd Vehicular Technology Conference (VTC Spring)*, pp. 1–5, Nanjing, China, May 2016.

[22] M. Marchetti and D. Stabili, "Anomaly detection of CAN bus messages through analysis of ID sequences," in *Proceedings of the 2017 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1577–1583, Angeles, CA, USA, June 2017.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the 2016 IEEE Conference On Computer Vision And Pattern Recognition (CVPR)*, pp. 770–778, Vegas, NV, USA, June 2016.

[24] S. Deng, N. Zhang, Z. Wen, J. Chen, J. Z. Pan, and H. Chen, "Knowledge-driven stock trend prediction and explanation via temporal convolutional network," in *Proceedings of the WWW'19: Companion Proceedings of the 2019 World Wide Web Conference*, pp. 678–685, San Francisco, CA, USA, May 2019.

[25] C. Lea, R. Vidal, A. Reiter, and G. D. Hager, "Temporal convolutional networks: a unified approach to action segmentation," *Lecture Notes in Computer Science*, Springer, Berlin, Germany, pp. 47–54, 2016.

[26] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proceedings of the*

Proceedings of International Conference on Machine Learning, pp. 807–814, Haifa, Israel, June 2010.

[27] R. Yang, D. Qu, S. Zhu, Q. Yekui, and T. Yongwang, “Anomaly detection for log sequence based on improved temporal convolutional network,” *Computer Engineering*, vol. 46, no. 8, pp. 50–57, 2020.

[28] K. He, X. Zhang, S. Ren, and J. Sun, “Delving deep into rectifiers: surpassing human-level performance on imagenet classification,” in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1026–1034, Santiago, Chile, December 2015.

Hindawi
Mobile Information Systems
Volume 2021, Article ID 7314508, 18 pages
https://doi.org/10.1155/2021/7314508

Research Article

Scrutinizing the Vulnerability of Ephemeral Diffie–Hellman over COSE (EDHOC) for IoT Environment Using Formal Approaches

Jiyoon Kim,[1] **Daniel Gerbi Duguma**,[1] **Sangmin Lee**,[1] **Bonam Kim**,[1] **JaeDeok Lim**,[2] **and Ilsun You**[1]

[1]*Dept. of Information Security Engineering, Soonchunhyang University, Asan 31538, Republic of Korea*
[2]*Electronics and Telecommunications Research Institute (ETRI), Daejeon 34129, Republic of Korea*

Correspondence should be addressed to Ilsun You; ilsunu@gmail.com

Received 9 July 2021; Revised 4 August 2021; Accepted 16 August 2021; Published 13 September 2021

Academic Editor: Zengpeng Li

Most existing conventional security mechanisms are insufficient, mainly attributable to their requirements for heavy processing capacity, large protocol message size, and longer round trips, for resource-intensive devices operating in an Internet of Things (IoT) context. These devices necessitate efficient communication and security protocols that are cognizant of the severe resource restrictions regarding energy, computation, communication, and storage. To realize this, the IETF (Internet Engineering Task Force) is currently working towards standardizing an ephemeral key-based lightweight and authenticated key exchange protocol called EDHOC (Ephemeral Diffie–Hellman over COSE). The protocol's primary purpose is to build an OSCORE (Object Security for Constrained RESTful Environments) security environment by supplying crucial security properties such as secure key exchange, mutual authentication, perfect forward secrecy, and identity protection. EDHOC will most likely dominate IoT security once it becomes a standard. It is, therefore, imperative to inspect the protocol for any security flaw. In this regard, two previous studies have shown different security vulnerabilities of the protocol using formal security verification methods. Yet, both missed the vital security flaws we found in this paper: resource exhaustion and privacy attacks. In finding these vulnerabilities, we leveraged BAN-Logic and AVISPA to formally verify both EDHOC protocol variants. Consequently, we described these security flaws together with the results of the related studies and put forward recommended solutions as part of our future work.

1. Introduction

IoT refers to a network environment in which all surrounding objects are connected to wired and wireless networks to interact and exchange information over the Internet. These objects (also referred to as "things") can range from a simple soil moisture sensor in a field to a complex implanted device in a human body. With continuous developments in low-cost electronics (such as sensors), fast progress in mobile communication (especially with the introduction of 5G), and significant advances in data analytics (e.g., machine learning and lightweight deep learning), IoT has become one of the most demanded technologies in our time [1, 2]. Currently, IoT serves as an instrumental platform to host many applications in manufacturing, healthcare, energy, cities, and many more. In the next four years (by 2025) only, the total market share of IoT can stretch to reach up to 3 trillion USD [3], while the number of devices operating in an IoT environment can cross 42 billion with over 73 ZB of generated data [4].

Despite the vast expansion of IoT-enabled devices and their widespread applications, IoT still has several challenges that needs to be tackled. Some of the issues are tightly related to the severe computing resource constraints concerning storage, processing, and communication [5–7]. Such tight requirements call for efficient mechanisms to enable devices operating within IoT environments to function through unstable channels with constrained bandwidth and varying topology [8]. To realize these stringent conditions, essential protocols, such as [9–11], have been standardized by IETF. In addition, because IoT devices transport several sensitive information, security problems can threaten the inability to provide services and the user's personal information. Some potential security attacks are device software malfunction, prying, malevolent code infusions, device tampering, and unauthorized access [12]. Furthermore, studies such as [13, 14] investigated the security issues of integrating LPWAN in the 5G ecosystem, as well as the practical evaluation of compression and fragmentation of standard protocols as applied to IoTs in LPWAN, respectively. Hence, IoT devices require more capable security schemes that work in tandem with the communication protocols to mitigate these security attacks.

Even though IoT applications anticipate solid security assurance, securing IoT frameworks is challenging. It is mainly because of their intrinsic nature of resource constraint and absence of "security aware" design. Although there are various security solutions designed for conventional networks, such as IKEv2 [15], TLS [16], and DTLS [17], they are not suitable for the IoT environment due to their high degree of processing power and memory space. For instance, the footprint in bytes for a DTLS is six times heavier than the EDHOC + CoAP (Constrained Application Protocol) [8]. Fortunately, there are now efforts in designing standard security protocols mainly intended to serve in IoT environments. One such application layer security protocol is the OSCORE [18]. The protocol is efficient for severely constrained networks as it maintains the minor communication overhead possible. Using OSCORE, however, requires preshared keys to establish a security context. For this purpose, the IETF is in the process of standardizing an authenticated Diffie–Hellman key exchange protocol known as EDHOC [19]. The protocol is aforesaid to provide essential security properties such as mutual authentication, perfect forward secrecy, identity protection, and cipher suite negotiation.

EDHOC will most certainly dominate IoT security once it becomes a standard, which is why it is critical to analyse it for security vulnerabilities thoroughly. Since its inception in March 2016, it passed through 26 different versions, among which only two of its versions ([20] in 2018 and [21] in 2020) were formally analysed by [22, 23] using ProVerif [24] and Tamarin [25], respectively. While these studies bring numerous essential security issues to light, there are still security flaws that they have not yet discovered. Furthermore, evaluating security protocols using several formal approaches increases our confidence in the protocols' resilience to various security threats since one can compensate for the weakness of the other. Accordingly, in this paper, we formally analysed both the symmetric and asymmetric variants of the EDHOC protocol by using BAN (Burrows, Abadi, and Needham)-Logic [26] and AVISPA (Automated Validation of Internet Security Protocols and Applications) [27] to uncover other security issues. The formal verification results indicate that the protocols suffer from resource exhaustion and privacy attacks. While the former vulnerability is related to a class of attacks known as (distributed) denial-of-service attacks, where excessive and unnecessary requests deplete a node's resource, the latter pertains to privacy violations due to *ID_PSK* and ID_CRED_R (in symmetric and asymmetric variations, respectively). Both security issues are described in detail in Section 4 of the paper. The core contributions of this paper are summarized as follows:

(i) We carried out a formal security verification of the asymmetric and symmetric variants of the EDHOC protocol using two formal approaches: BAN-Logic and AVISPA

(ii) We pointed out two novel potential security vulnerabilities that may lead to resource exhaustion and privacy attacks

(iii) We described a concise summary of the principal security threats found by former related studies together with those identified by us

The remainder of the paper is organized as follows. Section 2 describes the EDHOC protocol along with the related studies on its formal security analysis. The formal verification of the protocol and results, respectively, are presented in the subsequent two sections. Finally, Section 5 concludes the paper.

2. The EDHOC Protocol

2.1. Protocol Overview. The increasing usage of IoT devices in vertical applications, such as energy, smart factory, healthcare, and transportation, calls for more efficient approaches to power, communication, storage, and processing. Given their severe constraint concerning these requirements, it is not possible (or inefficient) to apply existing security protocols. The main reason is due to the heavy cryptography algorithms, message sizes, and total round trips involved with these schemes. Implementing security on the application layer of the IoT communication systems is especially beneficial when there is insufficient security at the transport layer or when considering the performance of the communication is required. To this point, there are fundamental advances in providing application-aware security solutions. Some of these schemes are the CoAP [9] and its lightweight extension to provide sufficient object security, OSCORE [18].

Another vital protocol that serves as a lightweight authenticated key exchange mechanism for OSCORE is the EDHOC. The EDHOC protocol provides session key establishment while supporting fundamental security properties like perfect forward secrecy and mutual authentication [19]. The protocol involves essential components like Elliptic Curve Diffie–Hellman (ECDH) for key exchange, CBOR (Concise Binary Object Representation) [10] for data encoding, COSE (CBOR Object Signing and Encryption) [28] for protecting the CBOR encoded messages, and CoAP for message transportation. In summary, the primary intent of EDHOC is to leverage the OSCORE initiated security so that the message footprints and the round trips are small. Figure 1 shows the IoT protocol stack with the EDHOC protocol located in the application layer.

2.2. Related Works. Formal security analysis of various authentication protocols has been performed to guarantee the resilience of different security schemes against numerous

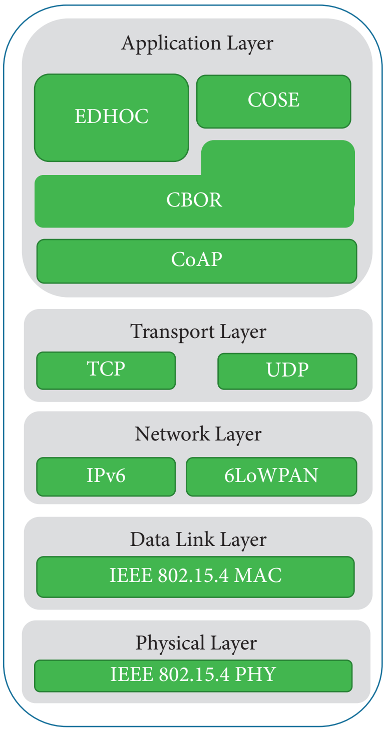


Figure 1: IoT protocol stack.

attacks. Concerning EDHOC, there are two significant studies that analysed the security of this protocol with a formal approach.

In [22], the authors formally analysed both symmetric-key and asymmetric-key options of the EDHOC protocol using ProVerif [24]. This research inspected the protocol against various security characteristics like identity protection against an active attacker, application data confidentiality and perfect forward secrecy, and robust authentication. Consequently, the authors highlighted the risk of leaking the responder's identity, although the initiator's identity is secured. Furthermore, by utilizing the same preshared key identifier *ID_PSK*, an attacker may link several sessions and launch various assaults to the symmetric variant of the protocol. Concerning the application data (AD_1 to AD_3), the authors also showed that only AD_3 (for both symmetric and asymmetric variants) satisfies secrecy, perfect forward secrecy, and integrity at both the time of message arrival and conclusion of the protocol.

Another paper [23] that analysed the EDHOC security using the Tamarin prover [25] verification tool found various improvement points. The authors, among other issues, identified the following flaws by the time they analysed the protocol: absence of nonrepudiation security property and lack of verification of ID_CRED_R of Msg2 by the initiator. The authors also showed that a security threat due to a prolonged metasession covering several sessions of the EDHOC protocol can happen when the responder rejects proposed cipher suites. The paper also recommended the use of a trusted execution environment (TEE) for security hardening.

2.3. Protocol Description. The initiator and responder of the EDHOC protocol can encrypt and protect the integrity of information communicated between them by following a similar construction as SIGMA-1 [29]. The initiator and responder exchange three messages to establish Diffie–Hellman's shared secrets and perform encryption using Authenticated Encryption with Associated Data (AEAD) [30]. Unique to EDHOC, however, new parameters like connection identifiers, transcript hashes, methods, and others exist. Moreover, EDHOC protocol works in two modes: asymmetric-key-based authentication technique that provides mutual authenticity via Diffie–Hellman shared ciphers and symmetric-key-based authentication that relies on preshared symmetric keys. Table 1 shows the parameters used in the EDHOC protocol.

2.3.1. Asymmetric-Key-Based EDHOC Protocol. The execution steps of an EDHOC protocol that uses asymmetric-keys are shown in Figure 2. Furthermore, to better understand and visualize the operations of both variants of the EDHOC protocol, we presented a state diagram as shown in Figures 3 and 4 . Take note that the figures show one session connection between the initiator and the responder.

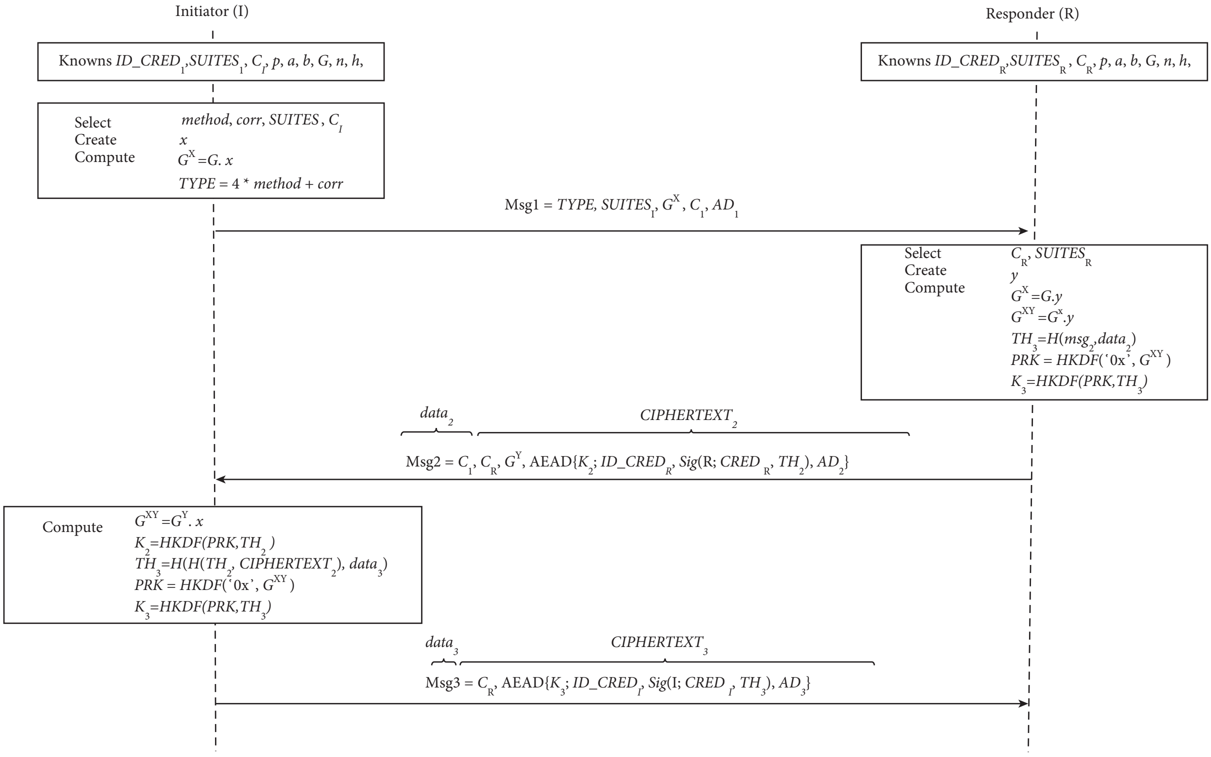
(1) Initiator ⟶ Responder. Before the commencement of the protocol, the initiator I stores the domain parameters for the agreed elliptic curve, ID_CRED_I, AD_1, and AD_2. Firstly, the caller generates a method that identifies the authentication method and the associated correlation (corr) of the transport mechanism. Here, "method" and "corr" take values from 0 to 3 as described in [19]. The initiator also chooses $\mathrm{SUITES_I}$ from the list of cipher suites that an EDHOC protocol recognizes and select the connection identifier C_I. It then picks a number x to serve as an ECDH private key. Once the preliminary information is ready, it computes the ECDH public key G_X ($=G.x$) and TYPE ($=4$ * method + corr). Finally, it constructs and sends $\mathrm{Msg_1}$ containing TYPE, $\mathrm{SUITES_I}$, G_X, C_I, and AD_1 to the responder. Note that AD_1, at this time, cannot guarantee security as it is transmitted in plaintext.

(2) Responder ⟶ Initiator. Once it receives Msg1, the responder selects a cipher suite SUITES_R and the connection identifier C_R, and calculates the ECDH public key G^Y ($=G.y$) the same way the initiator calculated G^X. It then calculates the ECDH shared key G^{XY} ($=G^X.y$). Subsequently, the transcript hash TH_2 is computed by hashing the received message Msg1 and $\mathrm{data_2}$, where $\mathrm{data_2}$ consists of the session identifiers C_I, C_R, and the ECDH public G^Y. The responder uses TH_2 for authentication. It then computes an encryption key K_2 (HKDF (PRK, G^{XY})) from the pseudorandom key PRK (HKDF ("0x," G^{XY})) and the transcript hash $\mathrm{TH_2}$. Next, the responder constructs Msg2 by concatenating $\mathrm{CIPHERTEXT_2}$ and $\mathrm{data_2}$.

The former message is formed by first signing $\mathrm{CRED_R}$ and $\mathrm{TH_2}$ with the responder's private key, followed by encrypting the signature, ID_CRED_R, and AD_2 with K_2. The latter message is simply the concatenation of C_I, C_R, and G^Y. Finally, the responder sends Msg2 to the initiator.

Table 1: Symbols and notations used in the EDHOC protocol.

Components	Description
Method	One of the four types of authentication methods agreed by the initiator and the responder.
Corr	One of the four types of correlation mechanisms provided by the transport path.
$SUITES\_I$, $SUITS\_R$	List of cipher suites (in order of preference) supported by the initiator and the responder, respectively.
x, y	The ECDH ephemeral private keys of the initiator and the responder, respectively.
G_X, G_Y	The ECDH ephemeral public keys of the initiator and the responder, respectively.
p	A prime number that states the size of the finite field.
a, b	The coefficients of the elliptic curve equation.
G	The generator (base point) of the subgroup.
h, n	The cofactor and order of the subgroup, respectively.
C_I, C_R	Connection identifiers for the initiator and responder, respectively, that are used to facilitate the retrieval of the protocol state.
AD	Application data (also known as external authorization data).
$CRED_I$, $CRED_R$	The credentials containing the public authentication keys of the initiator and the responder, respectively.
ID_CRED_I, ID_CRED_R	The identifiers for the credentials $CRED_I$ and $CRED_R$, respectively.
TH	Transcript hashes used for key derivation and additional authenticated data.
K	Session key.
PRK	Pseudorandom key.
PSK	Preshared key.
AEAD (K;)	Authenticated Encryption with Associated Data using a key K.
Sig (I; .), Sig(R; .)	Digital signatures made with the private authentication key of the initiator and the responder, respectively.



Initiator (I)
Responder (R)
Knowns $ID\_CRED_1, SUITES_1, C_I, p, a, b, G, n, h,$
Knowns $ID\_CRED_R, SUITES_R, C_R, p, a, b, G, n, h,$
Select
Create
Compute
$method, corr, SUITES, C_I$
x
$G^X = G.x$
$TYPE = 4 * method + corr$
Msg1 = $TYPE, SUITES_I, G^X, C_1, AD_1$
Select
Create
Compute
$C_R, SUITES_R$
y
$G^X = G.y$
$G^{XY} = G^x.y$
$TH_3 = H(msg_2, data_2)$
$PRK = HKDF('0x', G^{XY})$
$K_3 = HKDF(PRK, TH_3)$
$data_2$
$CIPHERTEXT_2$
Msg2 = C_1, C_R, G^Y, AEAD{K_2; ID_CRED_R, Sig(R; $CRED_R$, TH_2), AD_2}
Compute
$G^{XY} = G^Y.x$
$K_2 = HKDF(PRK, TH_2)$
$TH_3 = H(H(TH_2, CIPHERTEXT_2), data_3)$
$PRK = HKDF('0x', G^{XY})$
$K_3 = HKDF(PRK, TH_3)$
$data_3$
$CIPHERTEXT_3$
Msg3 = C_R, AEAD{K_3; ID_CRED_I, Sig(I; $CRED_I$, TH_3), AD_3}

Figure 2: Asymmetric-key-based EDHOC protocol.

(3) Initiator ⟶ Responder. When Msg2 reaches the initiator, it computes the ECDH shared key G^{XY} ($=G^{Y}.x$), PRK, TH_2, and K_2, like the responder. Then, it uses K_2 to decrypt CIPHERTEXT_2 and retrieve ID_CRED_I, the signature, and AD_2. It then validates the signature, and if the result succeeds, it generates TH_3 and K_3. The former is constructed by first hashing CIPHERTEXT_2 with TH_2 and then rehashing the result with C_R. The latter is computed by using HKDF (PRK, TH_3). Finally, the initiator constructs a message CIPHERTEXT_3 by signing $CRED_I$ and TH_3 with its private key and encrypting it together with ID_CRED_R and AD_3 using the computed session key K_3; it forms Msg3 by concatenating C_R and CIPHERTEXT_3 and sends it to the responder. The asymmetric-key-based EDHOC protocol

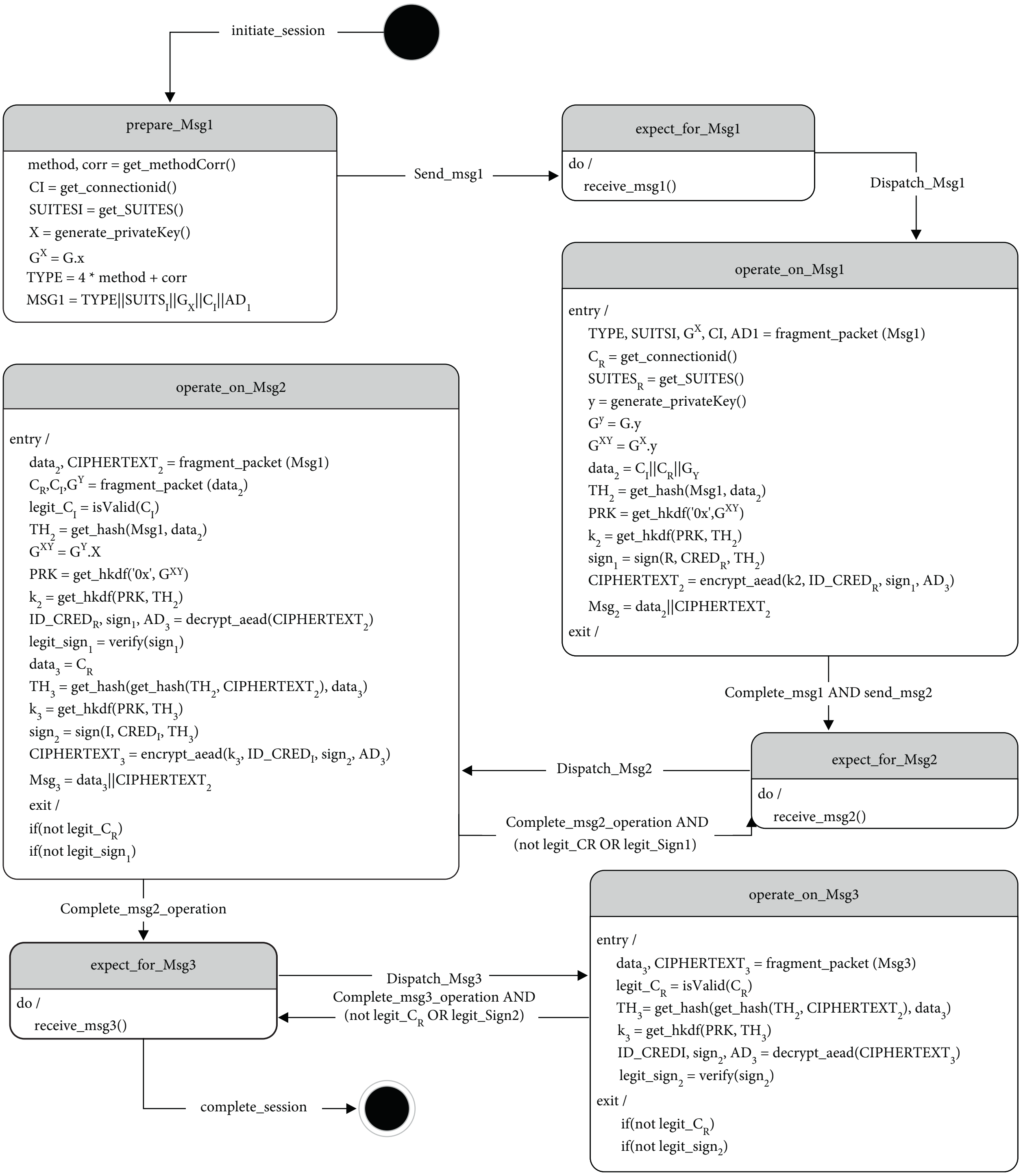

initiate_session
prepare_Msg1
method, corr = get_methodCorr()
CI = get_connectionid()
SUITESI = get_SUITES()
X = generate_privateKey()
G^X = G.x
TYPE = 4 * method + corr
MSG1 = TYPE||$SUITS_I$||G_X||C_I||AD_1
Send_msg1
expect_for_Msg1
do /
receive_msg1()
Dispatch_Msg1
operate_on_Msg1
entry /
TYPE, SUITSI, G^X, CI, AD1 = fragment_packet (Msg1)
C_R = get_connectionid()
$SUITES_R$ = get_SUITES()
y = generate_privateKey()
G^y = G.y
G^{XY} = G^X.y
$data_2$ = C_I||C_R||G_Y
TH_2 = get_hash(Msg1, $data_2$)
PRK = get_hkdf('0x',G^{XY})
k_2 = get_hkdf(PRK, TH_2)
$sign_1$ = sign(R, $CRED_R$, TH_2)
$CIPHERTEXT_2$ = encrypt_aead(k2, ID_$CRED_R$, $sign_1$, AD_3)
Msg_2 = $data_2$||$CIPHERTEXT_2$
exit /
Complete_msg1 AND send_msg2
expect_for_Msg2
do /
receive_msg2()
Dispatch_Msg2
operate_on_Msg2
entry /
$data_2$, $CIPHERTEXT_2$ = fragment_packet (Msg1)
C_R,C_I,G^Y = fragment_packet ($data_2$)
legit_C_I = isValid(C_I)
TH_2 = get_hash(Msg1, $data_2$)
G^{XY} = G^Y.X
PRK = get_hkdf('0x', G^{XY})
k_2 = get_hkdf(PRK, TH_2)
ID_$CRED_R$, $sign_1$, AD_3 = decrypt_aead($CIPHERTEXT_2$)
legit_$sign_1$ = verify($sign_1$)
$data_3$ = C_R
TH_3 = get_hash(get_hash(TH_2, $CIPHERTEXT_2$), $data_3$)
k_3 = get_hkdf(PRK, TH_3)
$sign_2$ = sign(I, $CRED_I$, TH_3)
$CIPHERTEXT_3$ = encrypt_aead(k_3, ID_$CRED_I$, $sign_2$, AD_3)
Msg_3 = $data_3$||$CIPHERTEXT_2$
exit /
if(not legit_C_R)
if(not legit_$sign_1$)
Complete_msg2_operation AND
(not legit_CR OR legit_Sign1)
Complete_msg2_operation
expect_for_Msg3
do /
receive_msg3()
Dispatch_Msg3
Complete_msg3_operation AND
(not legit_C_R OR legit_Sign2)
operate_on_Msg3
entry /
$data_3$, $CIPHERTEXT_3$ = fragment_packet (Msg3)
legit_C_R = isValid(C_R)
TH_3= get_hash(get_hash(TH_2, $CIPHERTEXT_2$), $data_3$)
k_3 = get_hkdf(PRK, TH_3)
ID_CREDI, $sign_2$, AD_3 = decrypt_aead($CIPHERTEXT_3$)
legit_$sign_2$ = verify($sign_2$)
exit /
if(not legit_C_R)
if(not legit_$sign_2$)
complete_session

FIGURE 3: State diagram for the asymmetric-key-based EDHOC protocol.

concludes by removing the Diffie–Hellman key pairs used to generate the encryption keys K_2 and K_3 to support perfect forward secrecy.

2.3.2. Symmetric-Key-Based EDHOC Protocol. The symmetric-key-based EDHOC protocol, shown in Figure 5, is very similar to the asymmetric-key-based protocol, with the following exceptions:

(1) The public key identifiers ID_CRED_I and ID_CRED_R are not used as part of the authenticated encryption

(2) Authentication happens via preshared key *PSK* (identified by *ID_PSK*) rather than the digital signatures used in the previous protocol

(3) The protocol session keys K_2 and K_3 are derived based on Diffie–Hellman shared keys, transcript hashes, and preshared keys *PSK*

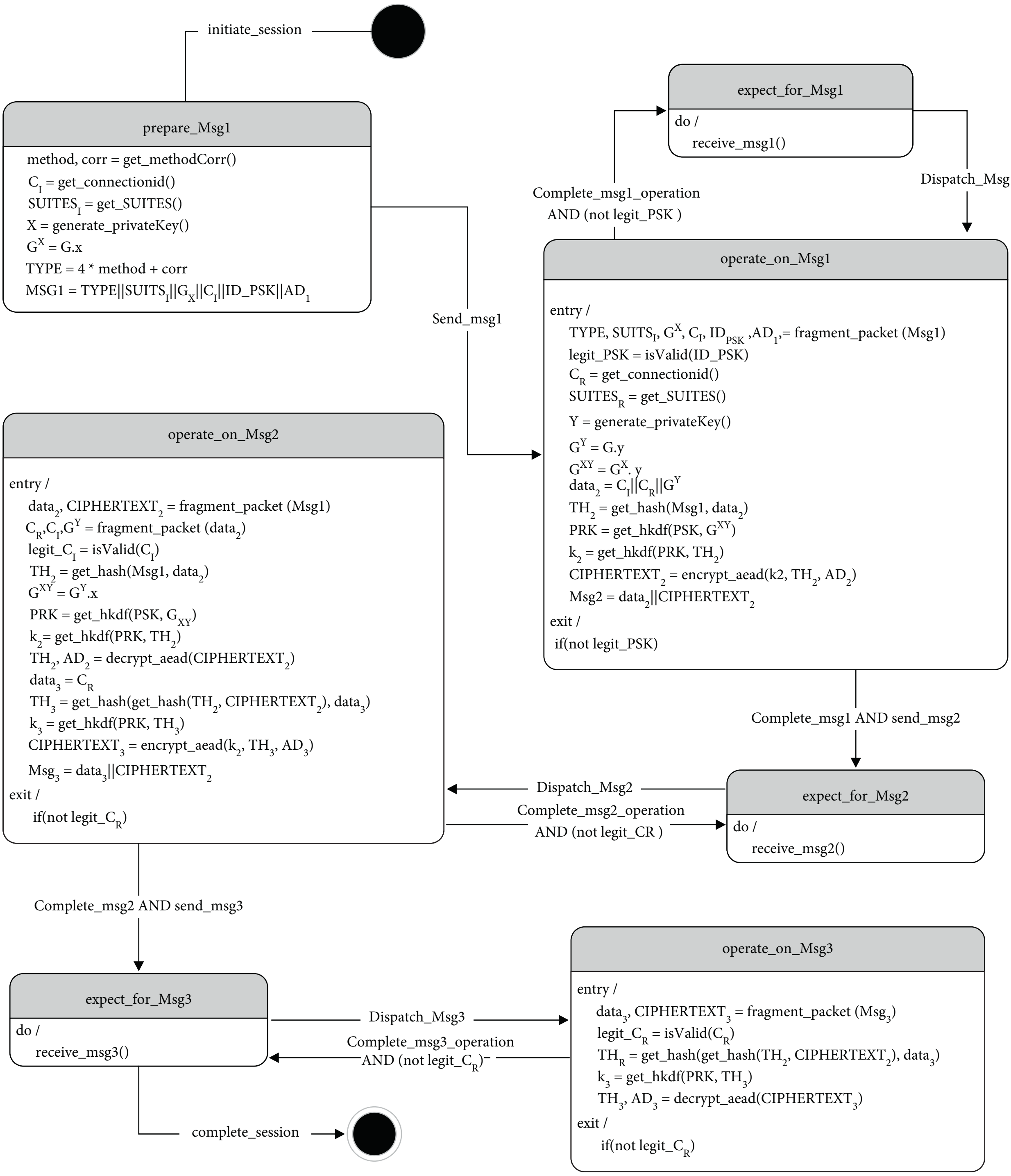

initiate_session
prepare_Msg1
method, corr = get_methodCorr()
C_I = get_connectionid()
$SUITES_I$ = get_SUITES()
X = generate_privateKey()
G^X = G.x
TYPE = 4 * method + corr
MSG1 = TYPE||$SUITS_I$||G_X||C_I||ID_PSK||AD_1
Send_msg1
expect_for_Msg1
do /
receive_msg1()
Complete_msg1_operation
AND (not legit_PSK)
Dispatch_Msg1
operate_on_Msg1
entry /
TYPE, $SUITS_I$, G^X, C_I, ID_{PSK} ,AD_1,= fragment_packet (Msg1)
legit_PSK = isValid(ID_PSK)
C_R = get_connectionid()
$SUITES_R$ = get_SUITES()
Y = generate_privateKey()
G^Y = G.y
G^{XY} = G^X. y
$data_2$ = C_I||C_R||G^Y
TH_2 = get_hash(Msg1, $data_2$)
PRK = get_hkdf(PSK, G^{XY})
k_2 = get_hkdf(PRK, TH_2)
$CIPHERTEXT_2$ = encrypt_aead(k2, TH_2, AD_2)
Msg2 = $data_2$||$CIPHERTEXT_2$
exit /
if(not legit_PSK)
operate_on_Msg2
entry /
$data_2$, $CIPHERTEXT_2$ = fragment_packet (Msg1)
C_R,C_I,G^Y = fragment_packet ($data_2$)
legit_C_I = isValid(C_I)
TH_2 = get_hash(Msg1, $data_2$)
G^{XY} = G^Y.x
PRK = get_hkdf(PSK, G_{XY})
k_2= get_hkdf(PRK, TH_2)
TH_2, AD_2 = decrypt_aead($CIPHERTEXT_2$)
$data_3$ = C_R
TH_3 = get_hash(get_hash(TH_2, $CIPHERTEXT_2$), $data_3$)
k_3 = get_hkdf(PRK, TH_3)
$CIPHERTEXT_3$ = encrypt_aead(k_2, TH_3, AD_3)
Msg_3 = $data_3$||$CIPHERTEXT_2$
exit /
if(not legit_C_R)
Complete_msg1 AND send_msg2
Dispatch_Msg2
Complete_msg2_operation
AND (not legit_CR)
expect_for_Msg2
do /
receive_msg2()
Complete_msg2 AND send_msg3
expect_for_Msg3
do /
receive_msg3()
Dispatch_Msg3
Complete_msg3_operation
AND (not legit_C_R)
operate_on_Msg3
entry /
$data_3$, $CIPHERTEXT_3$ = fragment_packet (Msg_3)
legit_C_R = isValid(C_R)
TH_R = get_hash(get_hash(TH_2, $CIPHERTEXT_2$), $data_3$)
k_3 = get_hkdf(PRK, TH_3)
TH_3, AD_3 = decrypt_aead($CIPHERTEXT_3$)
exit /
if(not legit_C_R)
complete_session

Figure 4: State diagram for the symmetric-key-based EDHOC protocol.

3. Formal Security Verification for EDHOC Protocol

This section describes the formal security verification of both variants of the EDHOC protocol. First, we leverage BAN-Logic to analyse any security flaw that may exist in the protocol. Next, to further strengthen the verification result and complement the weakness of the first approach, we will use the AVISPA tool.

3.1. BAN-Logic-Based Formal Verification. BAN-Logic is a modal logic of beliefs (proposed by Burrows, Abadi, and Needham) used to verify authentication protocols in a formal manner [26, 31]. The formal description of the authentication process, participants' knowledge, and beliefs serve as a foundation for analysing the changes at each level of the protocol. BAN-Logic is the most utilized approach for examining various security protocols due to its simplicity and robustness.

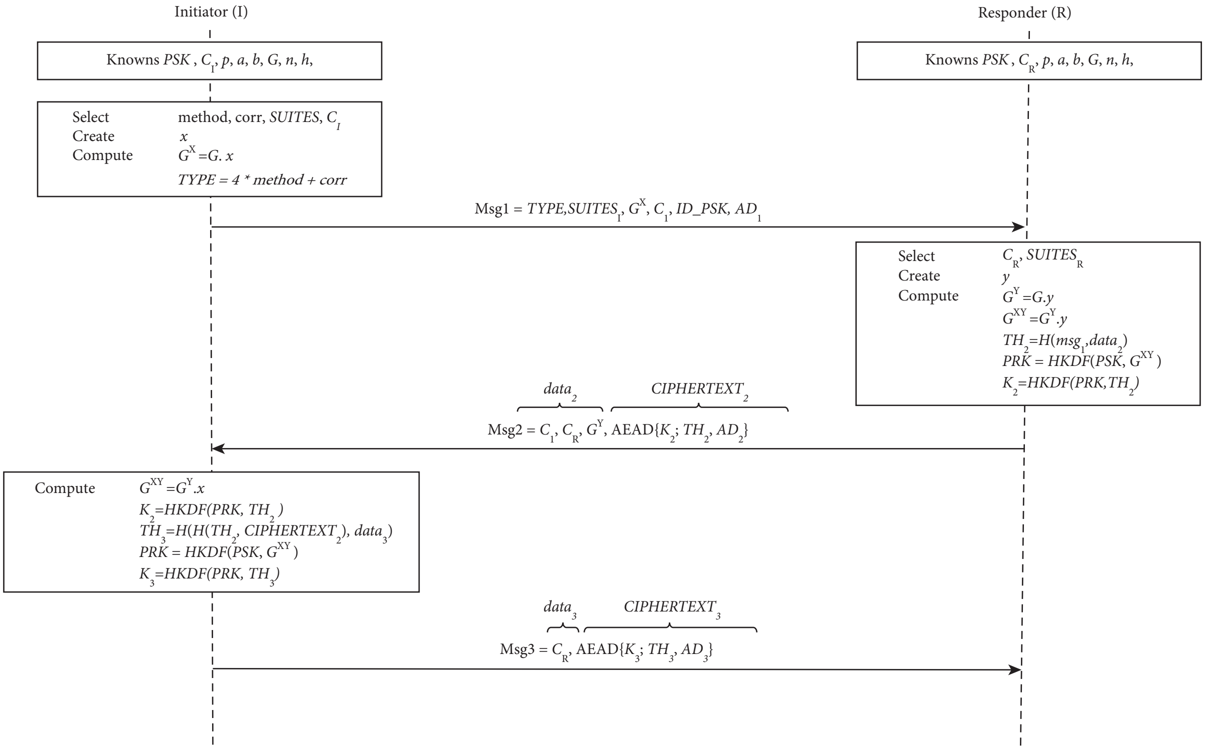


Figure 5: Symmetric-key-based EDHOC protocol.

Verification of security protocols using this method starts with converting the protocol into an idealized form through idealization. Here, only protected messages, traversing from one participant to another, are of interest. Then, realistic assumptions and security objectives that the protocol should guarantee proceed. Subsequently, the derivation of the security goals continues by applying different BAN-Logic rules, the premises, and the intermediate results of the derivation. Tables 2 and 3 describe the symbols and formulas used in the BAN-Logic formalization process, respectively.

3.1.1. The Asymmetric-Key Option

Idealization. An idealized version of the asymmetric form of the EDHOC protocol is shown below. Note that the idealized form only comprises encrypted (protected) communications, which is why Msg1 is left out.

$$R \longrightarrow I: \left\{ ID\_CRED_R, \left\{ CRED_R, \xrightarrow{G^Y} R, \xrightarrow{G^X} I, I \overset{G^{XY}}{\leftrightarrow} R \right\}_{R^{-1}}, AD_2 \right\}_{G^{XY}}, \tag{1}$$

$$I \longrightarrow R: \left\{ ID\_CRED_I, \left\{ CRED_R, \xrightarrow{G^Y} R, \xrightarrow{G^X} I, I \overset{G^{XY}}{\leftrightarrow} R \right\}_{R^{-1}}, AD_3 \right\}_{G^{XY}}. \tag{2}$$

Goals. The following objectives are established for verifying mutual authentication and key exchange between *I* and *R*. Consequently, while the goals in (5) and (6) show the beliefs *I* has on *R*'s trust concerning its credential identity and the associated data (respectively), (9) and (10) show the opposite. About key exchange, the goals in (3) and (4) show *I*'s belief on the session key, and (7) and (8) assert *R*'s belief on the same key.

Table 2: BAN-Logic notations.

Notation	Meaning
$R\,believes\,M$	R believes that message M is true
R sees M	R receives message M at any point in time
R said M	R previously sent message M
R controls M	R has jurisdiction over M
$Fresh\ (M)$	M is fresh
$R\overset{S}{\leftrightarrow} I$	S is a secret key shared between M and N
$\xrightarrow{S} R$	S is R's public key
$R\overset{S}{\Leftrightarrow} I$	S is a secret that R and I share
$\{M\}_K$	M is a message encrypted with a key K
M, V	M is combined with V

Table 3: BAN-Logic rules.

Rule name	Rule
Message meaning rule (MM)	$(R \text{ believes } R\overset{S}{\leftrightarrow} I,\ R \text{ sees } \{M\}_S\,/R \text{ believes } I \text{ said } M)$ $(R \text{ believes } R\overset{S}{\Leftrightarrow} I,\ R \text{ sees } M_S\,/R \text{ believes } I \text{ said } M)$ $(R \text{ believes } \xrightarrow{S} I,\ R \text{ sees } \{M\}_{S^{-1}}\,/R \text{ believes } I \text{ said } M)$
Nonce verification (NV) rule	$(R \text{ believes } \#\,(M), R \text{ believes } I \text{ said } M\,/R \text{ believes } I \text{ believes } M)$
Jurisdiction (JR) rule	$(R \text{ believes } I \text{ controls } K,\ R \text{ believes } I \text{ believes } K/R \text{ believes } K)$
Freshness (FR) rule	$(R \text{ believes fresh } (M)/R \text{ believes fresh } (M, Q))$
Decomposition (DR) rule	$(R \text{ sees } (M, Q)/R \text{ sees } M)$
Belief conjunction (BC) rule	$(R \text{ believes } M, R \text{ believes } Q\,/R \text{ believes } (M, Q))$ $(R \text{ believes } I \text{ believes } (M, Q)\,/R \text{ believes } I \text{ believes } M)$ $(R \text{ believes } I \text{ said } (M, Q)\,/R \text{ believes } I \text{ said } M)$
Diffie–Hellman (DH) rule	$(R \text{ believes } I \text{ said } \xrightarrow{G^M} I, R \text{ believes } \xrightarrow{G^Q} R\,/R \text{ believes } R\overset{g^{MQ}}{\leftrightarrow} I)$ $(R \text{ believes } I \text{ said } \xrightarrow{G^M} I, R \text{ believes } \xrightarrow{G^Q} R\,/R \text{ believes } R\overset{G^{MQ}}{\Leftrightarrow} G^{MQ} I)$

$$I \text{ believes } I\overset{G^{XY}}{\longleftrightarrow} R, \tag{3}$$

$$I \text{ believes } R \text{ believes } I\overset{G^{XY}}{\longleftrightarrow} R, \tag{4}$$

$$I\ believes\ R\ believes\ AD_2, \tag{5}$$

$$I\ believes\ R\ believes\ ID\_CRED_R, \tag{6}$$

$$R \text{ believes } I\overset{G^{XY}}{\leftrightarrow} R, \tag{7}$$

$$R \text{ believes } I \text{ believes } I\overset{G^{XY}}{\leftrightarrow} R, \tag{8}$$

$$R \text{ believes } I\ believes\ AD_3 \tag{9}$$

$$R\ believes\ I\ believes\ ID\_CRED_I. \tag{10}$$

Assumptions. There are some assumptions and hypotheses we need to set to derive the above goals. Accordingly, the assumptions in (11), (15), and (16) show R's belief in its ECDH public key, the long-term public key of I, and the freshness of its ECDH public key. On the other hand, while (12) and (14) point out I's belief in its ECDH public key and its freshness, (13) indicates the belief "I" has in R's long-term public key. Finally, the two hypotheses (17) and (18) imply that R trusts that I sent its ECDH public key and vice versa, respectively.

$$R \text{ believes } \xrightarrow{G^Y} R, \tag{11}$$

$$I \text{ believes } \xrightarrow{G^X} I, \tag{12}$$

$$I \text{ believes } \xrightarrow{P} U(R)\ R, \tag{13}$$

$$I \text{ believes } \# \xrightarrow{G^X} I, \tag{14}$$

$$R \text{ believes } \xrightarrow{P} U(I)\ I, \tag{15}$$

$$R \text{ believes } \#\left(\xrightarrow{G^Y} R\right), \tag{16}$$

$$R \text{ believes } I \text{ said } \xrightarrow{G^X} I, \tag{17}$$

$$I \text{ believes } R \text{ said } \xrightarrow{G^Y} R. \tag{18}$$

Derivations. As a final step of the formal analysis, derivation of goals proceeds. To do so, we leverage the BAN-Logic rules (shown in Table 3), idealizations, assumptions, and the intermediate results of the derivation process. Therefore, if all goals can be derived, the target protocol is considered secure. Otherwise, the protocol may be vulnerable to threats.

$$R \text{ believes } I \overset{G^{XY}}{\leftrightarrow} R \text{ by } (11),\ (17),\ DH, \tag{19}$$

$$I \text{ believes } I \overset{G^{XY}}{\leftrightarrow} R\, by\ (12),\ (18),\ DH, \tag{20}$$

$$I \text{ sees } \left\{ID\_CRED_R,\ \left\{CRED_R,\ \xrightarrow{G^Y} R\ ,\ \xrightarrow{G^X} I, I \overset{G^{XY}}{\leftrightarrow} R\right\}_{R^{-1}},\ AD_2\right\}_{G^{XY}} \text{from}\ (1), \tag{21}$$

$$I \text{ believes } R \text{ said} \left[\begin{array}{l} \qquad\qquad\qquad\qquad ID_{CREDR}, \\ \left\{CRED_R,\ \xrightarrow{G^Y} R\ ,\ \xrightarrow{G^X} I, I \overset{G^{XY}}{\leftrightarrow} R\right\}_{R^{-1}},\ AD_2 \end{array}\right] \text{by}\ (20),\ (21), MM, \tag{22}$$

$$I \text{ sees } \left\{CRED_R,\ \xrightarrow{G^Y} R\ ,\ \xrightarrow{G^X} I, I \overset{G^{XY}}{\leftrightarrow} R\right\}_{R^{-1}} \text{by}\ \ (22), \tag{23}$$

$$I \text{ believes } R \text{ said} \left[CRED_R,\ \xrightarrow{G^Y} R\ ,\ \xrightarrow{G^X} I, I \overset{G^{XY}}{\leftrightarrow} R\right] \text{ by } (23),\ (13), MM, \tag{24}$$

$$I \text{ belives } R \text{ believes } \left[CRED_R,\ \xrightarrow{G^Y} R\ ,\ \xrightarrow{G^X} I, I \overset{G^{XY}}{\leftrightarrow} R\right] \text{ by } (24),\ (14),\ FR,\ NV, \tag{25}$$

$$I \text{ believes } R \text{ believes } I \overset{G^{XY}}{\leftrightarrow} R \text{ by } (25),\ BC, \tag{26}$$

$$I \text{ believes } R \text{ believes } \xrightarrow{G^Y} R \text{ by } (25),\ BC, \tag{27}$$

$$I \text{ believes } R \text{ believes } ID\_CRED_R \text{ by } (22),\ (14),\ FR,\ NV,\ BC, \tag{28}$$

$$I \text{ believes } R \text{ believes } AD_2 \text{ by } (22),\ (14),\ FR,\ NV,\ BC, \tag{29}$$

$$R \text{ sees } \left\{ID\_CRED_I,\ \left\{CRED_I,\ \xrightarrow{G^Y} R\ ,\ \xrightarrow{G^X} I, I \overset{G^{XY}}{\leftrightarrow} R\right\}_{R^{-1}},\ AD_3\right\}_{G^{XY}} \text{from } (2), \tag{30}$$

$$R \text{ believes } I \text{ said} \left[ID_{CREDI},\ \left\{CRED_I,\ \xrightarrow{G^Y} R\ ,\ \xrightarrow{G^X} I, I \overset{G^{XY}}{\leftrightarrow} R\right\}_{I^{-1}},\ AD_3\right] \text{ by } (19),\ (30),\ MM, \tag{31}$$

$$R \text{ sees } \left\{CRED_I,\ \xrightarrow{G^Y} R\ ,\ \xrightarrow{G^X} I, I \overset{G^{XY}}{\leftrightarrow} R\right\}_{R^{-1}} \text{by } (31),\ BC, \tag{32}$$

$$R \text{ believes } I \text{ said} \left[CRED_I,\ \xrightarrow{G^Y} R\ ,\ \xrightarrow{G^X} I, I \overset{G^{XY}}{\leftrightarrow} R\right] \text{ by } (32),\ (15),\ MM, \tag{33}$$

$$R \text{ believes } I \text{ believes } \left[CRED_I,\ \xrightarrow{G^Y} R\ ,\ \xrightarrow{G^X} I, I \overset{G^{XY}}{\leftrightarrow} R\right] \text{ by } (33),\ (16),\ FR,\ NV, \tag{34}$$

$$R \text{ believes } I \text{ believes } I \overset{G^{XY}}{\leftrightarrow} R \text{ by } (34),\ BC, \tag{35}$$

$$R \text{ believes } I \text{ believes } \xrightarrow{G^X} I \text{ by } (34),\ BC, \tag{36}$$

$$R \text{ believes } I \text{ believes } ID\_CRED_I \text{ by } (31),\ (16),\ FR,\ NV, \tag{37}$$

$$R \text{ believes } I \text{ believes } AD_3 \text{ by } (31),\ (16),\ FR,\ NV, \tag{38}$$

Note that, without the two hypotheses in (17) and (18), this derivation should stop before (22). In other words, only if both hypotheses are true, the proposed protocol can achieve the goals in (3)~(10). Unfortunately, they cannot hold because the two parties have no trust in each other's ECDH public key. Therefore, we conclude that asymmetric-key option is not secure.

3.1.2. The Symmetric-Key Option

Idealization. The idealization forms of the symmetric-key option of EDHOC protocol are shown as follows:

$$R \longrightarrow I:\ \left\{AD_2,\ \xrightarrow{G^Y} R,\ \xrightarrow{G^X} I, I \overset{G^{XY}}{\leftrightarrow} R\right\}_{G^{XY}}, \tag{39}$$

$$I \longrightarrow R:\ \left\{AD_3,\ \xrightarrow{G^Y} R,\ \xrightarrow{G^X} I, I \overset{G^{XY}}{\leftrightarrow} R\right\}_{G^{XY}}. \tag{40}$$

Goals. In general, the goals involve the guarantee of secure key exchange and mutual authentication. In the former case, while (41) and (42) form the belief of I in the ECDH session key, (44) and (45) represent the same case for R. In the latter point, the goals in (43), (46), and (47) serve to verify the mutual authentication.

$$I \text{ believes } I \overset{G^{XY}}{\leftrightarrow} R, \tag{41}$$

$$I \text{ believes } R \text{ believes } I \overset{G^{XY}}{\leftrightarrow} R, \tag{42}$$

$$I \text{ believes } R \text{ believes } AD_2, \tag{43}$$

$$R \text{ believes } I \overset{G^{XY}}{\leftrightarrow} R, \tag{44}$$

$$R \text{ believes } I \textit{ believes } I \overset{G^{XY}}{\leftrightarrow} R, \tag{45}$$

$$R \text{ believes } I \text{ believes } AD_3, \tag{46}$$

$$R \text{ believes } I \text{ believes } ID\_PSK. \tag{47}$$

Assumptions. While the assumptions in (49) and (50) show I's belief concerning its ECDH public key and its freshness (respectively), (48) and (51) do the same for R (respectively). Moreover, the symmetric-key option of the EDHOC protocol also requires the same additional hypotheses in (52) and (53) as the asymmetric-key option.

$$R \text{ believes } \xrightarrow{G^Y} R, \tag{48}$$

$$I \text{ believes } \xrightarrow{G^X} I, \tag{49}$$

$$I \text{ believes } \#\left(\ \xrightarrow{G^X} I\ \right), \tag{50}$$

$$R \text{ believes } \#\left(\ \xrightarrow{G^Y} R\ \right), \tag{51}$$

$$R \text{ believes } I \text{ said } \xrightarrow{G^X} I, \tag{52}$$

$$I \text{ believes } R \text{ said } \xrightarrow{G^X} R. \tag{53}$$

Derivations. The derivations of this variant of the EDHOC protocol proceed as follows:

$$R \text{ believes } I \overset{G^{XY}}{\leftrightarrow} R \text{ by } (48),\ (52),\ DH, \tag{54}$$

$$I \text{ believes } I \overset{G^{XY}}{\leftrightarrow} R \text{ by } (49),\ (53),\ DH, \tag{55}$$

$$I \text{ sees } \left\{AD_2,\ \xrightarrow{G^Y} R,\ \xrightarrow{G^X} I, I \overset{G^{XY}}{\leftrightarrow} R\right\}_{G^{XY}} \text{from } (39), \tag{56}$$

$$I \text{ believes } R \text{ said } \left[AD_2,\ \xrightarrow{G^Y} R,\ \xrightarrow{G^X} I, I \overset{G^{XY}}{\leftrightarrow} R\right] \text{ by } (55),\ (56),\ MM, \tag{57}$$

$$I \text{ believes } R \text{ believes } \left[AD_2,\ \xrightarrow{G^Y} R,\ \xrightarrow{G^X} I, I \overset{G^{XY}}{\leftrightarrow} R\right] \text{ by } (57),\ (50),\ FR,\ NV, \tag{58}$$

$$I \text{ believes } R \text{ believes } I \overset{G^{XY}}{\leftrightarrow} R \text{ by } (58),\ BC, \tag{59}$$

$$I \text{ believes } R \text{ believes } AD_2 \text{ by } (58), BC, \tag{60}$$

$$R \text{ sees } \left\{AD_3,\ \xrightarrow{G^Y} R,\ \xrightarrow{G^X} I,\ I \overset{G^{XY}}{\leftrightarrow} R\right\}_{G^{XY}} \text{from } (40), \tag{61}$$

$$R \text{ believes } I \text{ said } \left[AD_3,\ \xrightarrow{G^Y} R,\ \xrightarrow{G^X} I, I \overset{G^{XY}}{\leftrightarrow} R\right] \text{ by } (54),\ (61),\ MM, \tag{62}$$

$$R \text{ believes } I \text{ believes } \left[AD_3,\ \xrightarrow{G^Y} R,\ \xrightarrow{G^X} I, I \overset{G^{XY}}{\leftrightarrow} R\right] \text{ by } (62),\ (51),\ FR,\ NV, \tag{63}$$

$$R \text{ believes } I \text{ believes } I \overset{G^{XY}}{\leftrightarrow} R \text{ by } (63),\ BC, \tag{64}$$

$$R \text{ believes } I \text{ believes } \xrightarrow{G_X} I \text{ by } (63),\ BC, \tag{65}$$

$$R \text{ believes } I \text{ believes } AD_3 \text{ by } (63),\ BC, \tag{66}$$

Similar to the asymmetric-key option, the symmetric-key option can achieve the goals in (41)~(47) in the case that the two hypotheses in (52) and (53) are true. Thus, the hypotheses, which cannot be proved to be true, show that this option fails to satisfy the goals in (41)~(47). On the other hand, the last goal (47) indicating the privacy property cannot be achieved because ID_PSK is sent without being encrypted in the first message as shown in Figure 5.

To realize mutual authentication between "I" and "R," the former must believe the latter's ECDH public key, and it also must believe that the latter believes this key, and vice versa. That is, the derivations [(54), (55), (61), and (65)] of the asymmetric-key option and [(54), (55), (61), and (65)] of the symmetric-key option need to be satisfied. However, since all these derivations are entirely dependent on the fact that "I" ("R") believes "R" ("I") sent the ECDH public keys, it automatically follows that mutual authentication can only be fulfilled when these hypotheses are satisfied.

Messages Msg2 and Msg3, as illustrated in the idealizations in (39) and (40), use the ECDH session key to derive the AEAD encryption keys K_2 and K_3. Consequently, this can only happen through the hypotheses in (52) and (53) and is illustrated in derivations (59) and (64). Consequently, it is impossible to conclude that the session key is successfully communicated in the present asymmetric-key form of the protocol. Similarly, the symmetric version of the EDHOC protocol also fails to successfully exchange the session key without the hypotheses in (52) and (53). Thus, the derivations in (54) and (55) for "I" and "R" to believe the session key, respectively, require the use of the hypotheses.

Perfect forward secrecy is a characteristic of robust protocols because it protects previous sessions from future key compromise attempts. Accordingly, the asymmetric variant of the EDHOC protocol leverages the unique generation of ECDH private keys for each session of the protocol run to realize perfect forward secrecy. Likewise, in the symmetric-key option of the EDHOC protocol, the generation of the secret keys K_2 and K_3 uses the nonstatic Diffie–Hellman session key between "I" and "R." Thus, the symmetric-key option of the EDHOC protocol also provides perfect forward secrecy.

For both symmetric and asymmetric alternatives of EDHOC protocol to provide confidentiality and integrity, secure session key exchange must be in place. However, "I" and "R" may fail to transfer this key securely, as described earlier. As a result, the protocol cannot guarantee both confidentiality and integrity security properties.

Finally, due to the absence of authentication for the initial message, the anonymity of the responder's identifier for the public authentication keys (ID_CRED_R, for asymmetric-key option) and preshared key (ID_PSK, for the symmetric-key option) can be exposed.

Table 4 summarizes the result of the BAN-Logic derivation process for both options of the EDHOC protocol. As illustrated in the table and explanation above, both options of the protocol are insecure.

3.2. AVISPA-Based Formal Verification.

AVISPA is an automation tool for modelling and analysing security protocols [27]. The description of the formal verification process using AVISPA proceeds as follows. First, we use a High-Level Protocol Specification Language (HLPSL) [32] to model the protocol. The HLPSL2IF component then converts the HLPSL-modelled protocols to Intermediate Format (IF). Finally, using the On-the-Fly Model-Checker (OFMC) [33], CL-based Attack Searcher (CL-AtSe) [34], SAT-based Model-Checker (SATMC) [35], and Tree-Automata-based Protocol Analyzer (TA4SP) [36], the IF is transformed to Output Format (OF). Figure 6 shows the general system

Table 4: Security property satisfaction.

No.	Security properties	Asymmetric-key option	Symmetric-key option
SP1	Mutual authentication	X	X
SP2	Secure key exchange	X	X
SP3	Perfect forward secrecy	✓	✓
SP4	Confidentiality	X	X
SP5	Integrity	X	X
SP6	Anonymity	X	X

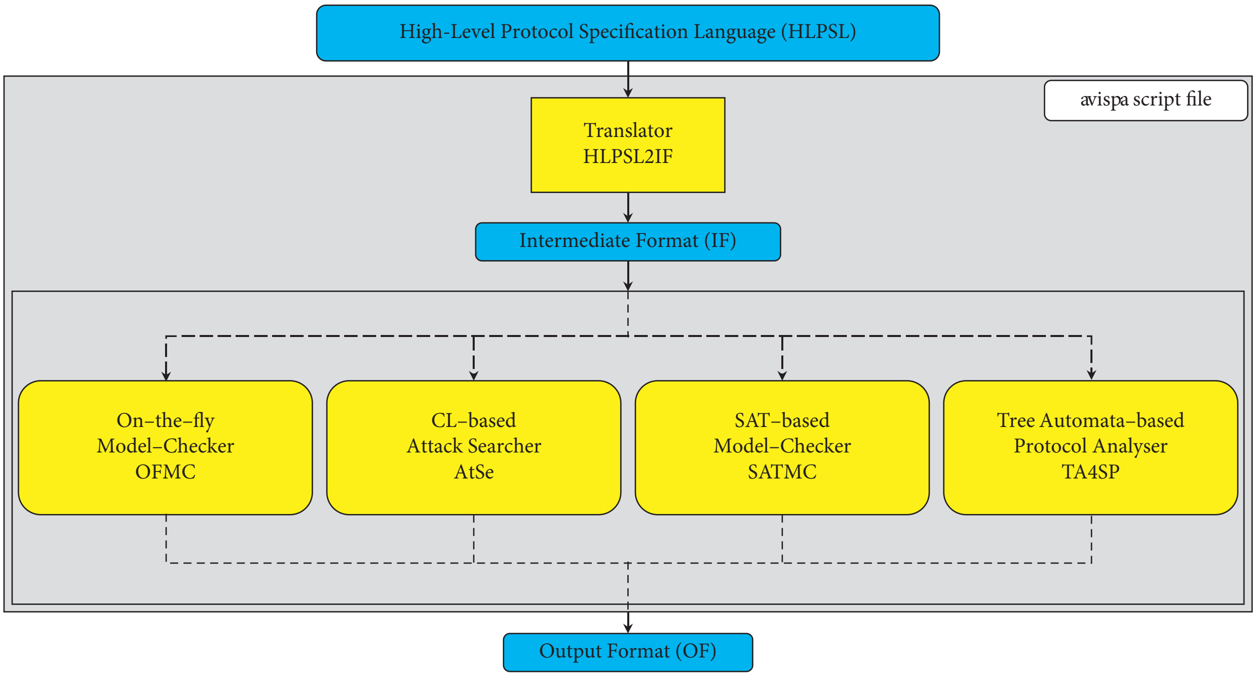


Figure 6: AVISPA system structure.

architecture of the tool, highlighting the main processes from HLPSL to OF.

HLPSL is composed of different roles such as Basic Role, Session Role, and Environment Role:

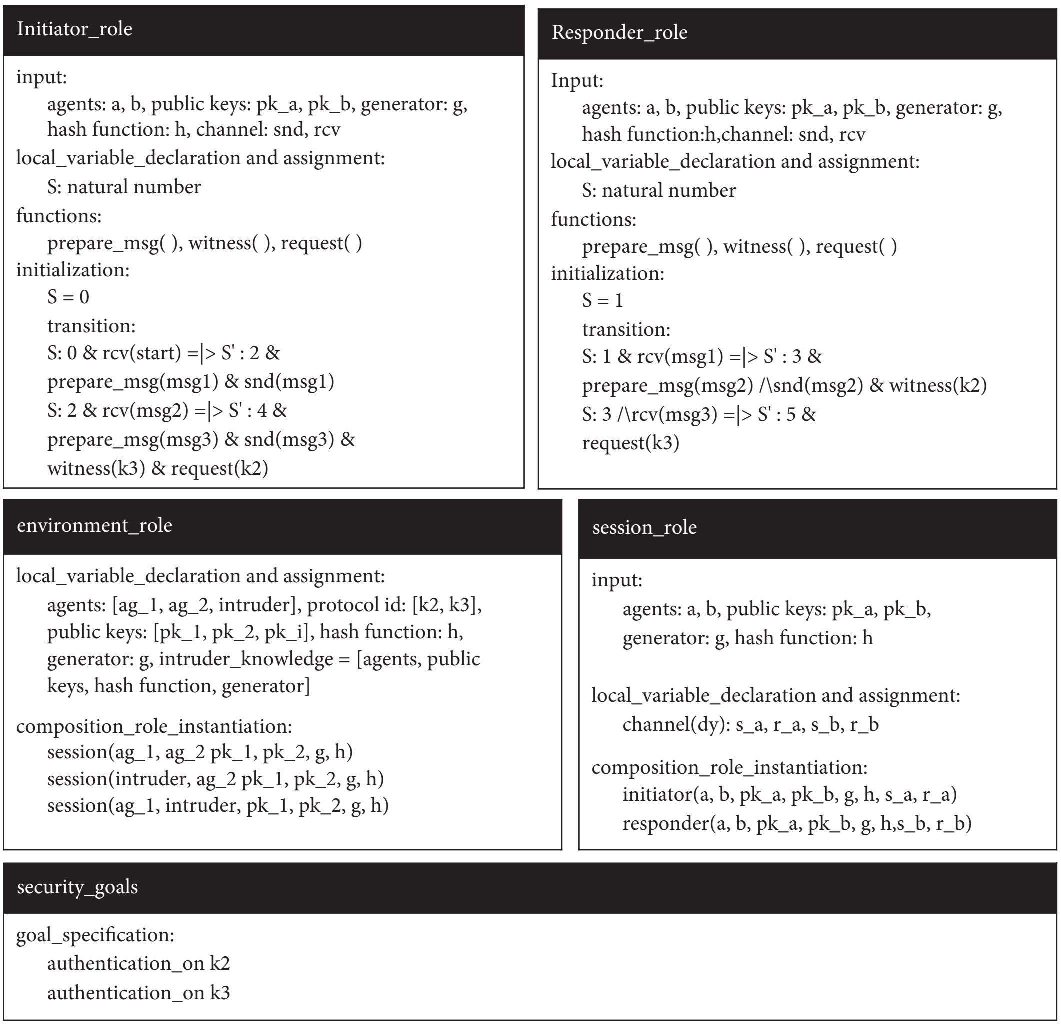
(i) *Basic Role*. This is a role that models protocol participants in a function with parameters. It consists of steps such as header expression, local variable declaration, and initialization. Additionally, it identifies communication modelling, which specifies the channel for communication between the modelled participants and indicates real-world protocol behaviour. It also defines, together with these parameters, transitions that denote message reception and the corresponding reply of the agent.

(ii) *Session Role*. This function receives the agents and other parameters to activate the previous role. It is executed via a composition to instantiate the parts in a parallel manner. /\ represents such parallel execution of prior roles.

(iii) *Environment Role*. It is a role that comprises global constants with the agents and sessions defined in the above two roles. In addition, it outlines an attacker's knowledge of the protocol's communication. The intruder's knowledge concerning the execution of the protocol is also defined. Like the Session Role, parallel execution of sessions executes with the intruder's information considered. Once the Environment Role completes, the security goals follow, and their verification proceeds with OFMC and CL-AtSe submodules.

3.2.1. *The Asymmetric-Key Option.*

At first, the asymmetric-key option of EDHOC protocol is modelled in HLPSL code. The code specifies the initiator and the responder roles with their security goals, Session Role to activate the basic roles, and finally the Environment Role. The source code for the AVISPA verification for both asymmetric and symmetric variants can be found in the Supplementary Materials (available here), while the pseudocodes are presented in Figure 7 (for asymmetric variant) and Figure 8 (for symmetric variant). The obtained verification results for the asymmetric option based on OFMC module and CL-AtSe module are also shown in Figure 9. The attack simulation of the asymmetric-key option of EDHOC protocol is illustrated in Figure 10.



```
Initiator_role
input:
    agents: a, b, public keys: pk_a, pk_b, generator: g,
    hash function: h, channel: snd, rcv
local_variable_declaration and assignment:
    S: natural number
functions:
    prepare_msg( ), witness( ), request( )
initialization:
    S = 0
    transition:
    S: 0 & rcv(start) =|> S' : 2 &
    prepare_msg(msg1) & snd(msg1)
    S: 2 & rcv(msg2) =|> S' : 4 &
    prepare_msg(msg3) & snd(msg3) &
    witness(k3) & request(k2)

Responder_role
Input:
    agents: a, b, public keys: pk_a, pk_b, generator: g,
    hash function:h,channel: snd, rcv
local_variable_declaration and assignment:
    S: natural number
functions:
    prepare_msg( ), witness( ), request( )
initialization:
    S = 1
    transition:
    S: 1 & rcv(msg1) =|> S' : 3 &
    prepare_msg(msg2) /\snd(msg2) & witness(k2)
    S: 3 /\rcv(msg3) =|> S' : 5 &
    request(k3)

environment_role
local_variable_declaration and assignment:
    agents: [ag_1, ag_2, intruder], protocol id: [k2, k3],
    public keys: [pk_1, pk_2, pk_i], hash function: h,
    generator: g, intruder_knowledge = [agents, public
    keys, hash function, generator]
composition_role_instantiation:
    session(ag_1, ag_2 pk_1, pk_2, g, h)
    session(intruder, ag_2 pk_1, pk_2, g, h)
    session(ag_1, intruder, pk_1, pk_2, g, h)

session_role
input:
    agents: a, b, public keys: pk_a, pk_b,
    generator: g, hash function: h

local_variable_declaration and assignment:
    channel(dy): s_a, r_a, s_b, r_b
composition_role_instantiation:
    initiator(a, b, pk_a, pk_b, g, h, s_a, r_a)
    responder(a, b, pk_a, pk_b, g, h,s_b, r_b)

security_goals
goal_specification:
    authentication_on k2
    authentication_on k3
```

Figure 7: A pseudocode for the AVISPA-based verification of asymmetric-key option of EDHOC protocol.

As shown in Figure 10, the attack simulation shows the asymmetric-key option of EDHOC protocol is vulnerable due to the fact that the message is sent without any verification of the sender. In other words, when the intruder sends the message of step 2, the responder should generate and calculate all the elements for communication without any proof to the user. It seems to be able to induce resource exhaustion attacks in R due to Msg2 created or modified by the attacker.

3.2.2. The Symmetric-Key Option. Like the previous case, once we translate the protocol into an HLPSL form, the AVISPA tool passes the code through the modules (such as CL-AtSe and OFMC) to check for any security flaws. Figure 11 presents the outcome of this process. According to the verification result, the symmetric-key option of the EDHOC protocol is unsafe. Figure 12 shows the possible attack simulation for the identified security flaw.

In Figure 12, when an intruder sends a message in step 2, the responder should create all the elements for communication without user authentication as Figure 10. This may deplete R's resource due to responses to numerous authentication requests from unauthorized users.

4. Results and Discussion

The results of the formal security analysis of both variants of the EDHOC protocol show some security-related shortcomings. In this section, we discuss these flaws.

The complete security analysis of the asymmetric-key EDHOC protocol depends on the assumption that the responder trusts the ephemeral ECDH public key G^X is from the initiator. Furthermore, the initiator also must believe that the responder sends the ephemeral ECDH public key G^Y. As shown in the BAN-Logic analysis, the hypotheses in (17) and (18) represent these two claims, respectively. Without these assumptions, mainly hypothesis (17), it is impossible to derive the goals we set. Moreover, it is worth mentioning that both hypotheses are merely there to complete the proof. Hence, it is crucial to realize them to guarantee the evidence.

Similarly, the AVISPA results for both asymmetric and symmetric variants of the protocol also show that an attack can happen (Figures 10 and 12). The responder's failure to

```
Initiator_role

Input:
    agents: a, b, symmetric key: psk, generator: g,
    hash function: h, channel: snd, rcv
local_variable_declaration and assignment:
    S: natural number
functions:
    prepare_msg( ), witness( ), request( )
initialization:
    S = 0
    transition:
    S: 0 & rcv(start) =|> S' : 2 &
    prepare_msg(msg1) & snd(msg1)
    S: 2 & rcv(msg2) =|> S' : 4 &
    prepare_msg(msg3) & snd(msg3) &witness(k3) & request(k2)
```

```
Responder_role

Input:
    agents: a, b, symmetric key: psk, generator: g,
    hash function: h, channel: snd, rcv
local_variable_declaration and assignment:
    S: natural number
functions:
    prepare_msg( ), witness( ), request( )
initialization:
    S = 1
    transition:
    S: 1 & rcv(msg1) =|> S' : 3 &
    prepare_msg(msg2) & snd(msg2) & witness(k2)
    S: 3 & rcv(msg3) =|> S' : 5 & request(k3)
```

```
environment_role

local_variable_declaration and assignment:
    agents: [ag_1, ag_2, intruder], symmetric key: psk,
    hash function: h, generator: g, protocol id: [k2, k3],
    intruder_knowledge = [agents, symmetric key,
    hash function, generator]
composition_role_instantiation:
    session(ag_1, ag_2, psk, g, h)
    session(intruder, ag_2, psk, g, h)
    session(ag_1, intruder, psk, g, h)
```

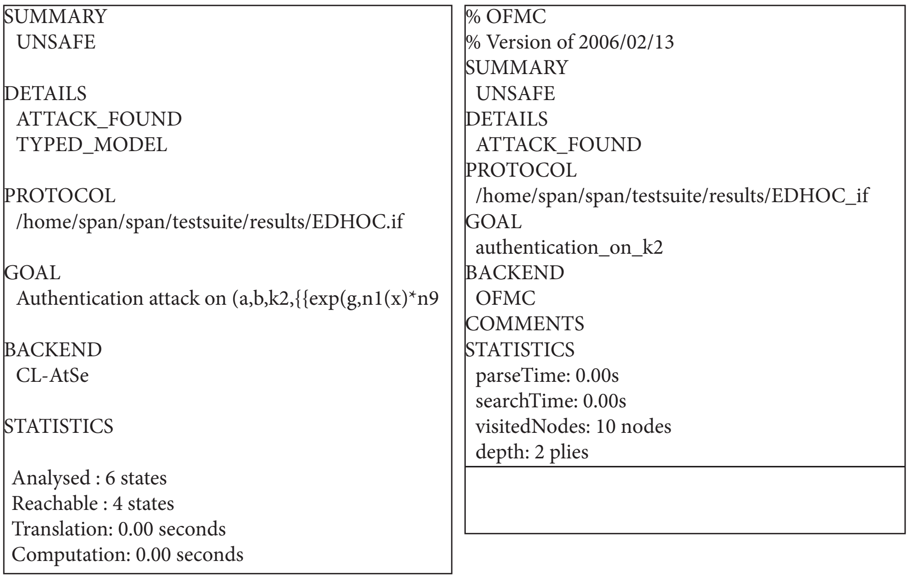
```
session_role

input:
    agents: a, b, symmetric key: psk, generator: g,
    hash function: h
local_variable_declaration and assignment:
    channel(dy): s_a, r_a, s_b, r_b
composition_role_instantiation:
    initiator(a, b, pk_a, pk_b, g, h, s_a, r_a)
    responder(a, b, pk_a, pk_b, g, h, s_b, r_b)
```

```
security_goals

goal_specification:
    authentication_on k2
    authentication_on k3
```

Figure 8: A pseudocode for the AVISPA-based verification of symmetric-key option of EDHOC protocol.



```
SUMMARY
  UNSAFE

DETAILS
  ATTACK_FOUND
  TYPED_MODEL

PROTOCOL
  /home/span/span/testsuite/results/EDHOC.if

GOAL
  Authentication attack on (a,b,k2,{{exp(g,n1(x)*n9

BACKEND
  CL-AtSe

STATISTICS

 Analysed : 6 states
 Reachable : 4 states
 Translation: 0.00 seconds
 Computation: 0.00 seconds
```

```
% OFMC
% Version of 2006/02/13
SUMMARY
  UNSAFE
DETAILS
  ATTACK_FOUND
PROTOCOL
  /home/span/span/testsuite/results/EDHOC_if
GOAL
  authentication_on_k2
BACKEND
  OFMC
COMMENTS
STATISTICS
  parseTime: 0.00s
  searchTime: 0.00s
  visitedNodes: 10 nodes
  depth: 2 plies
```

Figure 9: Verification result of the asymmetric-key option of EDHOC protocol.

ensure the integrity of Msg1 and the difficulty of the initiator in validating Msg2 are the significant reasons for this attack. Especially for the latter point, given that the generation of the secret key K_2 depends on the ECDH session key G^{XY}, the initiator has no option but to trust the responder's ECDH public key G^{Y} transmitted in plaintext to verify Msg2.

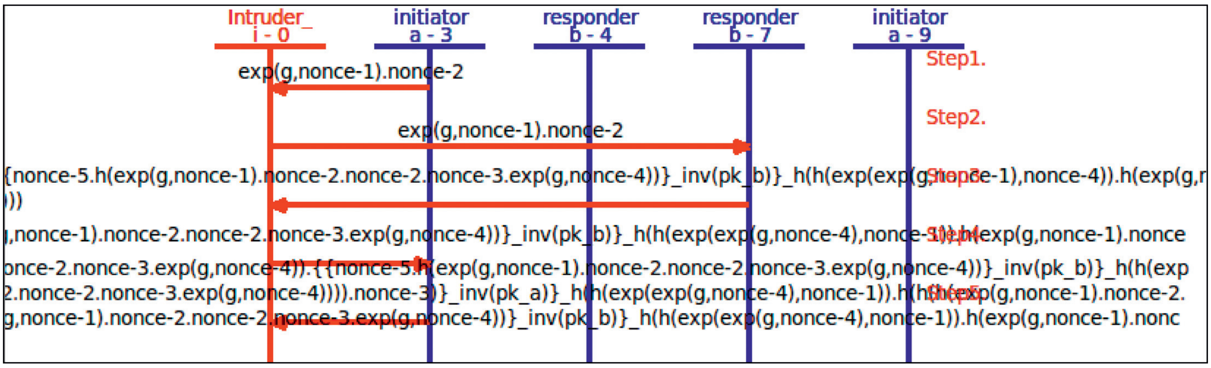
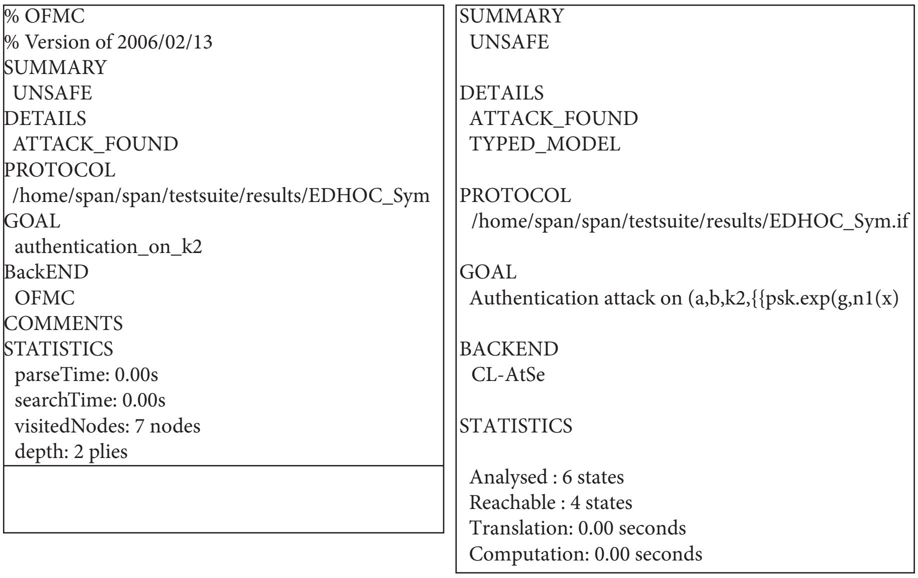

Intruder_
i - 0
initiator
a - 3
responder
b - 4
responder
b - 7
initiator
a - 9
Step1.
exp(g,nonce-1).nonce-2
Step2.
exp(g,nonce-1).nonce-2
Step3.
{nonce-5.h(exp(g,nonce-1).nonce-2.nonce-2.nonce-3.exp(g,nonce-4))}_inv(pk_b)}_h(h(exp(exp(g,nonce-1),nonce-4)).h(exp(g,r
))
Step4.
,nonce-1).nonce-2.nonce-2.nonce-3.exp(g,nonce-4))}_inv(pk_b)}_h(h(exp(exp(g,nonce-4),nonce-1)).h(exp(g,nonce-1).nonce
Step5.
once-2.nonce-3.exp(g,nonce-4)).{{nonce-5.h(exp(g,nonce-1).nonce-2.nonce-2.nonce-3.exp(g,nonce-4))}_inv(pk_b)}_h(h(exp
2.nonce-2.nonce-3.exp(g,nonce-4)))).nonce-3)}_inv(pk_a)}_h(h(exp(exp(g,nonce-4),nonce-1)).h(h(exp(g,nonce-1).nonce-2.
g,nonce-1).nonce-2.nonce-2.nonce-3.exp(g,nonce-4))}_inv(pk_b)}_h(h(exp(exp(g,nonce-4),nonce-1)).h(exp(g,nonce-1).nonc

Figure 10: Attack simulation of the asymmetric-key option of EDHOC protocol.



```
% OFMC
% Version of 2006/02/13
SUMMARY
 UNSAFE
DETAILS
 ATTACK_FOUND
PROTOCOL
 /home/span/span/testsuite/results/EDHOC_Sym
GOAL
 authentication_on_k2
BackEND
 OFMC
COMMENTS
STATISTICS
 parseTime: 0.00s
 searchTime: 0.00s
 visitedNodes: 7 nodes
 depth: 2 plies
```

```
SUMMARY
 UNSAFE

DETAILS
 ATTACK_FOUND
 TYPED_MODEL

PROTOCOL
 /home/span/span/testsuite/results/EDHOC_Sym.if

GOAL
 Authentication attack on (a,b,k2,{{psk.exp(g,n1(x)

BACKEND
 CL-AtSe

STATISTICS

 Analysed : 6 states
 Reachable : 4 states
 Translation: 0.00 seconds
 Computation: 0.00 seconds
```

Figure 11: Verification result of symmetric-key option of EDHOC protocol.

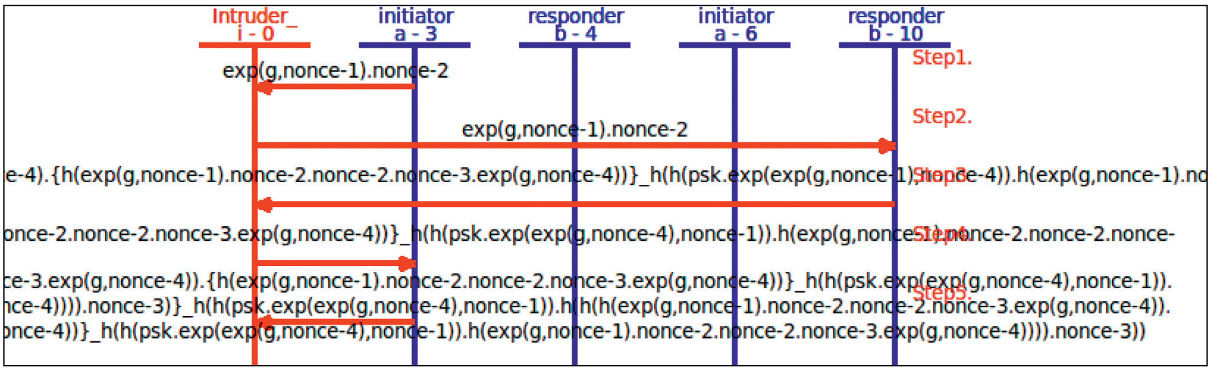

Intruder_
i - 0
initiator
a - 3
responder
b - 4
initiator
a - 6
responder
b - 10
Step1.
exp(g,nonce-1).nonce-2
Step2.
exp(g,nonce-1).nonce-2
Step3.
e-4).{h(exp(g,nonce-1).nonce-2.nonce-2.nonce-3.exp(g,nonce-4))}_h(h(psk.exp(exp(g,nonce-1),nonce-4)).h(exp(g,nonce-1).no
Step4.
once-2.nonce-2.nonce-3.exp(g,nonce-4))}_h(h(psk.exp(exp(g,nonce-4),nonce-1)).h(exp(g,nonce-1).nonce-2.nonce-2.nonce-
Step5.
ce-3.exp(g,nonce-4)).{h(exp(g,nonce-1).nonce-2.nonce-2.nonce-3.exp(g,nonce-4))}_h(h(psk.exp(exp(g,nonce-4),nonce-1)).
nce-4)))).nonce-3)}_h(h(psk.exp(exp(g,nonce-4),nonce-1)).h(h(h(exp(g,nonce-1).nonce-2.nonce-2.nonce-3.exp(g,nonce-4)).
once-4))}_h(h(psk.exp(exp(g,nonce-4),nonce-1)).h(exp(g,nonce-1).nonce-2.nonce-2.nonce-3.exp(g,nonce-4)))).nonce-3))

Figure 12: Attack simulation of the symmetric-key option of EDHOC protocol.

Another critical security threat refers to the denial-of-service attacks (more specifically, the resource exhaustion attack). Given the IoT devices' severe resource limitations concerning computation, storage, and communication, an attacker can send a significant amount of Msg1 to the responder. The responder then performs expensive operations such as encryption, signature, and key derivation functions for each of these messages before authenticating the initiator. Consequently, the responder can get easily overwhelmed by the traffic, deplete its energy, and finally cease communicating with the other end.

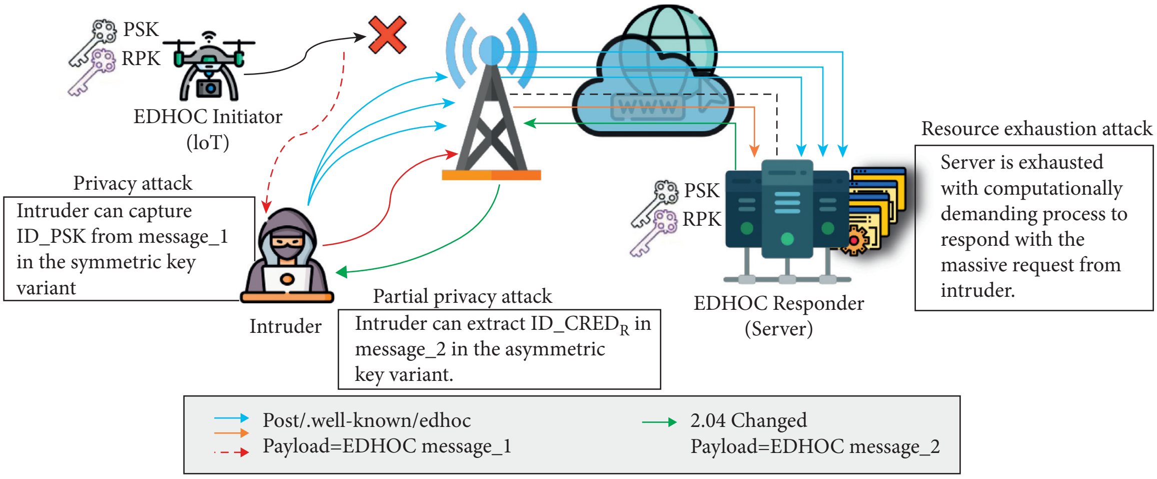

PSK
RPK
EDHOC Initiator
(loT)
Privacy attack
Intruder can capture ID_PSK from message_1 in the symmetric key variant
Intruder
Partial privacy attack
Intruder can extract ID_CRED$_R$ in message_2 in the asymmetric key variant.
PSK
RPK
EDHOC Responder
(Server)
Resource exhaustion attack
Server is exhausted with computationally demanding process to respond with the massive request from intruder.
Post/.well-known/edhoc
Payload=EDHOC message_1
2.04 Changed
Payload=EDHOC message_2

Figure 13: Attack simulation of the symmetric-key option of EDHOC protocol.

Table 5: Summary of related works.

Papers	Identified security issues	EDHOC version	Analysis tools used
[22]	(i) Disclosure of the responders identity in the asymmetric variant of the EDHOC protocol. (ii) An attacker can associate numerous sessions and perform attacks for the symmetric variant of the EDHOC protocol by using the same preshared key identifier. (iii) Only AD3 (for both symmetric and asymmetric variants) satisfies secrecy, perfect forward secrecy, and integrity at both the time of message arrival and the conclusion of the protocol.	Draft-selander-ace-cose-ecdhe-08 [20]	ProVerif [24]
[23]	(i) Absence of nonrepudiation security property. (ii) Lack of verification of ID_CRED$_R$ of Msg2 by the initiator. (iii) When the responder rejects recommended cipher suites, a security concern might arise because of a lengthy metasession spanning many EDHOC sessions.	Draft-selander-lake-edhoc-01 [21]	Tamarin [25]
Ours	(i) A resource exhaustion attack due to a significant amount of Msg1 sent to the responder. The responder does not authenticate Msg1 before computing expensive operations, hence depleting its resources. (ii) The responder's failure to ensure the integrity of Msg1 and the difficulty of the initiator in validating Msg2 threaten the security of the protocol. (iii) A partial privacy attack that exposes the responder's identity. Beside the mere violation of the secrecy of the responder's distinctiveness, it can enable the attacker to reduce the difficulty of stealing the public authentication keys by one step. Moreover, the privacy of ID_PSK, in symmetric-key option, is also violated as it is transmitted in plain text.	Draft-ietf-lake-edhoc-07 [19]	BAN-Logic and AVISPA [26, 27]

A second serious threat with the asymmetric-key-based EDHOC protocol, which we referred to as a partial privacy attack, is related to the privacy of *ID_CRED$_R$*. The access of the credentials containing the public authentication keys of both initiator and responder is via their identities. These identities (*ID_CRED$_I$* and *ID_CRED$_R$*), although they do not have any cryptographic purpose in the protocol, serve an essential purpose by facilitating the retrieval of the public authentication keys. Moreover, according to the standard, their privacy is protected by the session key computed by the initiator and responder. Thus, an attacker can easily break the privacy of *ID_CRED$_R$* as it can establish the session key K_2 with the responder. Therefore, it implies a privacy disclosure of one of the two identities, hence a partial privacy attack. Concerning the symmetric variant, a clear violation of privacy also happens as *ID_PSK* in Msg1 is transmitted in plain text.

It is important to note that if an attacker exploits these vulnerabilities, the results might be disastrous. For example, a medical IoT device attempting to obtain a remote service, perhaps for remote diagnostics, may fail owing to a resource depletion assault on the other end. Moreover, in cases where the responder is a sensitive medical IoT device, its identity (the identity of the credentials containing the public authentication keys) can be traced by an attacker that he/she may use to track and localize the patient eventually. Both resource exhaustion and privacy attacks (as part of transporting EDHOC via CoAP message exchanges) are shown in Figure 13. In addition, Table 5 summarizes the security issues identified by the related works together with the ones we identified.

It is critical to fix the highlighted security vulnerabilities before using EDHOC as a lightweight authenticated key exchange mechanism. Privacy-related threats are mainly initiated because the first message, from the initiator to the responder, is not authenticated. Hence, a preliminary authentication mechanism must be implemented. The responder and initiator can additionally guarantee the validity of Msg1 and Msg2 by using public-key certificates. In the case of a protracted metasession spanning several EDHOC sessions due to cipher suite rejection, the responder shall provide a mechanism that prevents the same initiator from resubmitting a new cipher suite proposal in the same session more than twice. Leveraging HMAC and timestamps can serve a good purpose in thwarting resource exhaustion attacks, as they let the responder first check the validity of the received message before performing computationally demanding instructions.

5. Conclusions

Although the rapid growth of the Internet of Things (IoT) technology is bringing a significant impact on society, efficient security protocols that are aware of the unique characteristics of IoT devices are still in their infant stage. With this regard, IETF is in progress to standardize one application layer protocol (known as EDHOC) that can assist secure communication across IoT devices while remaining lightweight. Consequently, in this paper, we formally analysed the security of this protocol using BAN-Logic and AVISPA to investigate its resilience to withstand attacks. The results show that both variants of the protocol have some serious security and privacy flaws. Primarily, a resource exhaustion attack that violates the availability of a responder's service by depleting its resources over expensive cryptographic operations such as encryption and signature can result. Next, an attacker can easily break the privacy of ID_CRED_R as it can establish the session key K_2 with the responder, which results in partial privacy disclosure of the responder's identity. A similar attack can happen when an attacker captures *ID_PSK* in the symmetric-key option of the protocol. Furthermore, an attacker can use the responder's failure to verify the integrity of Msg1 and the difficulties of the initiator in validating Msg2. Finally, we recommend that the protocol should consider authenticating the first message and provide a way to validate the second message while offering a solution to protect against resource exhaustion attacks. In future works, the authors would like to develop efficient solutions to mitigate these attacks while maintaining the lightweight nature of the EDHOC protocol.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was partly supported by the Institute of Information and Communications Technology Planning and Evaluation (IITP) grant funded by the Korea government (MSIT) (no. 2018-0-00231, Development of Context Adaptive Security Autonomous Enforcement Technology to Prevent Spread of the IoT Infrastructure Attacks, 95%) and the Soonchunhyang University Research Fund (5%).

Supplementary Materials

The AVISPA validation codes written in HLPSL for asymmetric and symmetric options are provided as separate files. (*Supplementary Materials*)

References

[1] V. Korzhuk, A. Groznykh, A. Menshikov, and M. Strecker, "Identification of attacks against wireless sensor networks based on behavior analysis," *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, vol. 10, no. 2, pp. 1–21, 2019.

[2] Z.-K. Zhang, M. C. Yi Cho, C.-W. Wang, C.-W. Hsu, C.-K. Chen, and S. Shieh, "IoT security: ongoing challenges and research opportunities," in *Proceedings of the 2014 IEEE 7th International Conference on IEEE*, pp. 230–234, Matsue, Japan, November 2014.

[3] Machina Research, *Press Release: Global Internet of Things Market To Grow to 27 Billion Devices, Generating USD3 Trillion Revenue in 2025*Machina Research, Stamford, CT, USA, 2021, https://bit.ly/3aHu1QG.

[4] IDC Corporate USA, *IoT Growth Demands Rethink of Long-Term Storage Strategies, Says IDC*, IDC Corporate USA, Needham, MA, USA, 2021, https://www.idc.com/getdoc.jsp?containerId=prAP46737220.

[5] J. Kim, J. Lee, J. Kim, and J. Yun, "M2M service platforms: survey, issues, and enabling technologies," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, pp. 61–76, 2014.

[6] M. Alizadeh, K. Andersson, and O. Schelén, "A survey of secure Internet of things in relation to blockchain," *Journal of Internet Services and Information Security*, vol. 10, no. 3, pp. 47–75, 2020.

[7] Y. M. Khamayseh, W. Mardini, M. Aldwairi, and H. T. Mouftah, "On the optimality of route selection in grid wireless sensor networks: theory and applications," *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, vol. 11, no. 2, pp. 87–105, 2020.

[8] B. Sim and D. Han, "A study on the side-channel analysis trends for application to IoT devices," *Journal of Internet Services and Information Security*, vol. 10, no. 1, pp. 2–21, 2020.

[9] Z. Shelby, K. Hartke, and C. Bormann, *The Constrained Application Protocol (CoAP)*, IETF RFC 7252, Fremont, CA, USA, 2014, https://datatracker.ietf.org/doc/html/rfc7252.

[10] C. Bormann and P. Hoffman, *Concise Binary Object Representation (CBOR)*, IETF RFC 8949, Fremont, CA, USA, 2020, https://datatracker.ietf.org/doc/html/rfc8949.

[11] A. Minaburo, L. Toutain, C. Gomez, D. Barthel, and JC. Zuniga, *SCHC: Generic Framework for Static Context Header Compression and Fragmentation*, IETF RFC 8724, Fremont, CA, USA, 2020, https://datatracker.ietf.org/doc/html/rfc8724.

[12] T. Alladi, V. Chamola, B. Sikdar, and K.-K. R. Choo, “Consumer IoT: security vulnerability case studies and solutions,” *IEEE Consumer Electronics Magazine*, vol. 9, no. 2, pp. 17–25, 2020.

[13] J. Sanchez-Gomez, D. Garcia-Carrillo, R. Sanchez-Iborra et al., “Integrating LPWAN technologies in the 5G ecosystem: a survey on security challenges and solutions,” *IEEE Access*, vol. 8, 2020.

[14] J. Sanchez-Gomez, J. Gallego-Madrid, R. Sanchez-Iborra, J. Santa, and A. Skarmeta, “Impact of SCHC compression and fragmentation in LPWAN: a case study with LoRaWAN,” *Sensors*, vol. 20, no. 1, p. 280, 2020.

[15] C. Kaufman, P. Hoffman, Y. Nir, P. Eronen, and T. Kivinen, *Internet Key Exchange Protocol Version 2 (IKEv2)*, IETF RFC 7296, Fremont, CA, USA, 2014, https://datatracker.ietf.org/doc/html/rfc7296.

[16] E. Rescorla, *The Transport Layer Security (TLS) Protocol Version 1.3*, IETF RFC 8446, Fremont, CA, USA, 2018, https://datatracker.ietf.org/doc/html/rfc8446.

[17] E. Rescorla, H. Tschofenig, and N. Modadugu, *The Datagram Transport Layer Security (DTLS) Protocol Version 1.3*, IETF Internet-Draft draft-ietf-tls-dtls13-43, Fremont, CA, USA, 2021, https://datatracker.ietf.org/doc/draft-ietf-tls-dtls13/.

[18] G. Selander, J. Mattsson, F. Palombini, and L. Seitz, *Object Security for Constrained RESTful Environments (OSCORE)*, IETF RFC 8613, Fremont, CA, USA, 2019, https://www.rfc-editor.org/rfc/rfc8613.html.

[19] G. Selander, J. Mattsson, and F. Palombini, *Ephemeral Diffie-Hellman over COSE (EDHOC)*, IETF Internet-Draft Draft-Ietf-Lake-Edhoc-07, Fremont, CA, USA, 2021, https://datatracker.ietf.org/doc/html/draft-ietf-lake-edhoc-07.

[20] G. Selander, J. Mattsson, F. Palombini, and L. Seitz, *Ephemeral Diffie-Hellman over COSE (EDHOC)*, IETF Internet-Draft Draft-Selander-ace-cose-ecdhe-08, Fremont, CA, USA, 2018, https://tools.ietf.org/pdf/draft-selander-ace-cose-ecdhe-08.pdf.

[21] G. Selander, J. Mattsson, F. Palombini, and L. Seitz, *Ephemeral Diffie-Hellman over COSE (EDHOC)*, IETF Internet-Draft Draft-Selander-Lake-Edhoc-01, Fremont, CA, USA, 2020, https://tools.ietf.org/pdf/draft-selander-lake-edhoc-01.pdf.

[22] A. Bruni, T. S. Jørgensen, T. G. Petersen, and C. Schürmann, “Formal verification of ephemeral Diffie-Hellman over COSE (EDHOC),” in *Proceedings of the 4th International Conference on Research in Security Standardisation*, pp. 21–36, Darmstadt, Germany, November 2018.

[23] K. Norrman, V. Sundararajan, and A. Bruni, “Formal analysis of EDHOC key establishment for constrained IoT devices,” 2020, https://arxiv.org/abs/2007.11427.

[24] B. Blanchet, B. Smyth, V. Cheval, and M. Sylvestre, “Proverif 2.00: automatic cryptographic protocol verifier, user manual and tutorial,” 2018, https://bensmyth.com/publications/2010-ProVerif-manual-version-2.00.

[25] S. Meier, B. Schmidt, C. Cremers, and D. Basin, “The TAMARIN prover for the symbolic analysis of security protocols,” *Computer Aided Verification*, vol. 8044, pp. 696–701, 2013.

[26] M. Burrows, M. Abadi, and R. Needham, “A logic of authentication,” *ACM Transactions on Computer Systems*, vol. 8, no. 1, pp. 18–36, 1990.

[27] A. Armando, D. Basin, Y. Boichut et al., “The AVISPA tool for the automated validation of Internet security protocols and applications,” *Computer Aided Verification*, vol. 3576, pp. 281–285, 2005.

[28] J. Schaad, *“CBOR Object Signing and Encryption (COSE)*, IETF RFC 8152, Fremont, CA, USA, 2017, https://datatracker.ietf.org/doc/html/rfc8152 accessed on.

[29] H. Krawczyk, “SIGMA: the “SIGn-and-MAc” approach to authenticated Diffie-Hellman and its use in the IKE protocols,” in *Proceedings of the 23rd Annual International Cryptology Conference (CRYPTO’03)*, vol. 2729, pp. 400–425, Santa Barbara, CA, USA, August 2003.

[30] P. Rogaway, “Authenticated-encryption with associated-data,” in *Proceedings of the 9th ACM Conference on Computer and Communication Security (CCS’02)*, pp. 98–107, ACM, Washington, DC, USA, November 2002.

[31] P. Syverson and I. Cervesato, “The logic of authentication protocols,” in *Foundations of Security Analysis and Design*, vol. 2171, pp. 63–137, Springer, Berlin, Germany, 2001.

[32] Y. Chevalier, L. Compagna, J. Cuellar et al., “A high level protocol specification language for industrial security-sensitive protocols,” in *Proceedings of the 2004 Workshop on Specification and Automated Processing of Security Requirements (SAPS’04)*, Austrian Computer Society, Linz, Austria, September 2004.

[33] D. Basin, S. Mödersheim, and L. Viganò, “OFMC: a symbolic model checker for security protocols,” *International Journal of Information Security*, vol. 4, no. 3, pp. 181–208, 2005.

[34] M. Turuani, “The CL-atse protocol analyser,” in *Lecture Notes in Computer Science*, vol. 4098, pp. 277–286, Springer, Berlin, Germany, 2006.

[35] A. Armando and L. Compagna, “SATMC: a SAT-based model checker for security protocols,” in *Logics in Artificial Intelligence*, vol. 3229, pp. 730–733, Springer, Berlin, Germany, 2004.

[36] Y. Boichut, P.-C. Héam, O. Kouchnarenko, and F. Oehl, “Improvements on the genet and Klay technique to automatically verify security protocols,” in *Proceedings of the 3rd International Workshop on Automated Verification of Infinite State Systems (AVIS’04)*, pp. 1–11, Barcelona, Spain, April 2004.

Hindawi
Mobile Information Systems
Volume 2021, Article ID 2807395, 15 pages
https://doi.org/10.1155/2021/2807395

Research Article

Designated-Verifier Anonymous Credential for Identity Management in Decentralized Systems

Xudong Deng, Chengliang Tian, Fei Chen, and Hequn Xian

College of Computer Science & Technology, Qingdao University, Qingdao 266071, China

Correspondence should be addressed to Chengliang Tian; tianchengliang@qdu.edu.cn

Received 26 May 2021; Revised 31 July 2021; Accepted 13 August 2021; Published 10 September 2021

Academic Editor: Vishal Sharma

Most of the existing identity management is the centralized architecture that has to validate, certify, and manage identity in a centralized approach by trusted authorities. Decentralized identity is causing widespread public concern because it enables to give back control of identity to clients, and the client then has the ability to control when, where, and with whom they share their credentials. A decentralized solution atop on blockchain will bypass the centralized architecture and address the single point of the failure problem. To our knowledge, blockchain is an inherited pseudonym but it cannot achieve *anonymity* and *auditability* directly. In this paper, we approach the problem of decentralized identity management starting from the *designated-verifier anonymous credential* (DVAC in short). DVAC would assist to build a new practical decentralized identity management with *anonymity* and *auditability*. Apart from the advantages of the conventional anonymous credential, the main advantage of the proposed DVAC atop blockchain is that the issued cryptographic token will be divided into shares at the issue phase and will be combined at the showing credential phase. Further, the smooth projective hash function (*SPHF* in short) is regarded as a designated-verifier zero-knowledge proof system. Thus, we introduce the *SPHF* to achieve the designated verifiability without compromising the privacy of clients. Finally, the security of the proposed DVAC is proved along with theoretical and experimental evaluations.

1. Introduction

Identity management is viewed as a tool for the protection of user identification and account privacy security, government enterprise management, and public service demand, or the security and economic needs of operators and providers. Before, the ID card system is succinct for the government to manage people's identity. With the rapid development of the Internet, a large number of identity management systems appear in our field of vision. They all have their merits for some special demand and also vulnerability for practice at the same time.

Typically, a trusted party certificate authority (CA) is used to manage and store identities for users. As far back as the late twentieth century, "Passport" is a unified identity management project based on the NET platform implemented in [1]. "Passport" provides great convenience to users by allowing them to authenticate with only one site. However, as its system architecture is centralized and coordinated, the problem follows. In practice, single points of failure are coming through a trusted party. Such as the DigiNotar incident [2], CA was held hostage by hackers, in which Google's certificate was published mistakenly. So, we need to effectively store a person's identity information and ensure its privacy and effectiveness on the Internet where threats and vulnerabilities are ubiquitous. And, how to protect the privacy of individuals and ensure their identity is verified without giving away their privacy is an important issue.

Bitcoin and blockchain had been proposed in 2008 [3]. The transaction of virtual currency built on the chain can guarantee the privacy and security of both parties. The natural decentralization and unchangeability of blockchain give us a new direction. With the rapid development of blockchain, there are more and more decentralized systems appearing based on blockchain. Furthermore, IBM and

Samsung have been working on an idea called ADEPT that uses blockchain technology to form a decentralized network of IoT devices. Simply, it is feasible to construct a relatively secure identity management system with blockchain because some security features on blockchain fully meet the requirements of the identity management system.

Recently, blockchain-based identity management has also had limited success, such as DAC [4] and DBLACR [5]. In these systems, users obtain information credentials from an authority (e.g., government) and upload their credentials to the blockchain. When an entity, such as a service provider, has requirements for its customers, users can prove those requirements for verification by the blockchain, which is used as a transparent infrastructure for authentication. Since we want to ensure the privacy of user information, we need to encrypt user information, but how to verify the encrypted information and verify it accurately and effectively is a problem (we certainly cannot provide authentication after decrypting user information, which violates the principle of privacy).

In this paper, we propose DVAC, a decentralized anonymous credential system to protect the privacy of the clients. In particular, interaction in the system is completely anonymous and the registered identity information is self-sovereign. Instead of traditional CA for the whole system level, DVAC employs a decision-making institution committee; the committee consists of an indefinite number of members. The function of the committee is the same as that of a traditional CA, which issues credentials to users, except that it requires the approval of its members when making important decisions. This can be seen as effectively diminishes the power of CA and avoiding the problem of single points of failure. Moreover, DVAC supports the change of membership in the committee. On the contrary, conflicts are inevitable between service providers and users because of the anonymity of users in the anonymous credential system. We need a reasonable and fair audit to protect the interests of both parties in the conflict.

To this end, we introduced proactive secret sharing. We use proactive secret sharing to distribute the private key of the committee to members, and no party in the committee can decide on its own. Moreover, proactive secret sharing can redistribute the secret key periodically according to the system conditions. In this way, members of the committee are prevented from being heavily bribed to ensure the correctness of the committee's decision-making. When a conflict occurs, members of the committee negotiate whether to conduct an audit, but only if the number of supporting reaches a certain threshold.

Contribution. Below, we conclude our main contribution along with the techniques' roadmap as follows:

(i) *A Neat Decentralized Anonymous Credential via Cost-Efficiency SPHF*. We construct a decentralized identity management system and describe each step of our scheme in detail. We use *SPHF* to realize anonymous authentication of the system, and we add an audit function to our system. It points us to a new way of identity management.

(ii) *A Privacy-Preserving* DVAC *Service via Our Designed* DAC. We design an application using our DVAC system for identity management.

(iii) *A Simply Prototype of* DVAC. We implement and evaluate our system and test its performance under different lengths of the secret key.

Organization. The rest of the paper is organized as follows: Section 2 shows the related work of DVAC. Section 3 presents our hardness assumptions and cryptographic building blocks. Our system model and security model are presented in Section 4. We reviewed previous scenarios for anonymous credentials in Section 5. In Section 6, we describe each step of the construction of our solution in detail and we have proved the correctness and security of our scheme. In Section 7, we briefly introduce an application of our scheme. In Section 8, we evaluate the performance of DVAC. Finally, we conclude our work in Section 9.

2. Related Work

DVAC is engaged in anonymous authentication and identity management of private data and privacy protection blockchains. DVAC uses a zero-knowledge proof scheme and proactive secret sharing to create a new decentralized anonymous identity management system, which achieved fast and provable correct queries.

2.1. Anonymous Credential (over Blockchain). In [6], a bilinear pair-based signature scheme was proposed, and based on the signature scheme, they constructed an efficient anonymous credential system. Au et al. proposed a constant-size anonymous authentication scheme [7], which use short group signature in [8]. The scheme can achieve multiple dynamic authentications, and the signature certificate length is constant, which makes the scheme have better efficiency. Additionally, Begum et al. [9] proposed another pairing-based anonymous credential system that also satisfies the constant size of the formula proof. In 2010, IBM proposed "Identity Mixer" [10], which can be used for anonymous authentication and the transfer of authentication attributes. It allows users to authenticate without collecting any other personal data. But the "Identity Mixer" has a defect, it does not provide identity tracking. In 2020, Li et al. [11] proposed a round-optimal asymmetric PAKE protocol, which could construct a new anonymous credential system. "DAC" [4] and "DBLACR" [5] are two decentralized anonymous credential systems based on blockchain. The anonymity of blockchain ensures that users' private information will not be disclosed. However, there must have a trusted party to verify the reasonableness of the user's identity information. There is still a risk of single points of failure.

2.2. Identity Management (over Blockchain). "Liberty Alliance" proposed a three-way interaction protocol based on users, identity provider, and service provider [12]. It has improved the issue of leakage of users' private information based on “Passport.” Subsequently, there comes out many identity management systems such as Tivoli Access Manager [13] and Central Authentication Service [14]. These systems provide a more secure privacy protection protocol. At the same time, these systems provide more complete basic functions and expand the application scope. But with the rapid popularization of electronic identity and the increasing demand of people, the centralized management scheme has become the most fundamental drawback [15]. CertCoin [16] was proposed by Conner Fromknecht et al. It constructs a distributed authentication system and maintains a common ledger for storing user IDs and public key information by using blockchain. After that, Blockstack decentralized PKI system [17] was proposed by Ali et al. Blockstack uses bitcoin’s working proof mechanism to maintain the system’s state consistency. In Coconut [18], Sonnino et al. proposed a new signature scheme based on Waters signature scheme, BGLS signature [19], and signature scheme of Pointcheval and Sanders [20]. Coconut enables general-purpose selective disclosure credentials to be efficiently used in settings with no natural single trusted third party to issue them. In addition, because of Shamir’s secret sharing [21], it will waste more time when adding and removing authorities. After Coconut, Nym credentials [22] were proposed by Halpin et al., namely, Nym credentials can be viewed as an improvement or extension of Coconut.

2.3. Public Distributed Ledger (atop Blockchain). In ZkLedeger [23], Narula et al. proposed the first privacy-protected, verifiable audit distributed ledger system. All information about the transaction is uploaded to a public ledger and publicly verifiable. Unlike zk-SNARKs, Zkledger provides efficient and fast audits through noninteractive zero-knowledge proof and it does not require a trusted setup. Furthermore, ZkLedger provides much different auditing queries and real-time feedback to the auditor. In PrETP [24], Balasch et al. proposed a new way in privacy-protection ETP system. It can protect the user’s privacy to the greatest extent and provide the correct audit. PrETP resolves the conflict between the user’s right to privacy and the service provider’s right of interest. After PrETP, Meiklejohn et al. proposed Milo [25]. Milo solves the problem of privacy leakage in the audit process in PrETP and provides a new way in preventing drivers from ganging up to cheat.

3. Preliminaries

Below, we will outline the cryptographic building blocks that will be used in the following sections.

Notation. Throughout this paper, let λ be the security parameter; then, we denote the vector (i.e., a) and the matrix using the lower letter and upper letter, respectively. In addition, let $\mathbb{G}$ be a group of prime order p with generator g.

Decisional Diffie–Hellman (DDH) Assumption is that given a group $G = \langle g \rangle$ of order q, distribution $(g^a, g^b, g^c | a, b, c \xleftarrow{\$} \mathbb{Z}_q)$, and $(g^a, g^b, g^{ab} | a, b \xleftarrow{\$} \mathbb{Z}_q)$ are indistinguishable for any polynomial time adversary.

Decisional Bilinear Diffie–Hellman (DBDH) Assumption is that no polynomial time algorithm can achieve non-negligible advantage in deciding the BDH problem; in other words, no adversary has the ability to distinguish the distribution $(g, g^a, g^b, g^c, e(g, g)^{abc})$ from $(g, g^a, g^b, g^c, e(g, g)^z)$.

Smooth Projective Hash Function (SPHF) was proposed by Cramer and Shoup [26]. SPHF gives us a method to achieve zero-knowledge proof for the specified verifier [27–29]. For a NP language L, a word $x \in L$ and a complete SPHF are defined as follows:

(i) $\mathrm{hk} \longleftarrow \mathit{SPHF.HashKG}(L)$: takes a language L as input and generates a hashing key hk.

(ii) $\mathrm{hp} \longleftarrow \mathit{SPHF.ProjKG}(\mathrm{hk}, L, x)$: for a word $x \in L$, use hk and L as inputs, and output the projective hashing key hp.

(iii) $H \longleftarrow \mathit{SPHF.Hash}(\mathrm{hk}, L, x)$: the algorithm’s inputs are same as *SPHF.ProjKG*; output a value H.

(iv) $H' \longleftarrow \mathit{SPHF.Proj}(\mathrm{hp}, L, x; w)$: w is a witness for that $x \in L$. This algorithm uses $(\mathrm{hp}, L, x; w)$ as inputs and outputs a value H'.
The SPHF contains two properties, one is *smoothness* and another *projective*.

(v) *Projective(Correctness).* For all $x \in L$ and its witness w, satify $\mathit{SPHF.Hash}(\mathrm{hk}, L, x) = \mathit{SPHF.Proj}(\mathrm{hp}, L, x; w)$.

(vi) *Smoothness.* For any $x \notin L$ and any parameters, the following distributions are statistically indistinguishable:

$\{(\mathrm{hp}, H)\colon \mathrm{hk} \longleftarrow \mathit{SPHF.HashKG}(L), \mathrm{hp} \longleftarrow \mathit{SPHF.ProjKG}(\mathrm{hk}, L, x), H \longleftarrow \mathit{SPHF.Hash}(\mathrm{hk}, L, x)\}$,
$\{(\mathrm{hp}, H)\colon \mathrm{hk} \longleftarrow \mathit{SPHF.HashKG}(L), \mathrm{hp} \longleftarrow \mathit{SPHF.ProjKG}(\mathrm{hk}, L, x), H \overset{\$}{\longleftarrow} \Theta\}$,where Θ is the set of hash values.

3.1. Pedersen Commitment. Let $h \in \mathbb{G}$ is a generator of the group $\mathbb{G}$, and we denote $x \in \mathbb{G}$ as the message. Randomly select a commitment parameter $r \in \mathbb{G}$. The commitment scheme is proceeded as follows:

(i) $c \longleftarrow \mathit{Com}(x, r)$ computes and outputs $c = g^x h^r$

(ii) $0/1 \longleftarrow \mathit{Open}(c, x, r)$

Pedersen commitments contain two important properties, one is perfectly hiding: the commitment **c** reveals nothing about the committed value x. Another is computationally binding: an adversary of *probabilistic polynomial time* cannot find a value x' for r such that $c = g^{x'} h^r$.

Linear Encryption (LE in short) is based on *Decision Linear problem* [8]. Below, we review the original scheme proposed by Boneh et al. [8]. For the public common parameters of LE scheme (G, G_1, p, g, g_1, e), the detailed construction is as follows:

(i) $(\mathrm{lsk}, \mathrm{lpk}) \longleftarrow \mathit{LE.KGen}(1^\lambda)$: lsk is the private key; lpk is the public key. We select $(x_1, x_2) \longleftarrow \mathbb{Z}_p$ randomly, and let $\mathrm{lsk} = (x_1, x_2)$ and $\mathrm{lpk} = (Y_1 = g^{x_1}, Y_2 = g^{x_2})$.

(ii) $c \longleftarrow LE.Enc(m; \mathrm{lpk})$: $m \in G$ is a message, $r = (r_1, r_2) \overset{\$}{\leftarrow} Z_p$ are random scalars, we use public key lpk, m, and r_1 and r_2 as input and output a ciphertext $c = (c_1 = Y_2^{r_1}, c_2 = Y_2^{r_2}, c_3 = g^{r_1+r_2} \cdot m)$.

(iii) $m \longleftarrow LE.Dec(c; \mathrm{lsk})$: we use ciphertext c and private key lsk as input and output plaintext $m = c_3/(c_1^{x_1^{-1}} c_2^{x_2^{-1}})$.

Waters Signature is an efficient signature scheme with some optimizations and follow-up works [30, 31]. In our solution, we only use the original one to assist DVAC. We define a generator $h \overset{\$}{\leftarrow} G$ and a vector $\mathbf{u} = (u_1, u_2, \ldots, u_k) \overset{\$}{\leftarrow} G^{k+1}$, which defines the *Waters hash* of a message $m = (m_1, m_2, \ldots, m_k) \in (0,1)^k$ as $F(m) = u_0 \prod_{i=1}^{k} u_i^{m_i}$. Below, we review the construction of Waters signature:

(i) $(\mathrm{wsk}, \mathrm{wpk}) \longleftarrow Waters.KGen(1^\lambda)$: wsk is the private key used for signing, and wpk is the public key used for public verification. $\mathrm{wsk} = h^z$ and $\mathrm{wpk} = g^z$ where $z \overset{\$}{\leftarrow} Z_p$.

(ii) $\sigma \longleftarrow Waters.Sign(m; wsk)$: we use a message m, private key wsk, and a random $s \overset{\$}{\leftarrow} Z_p$ as inputs and generate a signature $\sigma = (\sigma_1 = h^z \cdot F(m)^s, \sigma_2 = g^s)$ of m.

(iii) $(0,1) \longleftarrow Waters.Verify(m, \sigma; \mathrm{wpk})$: we use wpk, message m, and its signature σ as inputs. The verify algorithm will output a result that showed whether the signature $e(g, \sigma) = e(g^z, h) \cdot e(F(m), \sigma_2)$ is valid.

Decentralized Anonymous Credential (DAC in short) inherits the properties of anonymous credential [32, 33] except distributing the cryptographic token into a couple of token shares. Most of state-of-art of DAC are considering to instantiate in a decentralized way. Below, as shown in Figure 1, we adopt the definition of DAC discussed by Garman et al. [4] that contains seven PPT algorithms.

4. System and Threat Models

4.1. DVAC System Model Overview

4.1.1. System Participants. During our whole scheme, we separate five entities including user, committee, bulletin board (BB), service provider (SP), and auditor as illustrated in Figure 2 and optimized system model in distribution way in Figure 1.

User is required to register their information on the bulletin board and get a credential from the committee.

Committee is a group of members who can perform the same function as a certificate authority (CA) only under certain conditions.

Bulletin board is a distributed ledger that can be instantiated by the blockchain. The data stored on the BB pseudonymously are public and unchangeable, but only the client who has the secret key could read the data.

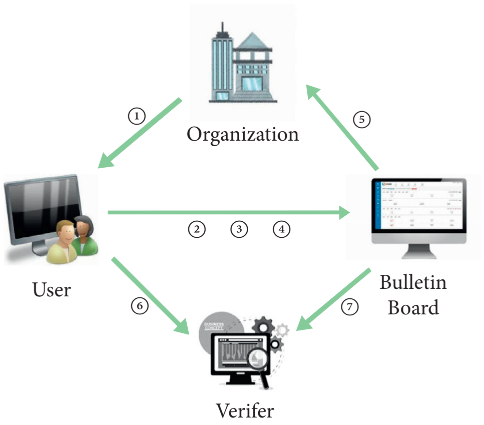


Figure 1: A brief illustration of DAC. ① user gets essential parameters of the system; ②③④ are executed offline by the user on the local side and finally uploads resulting values to the bulletin board; ⑤ the nodes of the organization verify the information from the bulletin board; ⑥ the user plays the role of a prover, and he sends proof of his credential with other auxiliary information to the verifier; ⑦ the verifier downloads the essential information from the bulletin board and validates the proof sent from the prover.

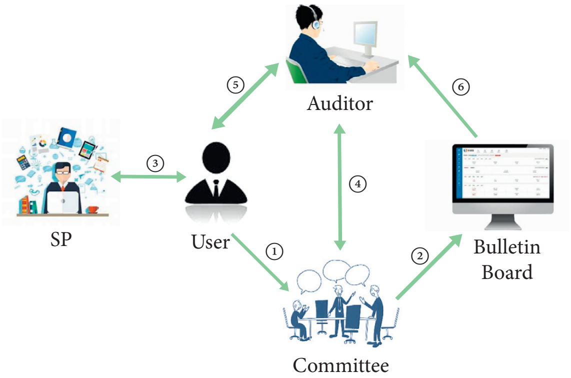


Figure 2: System model of DVAC. ① User send his information to committee. ② The committee upload user's information to blockchain and issue a credential to the user after verifying the information. ③ User shows his credential to the SP and acquires service. ④ The auditor requests the audit process by interacting with the committee. ⑤ The auditor interacts with the user to get the required information. ⑥ The auditor gets the information uploaded by the user at ① from the blockchain and returns the audit result.

Service provider is an organization or government that provides concrete services to clients, and SP determines whether the client has the qualification to access his services.

Auditor is an entity used to audit user information on the blockchain, which is typically acted by some of the members of the committee.

Figure 2 illustrates an example of the model of our scheme. The concept of this paper is that remove the single points of failure of CA and guarantee the privacy information of users during the certification process. We propose a decentralized self-sovereign credential management system. We use secret sharing, and blockchain provides the decentralized function and the immutability of data, and the

SPHF scheme guarantees zero-knowledge, and the blockchain could provide an environment which ensures anonymity. This means the security of users' private information is more strongly guaranteed when they obtain certifications. For simplicity, we have omitted the parameters of the entire system, and our approach proceeds as follows:

4.1.2. Initialization. First of all, the dealer generates system parameters and a pair of keys for the committee. The committee's public key is open for everyone and uses a secret share scheme to distribute the committee's secret key for people in this group. Meanwhile, the committee should set a threshold value for recovering the secret key. The committee can then perform the corresponding function (audit) only if the person who has reached this threshold agrees.

4.1.3. Registration. In registration, the user will get his unique id on the blockchain, which cannot be changed. A new client needs to register his identity information $info$ and his attribute $attr$ to the blockchain. He generates a commitment $\mathbf{c}$ for his identity information info and sends $(\mathrm{id}, \mathbf{c}, attr)$ to committee after signing by his secret key. The committee will first verify the validity of $(\mathrm{id}, \mathbf{c}, attr)$ after receiving it. If the verification passes, the committee will generate credentials for the user's attributions $attrs$ and upload the client's cryptographic information and attributions to the bulletin board.

4.1.4. Interactive Verification. After a client gets his credential, he is a legitimate user of the blockchain. He could interact with an SP or any other users as a prover. The process of interaction is anonymous; each presentation of a credential is anonymized. The prover proves that his identity info satisfies some requirement which comes from verifier through SPHF. Prover and verifier get the required parameters for proof and verification through an interaction. A verifier could verify the correctness of the user's credentials.

4.1.5. Secret Refresh. In our system, the group of the committee is not fixed and we should keep the number of members in a stable state. There must have members quit or join. For example, members of the committee may be bribed or some user wants to be a member of this committee. Although a few bribes do not affect the overall situation, the committee must regularly check and weed out those who have been bribed. So, we need to satisfy (1) security: let the sharing in the hands of the person weed out to be invalid, that is, he can no longer participate in the decision-making and his share cannot use to recover the secret key of the committee. (2) Fairness: provide regular rights of membership for new members.

To ensure the security and fairness of the system, we must refresh the secret key held by the members of the committee. After the committee's secret key had been shared, the $Refresh$ algorithm will be recalled after some time (periodic testing by the committee) or some special change of participants. We will describe the details in Section 6.

4.1.6. Audit. Consider that the user can provide attributes that he does not have to get the committee to sign or the member of the committee falsified. If an SP doubts the truth of a user's credential, SP could apply an audit for the user's identity. When the member in the committee who has reached the threshold agrees, the committee will generate an audit request for the client. This request will ask the user to open the commitment of his identity and compare it with the credential issued by the committee. If the user agreed, he should open the commitment of his information. At the end of the audit, the auditor returns a list of dishonest users and the committee will cancel these users' id on the blockchain. If he refused, the committee will adopt other means of restriction for him.

4.2. Threat Model. In DVAC, it is assumed that the members of the committee will be bribed. We assume that no more than a third of the members will be malicious in each period of a secret refresh. In terms of privacy, we need to protect against a malicious SP speculate the identity of a user in the course of interacting with the user, even if SPs arbitrarily collude. We assume that there are dishonest users who try to trick SP into providing services without the relevant credentials. In DVAC, auditing is only used as a solution for resolving conflicts when they occur. There must have been information leaked to the auditor by an audit process. Users' privacy might be hard to protect if frequent audits are carried out.

4.3. System Goals. As an identity management system, DVAC should achieve the following goals:

(i) *Completeness.* A prover who shows his credentials correctly will surely pass the verification of the verifier.

(ii) *Anonymity.* Private information of each user can only be read by oneself. It means that a verifier does not know the prover's other information except the validity of the credential.

(iii) *Unforgeability.* Prover can generate a valid verifier convinced proof if and only if it has a valid credential and the information contained in the certificate meets the security policy.

(iv) *Unlinkability.* In any two credential showing processes, or in the credential issuance process and credential showing process, an adversary verifier has only one negligible advantage linking them.

(v) *Decentralized Auditability* (D-Auditability). If there is a dispute between SP and user in the process of interaction between the two parties, third parties will intervene and audit. But third parties must reach a consensus (majority rule) to avoid a single point of failure on the part of the auditor.

5. Neat Decentralized Anonymous Credential from SPHF

5.1. Review Decentralized Anonymous Credential. Below, we recall the generic construction of decentralized anonymous credential (DAC) that follows the solution of Garman et al. [4]. In a nutshell, they adopt the following:

(i) $param_{\mathrm{DAC}} \longleftarrow Setup(1^{\lambda})$ takes as input the security parameter λ and then proceeds the following computations:

  (1) $param_{\mathrm{ACC}} \longleftarrow AccSetup(1^{\lambda})$, where $param_{\mathrm{ACC}} = (N, u, p, q, g_0, \ldots, g_n)$ are parameters of RSA, p and q are primes such that $p = 2^w q + 1$ for $w \geq 1$, and $g_0, \ldots, g_n$ are random generators of a group $\mathbb{G}$

  (2) Output $param_{\mathrm{DAC}} = param_{\mathrm{ACC}}$

(ii) $msk \longleftarrow MskGen(param_{\mathrm{DAC}})$ generates a main secret key msk for the user U, while the msk is the only one that is bound to the user, and the detailed procedures are as follows:

  (1) $msk \longleftarrow Z_q$, where msk is randomly selected, and it will be used in the credential mint phase

  (2) Output msk

(iii) $(\mathrm{nym}_{\mathrm{U}^{\mathrm{O}}}, msk_{\mathrm{nym}_{\mathrm{U}^{\mathrm{O}}}}) \longleftarrow NymGen$ $(param_{\mathrm{DAC}}, msk, \mathrm{U}, \mathrm{O})$. The pseudonym $\mathrm{nym}_{\mathrm{U}^{\mathrm{O}}}$ (from the user U to the origination O) is generated by U who proceeds under the master secret key msk as follows:

  (1) $msk_{\mathrm{nym}_{\mathrm{U}^{\mathrm{O}}}} \longleftarrow NymMskGen(1^{\lambda})$, picks up a randomness r_{msk} from Z_q, and sets $msk_{\mathrm{nym}_{\mathrm{U}^{\mathrm{O}}}} = r_{msk}$

  (2) $\mathrm{nym}_{\mathrm{U}^{\mathrm{O}}} \longleftarrow BuildNym(param_{\mathrm{DAC}}, msk_{\mathrm{nym}_{\mathrm{U}^{\mathrm{O}}}, msk})$ and computes $\mathrm{nym}_{\mathrm{U}^{\mathrm{O}}} = g_0^{r_{msk}} g_1^{msk}$ for an organization O

  (3) Output $(\mathrm{nym}_{\mathrm{U}^{\mathrm{O}}}, msk_{\mathrm{nym}_{\mathrm{U}^{\mathrm{O}}}})$

(iv) $(\sigma, sk_{\mathrm{U}}, \pi_{\mathrm{U}}) \longleftarrow CredMint(msk, \mathrm{nym}_{\mathrm{U}^{\mathrm{O}}}, msk_{\mathrm{nym}_{\mathrm{U}^{\mathrm{O}}}, \mathrm{att}, \mathrm{aux}})$ is executed by the user U to generate a credential σ which is essential to a Pedersen commitment for his corresponding attributes $\mathrm{att} = (a_0, a_1, \ldots, a_m) \in Z_q$, i.e., $\overline{\mathrm{c}} = g_0^{r_{\mathrm{ped}}} g_1^{msk} \prod_{i=0}^{m} g_{i+2}^{a_i}$ for its randomness $r_{\mathrm{ped}} \in Z_q$ and msk, along with a secret key sk_{U} for the user U and a zero-knowledge proof

  (1) $\overline{\mathrm{c}} \longleftarrow Ped.Com(param, msk, \mathrm{att})$, and it takes as input a selected random number $r_{\mathrm{ped}} \in Z_q$ and sets $\mathrm{sk}_{\mathrm{U}} = r_{\mathrm{ped}}$; then, the user calculates

$$\sigma \coloneqq \overline{\mathrm{c}} = g_0^{r_{\mathrm{ped}}} g_1^{msk} \prod_{i=0}^{m} g_{i+2}^{a_i}. \quad (1)$$

  (2) $\pi_{\mathrm{U}} \longleftarrow ZK.Prove(msk, r_{msk}, r_{\mathrm{ped}}, \mathrm{aux})$, and the user proves that
    the commitment $\overline{\mathrm{c}}$ and the pseudonym nym
  $_{\mathrm{U}^{\mathrm{O}}}$ contain the same master secret key msk and
    the attributes are in some allowed set

  (3) Output $(\overline{\mathrm{c}}, \mathrm{sk}_{\mathrm{U}}, \pi_{\mathrm{U}})$

After generating a credential of attributes, user submits $(\overline{\mathrm{c}}, \pi_{\mathrm{U}}, \mathrm{att}, \mathrm{nym}_{\mathrm{U}^{\mathrm{O}}})$ with auxiliary data aux to blockchain.

(i) $\{0, 1\} \longleftarrow MintVerify(param_{\mathrm{DAC}}, \mathrm{att}, \sigma, \mathrm{nym}_{\mathrm{U}^{\mathrm{O}}}, \mathrm{aux}, \pi_{\mathrm{U}})$. This algorithm is run by nodes in the organization O. The credential's legality is accepted to the blockchain if and only if this algorithm returns 1.

  (1) $\{0, 1\} \longleftarrow ZK.Verify(param_{\mathrm{DAC}}, \mathrm{att}, \sigma, \mathrm{nym}_{\mathrm{U}^{\mathrm{O}}}, \mathrm{aux}, \pi_{\mathrm{U}})$

  (2) Output 1 if verification is successful, otherwise 0

If user's credential through verification of organization's nodes, he could show others a proof that he is a legal user and satisfy some attributes without revealing his credential. This process is noninteractive and anonymous.

(ii) $\pi_s \longleftarrow CredShow(msk, \overset{\mathrm{UV}}{\mathrm{nym}}, msk_{\overset{\mathrm{UV}}{\mathrm{nym}}}, \sigma, \mathrm{sk}_{\mathrm{U}}, C_{\mathrm{O}})$. In this phase, the user shows a proof to a verifier, where a verifier is either an organization O or a third verifier.

  (1) $msk_{\overset{\mathrm{UV}}{\mathrm{nym}}} \longleftarrow NymMskGen(1^{\lambda})$ and outputs $msk_{\overset{\mathrm{UV}}{\mathrm{nym}}} = r'_{msk}$, where $r'_{msk} \in Z_q$

  (2) $\overset{\mathrm{UV}}{\mathrm{nym}} \longleftarrow BuildNym(msk_{\overset{\mathrm{UV}}{\mathrm{nym}}}, msk_{\overset{\mathrm{UV}}{\mathrm{nym}}}, msk)$ and outputs $\overset{\mathrm{UV}}{\mathrm{nym}} = g_0^{r'_{msk}} g_1^{msk}$ for a verifier V

  (3) $A = Acc(param)_{\mathrm{ACC}}, C_O$, on inputs $param(N, u)$ is executed by the user U and computes $A = u^{\overline{\mathrm{c}} \cdot \overline{\mathrm{c}}_1 \cdot \overline{\mathrm{c}}_2, \ldots \cdot \overline{\mathrm{c}}_{\mathrm{n}}}$ where $C_O = \{\overline{\mathrm{c}}, \overline{\mathrm{c}}_1, \overline{\mathrm{c}}_2, \ldots, \overline{\mathrm{c}}_n\}$ are a set of credentials fetched from the blockchain

  (4) $w = WitGen(param, c, C_O)$, on inputs $para\underline{m}(N, u)$; then, the user computes a witness $w = u^{\overline{\mathrm{c}}_1} \cdot \overline{\mathrm{c}}_2, \ldots, \overline{\mathrm{c}}_n$

  (5) $\pi_s \longleftarrow ZK.Prove(msk, w, r'_{msk}, \overline{\mathrm{c}}, r_{\mathrm{ped}}, \mathrm{nym}_U^V)$:

$$\begin{aligned} & AccVer(param_{\mathrm{ACC}}, A, \overline{\mathrm{c}}, w) = 1 \\ & \qquad \wedge \overline{\mathrm{c}} = g_0^{r_{\mathrm{ped}}} g_1^{msk} \prod_{i=0}^{m} g_{i+2}^{a_i} \\ & \qquad \wedge \mathrm{nym}_U^V = g_0^{r'_{msk}} g_1^{msk}. \end{aligned} \quad (2)$$

  (6) Outputs a proof π_s

At first, the user should generate a new nym $\overset{\mathrm{UV}}{\mathrm{nym}}$ and its secret key $msk_{\mathrm{nym}_U^V}$ for this verifier. Then, the user calculates a proof that

(i) He knows a credential on the blockchain from organization O

(ii) The credential opens to the same msk pseudonym, and

(iii) It has some attributes

At the end of this phase, the prover sends π_s to verifier through his nym nym_U^V.

(i) $\{0, 1\} \longleftarrow ShowVerify(param_{\mathrm{DAC}}, \mathrm{nym}_U^V, \pi_s, C_O)$. Upon receiving the proof π_s from $\overset{\mathrm{UV}}{\mathrm{nym}}$, the verifier executes the verify algorithm.

(1) $A = Acc(param)_{\mathrm{ACC}}, C_O$, and verifier computes $A = u^{\bar{\mathrm{c}}\cdot\bar{\mathrm{c}}_1\cdot\bar{\mathrm{c}}_2,\ldots,\bar{\mathrm{c}}_n}$

(2) $\{0,1\} \longleftarrow ZK.Verify(param_{\mathrm{DAC}}, \mathrm{nym}_U^V, \pi_s, A,)$, and verify that π_s is the aforementioned proof of knowledge on $(\bar{\mathrm{c}}, C_O)$ and nym_U^V using the known public values

(3) Output 1 if verification is successful, otherwise 0

5.2. Review (Interactive) Anonymous Credential via SPHF. Below, we recall the generic construction of *SPHF*-based anonymous credential that follows the solution of Blazy et al. [34]. In a nutshell, they adopt the

(i) $param_{\mathrm{BPV}} \longleftarrow \mathrm{BPV}.\,Setup(1^\lambda)$ is performed by the *credential issuer*, and it takes as input the security parameter λ and then proceeds the following computations:

(1) $param_{\mathrm{Waters}} \longleftarrow \mathrm{Waters}.Setup(1^\lambda)$, where $param_{\mathrm{Waters}}(p, G, g, G_t, e)$ are parameters of a bilinear map, $h \in G$

(2) $param_{\mathrm{LE}} \longleftarrow \mathrm{LE}.Setup(1^\lambda)$, where (p, G, g, G_t, e) are parameters of a bilinear map

(3) Output $param_{\mathrm{LZ}} = (param_{\mathrm{LE}}, param_{\mathrm{Waters}})$

(ii) $((\mathrm{ek}, \mathrm{dk}), (\mathrm{sk}, \mathrm{vk})) \longleftarrow \mathrm{LZ}.KGen(param_{\mathrm{LZ}})$ generates key pair $(\mathrm{sk}, \mathrm{vk})$ of Waters signature for the *credential issuer* while issuing secret and public key pair $(\mathrm{ek}, \mathrm{dk})$ of linear encryption to the *user*, and the detailed procedures are as follows:

(1) $(\mathrm{sk}, \mathrm{vk}) \longleftarrow Waters.KGen(param)$, where $\mathrm{sk} = h^z$ and $\mathrm{vk} = g^z, z \in Z_p$

(2) $(\mathrm{ek}, \mathrm{dk}) \longleftarrow LE.KGen(param)$, where $\mathrm{ek} = (\mathrm{ek}_1 = g^{y_1}, \mathrm{ek}_2 = g^{y_2})$ and $\mathrm{dk} = (y_1, y_2)$, y_1 and $y_2 \in Z_p$

(iii) $\sigma \longleftarrow CredMint(\mathrm{sk}, M)$ is executed by the *credential issuer* to generate a *cryptographic credential token* that is essential to a signature of Waters $\sigma \longleftarrow Waters.Sign(\mathrm{sk}, M)$ by inputting attributes M of the user under sk. In more detail, the *credential issuer* proceeds as follows:

(1) Takes as input a selected random number $s \xleftarrow{\$} Z_p^*$ and calculates

$$\begin{aligned} \sigma_1 &= \mathrm{sk} \cdot F(M)^s \\ \sigma_2 &= g^s, \end{aligned} \tag{3}$$

where $F(M) = u_0 \prod_{i=1}^{k} u_i^{m_i}$ for a vector $u = (u_0, u_1, \ldots, u_k) \in \mathbb{G}^{k+1}$ and an attribute set $M = (m_1, m_2, \ldots, m_k) \in \{0,1\}^k$.

(2) Outputs $\sigma = (\sigma_1, \sigma_2)$

Remarkably, conventional DAC eliminates the centralized origination to issue the cryptographic credential token in a noninteractive approach; however, the simple AC should depend on a centralized part to issue the credential with an associated knowledge proof. Hence, compared with the syntax of DAC, in AC, there is no legal verification after issuing the credential by the origination.

(i) $CredShow(Prover(\mathrm{vk}, \mathrm{ek}, \sigma, M), Verifier(\mathrm{vk}, M))$. This algorithm is played by the prover and the verifier, and we regard the user as a prover for simplicity of illustration, where the witness of the prover is denoted as the randomness for the linear encryption r_1 and r_2.

(1) *Prover* proceeds as follows

(a) $\sigma' \longleftarrow Blind(\sigma)$, selects a randomness $s' \xleftarrow{\$} Z_p$ to blind the issued credential σ from the centralized credential issuer (*a.k.a*, certification authority), where $\sigma \longleftarrow \mathrm{Waters.Sign}(\mathrm{sK}, M) = (\sigma_1 = \mathrm{sk} \cdot F(M)^s$, $\sigma_2 = g^s)$, and outputs $\sigma' = (\sigma_1', \sigma_2')$, i.e.,

$$\left(\sigma_1' = \sigma_1 \cdot F(M)^{s\prime}, \sigma_2' = \sigma_2 \cdot g^{s'}\right). \tag{4}$$

(b) $ct \longleftarrow LE.Enc(\mathrm{ek}, \sigma')$, selects two different random numbers $(r_1, r_2) \xleftarrow{\$} Z_p$, encrypts σ' under ek, and outputs the ciphertext $ct = (c_1 = \mathrm{ek}_1^{r_1}, c_2 = \mathrm{ek}_2^{r_2}, c_3 = g^{r_1+r_2} \cdot \sigma_1', \sigma_2')$
At the end of this phase, the prover sends $ct = (c_1 = \mathrm{ek}_1^{r_1}, c_2 = \mathrm{ek}_2^{r_2}, c_3 = g^{r_1+r_2} \cdot \sigma_1', \sigma_2')$ to verifier

(2) Upon receiving the ciphertext, the *verifier* proceeds as follows

(a) $\mathrm{hk} \longleftarrow SPHF.HashKG(param, L)$ and outputs the hashing key $\mathrm{hk} = (x_1, x_2, x_3) \in Z_p^3$ by picking up three randomnesses x_1, x_2, and x_3 from Z_p

(b) $\mathrm{hp} \longleftarrow SPHF.ProjKG(param), L, \mathrm{hk}, \mathrm{ek}$ and outputs the projective hashing key $\mathrm{hp} = (\mathrm{hp}_1 = \mathrm{ek}_1^{x_1} g^{x_3}, \mathrm{hp}_2 = \mathrm{ek}_2^{x_2} g^{x_3})$

(c) $v \longleftarrow SPHF.Hash(param, \mathrm{hk}, (L, M), ct)$ and outputs v as $c_1^{x_1} c_2^{x_2} (c_3/M)^{x_3}$; particularly,

$$e(c_1, g)^{x_1} e(c_2, g)^{x_2} \left(\frac{e(c_3, g)}{e(h, \mathrm{vk}) e(F(M), \sigma_2')}\right)^{x_3}. \tag{5}$$

(d) Chooses a random session key K and a random challenge $r \in \{0,1\}^n$, and computes $Q = K \oplus \mathrm{KDF}(v)$ and $W = K \oplus r$.
Finally, the *verifier* sends (hp, Q, W) to the *prover.*

(3) Upon receiving (hp, Q, W), the *prover* proceeds as follows

(a) $v' \longleftarrow SPHF.Proj(param, \mathrm{hp}, (L, M), ct; w)$ and computes $v' = (\mathrm{hp}_1^{r_1} \mathrm{hp}_2^{r_2}, g)$, where $w = (r_1, r_2)$ is a witness owned by user privately. Particularly,

$$e(\mathrm{hp}_1^{r_1} \mathrm{hp}_2^{r_2}, g) = e(g, g)^{(y_1 r_1 x_1 + y_2 r_2 x_2 + (r_1 + r_2) x_3)}. \tag{6}$$

(b) Computes a session $K' = Q \oplus \mathrm{KDF}(v')$ and a random challenge $r' = K' \oplus W$.

At the end of this phase, the *prover* sends r' to verifier. Finally, the *verifier* returns 1 if r' is valid, 0 otherwise.

6. Our Construction: Self-Sovereign Decentralized Anonymous Credential Management System

Below, we will specifically describe our solution that contains four phases, i.e., *Initialization*, *Registration*, *Show Credential*, and *Audit*, as illustrated in Figure 3.

Initialization phase is performed by *Committee* members, and at the end of this phase, all public parameters of DVAC are generated. The committee specifies a language L: $\mathrm{WLin}(wpk, \mathrm{lpk}, M)$ and proceeds the following steps with SPHF over L:

(i) Define a vector $\mathbf{u} = (u_0, \ldots, u_k) \overset{\$}{\leftarrow} G^{k+1}$

(ii) A function $F(M) = u_0 \prod_{i=1}^{k} u_i^{m_i}$ for a vector $u = (u_0, u_1, \ldots, u_k) \in \mathbb{G}^{k+1}$, where m_i is an attribute set $m_i \in \{0,1\}^k, i \in (1, k)$

(iii) A hash function $H(\cdot)$

Then, the committee proceeds as follows:

(1) $param_{\mathrm{DVAC}} \longleftarrow Setup(\lambda)$ generates global parameters $param_{\mathrm{DVAC}} = (\mathrm{G}, \mathrm{G}_t, g, g_t, p, e)$ for the whole system, where $G \,\mathrm{and}\, G_t$ are two circle groups of order p, g is a generator of group G, g_t is a generator of group G_t, and e: $G \times G \longrightarrow G_t$, $g_t = e(g, g)$

(2) Process of key generation

   (i) $(vsk, vpk) \longleftarrow EGSign.KGen(1^\lambda)$ and generate a keypair of ElGamal signature $((vsk = x_v, vpk = g_v^{x_v}))$, where $x_v \overset{\$}{\leftarrow} G$ and g_v is a generator of group G

   (ii) $(wsk, wpk) \longleftarrow Waters.KGen(1^\lambda)$, select a random generator $h_w \overset{\$}{\leftarrow} G$, and generate a keypair of Waters signature $(wsk = h_w^{x_w}, wpk = g^{x_w})$, where $x_w \overset{\$}{\leftarrow} G$

   (iii) $(usk, upk) \longleftarrow EG.KGen(1^\lambda)$ generate a keypair of ElGamal $(usk = x_m, upk = g_m^{x_m})$, where $x_m \overset{\$}{\leftarrow} G$ and g_m is a generator of group G

(3) The committee divides vsk into n shares $(s_1, s_2, \ldots, s_n) \longleftarrow SS.Share(vsk)$, and the committee selects a polynomial $p(X)$ of order n:

$$p(X) = x_v + a_1 X + a_2 X^2 + \cdots + a_{t-1} X^{t-1}$$
$$\wedge\, p(0) = x_v = \lambda_1 s_1 + \lambda_2 s_2 + \cdots + \lambda_t s_t, \qquad (7)$$

where n is the number of members of committee, λ_j is publicly constructible Lagrange coefficient, and t is the threshold value, $t < n$.

(4) Outputs $param_{\mathrm{DVAC}}$, (vsk, vpk), (wsk, wpk), (usk, upk), and $(s_1, s_2, \ldots, s_n)$

Finally, $(s_1, s_2, \ldots, s_n)$ is distributed by the dealer to all members of committee randomly.

(i) *Registration.* A new client runs Request($\cdot$) algorithm to send a registration request to the committee. The committee runs an Authenticate($\cdot$) algorithm to issue credentials to a new client. The entire process we have illustrated by the first part of Figure 3. When a new client joins this system, he will get his unique id on the blockchain.

(1) *New Client* generates his public key and private key locally and proceeds as follows:

(a) $(\mathrm{esk}, \mathrm{epk}) \longleftarrow EG.KGen(1^\lambda)$, where $\mathrm{esk} = x_e, \mathrm{epk} = g_e^{x_e}, x_e \overset{\$}{\leftarrow} G$, and g_e is a random generator in group G. The keypair $(\mathrm{esk}, \mathrm{epk})$ is used to sign and encrypt message at the registration phase.

(b) $(\mathrm{lsk}, \mathrm{lpk}) \longleftarrow LE.KGen(1^\lambda)$, where $\mathrm{lsk} = (\mathrm{lsk}_1 = x_1, \mathrm{lsk}_2 = x_2), \mathrm{lpk} = (\mathrm{lpk}_1 = g^{x_1}, \ \mathrm{lpk}_2 = g^{x_2})$, and x_1 and $x_2 \overset{\$}{\leftarrow} G$. The key pair $(\mathrm{lsk}, \mathrm{lpk})$ will be used in the second phase.

(c) $\mathrm{ck} \longleftarrow Ped.KGen(1^\lambda)$ and output a commitment parameter $\mathrm{ck} \overset{\$}{\leftarrow} G$.

(d) $\mathrm{attrs} \longleftarrow f(\mathrm{info})$, call attribution extraction function $f(\cdot)$, extract his attributes $attrs$ from his private information $info$, and output $attrs$.

(e) $I \longleftarrow EG.Enc(\mathrm{info}, vpk)$ and output a ciphertext I by calculating $I = g_v^{r_1}, vpk^{r_1} \cdot \mathrm{info}$, where $r_1 \overset{\$}{\leftarrow} G$.

(f) $C_I \longleftarrow Ped.Com(I, \mathrm{ck})$, take inputs as commitment parameter ck, and output a commitment $C_I = g_c^I h_c^{\mathrm{ck}}$ for ciphertext I, where g_c and h_c are two random generator of group G.

(g) $m \longleftarrow EG.Enc(C_I, \mathrm{attrs}, \mathrm{id}; upk)$ and output a ciphertext m by calculate $m = (g_m^{r_2}, upk^{r_2} \cdot (C_I \| \mathrm{attrs} \| \mathrm{id}))$, where $r_2 \overset{\$}{\leftarrow} G$.

(h) $s_m \longleftarrow EGSign.Sign(m, \mathrm{esk})$, sign for ciphertext m, and output a signature $s_m = (g_e^{k_e}, k_e^{-1}(H(m) - t x_e n g_e^{k_e}))$ by esk_i, where g_e is a random generator of group G and $k_e \overset{\$}{\leftarrow} G$. Finally, the *New Client* sends (m, s_m) to the committee. The request part is all done offline by the *New Client*.

(2) After receiving the request from *New Client*, the *Committee* proceeds as follows:

(a) $(0, 1) \longleftarrow EGSign.Verify(m, s_m, \mathrm{epk})$ and verify the validity of m by epk at first. Calculate $g_e^{H(m)} \overset{?}{=} (g_e^{x_e})^{g_e^{k_e}} (g_e^{k_e})^{k_e^{-1}(H(m) - x_e g_e^{k_e})}$. Continue if the output is 1, otherwise return false to New Client.

(b) $(C_I, \mathrm{attrs}, \mathrm{id}) \longleftarrow EG.Dec(m, usk)$ and output $(C_I, \mathrm{attrs}, \mathrm{id})$ by calculating $(\mathbf{C}_I \| \mathrm{attrs} \| \mathrm{id}) = ((upk)^{r_2} \cdot (\mathbf{C}_I \| \mathrm{attrs} \| \mathrm{id})) \cdot ((g_m^{r_2})^{usk})^{-1}$.

(c) $\sigma \longleftarrow Waters.Sign(\mathrm{attrs}, wsk)$, take as input a selected random number $r_w \overset{\$}{\leftarrow} G$, and calculate and output $\sigma = (\sigma_1 = h_w^{x_w} \cdot F(\mathrm{attrs})^{r_w}, \sigma_2 = g^{r_w})$.

(d) Upload (id: C_I, epk) to bulletin board. This information will not be changed forever. Finally, the *Committee* returns σ to *New Client*.

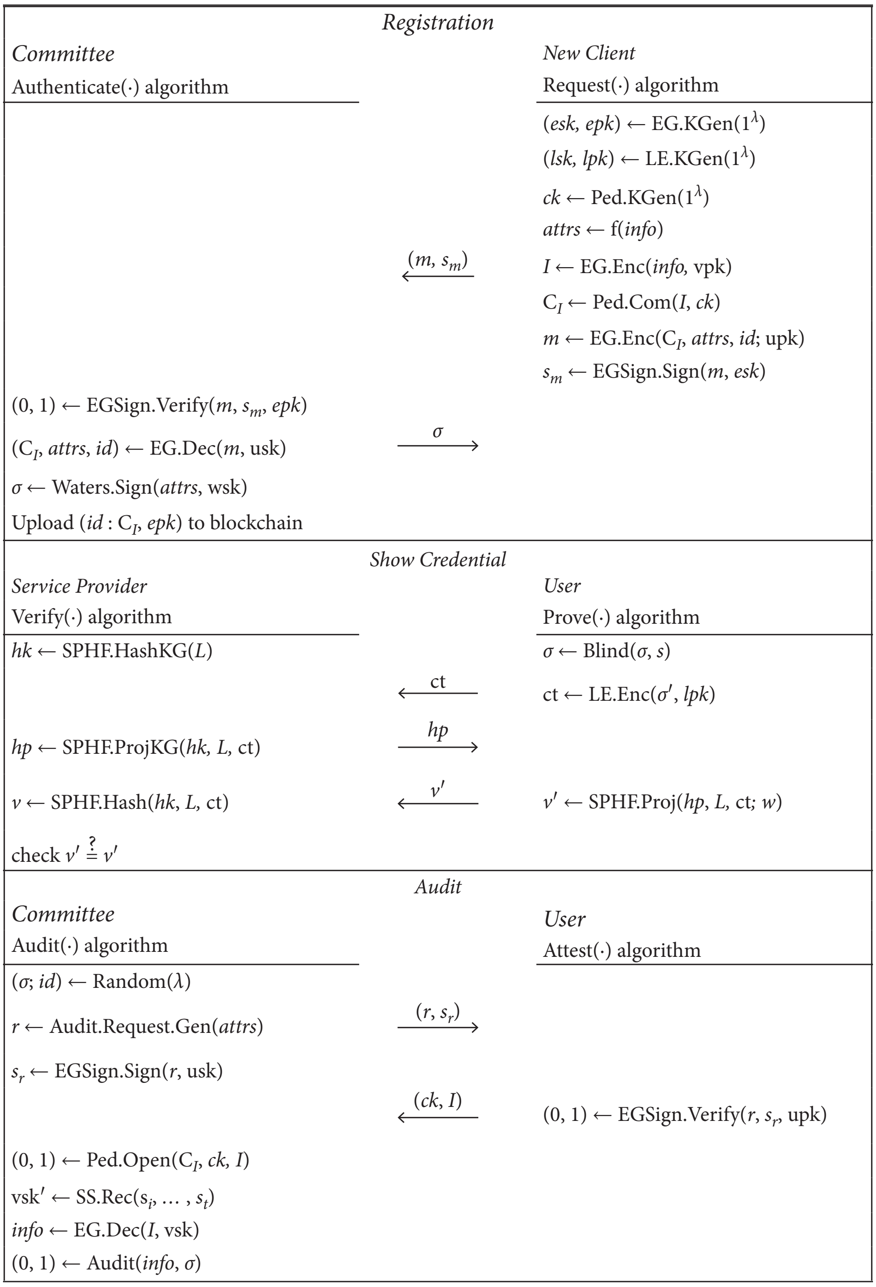

Registration
Committee
Authenticate(·) algorithm
New Client
Request(·) algorithm
(esk, epk) ← EG.KGen(1^λ)
(lsk, lpk) ← LE.KGen(1^λ)
ck ← Ped.KGen(1^λ)
attrs ← f(info)
(m, s_m)
I ← EG.Enc(info, vpk)
C_I ← Ped.Com(I, ck)
m ← EG.Enc(C_I, attrs, id; upk)
s_m ← EGSign.Sign(m, esk)
(0, 1) ← EGSign.Verify(m, s_m, epk)
σ
(C_I, attrs, id) ← EG.Dec(m, usk)
σ ← Waters.Sign(attrs, wsk)
Upload (id : C_I, epk) to blockchain
Show Credential
Service Provider
Verify(·) algorithm
User
Prove(·) algorithm
hk ← SPHF.HashKG(L)
σ ← Blind(σ, s)
ct
ct ← LE.Enc(σ′, lpk)
hp
hp ← SPHF.ProjKG(hk, L, ct)
v′
v ← SPHF.Hash(hk, L, ct)
v′ ← SPHF.Proj(hp, L, ct; w)
check v′ ?= v′
Audit
Committee
Audit(·) algorithm
User
Attest(·) algorithm
(σ; id) ← Random(λ)
(r, s_r)
r ← Audit.Request.Gen(attrs)
s_r ← EGSign.Sign(r, usk)
(ck, I)
(0, 1) ← EGSign.Verify(r, s_r, upk)
(0, 1) ← Ped.Open(C_I, ck, I)
vsk′ ← SS.Rec(s_i, … , s_t)
info ← EG.Dec(I, vsk)
(0, 1) ← Audit(info, σ)

Figure 3: An illustration of DVAC.

(ii) *Show Credential.* After a new client gets his credentials from the committee, it means that he is a legal user of blockchain. When a user wants to obtain a service from SP, he needs to provide the corresponding credentials. And, this process is anonymous for SP. This means that the SP can only provide the appropriate service based on the credentials, and it can learn no information from the user. This phase is shown in the second part of Figure 3. User runs Prove(·) algorithm and SP run Verify(·) algorithm.

(1) User proceeds as follows:

(a) $\sigma' \longleftarrow Blind(\sigma, s)$, select a randomness $s \overset{\$}{\leftarrow} Z_p$ to blind the issued credential σ from the centralized credential issuer (*a.k.a*, certification authority), where $\sigma \longleftarrow \mathrm{Water.Sign}(\mathrm{wsk}, \mathrm{attrs}) = (\sigma_1 = \mathrm{wsk} \cdot F(\mathrm{attrs})^{r_w}, \sigma_2 = g^{r_w})$, and output $\sigma' = (\sigma_1', \sigma_2')$, i.e., $(\sigma_1' = \sigma_1 \cdot F(M)^s, \sigma_2' = \sigma_2 \cdot g^s)$

(b) $ct \longleftarrow LE.Enc(\sigma', \mathrm{lpk})$, select two different random numbers $(r_1, r_2) \overset{\$}{\leftarrow} Z_p$, encrypt σ' under lpk, and output the ciphertext $ct = (c_1 = \mathrm{lpk}_1^{r_1}, c_2 = \mathrm{lpk}_2^{r_2}, c_3 = g^{r_1+r_2} \cdot \sigma_1', \sigma_2')$

At the end of this phase, the prover sends $ct = (c_1 = \mathrm{lpk}_1^{r_1}, c_2 = \mathrm{lpk}_2^{r_2}, c_3 = g^{r_1+r_2} \cdot \sigma_1', \sigma_2')$ to SP.

(2) Upon receiving the ciphertext, the SP proceeds as follows:

(a) $\mathrm{hk} \longleftarrow SPHF.HashKG(param_{\mathrm{DVAC}}, L)$ and output the hashing key $\mathrm{hk} = (x_1, x_2,$

$x_3) \in Z_p^3$ by picking up three randomnesses x_1, x_2, and x_3 from Z_p

(b) $\text{hp} \longleftarrow \mathit{SPHF.ProjKG}(\mathit{param}_{\text{DVAC}}, L, \text{hk}, \text{lpk})$ and output the projective hashing key $\text{hp} = (hp_1 = \text{lpk}_1^{x_1} g^{x_3}, \text{hp}_2 = \text{lpk}_2^{x_2} g^{x_3})$

(c) $v \longleftarrow \mathit{SPHF.Hash}(\mathit{param}_{\text{DVAC}}, \text{hk}, (L, \text{attrs}), ct)$ and output v $\text{as}c_1^{x_1} c_2^{x_2} (c_3/\text{attrs})^{x_3}$; particularly,

$$e(c_1, g)^{x_1} e(c_2, g)^{x_2} \left(\frac{e(c_3, g)}{e(h, wpk) e(F(\text{attrs}), \sigma_2')} \right)^{x_3}. \quad (8)$$

Finally, the SP sends hp to the User.

(3) $v' \longleftarrow \mathit{SPHF.Proj}(\mathit{param}_{\text{DVAC}}, \text{hp}, (L, M), ct; w)$, and upon receiving hp, the User computes $v' = (\text{hp}_1^{r_1} \text{hp}_2^{r_2}, g)$, where $w = (r_1, r_2)$ is a witness owned by user privately.

At the end of this phase, the User sends v' to SP. Finally, the *verifier* returns 1 if r' is valid, 0 otherwise.

(i) *Audit.* We denote the algorithm implemented by the committee and user, respectively, as Attest(·) and Audit(·). This phase is shown in the third part of Figure 3.

(1) The *Committee* first performs the following actions:

(a) $(\sigma; \text{id}) \longleftarrow \mathit{Random}(\lambda)$, and when running the Audit(·) algorithm, the committee randomly selects id on blockchain and the corresponding credentials committee had issued.

(b) $r \longleftarrow \mathit{Audit.Request.Gen}(\text{attrs})$, and the committee generates a request message r based on the relevant credentials.

(c) $s_r \longleftarrow \mathit{EGSign.Sign}(r, usk)$, and generate a signature of the request message r, and the committee calculates $s_r = (g_m^{k_m}, k_m^{-1}(H(r) - t\ x_m n g_m^{k_m}))$, where $k_m \overset{\$}{\leftarrow} G$.

After that, the *Committee* sends (r, s_r) to *User.* When the user receives the committee's audit request, the user should decide whether he is audited. If the user refuses, the committee will add a suspect tag to the blockchain. The user's behavior in the future will be affected. If the user accepts, he/she will do the following:

(2) $(0,1) \longleftarrow \mathit{EGSign.Verify}(r, s_r, upk)$, and verify the validity of the audit request by computing $g_m^{H(r)} \overset{?}{=} (g_m^{x_m})^{g_m^{k_m}} (g_m^{k_m})^{k_m^{-1}(H(r) - x_m g_m^{k_m})}$.

Then, the *User* sends his opening ck and I to *Committee.*

(3) Upon reception of opening ck and I, the committee executes as follows:

(a) $(0,1) \longleftarrow \mathit{Ped.Open}(C_I, \text{ck}, I)$, and the committee opens the user's commitment which uploads to blockchain in the registration phase and computes $g_c^I h_c^{ck} \overset{?}{=} C_I$; return true if output 1; otherwise, return false to the user.

(b) $vsk' \longleftarrow \mathit{SS.Rec}(s_i, \ldots, s_t)$, and vsk' is a temporary private key related to vsk, and it can only be used once for each user. The committee recovers temporary private key by calculate $vsk' = (g_v^{r_1})^{\lambda_1 s_1}, \ldots, (g_v^{r_1})^{\lambda_t s_t}$ and then output vsk'.

(c) $\text{info} \longleftarrow \mathit{EG.Dec}(I, vsk)$, and the committee calculates $\text{info} = (vpk)^{r_1} \cdot \text{info} \cdot (vsk\prime)^{-1}$

(d) $(0,1) \longleftarrow \mathit{Audit}(\text{info}, \sigma)$, and the committee verifies whether σ is generated by info. If the audit fails, the committee will remove the user's information on the blockchain; otherwise, return true.

(ii) *Secret Refresh.* This phase usually has a fixed amount of time to run within a period, unless something special happens to trigger it (such as a sudden occurrence). Same as the *Audit* phase, there must be approval by threshold members of the committee in the current period, and the *Secret Refresh* phase will be run.

(1) $(m, k, l) \longleftarrow \mathit{SS.RefSetup}(\lambda)$, and all members of the committee decide the new number of members m, the new threshold k, and the new time of interval l.

(2) $u_{i,j} \longleftarrow \mathit{SS.RefCompute}(n, m)$, and each of them constructs a random polynomial of the form $\delta_i(z) = \delta_{i,1} z^1 + \delta_{i,2} z^2 + \cdots + \delta_{i,k} z^{k-1}$, where $\delta_{i,k}$ is a coefficient of polynomial $\delta_i(z)$. Then, they compute and send all other players $u_{i,j} = \delta_i(j)$, where $i \in (0, n)$ and $j \in \{0, 1, \ldots, m\}$.

(3) $s_i^{l+1} \longleftarrow \mathit{SS.Rec}(s_i, u_{1,j}, \ldots, u_{n,j})$, and each of them updates their share by $s_i^{l+1} = s_i^l + u_{1,j}^l + \cdots + u_{n,j}^l$.

$$\text{DVAC}\mathit{Adv}_{\mathscr{A}}^{\text{sem}} \leq 2 \cdot \text{DDH}\mathit{Adv}_{B_{\text{DDH}}} + 2 \cdot \text{DL}\mathit{Adv}_{B_{\text{DL}}}. \quad (9)$$

Theorem 1 (correct). *DVAC is a scheme which satisfies soundness.*

Proof. The soundness of *Show Credential* phase relies on the correctness of *SPHF*, $\mathit{SPHF.Hash}(\text{hk}, L, ct) = \mathit{SPHF.Proj}(\text{hp}, L, ct; w)$, and $v' = v$. □

Proof. The soundness of the Audit phase relies on the binding of Pedersen commitment and the correctness of Shamir's Secret Sharing, $g_c^I h_c^{ck} = C_I, vsk\prime == (g_v^{r_1})^{\lambda_1 s_1}, \ldots, (g_v^{r_1})^{\lambda_t s_t}$, thus $\text{info} = (vpk)^{r_1} \cdot \text{info} \cdot (vsk\prime)^{-1}$. □

Theorem 2 (semantically secure). *The DVAC is Semantically Secure if DDH holds for G, and the commitment scheme is perfectly hiding:*

Proof. We assume that an adversary $\mathscr{A}$ against the semantic security of our scheme with advantage ϵ. We start from this initial security game. $\mathscr{G}_0$ is consistent with the situation in a real attack.

Game $\mathscr{G}_0$. Let us emulate this security game:

(1) $\mathscr{B}$ emulates the initialization of the system: it runs $EG.KGen(1^\lambda)$, generates (vpk, upk), and sends (vpk, upk) to adversary $\mathscr{A}$

(2) $\mathscr{B}$ simulates oracles of the transcripts of protocol:

   (i) Runs $EG.Enc(M, vpk)$ for a message M and outputs I

   (ii) Runs $Ped.Com(I, \mathrm{ck})$ for a I and outputs C_I

   (iii) Runs and outputs m

   (iv) Runs $EG.Enc(C_I, \mathrm{attrs}, \mathrm{id}; upk) EGSign.Sign(m, \mathrm{esk})$ for an I and outputs s_m

(3) $\mathscr{A}$ generates two inputs (M_0, M_1) and sends to $\mathscr{B}$

(4) $\mathscr{B}$ chooses a random bit $b \longleftarrow \{0, 1\}$ and simulates the protocol for M_b, and then, $\mathscr{B}$ sends m and s_m to adversary $\mathscr{A}$

(5) $\mathscr{A}$ outputs a bit b' and sends b' to $\mathscr{B}$

In this game, $\mathscr{B}$ only plays the role of a challenger; in $\mathscr{A}'s$ perspective, he/she is still interacting with the real DVAC. We then modify the challenger to obtain Games $\mathscr{G}_1$ and $\mathscr{G}_2$. In each game, b denotes the random bit chosen by the challenger $\mathscr{B}$, while b' denotes the bit output by $\mathscr{A}$. Also, for $j = 0$, 1, and 2, we define W_j to be the event $\mathscr{A}$ win this game for $b' = b$. By definition, we have

$$\mathrm{DVAC}Adv_{\mathscr{A}}^{\mathrm{sem}} = \left|\Pr[W_0] - \frac{1}{2}\right|. \tag{10}$$

Game $\mathscr{G}_1$. In this game, the challenger is as $\mathscr{G}_0$, except a little modifications as follows:

(i) $\mathscr{B}$ uses a randomness $\gamma \overset{\$}{\leftarrow} G$ such that $m = (g_m^{r_2}, g_m^{\gamma} \cdot (C_I \| \mathrm{attrs} \| \mathrm{id}))$.

Adversary $\mathscr{B}$ plays attack game against challenger of DDH and plays the role of challenger to $\mathscr{A}$ as $\mathscr{G}_0$. When $\mathscr{A}$ outputs b', if $b = b'$, then $\mathscr{B}$ outputs 1; else, outputs 0. So, we have

$$\mathrm{DDH}Adv_B^{\mathrm{sem}} = \left|\Pr[W_0] - \Pr[W_1]\right|. \tag{11}$$

Game $\mathscr{G}_2$. In this game, the challenger is as $\mathscr{G}_1$, except a little modifications as follows:

(ii) $\mathscr{B}$ chooses a random bit $b \in \{0, 1\}$, sends to challenger of DL, and then obtains C_b'. Finally, $\mathscr{B}$ sends C_b' to adversary $\mathscr{A}$.

In more detail, adversary $\mathscr{B}$ plays attack game against challenger of DL and plays the role of challenger to $\mathscr{A}$. When $\mathscr{A}$ outputs b', if $b = b'$, then $\mathscr{B}$ outputs 1; else, outputs 0. Based on DL assumption, the adversary $\mathscr{A}$ has only 1/2 probability win this game. So, we have

$$\mathrm{DL}Adv_B^{\mathrm{sem}} = \left|\Pr[W_1] - (1/2)\right|. \tag{12}$$

Combining 2 and 3, yields 1, which completes the proof of the theorem. □

7. Self-Sovereign Decentralized Identity Management via DVAC

7.1. Application of Identity Management. In this section, we discuss several of the applications by DVAC. We consider the scenario where a user wants to register a long-term identity credential on the blockchain. This credential enables repeated presentation to any third parties several times without revealing extrainformation. A third party could verify the validity (whether it has the corresponding attributes) of credentials and provide related services for the user. All users on blockchain will be managed by the miner nodes which we called members of the committee. Miner nodes are not permanent, and each user could be a miner node. In addition to maintaining the system, these nodes are responsible for auditing users with a decentralized operation. This application extends the work of DAC [4] which does not consider audit and the issuance of a decentralized credential.

Our identity management system is based on blockchain and cryptocurrency. We use a public ledger to record the credentials of users with their other information, such as the bulletin board in DAC. There are three types of parties except for the public ledger: a group of credentials issuance and audit *Committee*, a set of SP, and a set of *User*.

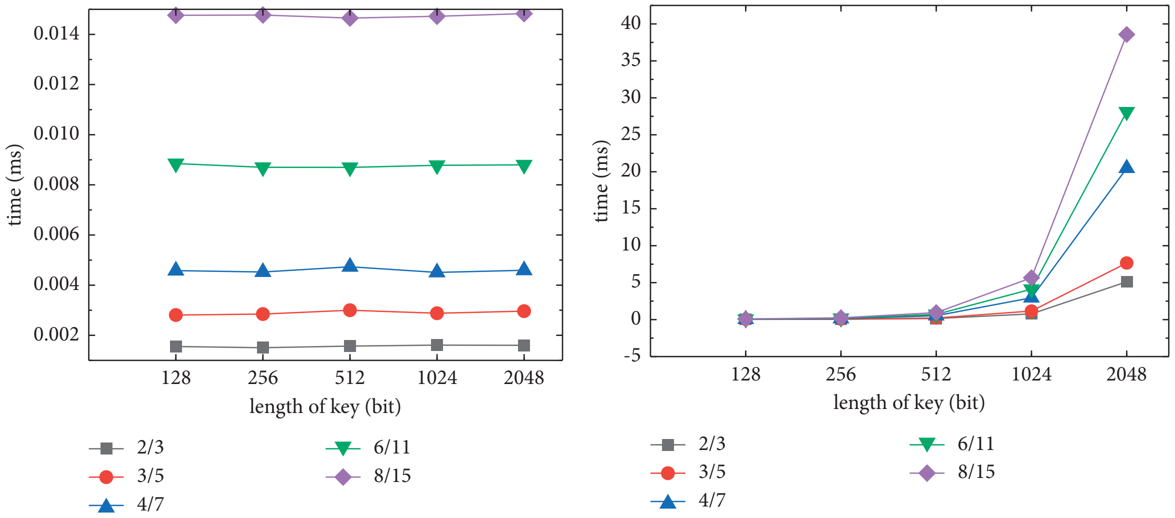
Before the system initializes, the first batch of members of *Committee* should be specified. As shown in Figure 3, the system parameters have been setup at the initialization phase. A user sends a registration request, a commitment of encrypted privacy information, and some information needed to prove his identity attributes to *Committee*. These attributes could be age, address, or credibility. The *Committee* checks the user's information and issues a long-term credential (signature) on its sign private key *wsk*. This credential can be obtained only once and will not gonna change.

When a user wants to get service from a SP, he/she needs to show his attribute that satisfies the request of the SP. He/she should show the corresponding credential (age, sex, or property) through a zero-knowledge proof for a specified verifier. To prevent SP replay attacks, he/she blinds this credential at first. Then, he/she follows the WLin language interactive with SP. This process could not reveal any other information except the user's attributes.

Committee could doubt attributes' authenticity of users and combine with an audit algorithm to judge it periodically. At first, the *Committee* selects users on blockchain randomly. Then, the *Committee* sends an audit request to these users. It means that the right to accept the audit or not is in the hands of the users. If the user refuses, the *Committee* will mark him and his reputation will be impacted. If users accept, they send commitment openings (*ck*, *I*) to *Committee*. The *Committee* opens the commitment that is sent by

Table 1: Comparison of the identity management system.

Scheme	Passport [1]	CertCion [16]	DAC[4]	DBLACR [5]	DVAC
Unlinkability	✗	✗	✓	✓	✓
Unforgability	✗	✗	✓	✓	✓
Anonymity	✗	✓	✓	✓	✓
D-Auditability	✗	✗	✗	✗	✓



(a)

(b)

Figure 4: (a) Time-consuming of secret distribution in different thresholds (left). (b) Time-consuming of secret reconstruction (right).

the user in the registration phase. Finally, the *Committee* recovers the corresponding secret key and audits the privacy information of the user. If the result of the audit is right, return true; otherwise, the user will be removed from the blockchain and investigated relevant legal liabilities.

In Table 1, we compare our construction DVAC with other systems. Although DVAC has a slight deficiency in performance because of the secret sharing, it has a sufficient guarantee in security and privacy.

7.2. Future Work. DVAC's focus is on adding audit capability to a decentralized identity management system and addressing the single point of failure that can occur during the audit process. However, it is still an interactive proof system between the user and the SP, which will incur unnecessary waiting time loss. If in an environment with high network latency, it is likely to cause network congestion. DVAC also does not support forward-secure for audit secret key. The future work of DVAC is to provide a forward-secure audit and noninteractive proof system.

8. Evaluation

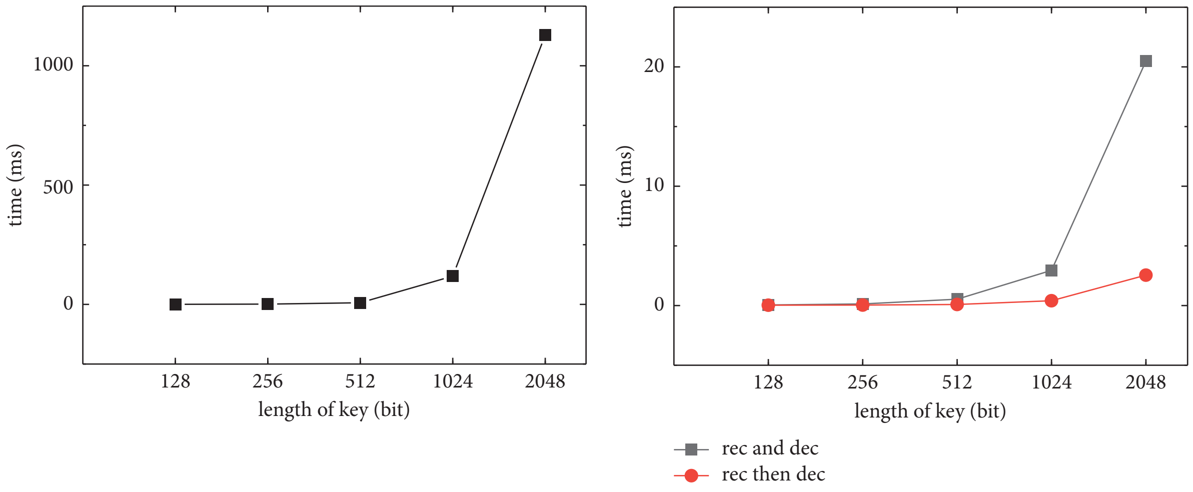
To evaluate DVAC, we implemented each step of DVAC with C++. Our essential environment is based on GMP and PCB library. We run microbenchmarks on a 4 core Intel machine with i7-8500T 2.1 GHz CPU and 8 GB of RAM, running 64-bit Windows 10. We measured the time-consuming of the key generation process, secret sharing process, and secret reconstruct process for secret keys of different bit lengths because the real interactive system is related to the speed of network transmission and is not stable. Let us assume that the network transfer does not take time and only measures the time consumption in the local calculation.

As shown in Figure 4(a), we compare the time-consuming of different key lengths. At the secret key distribution stage, there is no time consumption of secret keys of different lengths is no significant change. The timing of the distribution of the secret keys is only related to the number of participants involved in the distribution.

Figure 4 rightly depicts the time required for the auditor to recover the user information for different thresholds and different secret key lengths during the audit phase. The values of the points are very close to each other when the secret key size is less than 1024 bits. Only when the secret key size is 1024 bit and 2048 bit, the time difference required by different threshold sizes are relatively significant changes.

As shown in Figure 5(a), we can find that, with the increase of the length of the secret key, the generation time of the secret key increases exponentially. However, this increase in time is not something we should care about because the entire initialization phase is done entirely offline.

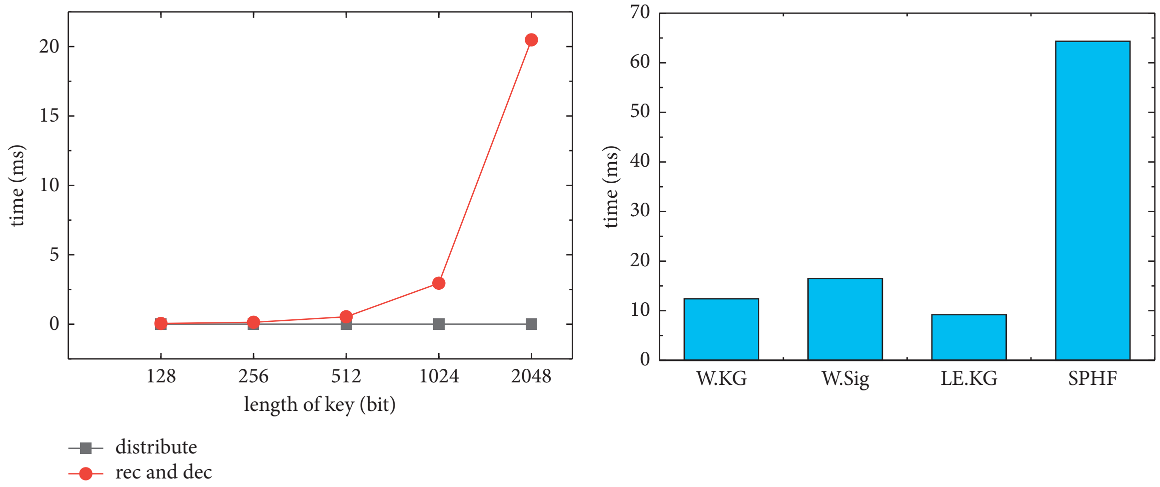
To evaluate the time-consuming during the secret reconstruction phase, as shown in Figure 5(b) and Figure 6(a), we select the threshold as 4 and evaluate the 4 out of 7 setting. The results imply that the time consumption in the



(a)

(b)

Figure 5: (a) Time-consuming of key generation (left). (b) Time-consuming of secret reconstruction by different methods (right).



(a)

(b)

Figure 6: (a) Time-consuming of secret distribution and secret reconstruction (left). (b) Time-consuming in Waters key-generation (W.KG), Waters signature (W.Sig), linear encryption key-generation (LE.KG), and entire interaction process of SPHF (right).

audit stage is much greater than that in the secret key distribution stage when the length of the secret key is greater than 1024 bits. Further, comparing with traditional "reconstruction then decryption," the approach of "reconstruction and decryption" is more time-consuming. This is due to the multiple modular exponentiations. When the length of the secret key is longer, its time grows faster. This is the main disadvantage in realization.

As shown in Figure 6(b), we compare the time-consuming verification algorithms including key-generation and signature of Waters scheme, key-generation of linear encryption, and total verification of SPHF.

9. Conclusion

The existing anonymous credential identity management system usually exposes two shortcomings when it is implemented. One is that the correctness of identity information cannot be guaranteed when the privacy of the user's identity information is guaranteed. Second, its centralized management will lead to a single point of the failure problem. In this paper, we propose DVAC, a designated-verifier anonymous credential in self-sovereign decentralized identity management. We added the audit function to solve the problem that the correctness of user identity information could not be guaranteed. We also solved the problem of single-point failure of the centralized management system to some extent through secret sharing and realized decentralized identity management. We provide the detailed step of the DVAC system and the cryptographic primitives underlying DVAC. We also implement and evaluate DVAC and application of identity management. DVAC provides a new way for designing an identity management system.

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 National Natural Science Foundation of China (61702294), Applied Basic Research Project of Qingdao City (17-1-1-10-jch).

References

[1] R. O. Microsoft, "Net passport: a security analysis," *Computer*, vol. 36, no. 7, pp. 29–35, 2003.

[2] D. Fisher, "Final report on diginotar hack shows total compromise of ca servers," 2012.

[3] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008.

[4] C. Garman, M. Green, and I. Miers, "Decentralized anonymous credentials," in *Proceedings of the 21th NDSS 2014*, February 2014.

[5] R. Yang, M. H. Au, Q. Xu, and Z. Yu, "Decentralized blacklistable anonymous credentials with reputation," *Computers & Security*, vol. 85, pp. 353–371, 2019.

[6] J. Camenisch and A. Lysyanskaya, "Signature schemes and anonymous credentials from bilinear maps," in *Proceedings of the 24th CRYPTO 2004*, vol. 3152, pp. 56–72, Springer, Santa Barbara, CA, USA, August 2004.

[7] M. H. Au, W. Susilo, Y. Mu, and S. S. M. Chow, "Constant-size dynamic k-times anonymous authentication," *IEEE Systems Journal*, vol. 7, no. 2, pp. 249–261, 2013.

[8] D. Boneh, X. Boyen, and H. Shacham, "Short group signatures,"vol. 3152, pp. 41–55, in *Proceedings of the 24th CRYPTO 2004*, vol. 3152, Springer, Santa Barbara, CA, USA, August 2004.

[9] N. Begum, T. Nakanishi, and N. Funabiki, "Efficient proofs for cnf formulas on attributes in pairing-based anonymous credential system," *ICE Transactions on Fundamentals of Electronics Communications and Computer ences*, vol. E96-A, no. 12, pp. 495–509, 2012.

[10] J. Camenisch, S. Mödersheim, and D. Sommer, "A formal model of identity mixer," in *Proceedings of the 15th FMICS 2010*, vol. 6371, pp. 198–214, Springer, Antwerp, Belgium, September 2010.

[11] Z. Li, D. Wang, and E. Morais, "Quantum-safe round-optimal password authentication for mobile devices," *IEEE Transactions on Dependable and Secure Computing*, 2020.

[12] C. Buchholz, "Liberty alliance project-gemeinschaftliche identitätsverwaltung," *Datenschutz und Datensicherheit*, vol. 27, no. 9, 2003.

[13] G. Karjoth, "Access control with IBM Tivoli access manager," *ACM Transactions on Information and System Security*, vol. 6, no. 2, pp. 232–257, 2003.

[14] W. Stackpole, "Centralized authentication services," in *Encyclopedia of Information Assurance*Taylor & Francis, Boca Raton, FL, USA, 2011.

[15] Z. Li, C. Ma, and H.-S. Zhou, "Multi-key FHE for multi-bit messages," *Science China Information Sciences*, vol. 61, no. 2, pp. 029101:1–029101:3, 2018.

[16] C. Fromknecht, D. Velicanu, and S. Yakoubov, "A decentralized public key infrastructure with identity retention. IACR Cryptol," 2014.

[17] M. Ali, J. C. Nelson, R. Shea, and M. J. F. Blockstack, "A global naming and storage system secured by blockchains," in *Proceedings of the USENIX ATC 2016*, pp. 181–194, USENIX Association, Berkeley, CA, USA, June 2016.

[18] A. Sonnino, M. Al-Bassam, S. Bano, S. Meiklejohn, and G. Danezis, "Coconut: threshold issuance selective disclosure credentials with applications to distributed ledgers," 2019.

[19] D. Boneh, C. Gentry, B. Lynn, and H. Shacham, "Aggregate and verifiably encrypted signatures from bilinear maps," in *Proceedings of the EUROCRYPT 2003*, vol. 2656, pp. 416–432, Springer, Warsaw, Poland, May 2003.

[20] D. Pointcheval and O. Sanders, "Short randomizable signatures," in *Proceedings of the CT-RSA 2016*, vol. 9610, pp. 111–126, Springer, San Francisco, CA, USA, February 2016.

[21] A. Shamir, "How to share a secret," *Communications of the ACM*, vol. 22, no. 11, pp. 612-613, 1979.

[22] H. Halpin, "Nym credentials: privacy-preserving decentralized identity with blockchains," in *Proceedings of the CVCBT 2020*, pp. 56–67, IEEE, Rotkreuz, Switzerland, June 2020.

[23] N. Narula, W. Vasquez, and M. Virza, "Zkledger: privacy-preserving auditing for distributed ledgers," in *Proceedings of the 15th NSDI 2018*, pp. 65–80, USENIX Association, Renton, WA, USA, April 2018.

[24] J. Balasch, A. Rial, C. Troncoso, B. Preneel, I. Verbauwhede, and C. Geuens, "Pretp: privacy-preserving electronic toll pricing," in *Proceedings of the 19th USENIX 2010*, pp. 63–78, USENIX Association, Washington, DC, USA, August 2010.

[25] S. Meiklejohn, K. Mowery, S. Checkoway, and H. Shacham, "The phantom tollbooth: privacy-preserving electronic toll collection in the presence of driver collusion," in *Proceedings of the 20th USENIX 2011*, USENIX Association, Berkeley, CA, USA, August 2011.

[26] R. Cramer and V. Shoup, "Universal hash proofs and a paradigm for adaptive chosen ciphertext secure public-key encryption," in *Proceedings of the EUROCRYPT 2002*, vol. 2332, pp. 45–64, Springer, Amsterdam, The Netherlands, April 2002.

[27] Z. Li and D. Wang, "Two-round PAKE protocol over lattices without NIZK," in *Information Security and Cryptology - 14th International Conference, Inscrypt 2018, Fuzhou, China, December 14-17, 2018, Revised Selected Papers, Volume 11449 of Lecture Notes in Computer Science*, pp. 138–159, Springer, Berlin, Germany, 2018.

[28] Z. Li and D. Wang, "Achieving one-round password-based authenticated key exchange over lattices," *IEEE Transactions on Services Computing*, 2019.

[29] Z. Li, Z. Yang, P. Szalachowski, and J. Zhou, "Building low-interactivity multi-factor authenticated key exchange for industrial internet-of-things," *IEEE Internet of Things Journal*, vol. 8, no. 2, pp. 844–859, 2021.

[30] O. Blazy, G. Fuchsbauer, D. Pointcheval, and D. Vergnaud, "Signatures on randomizable ciphertexts," in *Proceedings of the 14th PKC 2011*, vol. 6571, pp. 403–422, Springer, Taormina, Italy, March 2011.

[31] B. Waters, "Efficient identity-based encryption without random oracles," in *Proceedings of the 24th EUROCRYPT 2005*,

vol. 3494, pp. 114–127, Springer, Aarhus, Denmark, May 2005.

[32] M. Blanton, “Online subscriptions with anonymous access,” in *Proceedings of the ACM ASIACCS 2008*, pp. 217–227, ACM, Tokyo, Japan, March 2008.

[33] M. Blanton and W. M. P. Hudelson, “Biometric-based non-transferable anonymous credentials,” in *Proceedings of the 11th ICICS 2009*, vol. 5927, pp. 165–180, Springer, Beijing, China, December 2009.

[34] O. Blazy, D. Pointcheval, and D. Vergnaud, “Round-optimal privacy-preserving protocols with smooth projective hash functions,” in *Proceedings of the 9th TCC 2012*, vol. 7194, pp. 94–111, Springer, Sicily, Italy, March 2012.

Hindawi
Mobile Information Systems
Volume 2021, Article ID 9926106, 11 pages
https://doi.org/10.1155/2021/9926106

Research Article

Research on Accounting Information Security Management Based on Blockchain

Huaqing Shao[1], **Zongli Zhang**[1], **and Bin Wang**[2]

[1] *College of Economics and Management, Jiamusi University, Jiamusi, China*
[2] *College of Information Science and Electronic Technology, Jiamusi University, Jiamusi, China*

Correspondence should be addressed to Zongli Zhang; jeanneettee@163.com

Received 6 March 2021; Revised 10 May 2021; Accepted 25 July 2021; Published 18 August 2021

Academic Editor: Zengpeng Li

At present, accounting information presents various and complex characteristics, which leads to the decline in the comprehensive scheduling level of accounting information security management system. For this problem, a blockchain-based accounting information security management information model is designed. This paper constructs the blockchain accounting information security association blockchain Big Data analysis model and processes the sample data, uses the semantic rough feature matching method to decompose the characteristics of blockchain accounting information, realizes the feature information fusion and autocorrelation feature matching and finally reorganizes and manages the blockchain accounting information security. The simulation results show that this method has better comprehensive scheduling ability, information fusion scheduling ability is greater than 92%, convergence is greater than 91.8%, feature recognition rate is greater than 90.1%, and management accuracy is greater than 95.6%. The design method can effectively improve the security and stability of accounting information storage and management.

1. Introduction

With the development of Big Data information processing and cloud computing technology, it makes the current accounting information security data present diverse, complex, and massive characteristics. In this background, methods concerning the improvement of accounting information security management capabilities have received much attention, while the study of accounting information security management methods is also of great significance in promoting the secure integration and scheduling of accounting information [1].

Privacy and security issues are involved in accounting information databases, and data anonymity can protect the security of accounting information data. In this area, many techniques on data security and privacy protection have been proposed by researchers, such as homomorphic encryption and attribute-based encryption schemes [2, 3]. In recent years, with the development of cloud storage technology, researchers have proposed a cloud technology-based storage service, which achieves the purpose of sharing accounting information data through the control of access rights. Esposito et al. proposed a data sharing model using cloud storage technology in the context of accounting information and enumerated the possible challenges of using blockchain technology in accounting data sharing. However, these accounting information networks rely on a role that is trusted by both parties to the transaction, that is, the use of a trusted third party (TTP) to guarantee the proper conduct of the transaction. This requires the third party to be absolutely trusted and not subject to cyber-attack. However, such an ideal network environment is almost impossible to achieve, so traditional healthcare information solutions are not a good solution.

The research on accounting information security management methods by Zhao and Cheng [4] is based on the Big Data fusion and characterization of blockchain accounting information. The method uses similarity information feature decomposition and quantitative parameter regression analysis methods for internal control and quantitative parameter analysis for accounting information security management and uses local parameter search control for accounting information security management. Among the traditional methods, there are mainly accounting information security management methods of fuzzy information feature detection, accounting information security management methods based on similarity feature analysis, and accounting information security management methods based on elastic template feature matching. These methods construct the elastic Big Data feature analysis model for accounting information security management and perform accounting information security management through fuzzy similarity feature decomposition [5]. Massicotte and Henri [6] discuss how management accounting information is used to monitor strategy implementation in the context of corporate governance. By establishing theoretical attributes and proposing a measurement model, the model captures the board's use of budget, financial, and nonfinancial performance indicators to monitor strategic plans.

However, the adaptability of this method to accounting information security management is not strong, and the level of feature recognition is not high. Pérez-González et al. [7] verified the information security management performance model by collecting data through questionnaire surveys. The results show that information security knowledge sharing, information security education, information security visibility, and security organization practice have a positive impact on information security management performance. But this method for accounting information security management, feature recognition level is low, information clustering is poor. Chen et al. [8] construct an application-oriented quantitative evaluation method of urban security. A new evaluation concept of "comprehensive screening, key analysis, and comprehensive evaluation" is put forward. However, this method has great variability and poor convergence in accounting information management. Mehedi et al. [9] proposed the security management of Ethereum transaction Internet-of-things infrastructure based on blockchain. This method points out that blockchain technology is a luxury technology, which will bring high bandwidth, extended time, and memory cost incompatible with IOT devices. Using terminal equipment as network technology and Ethereum as the blockchain platform, it can produce a back-end system to ensure high availability, improved security and privacy, and replace the traditional back-end system. Xu et al. [10] proposed the integrated application of blockchain in power information management system. The method points out that blockchain technology has been applied in many fields to improve the management and data security of information systems. This paper introduces the application of blockchain technology in power management information system.

First, the composition and structure of blockchain framework are introduced. Then, the blockchain-based authentication application is studied to realize the integration with the existing IT infrastructure. Finally, the advantages and limitations of the integration framework are analyzed. Based on this analysis, it can be seen that the application of blockchain technology in accounting information security management has certain effectiveness. Datta et al. [11] solved the problems in the process of dealing with network attacks based on pin security system and proposed a module to help the secure transmission of sensitive data by encrypting images and other files. However, this method has not been implemented in specific enterprises to verify its effectiveness. Patel et al. [12] proposed a hybrid anomaly detection method in order to solve the problem of consumer network attack under the condition of limited resources. This method only uses basic network information, such as packet size, source port, and target port, time between subsequent packets, transmission control protocol (TCP) flag, and so on. However, this method is difficult to distinguish sensitive data and easy to cause processing error.

In order to solve the problems of poor adaptability and low-level feature identification in the existing accounting information management, this paper proposes a blockchain-based accounting information security management model. Through the feature decomposition of accounting information in blockchain, the features of accounting information are matched by autocorrelation features to realize the security reorganization and management of accounting information in blockchain. The simulation test shows the superiority of the proposed method.

2. Statistical Analysis and Feature Extraction for Accounting Information Security Management

Before designing the accounting information security management model based on blockchain technology in this paper, we first need to statistically analyze the accounting information security management information, extract features, and perform feature fusion processing on this basis. During the risk identification phase of accounting information security management, the focus should be consciously expanded to focus on learning activities in an effort to find gaps between the system and the environment. These gaps can bring reverse effects and threats to the security of accounting information systems. Accounting information security risk identification is the need of accounting information security risk strategy. Risk identification identifies, classifies, and prioritizes the accounting information and accounting-related information in an enterprise to understand which accounting information in an enterprise is the target of various threats and threat tactics, with the goal of protecting this accounting information from threats. The details are shown below.

2.1. Statistical Analysis of Accounting Information Security Management. Due to the large amount of accounting information security management data, which leads to the problem of large errors when building accounting information security management models, the method of fuzzy information feature detection and correlation fusion is used to realize the sample clustering processing of blockchain accounting information. The block-link regression analysis method is used to obtain the random neighbor characteristic parameter analysis model $\phi\ (x_i)$ for accounting information security management. Under the condition of ambiguity information fusion, using semantic combination control, the adaptive quantitative parameter adjustment model for accounting information security management is obtained:

$$G = R^2 + A\sum_i \frac{\phi\ (x_i)}{\xi_i},\qquad \text{s.t: } \phi\ (x_i) \le R^2 + \xi_i. \tag{1}$$

In formula (1), ξ_i is the constraint index parameter set of accounting information security management and R^2 is the random characteristic parameter distribution set of accounting information security management. Through the group regression test analysis method, the random cluster distribution binomial parameter analysis model of accounting information security management indicators is constructed, and the quantitative parameter analysis of accounting information security management is carried out. Through autocorrelation information fusion, the sample parameter test analysis model of accounting information security management is constructed, which is expressed as

$$K = G\sum_i\sum_j \alpha\left(x_i, x_j\right) + \frac{\omega_{\max} - \omega_{\min}}{\lambda}. \tag{2}$$

In formula (2), ω is the adjustment coefficient of accounting information security management, λ is the integration scale of accounting information security management, α is the control factor of accounting information security management, and (x_i, x_j) is expressed as the sample parameter coordinates. Through correlation dimension analysis, using the embedded scheduling method, construct the variable parameter fusion model of accounting information security management, which is expressed as

$$S = \begin{cases} \alpha K_i, & \text{if } i = 1,\\ \lambda\,\mathrm{New}_i, & \text{otherwise.}\end{cases} \tag{3}$$

In formula (3), New_i represents the block scheduling parameter set for accounting information security management. The method of fuzzy information feature detection and correlation fusion is used to realize the sample clustering processing of blockchain accounting information and complete the statistical analysis of accounting information security management.

2.2. Analysis of the Characteristics of Accounting Information Security Management. On the basis of the above statistical analysis, in order to improve the accounting information security management, it is necessary to establish the feature integration model of blockchain accounting information for feature analysis using fuzzy extended sample regression analysis method. In this paper, segmented sample detection and quantile regression analysis methods are applied to accounting information security management, and the dynamic fusion parameter matching set is obtained as

$$U = \beta\int \ln\left(1 + \phi\ (x_i) \times p\right)\mathrm{d}i. \tag{4}$$

In formula (4), β represents the relevant characteristic value of accounting information security management. Under the constraint of elasticity law, the high-order statistical distribution sequence associated with accounting information is $r = r(1), r(2), \ldots, (n)$, and the sampling interval t of random samples of accounting information is obtained. The constraint quantitative index parameter set of information security management is expressed as

$$C = U\sum_{n=1} r(n) + \varphi + t\sum_j \alpha\left(x_i, x_j\right). \tag{5}$$

In formula (5), φ is expressed as the detection statistical characteristic value of the distribution of accounting information blockchain. The larger the detection statistical characteristic value, the higher the degree of restraint of accounting information security management. Thus, the subset of accounting information security distribution constraint parameters is as follows:

$$B = \arg\min\left(\max\left|\sum_{i=1} C \times x_i\right|\right). \tag{6}$$

Based on the aforementioned analysis, constrained regression analysis model is constructed for accounting information security management, which is expressed as $y(t)$. Through the variance fusion of accounting information and the regression analysis results, the detection statistical characteristic value is obtained:

$$\varphi = B\left[y^2(t) - A\sum_i \xi_i\right]. \tag{7}$$

On the basis of the detection statistical characteristic values, in order to achieve balanced scheduling, combined with the blockchain fusion distribution of accounting information security management, the calculation of the constraint object distribution complex envelope $s_i(t)$ of accounting information security management is as follows:

$$s_i(t) = \varphi\int y^2(t) + \frac{\left(p/R^2\right)}{\omega}\mathrm{d}t. \tag{8}$$

In Equation (8), p is denoted as the equilibrium scheduling channel noise. Using random cluster analysis, the blockchain distribution domain A of accounting information is divided into the number $W \times L$ of $(\sqrt{2}/2)R^2 \times (\sqrt{2}/2)R^2$ block-matching regions, and the feature integration model of blockchain accounting information is obtained by fuzzy extended sample regression analysis method as

$$Z = s_i(t) \times X^N + \chi. \tag{9}$$

In equation (9), X^N is the fuzzy component of accounting information, and χ is the fusion coefficient of random feature parameters. Through the aforementioned study, feature decomposition and information fusion are performed using segmented sample detection method [6], so that the accounting information security management feature extraction and fusion processing are completed.

3. Blockchain Accounting Information Security Management

On the basis of the aforementioned fusion treatment of blockchain accounting information security management features, a blockchain security management model is constructed to improve accounting information security management capabilities. China's relevant system of accounting information security is not perfect, the qualities of accounting personnel themselves have serious defects, the management does not pay enough attention to accounting personnel, and so on. Most enterprises will choose to buy more advanced and efficient machines and equipment, or choose to buy more secure and reliable systems or other equipment matching with them. However, this method ignores the subjective and objective factors of the accounting security system, that is, the employees in the accounting positions, and such neglect has laid a hidden danger to the security and stability of the enterprise accounting information system. Even if the enterprise acquires more advanced and sophisticated equipment and adopts more strict accounting system, if the staffs in accounting positions do not have cautious awareness of accounting information security and subconsciously leak out the accounting information, it will bring many insecurity factors to the enterprise, and even directly lead to the business closure or even collapse of the enterprise. Therefore, in order to realize the security of accounting information system, it is especially necessary to realize the security of accounting system, starting with the accounting post staff.

3.1. Blockchain Integration of Accounting Information. The accounting information security management features have been extracted through the content of part 1.2, and this part will realize the blockchain accounting information feature decomposition by semantic rough feature matching method. The security of accounting information involves the security of servers, storage devices, network devices, and users. According to the current accounting information security management needs of enterprises, it is very necessary to establish a practical accounting information security management system. After enterprises choose cloud accounting, cloud accounting service providers focus on the security of hardware and network infrastructure, while enterprises focus their accounting information security efforts on the security management of users. Using the method of high-dimensional feature information space reconstruction and information fusion [13], the fuzzy distribution set of blockchain accounting information fusion is established as

$$D(i) = \frac{Z(i)}{\eta \exp[\ln Z]}. \tag{10}$$

In equation (10), η is the blockchain accounting information fusion degree. The dummy variables of organizational nature are constructed, and the neighborhood equilibrium scheduling method is used to obtain the blockchain accounting information characteristics game parameters using the Big Data fusion scheduling method as

$$k = \exp\left(k^{1/i}\right) - \ln D. \tag{11}$$

Through the method of cooperative innovation and game equilibrium control, the parameter regression analysis model of blockchain combination scheduling to obtain accounting information is

$$N = a\phi\left(x_i\right) + k. \tag{12}$$

In Equation (12), the coefficient $a \geq 1$. Combining the blockchain fuzzy constraint control method of accounting information [14], a regression analysis and constraint evolution model of accounting information association is constructed, which is expressed as

$$Y = \kappa X^N + V^i, \tag{13}$$

$$V^i = N + \sum_{i=1} \lambda\left(p - x_i\right). \tag{14}$$

Equation (13) is associated with the regression model accounting information, and equation (14) is associated with accounting information constraints evolutionary model. κ is expressed as the dynamic parameter distribution set of accounting information security management. The decision-making quantitative set for constructing accounting information security management is M. The adaptive equilibrium control method is adopted to control the associated constraint of accounting information security management, and the blockchain fusion function of accounting information is obtained as

$$H = \alpha Y + v, \tag{15}$$

$$v = \omega V^i \int_{i=1} \kappa\left(G - x_i\right) \mathrm{d}i. \tag{16}$$

Equation (15) is the autocorrelation resolution function of accounting information security management, and equation (16) is the blockchain fusion function of accounting information. Blockchain accounting information feature decomposition by semantic rough feature matching method. The internal control and prudential control analysis model of accounting information is calculated by high-dimensional feature information space reconstruction [15], and the block fusion processing of accounting information is completed up to this point.

3.2. Blockchain Accounting Information Management Optimization. In order to further improve the ability of accounting information security management, this paper adopts the method of random discrete combination control to construct the information fusion and feature reorganization model of blockchain accounting information management [16]. Let F be the covariance fusion model of the internal control of accounting information, and the distributed combination control parameters of blockchain accounting information is obtained as

$$f = v\sqrt{s_i(t)} \times \Delta t. \tag{17}$$

A real-time data clustering analysis model is established for accounting information management, and the window function $h(t)$ for the distribution of blockchain accounting information, thereby obtaining a random probability distribution model for blockchain accounting information reorganization and security management:

$$q(x_i) = \begin{cases} \dfrac{x_i}{f}\exp[v - h(t)], & x_i \geq 0, \\ 0, & x_i < 0. \end{cases} \tag{18}$$

In the case of $f = 1$, the segmented information fusion is realized for the accounting information characteristic parameters, and the detection statistics of the blockchain accounting information management are obtained as

$$E = \left(2 - \frac{\pi}{2}\right)f^2 - \sqrt{q(x_i)}. \tag{19}$$

On the basis of the detection statistics derived from equation (19), information fusion and autocorrelation feature matching processes are performed on the features in the process of blockchain accounting information security management. According to the information feature matching results, the real-time data parameter association knowledge set of accounting information management is established, and the blockchain fusion and factor analysis of accounting information is combined with the multivariate linear fusion method, and the joint feature distribution set of relevant parameters for blockchain fusion and security association of accounting information is obtained as

$$I_{N\times 1} = E \times R_L. \tag{20}$$

The negative binomial regression model is constructed, and the quantile regression test analysis method of aggregation coefficients is used to achieve the information security management of blockchain accounting, and the optimal decision model for the information security management of blockchain accounting is expressed as

$$y_i = \begin{cases} 0, & M - I_{N\times 1}\sum\limits_{j=1} x_j \leq 0, \\ 1, & M - I_{N\times 1}\sum\limits_{j=1} x_i > 0. \end{cases} \tag{21}$$

Using the methods of rough set feature matching and nearest neighbor parameter analysis, the nonlinear constraint statistical feature Q^N of blockchain accounting information security management is obtained, which satisfies the correlation distribution relationship:

$$u\left(Q^N\right) = \text{angle}(y_i) - \varphi + \text{mod}(2\pi). \tag{22}$$

A spatial parameter-matching model for information security management of blockchain accounting is established. According to the information feature-matching results, the method of similarity feature decomposition [17] is used to realize the information security reorganization and management of blockchain accounting information, and the realization process is shown in Figure 1.

As can be seen from Figure 1, the implementation process of blockchain accounting information security management is mainly divided into the following steps:

Step 1: Sample regression analysis of accounting information data security.

Step 2: According to the results of sample regression analysis, build the blockchain Big Data analysis model of accounting information security, and use the accounting information data.

Step 3: Feature analysis and segment fusion of the data in the model.

Step 4: Make linear prediction for the characteristics of accounting information after fusion.

Step 5: Divide the prediction results into training set and sample set, and carry out adaptive learning on the training set.

Step 6: Judge the output result of Step 5, output the accounting information security management data if the threshold is set, and return to Step 1 if the threshold is not set.

3.3. Accounting Information Security Relationship System Requirements. The main users of the system are identified through business process analysis as the main user roles of the system, which are company financial staff, system administrators, and company leaders. The accounting system has six major functional modules: account information management, accounting management, data management, report statistics, decision support, and system management. The functions required by the company's financial staff include adding, modifying, querying, and deleting information in the accounting information management module. The finance staff can add, modify, query, and delete information in the accounting management module, and submit printouts of relevant data. In the data management module, finance staff can add, modify, delete, and query the information in it. In the report statistics module, finance staff and company leaders are mainly able to query and print reports. In the decision support module, company leaders can get data from the data analysis submodule to support decision-making. In the system management module, the administrator users can add, modify, delete, and query user information in the user management submodule. In the function module permission management submodule, users are able to set their operation permissions. In the company announcement management submodule, the company announcement information can be added, modified, and deleted. In the log management submodule, the system administrator can also view and delete the operation log information of the system.

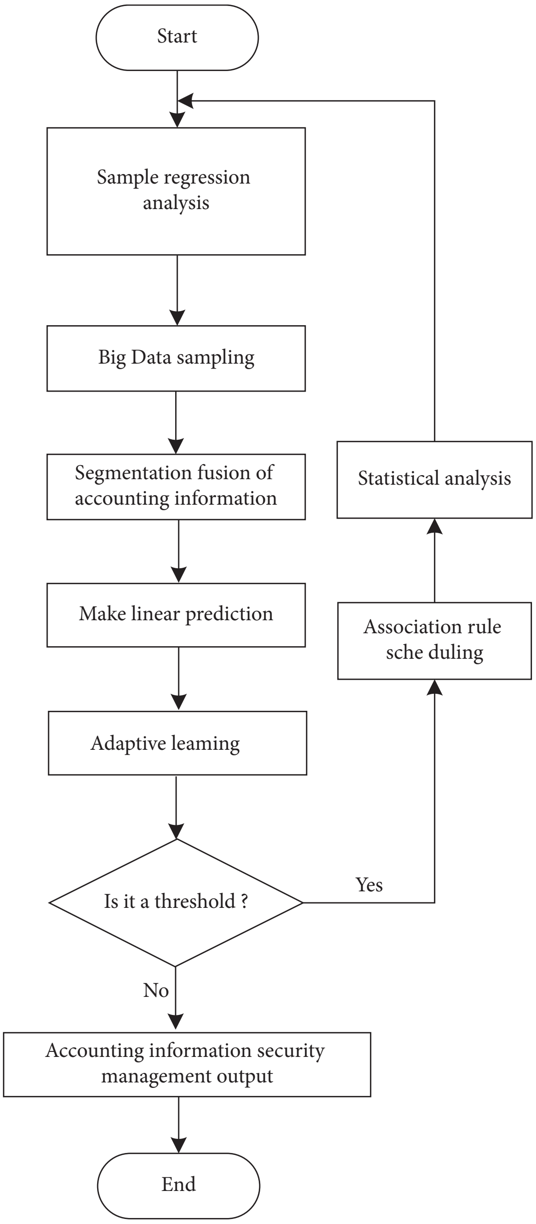

Start
Sample regression analysis
Big Data sampling
Segmentation fusion of accounting information
Make linear prediction
Adaptive leaming
Is it a threshold ?
Yes
No
Statistical analysis
Association rule sche duling
Accounting information security management output
End

Figure 1: Blockchain accounting information security management implementation process.

(1) Account information management: account information management mainly includes several submodules, such as advance payment management, staff advance management, cargo information management, carrier vehicle advance management, advance receipt management, and transaction order management

(2) Accounting management: accounting management mainly includes several submodules, including reimbursement order management, consignee checkout management, supplier checkout management, and freight payment management

(3) Data management: data management mainly includes goods classification management, sales contract management, procurement contract management, staff management, cargo information management, and carrier information management submodule

(4) Statistical report: the statistical report function module mainly provides the query and printing function for the aforementioned reports

(5) Decision support: it generates various kinds of data information, which is an important data source for the company's accounting information and can be used as a basis for decision support

(6) System management: system management mainly includes user management, functional module authority management, company announcement management, and log management [18]

The the use case model for the blockchain accounting information security management function is shown in Figure 2.

3.4. Definition of Accounting Information Security. The purpose of accounting information security management is to ensure the integrity, availability, and ease of use of accounting information; that is, accounting data can only be disclosed to the right to know, accounting data can only be modified within the scope of authorization and accounting information system can only be used when necessary.

Investors and lenders use the accounting information of the enterprise to make investments and operation decisions, evaluate the enterprise value according to the accounting information, and predict the future cash flow of the enterprise. At the same time, the relevant government departments carry out macro-control on the market according to the relevant indicators provided by accounting information, so as to improve and strengthen the enterprise management. Therefore, it is of great significance to study blockchain-based accounting information security management.

4. Simulation and Result Analysis

In order to verify the performance of the application of this paper's method in implementing blockchain accounting information security management, SPSS statistical analysis and Matlab simulation software are used for simulation experiments, and the simulation platform is built with Intel (R) Core (TM) i7-47 70 CPU, 16 GB of memory, and Windows 1064 bit operating system. Based on the aforementioned parameters, the regression analysis value of accounting safety management evaluation is obtained, which is reflected by objective function. Figure 1 shows the convergence curve of the optimal objective function of the model under different iterations and different calculation times.

Figure 3 shows that the function curve of the method in this paper is closer to the standard convergence curve than the curve of the statistical analysis method, which improves the process convergence of the evaluation, thus it can be seen that the method in this paper can effectively realize the

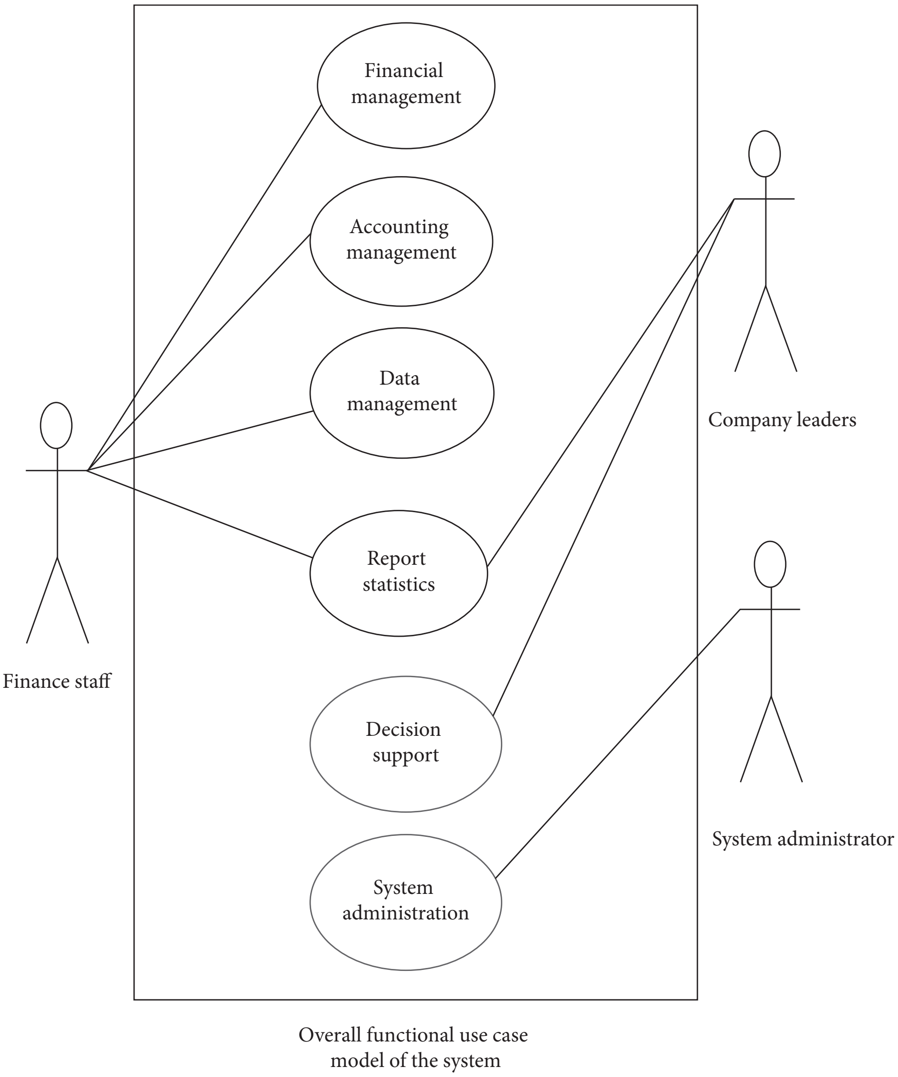


Figure 2: Blockchain accounting information security management functional use case model.

research credit evaluation, and the optimization of its objective function is better.

The boundary constraint coefficient of present accounting information management is set as 0.36, the autocorrelation statistical feature component is 0.67, the sample number of Big Data information sampling for the present accounting information management is 1200, the test set is 120, the data set of semantic ontology information distribution is 60, and the descriptive statistical distribution set of accounting information security association is shown in Table 1.

Based on the results of the aforementioned descriptive statistical analysis of the parameters of accounting information security management, in order to visualize the results of the statistical analysis of accounting information, the data in Table 1 are transformed as shown in Figure 4.

Based on Figure 4, taking test object set 10 as the research object, the accounting information security management method based on Massicotte and Henri's study [6] enterprise computerization quantitative feature analysis, the accounting information security feature matching method based on Pérez-González et al.'s study [7], and the method in this paper mentioned in the introduction are respectively used to fuse and schedule the blockchain accounting information, and the boundary constraint coefficient of the current accounting information management is set as 0.36. The autocorrelation statistical characteristic component is 0.67, and the comparison results of the same test set under different similarity coefficients are obtained, as shown in Figure 5.

Figure 5 shows that compared with the method by Massicotte and Henri [6] and Pérez-González et al. [7], the method in this paper can effectively realize the security management of accounting information, the information fusion and scheduling ability is above 92%, and the statistical analysis results are accurate and reliable. The main reason is that the random nearest neighbor characteristic parameter analysis model of accounting information security management is obtained using the block-link regression analysis

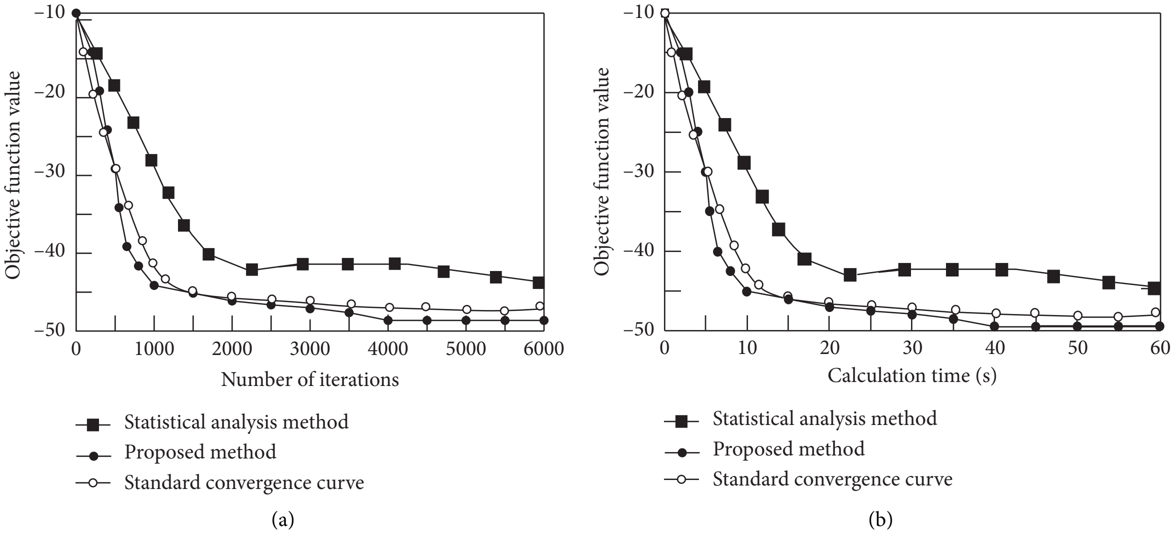


Figure 3: Convergence curve of the optimized objective function: (a) convergence curves of the objective function optimized under different number of iterations and (b) convergence curves of the objective function for optimization at different times.

Table 1: Results of descriptive statistical analysis.

Test object set	Similarity coefficient	Test regression analysis level (%)	Variance
1	0.388	43.53	0.388
2	0.546	42.46	0.655
3	0.677	34.32	0.554
4	0.366	53.56	0.454
5	0.434	54.58	0.654
6	0.143	53.64	0.678
7	0.342	35.53	0.435
8	0.324	64.56	0.457
9	0.445	43.54	0.544
10	0.532	56.24	0.567
11	0.443	43.63	0.544
12	0.432	32.56	0.565
13	0.435	65.65	0.655

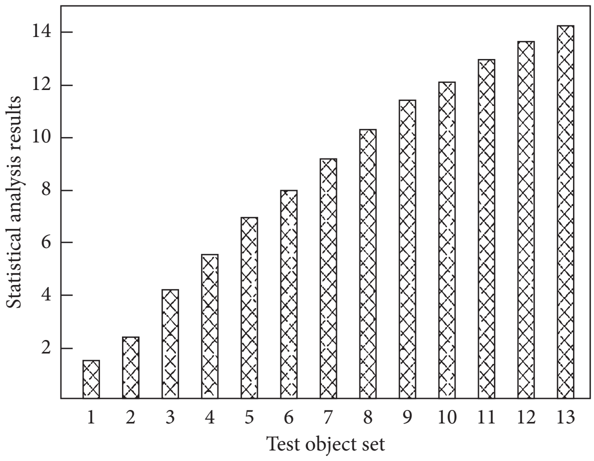


Figure 4: Statistical analysis results.

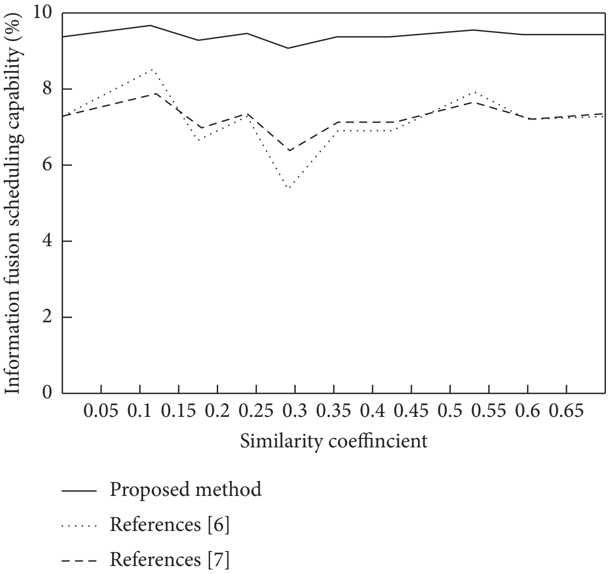


Figure 5: Comparison results of blockchain accounting information fusion and scheduling.

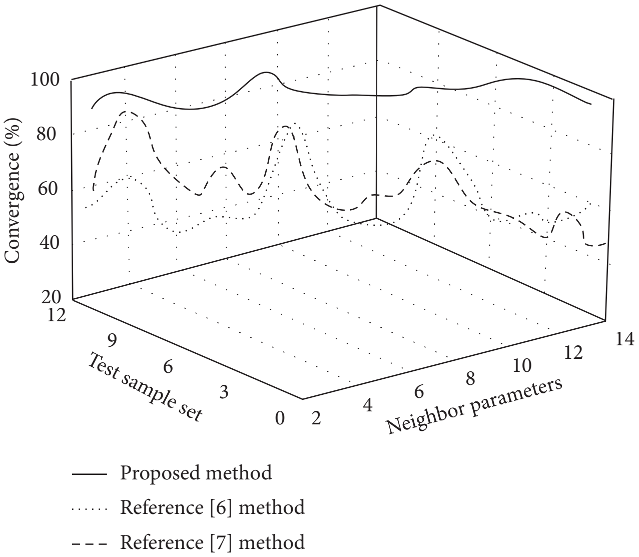


Figure 6: The convergence surface of accounting information security management.

method. The model is used for data management and analysis to improve the reliability of the final results.

In order to further verify the effectiveness of the proposed method, the unified similarity coefficient is 0.445, and other conditions remain unchanged. The convergence of accounting information security management is obtained by testing the fusion level of different methods, as shown in Figure 6.

Figure 6 shows that the convergence of the method in this paper is better compared to the method by Massicotte and Henri [6] and the method by Pérez-González et al. [7] for accounting information security management, and the convergence is above 91.8%. The convergence is mainly manifested in the flexible scheduling ability of the model. This method uses the joint characteristic distribution set of the relevant parameters of the blockchain fusion and security association of accounting information to improve the flexibility of accounting information security management and further has a better optimization effect on the convergence.

The feature recognition rate of accounting information security management is tested and the comparison results are obtained as shown in Figure 7. Analyzing Figure 7, we know that the method of this paper performs accounting information security management with a higher level of feature recognition, which is due to the fact that the method of this paper extracts features and performs feature fusion processing on the basis of statistical analysis of accounting information security management information, which improves the level of feature recognition above 90.1%. The

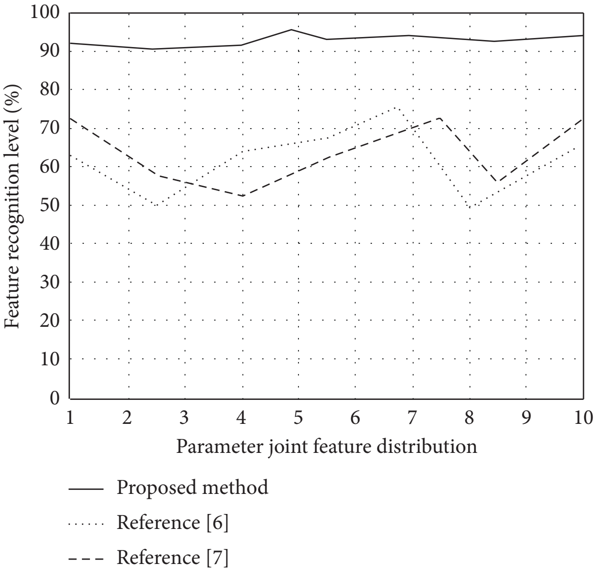


Figure 7: Comparison of feature recognition rate of accounting information security management.

Table 2: Comparison of accuracy of quantitative assessment of security posture of dual-channel wireless blockchain networks.

Dynamic fusion parameters	Proposed method	Massicotte and Henri's method [6]	Pérez-González et al.'s method [7]
0.2	0.956	0.824	0.835
0.4	0.978	0.846	0.867
0.6	0.997	0.859	0.893
0.8	0.999	0.912	0.912

main reason is that this paper uses blockchain technology to establish blockchain fusion function of accounting information and improves the ability of accounting information feature fusion.

Test the accuracy of accounting information security management and get the comparison results in Table 2. From Table 2, we know that the accuracy of accounting information security management by the method of this paper is higher than 95.6%. The reason for the high accuracy is to obtain the optimal decision model of the blockchain accounting information security management and realize the blockchain accounting information security management.

5. Conclusion

This paper proposes a model of blockchain-based accounting information security management, which improves the ability of accounting information security management.

(1) In this paper, the fuzzy extended sample regression analysis method is used to establish the feature integration model of blockchain accounting information, calculate the internal control and prudent control analysis model of accounting information, and improve its security management ability.

(2) The experimental results show that the information fusion and scheduling ability of this method is more than 92%, the convergence is more than 91.8%, the feature recognition level is more than 90.1%, and the management accuracy is more than 95.6%. The comprehensive experimental results show that this method has certain effectiveness.

(3) This study also has some shortcomings, mainly in the absence of specific accounting information research and investigation, which will be taken as the next research direction to further enhance the practicability of this paper.

Data Availability

The data used to support the findings of this study are included within the article.

Disclosure

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the Planning Project of Philosophy and Social Sciences in Heilongjiang Province (grant no. 19GLB086) and the Natural Science Fund of Heilongjiang Province (grant no. LH2020F006).

References

[1] Y. Joyce, "Building trust in crisis management: a study of insolvency practitioners and the role of accounting information and processes," *Contemporary Accounting Research*, vol. 37, no. 3, pp. 1622–1657, 2020.

[2] Z. Li, C. Ma, and D. Wang, "Achieving multi-hop PRE via branching program," *IEEE Transactions on Cloud Computing*, vol. 8, no. 1, pp. 45–58, 2020.

[3] Z. Li, C. Ma, and H.-S. Zhou, "Multi-key FHE for multi-bit messages," *Science China Information Sciences*, vol. 61, no. 2, pp. 029101:1–029101:3, 2018.

[4] J. Zhao and C. Cheng, "Dynamic cooperative random drift particle swarm optimization algorithm assisted by evolution information," *Journal of Computer Applications*, vol. 40, no. 11, pp. 3119–3126, 2020.

[5] M. J. Ershadi and M. Forouzandeh, "Information security risk management of research information systems: a hybrid approach of fuzzy FMEA, AHP, TOPSIS and shannon entropy," *Journal of Digital Information Management*, vol. 17, no. 6, pp. 321–336, 2019.

[6] S. Massicotte and J. F. Henri, "The use of management accounting information by boards of directors to oversee strategy implementation," *The British Accounting Review*, vol. 53, no. 3, Article ID 100953, 2020.

[7] D. Pérez-González, S. Preciado, and P. Solana-González, "Organizational practices as antecedents of the information security management performance," *Information Technology and People*, vol. 32, no. 5, pp. 1262–1275, 2019.

[8] G. Chen, Q. Yang, X. Chen et al., "Methodology of urban safety and security assessment based on the overall risk management perspective," *Sustainability*, vol. 13, 2021.

[9] S. K. T. Mehedi, A. A. M. Shamim, and M. B. A. Miah, "Blockchain-based security management of IoT infrastructure with Ethereum transactions," *Iran Journal of Computer Science*, vol. 2, no. 3, pp. 189–195, 2019.

[10] C. Xu, Y. Fang, and Y. Ma, "Integrated application of blockchain in the electric information management system," *Procedia Computer Science*, vol. 162, pp. 88–93, 2019.

[11] D. Datta, L. Garg, K. Srinivasan et al., "An efficient sound and data steganography based secure authentication system," *Computers, Materials and Continua*, Cmc -Tech Science Press-, vol. 67, no. 1, pp. 723–751, 2021.

[12] D. Patel, K. Srinivasan, C. Y. Chang et al., "Network anomaly detection inside consumer networks - a hybrid approach," *Electronics*, vol. 9, no. 6, pp. 1–12, 2020.

[13] Z. Wang, N. Wang, X. Su, and S. Ge, "An empirical study on business analytics affordances enhancing the management of cloud computing data security," *International Journal of Information Management*, vol. 50, no. Feb, pp. 387–394, 2020.

[14] M. Mirtsch, J. Kinne, and K. Blind, "Exploring the adoption of the international information security management system standard ISO/iec 27001: a web mining-based analysis," *IEEE Transactions on Engineering Management*, vol. 68, no. 99, pp. 1–14, 2020.

[15] A. S. Al-Delawi and W. M. Ramo, "The impact of accounting information system on performance management," *Polish Journal of Management Studies*, vol. 21, no. 2, pp. 36–48, 2020.

[16] A. Naseer, H. Naseer, A. Ahmad, S. B. Maynard, and A. Masood Siddiqui, "Real-time analytics, incident response process agility and enterprise cybersecurity performance: a contingent resource-based analysis," *International Journal of Information Management*, vol. 59, no. 8, Article ID 102334, 2021.

[17] H. O. Kadhim and A. Z. Latif, "The impact of supply chain accounting information systems harmonization on creating a competitive advantage for the Iraqi general commission taxation[J]," *Journal of Supply Chain Management*, vol. 8, no. 2050-7399, pp. 448–452, 2019.

[18] M. S. Alathamneh, "The impact of accounting information systems reliability on enhancing the requirements of planning process at Jordanian commercial banks," *Management Science*, vol. 10, no. 5, pp. 1043–1050, 2019.

Hindawi
Mobile Information Systems
Volume 2021, Article ID 7611619, 16 pages
https://doi.org/10.1155/2021/7611619

Research Article

Hierarchical Hybrid Trust Management Scheme in SDN-Enabled VANETs

Ming Mao, Peng Yi, Tao Hu, Zhen Zhang, Xiangyu Lu, and Jingwei Lei

People's Liberation Army Strategic Support Force Information Engineering University, Zhengzhou 450001, China

Correspondence should be addressed to Ming Mao; maoming12345@163.com

Received 20 May 2021; Revised 13 July 2021; Accepted 6 August 2021; Published 17 August 2021

Academic Editor: Vishal Sharma

One of the principal missions of security in the Internet of Vehicles (IoV) is to establish credible social relationships. The trust management system has been proved to be an effective security solution in a connected vehicle environment. The use of trust management can play a significant role in achieving reliable data collection and dissemination and enhanced user security in the Internet of Vehicles. However, due to a large number of vehicles, the limited computing power of individuals, and the highly dynamic nature of the network, a universal and flexible architecture is required to realize the trust of vehicles in a dynamic environment. The existing solutions for trust management cannot be directly applied to the Internet of Vehicles. To ensure the safe transmission of data between vehicles and overcome the problems of high communication delay and low recognition rate of malicious nodes in the current trust management scheme, an efficient flow forwarding mechanism of the RSU close to the controller in the Software-Defined Vehicular Network is used to establish a hierarchical hybrid trust management architecture. This method evaluates the dynamic trust change of vehicle behavior based on the trust between vehicles and the auxiliary trust management of the infrastructure to the vehicle, combined with static and dynamic information and other indicators. The proposed trust management scheme is superior to the comparative schemes in resisting simple attacks, selective misbehavior attacks, and time-dependent attacks under the condition of ensuring superior real-time performance. Its overall accuracy is higher than that of the baseline scheme.

1. Introduction

Software-Defined Network (SDN) adopts the idea of separation of control plane and data plane, and through the use of perfect interfaces (such as the southbound interface of OpenFlow protocol), it has played a great role in the increasingly complex structure of data center and wired network. At the same time, in the wireless and mobile network-related fields, research on Software-Defined Wireless Network (SDWN) [1] has also made progress. Researchers adjust and expand the SDN and SDWN architecture and related concepts to build Software-Defined Vehicular Network (SDVN) to meet the exclusive characteristics of VANETs (Vehicular Ad Hoc Networks) and improve the performance of vehicle communication networks. Jiacheng et al. [2] pointed out that SDN is a powerful innovative solution, which improves the dynamic characteristics of VANET and ITS (Intelligent Transport System) applications by encouraging the flexibility of network management and the large-scale unified optimization of abstraction. In the future, innovative development of 5G VANET must rely on cloud computing, SDN, and fog computing to meet the new requirements of the continuous development and change of ITS.

As shown in Figure 1, in SDVN architecture, the control layer uses the northbound interface (NBI) to connect with the application layer, and the application layer implements services such as traffic management, location prediction, and security. The SDN controller tracks the status of the data plane elements and programs the southbound interface (SBI) through its predefined application to inject forwarding rules into the data plane. The most commonly used SBI is OpenFlow. The data plane consists of an upper data plane and a lower data plane. The upper data plane includes

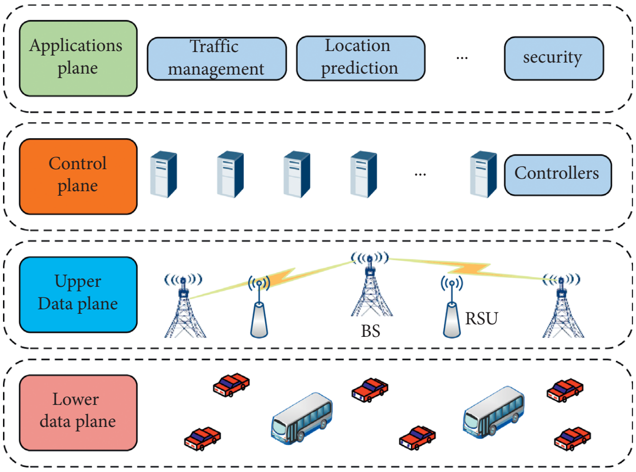


Figure 1: Software-Defined Vehicular Network.

OpenFlow switches, routers, and wireless access infrastructure, namely, roadside unit (RSU), base station (BS), etc. The lower data plane is composed of onboard units (OBUs), that is, the vehicle is equipped with OBU as the terminal user. In this structural system, the specific decisions of the control plane can be conveyed to a single OBU, which promotes fine-grained control, greater scalability, and programmability.

Vehicle communication security issues have consistently been the focus of the SDVN, including availability, authenticity, confidentiality, integrity, and non-repudiation. For example, if a vehicle sends a message that there is congestion somewhere on the road, should other vehicles consider this information to be correct and take corresponding measures? To meet the above requirements, with the help of cryptographic methods, many mechanisms have been proposed to prevent VANET from security attacks. The management scheme based on cryptography has been applied to VANET's message authentication [3, 4]. Although the cryptography-based management scheme has numerous advantages, due to the limited computing power of the OBU, cryptography-based methods are prone to introduce excessive delays to complete all necessary checks. In addition, the verification of messages from unknown vehicles involves the exchange of public certificates, which results in higher message overhead. These methods mainly rely on traditional cryptography-based solutions and have not yet fully resolved the dynamic and distributed behavior of vehicle networks. In addition, encryption technology cannot deal with internal attackers. It is obvious that in a VANET environment, it may be extremely challenging to reduce network management overhead, protect privacy, and implement low-latency communication and intelligent resource management.

Compared with cryptographic methods, the solution architecture based on the trust model (TM) is semi-centralized or distributed. Therefore, it can work independently of the data exchange center in the case of high-mobility network. Trust metric is described as the confidence coefficient that a when node performs certain operations to another node [5]. This operating information is based on information about events (for instance, accidents) between two vehicles and is exchanged through two communication modes, namely, vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication. In critical applications such as hazard warnings, receiving nodes need to ensure their authenticity and trustworthiness before responding to received messages. Once the information is received, trust is calculated based on various factors, including previous interactions, neighbors' suggestions, and statistics related to the history of the event. However, since VANET involves highly maneuverable and diverse vehicles and very frequent topology changes, trust between adjacent vehicles is created in a very short time interval [6]. Therefore, it is also very challenging and difficult to calculate and evaluate trust based on various factors within a limited time.

The current trust management architecture mainly includes infrastructure-based shared management and vehicle self-organization management (as shown in Figure 2). Infrastructure-sharing trust management systems [7, 8] usually deploy vehicle trust management structures above the infrastructure. It realizes the sharing and management of trust information through infrastructure, and it usually needs to set up certificate authorities (CAs) to realize vehicle certification by satisfying a series of trust requirements, including certification, integrity, non-repudiation, etc. The disadvantage of this architecture system is that CAs must be completely credible, and in the event of malicious attacks, they may combine with malicious vehicles to deceive honest vehicles. In addition, the architecture must ensure that all vehicles are within the coverage of the RSU to guarantee the real-time transmission of trust information. In the vast rural areas and suburbs, it is difficult to ensure that vehicles can always meet the RSU coverage service.

Another trust management architecture is a self-organizing vehicle distributed trust management scheme [9–11]. This scheme can realize the trust management of the vehicle through the trust information interaction between the vehicle itself and the vehicle without considering the central authorization and certification CAs. The advantage of self-organizing trust management architecture is that it can acquire trust value in a short time because the trust knowledge it acquires comes from its own and neighbor vehicles' recommendations. Therefore, it can adapt to the highly dynamic changes of the VANET architecture. Its disadvantages are as follows. (1) Due to the dynamic inherent high variability of the VANET structure, similar to a social network, it cannot completely rely on its own existing trust to obtain accurate trust management for new requests from existing vehicles. (2) Since the vehicle adopts a self-organizing trust management method, it is unable to obtain comprehensive trust information, so the trust result obtained may be one-sided and sometimes even wrong. VANET is a decentralized open system. If it does not rely on the infrastructure, peers can join or exit the network at any time. If the neighbor is interacting with the vehicle now, there is no guarantee of interacting with the same vehicle in the future.

The main contributions of this paper are as follows.

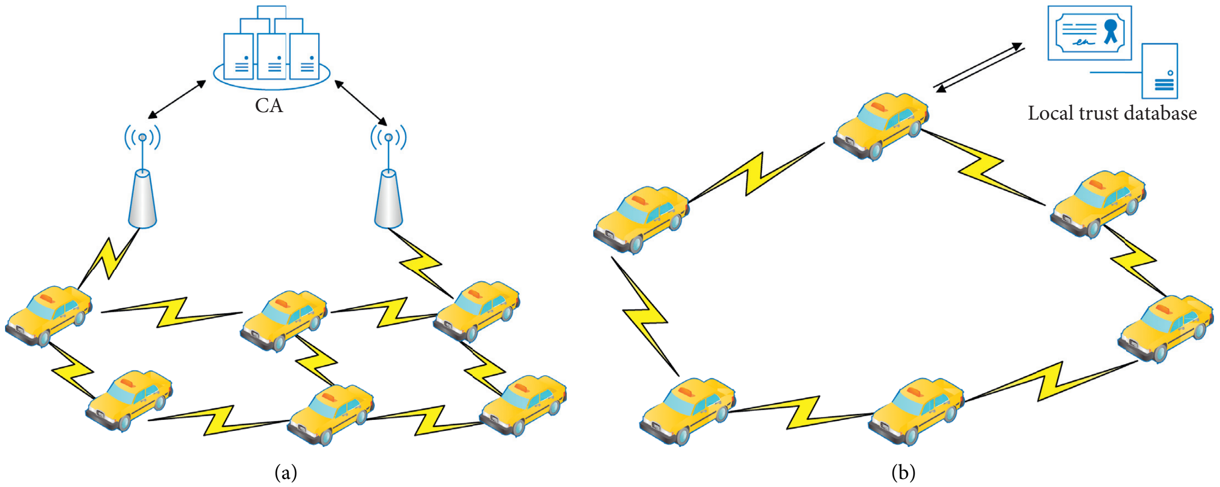


Figure 2: (a) Infrastructure-based trust management architecture. (b) Vehicle-based distributed trust management architecture.

First, a hierarchical hybrid trust management system (HHTM) is proposed, which can conduct a wide range of trust management assessments according to the different environments in which the vehicle is located. If the vehicle is within the coverage area of the RSU, it performs a hybrid trust management evaluation. If the vehicle is not within the coverage area of the RSU, it can still perform distributed trust management evaluation to realize the trust management of the vehicle.

Secondly, according to the characteristics of high mobility of vehicles, the subjective trust between vehicles is calculated according to the local trust database of vehicles, and the recommended trust between vehicles is calculated according to the interactive information between the vehicles and neighbors, to complete the calculation of the trust metric between vehicles. The concept of similarity is utilized to calculate the similarity between the vehicle information in the infrastructure trust table and the message sending vehicle, and the calculation of the infrastructure trust value is realized by combining the distance coefficient of the infrastructure. Meanwhile, we design algorithms to realize the calculation of vehicle hybrid trust value.

Finally, a dynamic simulation environment is established for extensive simulation experiment analysis, which verifies that the robustness of the proposed scheme against node attacks is significantly better than that of the existing schemes.

2. Related Work

The existing literature proposes various solutions to realize trust management and evaluate the trustworthiness and authenticity of the transmitted messages in VANET. A vehicle's trust in information can be calculated based on various factors, including the neighbours' opinions, the reputation of the vehicle, and their past interactions with communication vehicles [12]. Based on the above goals, trust management models are roughly divided into three categories, namely, data-oriented, entity-oriented, and combined trust models [13, 14].

2.1. Data-Oriented Trust Models (DTMs). In this model, "data" are regarded as an important part of the TM, where the trust in the message (data) is calculated based on the opinions generated by neighboring vehicles or the historical interactions between peers.

Raya et al. [15] proposed a DTM that uses Bayesian Inference (BI) and Dempster–Shafer Theory (DST) to evaluate evidence about events received from the neighborhood. The TM consists of three main stages. Firstly, the evaluator node (EvN) accumulates reports generated by neighboring vehicles. Secondly, EvN assigns weights to the received reports according to the spatiotemporal characteristics of the event. Finally, EvN forwards these reports to the decision logic module and uses BI and DST for trust calculation. The limitation of this technology is that it calculates trust based on the received data of EvN, which is inefficient for high-mobility networks.

Gazdar et al. [16] adopted a layer-based analysis method. The vehicle continuously evaluates the credibility of the received data based on its direct experience. In this TM, each participating vehicle is evaluated for trust, and its main purpose is to identify highly trusted vehicles and dishonest vehicles based on the exchanged data. Each vehicle maintains a trust table for its neighbors. The trust value of messages received from trusted vehicles will increase, while for malicious vehicles, it will decrease. Since this technology only involves the direct experience of participating vehicles, it is very effective in identifying malicious vehicles.

Wu et al. [17] proposed a centralized trust modeling framework for data evaluation by taking advantage of the RSU. On RSU, trust is calculated based on two factors: (1) observation and (2) feedback. The vehicle generates observation results for detected events and their credibility. The credibility depends on the distance to the event, the maximum message detection rate, and the number of sensors that detect the event. Then, the observation results are shared with the RSU, which updates the list of recently observed events. RSU evaluates the credibility of the received observations by using the ant colony optimization algorithm to perform trust calculations on the received observations.

Updated trust information is distributed by RSU together with nearby vehicles. Because this method relies too much on infrastructure, it cannot be applied in suburban and remote rural areas.

Gurung et al. [18] proposed an information-oriented trust model that enables each individual vehicle to assess the credibility of a potentially large number of messages received in VANET without relying on any infrastructure support, such as RSUs or central servers. The proposed trust model "RMCV" takes into account several factors that affect message credibility, including message content similarity, content conflict, and message routing path similarity. The RMCV scheme consists of two main parts: (1) message classification and (2) information-oriented trust pattern.

2.2. Entity-Oriented Trust Models (ETMs). In ETM, the credibility of the entity (vehicle) is evaluated. This method relies on providing recommendations from the sender to EvN's neighbors to identify dishonest nodes in the legitimate vehicle pool.

Khan et al. [19] made extensive use of the cluster-based technology and first chose the cluster head (CH), and it is responsible for evaluating the trust in the network. In this TM, CH implements a watchdog mechanism in which nearby vehicles will provide reports on vehicles that behave abnormally. If such vehicles are detected, CH will notify the trusted authority (TA) responsible for revoking these vehicles to maintain the trusted network. The disadvantage of this scheme is that the communication overhead caused by the message exchange by the CH reduces the efficiency of the entire network.

Yang [20] proposed a TM by using the similarity mining method to calculate the trust degree. After receiving the message from the vehicle, EvN calculates the similarity between the received messages based on the Euclidean distance and the trust of the sending vehicle. Since trust is obtained by EvN using Euclidean local distance, this TM cannot provide any global information about message similarity.

In order to quickly and accurately distinguish malicious or selfish nodes that spread false or fake messages throughout the network, Mármol and Pérez [21] proposed an infrastructure-based trust and reputation model, namely, TRIP. The model calculates reputation scores based on recommendations given by other vehicles and RSU. The decision in this model is based on fuzzy logic and probability.

2.3. Combined Trust Models (CTMs). CTM aggregates the attributes of entity-oriented and data-oriented trust management schemes, where node trust is calculated based on the trust evaluation of the received message.

Ahmed and Tepe [22] proposed a CTM whose logic-based trust calculation is used to identify nodes that inject false information into the network. In this TM, when neighboring vehicles share messages, EvN can identify the credibility of the event. Once the true event is determined, this information is used to classify the behavior of the sender node as legitimate or malicious. EvN calculates trust through weighted voting and a logical trust function. The TM can effectively identify dishonest vehicles that spread false information. However, the main limitation of this TM is its reliance on weighted voting, which may be biased when dishonest vehicles are in the majority.

To enhance user privacy in the network, Chen and Wei [23] proposed a beacon-based CTM that combines the characteristics of ETM and DTM. The trust level is calculated in two steps. First, it establishes entity trust based on the received beacon. Then, the data trust will be calculated based on various reasonableness checks to identify and revoke dishonest vehicles and their malicious content. The TM is highly dependent on the Public Key Infrastructure (PKI) and the central authority for trust evaluation, and adding it to each forwarded message will cause greater overhead.

Shrestha and Nam [24] proposed a CTM to calculate the trust in vehicles in a completely distributed manner. First, it evaluates the vehicle's trustworthiness and then calculates the trustworthiness of the information. The model uses a clustering algorithm to achieve trust. In this algorithm, honest and dishonest vehicles are divided into two separate groups to identify the credibility of neighboring nodes. Next, the modified threshold random walk algorithm is used to evaluate EvN's trust in the received message. The main disadvantage of this scheme is that it assumes that the distribution of dishonest nodes in the network is even. In VANET, malicious tools are randomly distributed throughout the network. This assumption may be incorrect.

The core element of the IoV is the vehicle, and trust management is based on data interaction. The credible data transmission between vehicles is carried out by the vehicle as a relay. Therefore, trust management around data and the vehicle is inseparable. We propose a hierarchical hybrid trust management mechanism (HHTM), which includes management of vehicle trust information shared between infrastructures and the management of trust information between vehicles. Because this method not only uses the infrastructure to share the management trust value but also takes into account the calculation of the self-organizing management trust value between vehicles, the flexibility of this structure allows it to overcome the low accuracy and the real-time problem of trust information that the vehicle may encounter during trust management.

3. Vehicle Trust Management Model

Alioua et al. [25] pointed out that to ensure that the installation time of flow rules can meet the low-latency requirements of applications for most vehicles' safety in dense networks like HetVNet (Heterogeneous Vehicular Network), the SDN controller must be installed at the edge of the network, the closest network location to the vehicle. Since the RSU acts as an OpenFlow switch in some locations, this trust solution uses the centralized and efficient flow forwarding mechanism of the RSU and the control plane to set the upper trust management plane at the RSU layer. The upper trust management plane includes trust query, trust calculation, trust update, and blacklist upgrade functions (as shown in Figure 3). The lower trust management plane depends on the trust management of the vehicle itself, which makes full use of the characteristics of the vehicle's high mobility to ensure that trust information is updated in real time.

The main abbreviations used in this paper are summarized in Table 1. The complete hierarchical hybrid trust management model and its calculation process are shown in Figure 4. After vehicle *i* receives the message from vehicle *j*, it calculates the trust between vehicles and inquires the trust based on the infrastructure of the vehicle respectively. The trust calculation between vehicles is divided into two parts: ST and RT, and the trust opinion of infrastructure is IT. After calculating the above value, we can get the hybrid trust value HT. Finally, the system judges whether the value meets the evaluation criteria that have been set, so as to determine whether the node is credible.

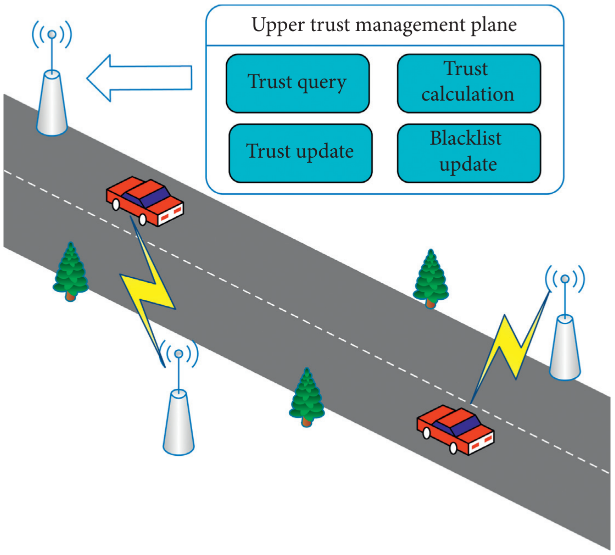


Figure 3: Upper trust management plane.

Table 1: Abbreviations.

Abbreviation	Description
RSU	Roadside unit
BS	Base station
OBU	Onboard unit
TM	Trust model
CA	Certificate authority
DTM	Data-oriented trust model
ETM	Entity-oriented trust model
CTM	Combined trust model
EvN	Evaluator node (information receiving vehicle)
ST	Inter-vehicular subjective trust
STW	Subjective trust weight
OTV	Original trust of vehicle
RT	Inter-vehicular recommendation trust
RW	Role-based trust weight
NT	Neighbor trust
SP	Trust opinion of neighbors
DC	Distance coefficient of RSU
sim_i^j	Similarity between node *i* and node *j*
IT	Infrastructure trust
HT	Hybrid trust

3.1. Inter-Vehicular Trust Calculation.

3.1. Inter-Vehicular Trust Calculation. The trust between vehicles includes subjective trust and recommendation trust. The subjective trust is determined by the vehicle's existing social knowledge to calculate the vehicle's subjective trust value of the vehicle that receives the interactive information, and the recommendation trust requires the information receiving vehicle (EvN) to calculate the vehicle recommendation trust based on the interactive information from the message sending vehicle.

3.1.1. Inter-Vehicular Subjective Trust (ST). The subjective trust model is concentrated on the social relationship between vehicles. The EvN calculates the trust value of the vehicle network by applying existing trust rules created based on social relationships. To quantify this social relationship, the following two main social indicators are used.

(1) Inter-Vehicular Subjective Trust Weight (STW). Existing vehicle information mainly communicates to cloud services and RSUs. Since the transmission distance of the vehicle is identified, we believe that the EvN has low credibility for receiving messages sent from vehicles over a long distance crossing multiple RSU coverage. The distance between vehicles can be used to intuitively determine the weight of the subjective credibility of vehicles. We use DIS_{ij} to denote the Euclidean distance between node *i* and node *j* and utilize COV_{RSU} to denote the coverage radius of RSU. We define the subjective trust weight (STW) of a vehicle based on the distance between vehicles as follows:

$$\text{STW} = \begin{cases} 1, & \text{if } 0 < \text{DIS}_{ij} \leq 2\text{COV}_{\text{RSU}}, \\ 0.75, & \text{if } 2\text{COV}_{\text{RSU}} < \text{DIS}_{ij} \leq 3\text{COV}_{\text{RSU}}, \\ 0.5, & \text{if } 3\text{COV}_{\text{RSU}} < \text{DIS}_{ij} \leq 4\text{COV}_{\text{RSU}}, \\ 0.25, & \text{if } \text{DIS}_{ij} > 4\text{COV}_{\text{RSU}}. \end{cases} \tag{1}$$

(2) Original Trust of Vehicle (OTV). In the decentralized trust system, each vehicle stores a local trust database (LTD), which records the trust information generated by the original vehicle interaction data. It also contains legitimate and illegitimate interaction information generated by vehicle interaction. First, we define LEG_{ij} to represent the number of legitimate interaction messages from vehicle *j* that vehicle *i* has received and MAL_{ij} to represent the number of illegitimate interaction messages from vehicle *j* that vehicle *i* has received. Then, the original trust of the vehicle OTV_{ij} can be expressed as follows:

$$\text{OTV}_{ij} = \frac{\text{LEG}_{ij}}{\text{LEG}_{ij} + \text{MAL}_{ij}} * \left(1 - \frac{1}{\text{LEG}_{ij} + 1}\right). \tag{2}$$

The subjective trust of a vehicle is a trust association established on a social basis. After the vehicle receives the sender's information, it first queries whether the trust information of the sending vehicle exists in the LTD. If it

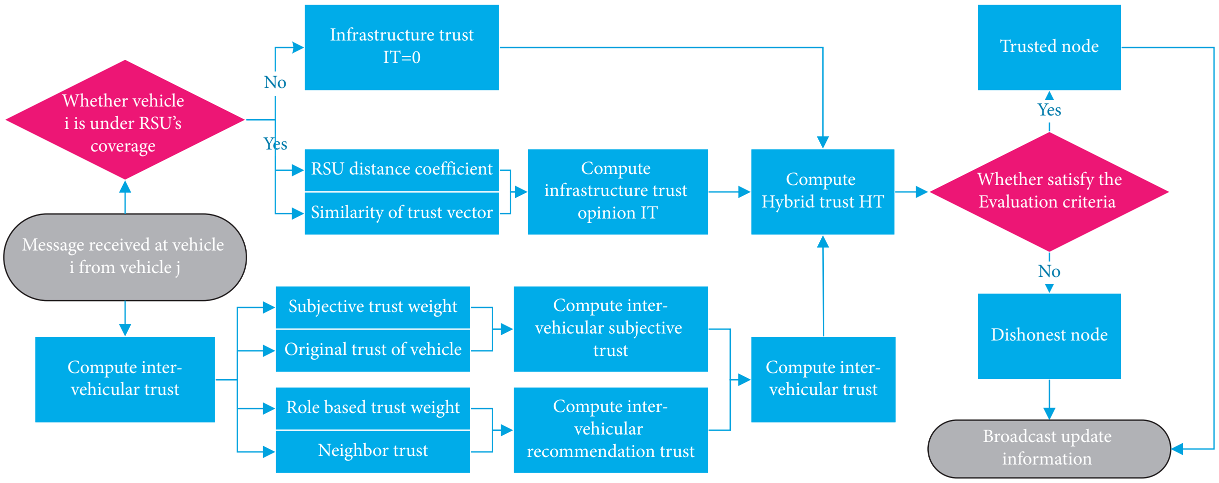


Figure 4: Hierarchical hybrid trust management model.

exists, it directly calculates the OTV. If it does not exist, then assign an initial value OTV_{ini} to it and update this value as the original trust in the vehicle trust table.

(3) Subjective Trust Calculation. The subjective trust of the vehicle can be obtained by multiplying the STW of the vehicle with the OTV:

$$\text{ST} = \text{STW} * \text{OTV}_{ij}. \tag{3}$$

3.1.2. Inter-Vehicular Recommendation Trust (RT). Due to the high mobility of VANET, the two vehicles cannot always maintain direct communication. Therefore, the trust between the vehicles must be obtained indirectly by relying on the cognition of the data and information of other neighboring vehicles. If the vehicle has never interacted with the information sending vehicle before, then the trust suggestions received by the vehicle from other neighboring vehicles will become the only evaluating variable for evaluating the trust value of the information sending vehicle.

(1) Role-Based Trust Weight (RW). According to the different social attributes of the vehicle, the trust basis of the vehicle is also different. Based on the social role to which the vehicle belongs, we divide the role of vehicle (RV) as follows:

RW_1: authoritative vehicles, such as law enforcement department, prisons, police, and so on.

RW_2: vehicles of specific companies, such as TV stations, newspapers, banks, and so on.

RW_3: local vehicles familiar with traffic conditions, such as freight drivers on fixed routes, commuters, taxi drivers, and so on.

RW_4: ordinary roles (all roles except the above three roles).

Assign the corresponding trust weight to each role type:

$$\text{RW} = \begin{cases} 1, & \text{if } \text{RV}_i \in \text{RW}_1, \\ 0.9, & \text{if } \text{RV}_i \in \text{RW}_2, \\ 0.8, & \text{if } \text{RV}_i \in \text{RW}_3, \\ 0.7, & \text{if } \text{RV}_i \in \text{RW}_4. \end{cases} \tag{4}$$

(2) Neighbor Trust Calculation (NT). The trustworthiness of the neighbor depends on the trust opinions of the neighboring vehicles of vehicle i on vehicle j. Define the trustworthiness of the neighbor of vehicle i to vehicle j as follows:

$$\text{NT}_{ij} = \left[\prod_{N}^{\forall k \in \text{Neigh}(i)} \left(\text{OTV}_{ik} * \text{SP}_{kj} \right)^{1/2} \right]^{1/N}. \tag{5}$$

The trustworthiness calculation of neighbor j includes the trust score of vehicle j by vehicle k in the one-hop neighbor node set Neigh (i). Among them, OTV_{ik} is the original trust of vehicle k in vehicle i, and SP_{kj} is the indirect score of vehicle k on vehicle j. The following describes how to obtain the recommender j's score.

(3) Trust Opinion of Neighbors (SP). For the EvN, the more the neighbor nodes receive the message of vehicle j, the higher the credibility of the message. At the same time, the packet delivery rate reflects the reliability of information transmission, so a high delivery rate can also enhance the trust of the EvN to the information sending vehicle. Therefore, we should increase the trust score for vehicles that meet the conditions and reduce the trust score for vehicles that do not meet the conditions. SP_{ij} consists of two parts: the proportion of the number of vehicles in the neighboring vehicles that received messages of vehicle j and the packet delivery rate of the message sending vehicle to the neighboring vehicles. Use ROV_{ij} to denote the proportion of the number of the neighbor vehicle set Neigh (i) receiving vehicle j's messages. If it is greater than or equal to the threshold ROV_{thre}, increase the reward factor λ; otherwise,

subtract the penalty factor μ. QOD_{ij} represents the packet delivery rate to node i during the packet transmission process of node j. If it is greater than or equal to the delivery rate threshold $\mathrm{QOD}_{\mathrm{thre}}$, the reward factor λ is increased; otherwise, the penalty factor μ is subtracted. It can be seen that the range of ROV_{ij} and QOD_{ij} is from 0 to 1. The definition of SP_{ij} is as follows:

$$\mathrm{SP}_{ij} = \frac{1}{2} * \frac{\mathrm{ROV}_{ij} * \mathrm{QOD}_{ij}}{\mathrm{ROV}_{ij} + \mathrm{QOD}_{ij}}. \tag{6}$$

(4) Recommendation Trust Calculation. We can get inter-vehicular recommendation trust as follows:

$$\mathrm{RT} = \mathrm{RW} * \mathrm{NT}_{ij}. \tag{7}$$

We use Algorithm 1 to calculate the trust value between vehicles.

3.2. Calculation of the Infrastructure Trust Opinion

3.2.1. Distance Coefficient of RSU (DC). The trust management mechanism of the RSU has higher requirements for the time delay. The closer the RSU to the vehicle that sent the original message, the more detailed the vehicle message that can be obtained. Therefore, the distance between the sending vehicle and the RSU for trust management has also become an important criterion. The distance coefficient (DC) is expressed as follows:

$$\mathrm{DC}_{ij} = \frac{\sum_{r=1}^{|R|} \mathrm{DIS}_{rj} - \mathrm{DIS}_{ij}}{\sum_{r=1}^{|R|} \mathrm{DIS}_{rj}}, \tag{8}$$

where R represents the set of RSUs that received the original message of sending vehicle j. It can be seen that the closer the estimated RSU is to the sending vehicle, the greater the DC of the RSU to vehicle j is.

3.2.2. Similarity Calculation. To improve the trustworthiness accuracy of the infrastructure to the message sending vehicle, the concept of similarity metrics is used to measure the trust measurement opinion of the RSU to the vehicle [26]. Reference [27] uses cosine-based similarity to judge the similarity of two vectors. The cosine similarity or cosine metric calculates the similarity between two vectors in the inner product space by determining the cosine of the angle between them. This index is widely used for information retrieval and text mining [28]. Each trust level can be regarded as a vector in a k-dimensional space. If the node does not evaluate other nodes, the default rating is used. We define the similarity measure as sim_i^j. Assuming it is an n-dimensional normalized vector, we express the similarity as follows:

$$\mathrm{sim}_i^j = \frac{\sum_{k=1}^{n} \mathrm{TV}_k^i * \mathrm{TV}_k^j}{\sqrt{\sum_{k=1}^{n} \left(\mathrm{TV}_k^i\right)^2} * \sqrt{\sum_{k=1}^{n} \left(\mathrm{TV}_k^j\right)^2}}, \tag{9}$$

where TV_k^i and TV_k^j represent the kth dimension of the normalized vector of node i and node j, respectively. Since the value of this vector cannot be negative, the similarity value range is between 0 (dissimilar) and 1 (completely similar). After the infrastructure receives the similarity calculation instruction, it will calculate the similarity of the interest preferences of the vehicle j that sends the message and the vehicle i in the trust table of the infrastructure. The greater the similarity value is, the closer the interest preferences are between them and the more likely it is to be accepted as a trusted node.

3.2.3. Infrastructure Trust Value (IT). The trust value management of IT can be realized between the infrastructure RSUs, and the trust upgrade information about the upper trust management plane can be updated synchronously. By combining the RSU distance coefficient and the similarity between the computing nodes, the trust calculation of the infrastructure for the vehicle can be expressed by the following formula:

$$\mathrm{IT} = \mathrm{DC}_{ij} * \mathrm{sim}_i^j. \tag{10}$$

3.3. Hybrid Trust Calculation (HT). In VANET, the ultimate global hybrid trust calculation should include the trust between vehicles and the trust between vehicles and infrastructure. Owing to the complementary role of infrastructure in trust management, trust between vehicles is more important than trust in infrastructure. Therefore, if n is used to represent the number of vehicle interactions and $1/(n+1)$ is used as the adjustment factor, it can ensure that the trust between vehicles gets more weight. Then, the hybrid trust can be obtained by the following formula:

$$\mathrm{HT} = \left[\left(1 - \frac{1}{n+1}\right) * \sqrt{\mathrm{ST} * \mathrm{RT}}\right] + \left[\frac{1}{n+1} * \mathrm{IT}\right]. \tag{11}$$

We use Algorithm 2 to represent the complete vehicle hybrid trust calculation process.

4. Simulation and Performance

To test the performance of the proposed scheme, in this section, we first introduce the relevant attack models and explain the tools and parameter settings used in the simulation environment. Secondly, we define evaluation indicators to evaluate the accuracy of HHTM, and finally, we carry out the comparative analysis of experimental results under different schemes.

4.1. Attack Models. The trust management model is mainly to spread trustworthy information in the IoV, so this paper mainly notes the following malicious attacker model to evaluate the performance of HHTM.

4.1.1. Simple Attacks (SAs). The attacker acts as a receiver where messages are deliberately discarded or delayed, thereby preventing legitimate vehicles from receiving safety

```
Input LTD, OTV, vehicle ID, $ROV_{ij}$, $QOD_{ij}$
Output ST, RT
if vehicle $i$ receives interactive information from vehicle $j$ then
  Calculate the distance between nodes and obtain the STW value according to equation (1)
  Check the local trust database (LTD) of vehicle $i$
  if $ID_j \in LTD_i$ then
    Search $OTV_{ij}$
  else
    Take $OTV_{ini}$ as the OTV value of vehicle $i$ to vehicle $j$
  end if
  Upgrade the OTV information of vehicle $i$
  Calculate the ST of vehicle $j$ using equation (3)
  if $ROV_{ij} \geq ROV_{thre}$ then
    $ROV_{ij} \longleftarrow ROV_{ij} + \lambda$
  else
    $ROV_{ij} \longleftarrow ROV_{ij} - \mu$
    if $QOD_{ij} \geq QOD_{thre}$ then
      $QOD_{ij} \longleftarrow QOD_{ij} + \lambda$
    else
      $QOD_{ij} \longleftarrow QOD_{ij} - \mu$
    end if
  end if
  Use equation (6) to calculate the trust score SP between vehicles
  Calculate the value of NT according to the value of $SP_{ij}$ using equation (5)
  Determine the role type of vehicle $j$ and use equation (4) to find RW
  Use equation (7) to calculate the RT of vehicle $i$ to vehicle $j$
end if
```

ALGORITHM 1: Inter-vehicular trust calculation.

```
Input DIS, TV, $HT_{thre}$
Output HT
if vehicle $i$ receives interactive information from vehicle $j$ then
  Use Algorithm 1 to solve the ST and RT of the vehicle
  if Vehicle $i$ is within the coverage of RSU then
    Calculate $DC_{ij}$ using equation (8)
    Calculate $sim_i^j$ using equation (9)
    Calculate IT using equation (10)
  else
    $IT = 0$
  end if
  Calculate HT using equation (11)
end if
if $HT \geq HT_{thre}$ then
  Confirm that vehicle $j$ is a trusted vehicle
  Continue to receive interactive information from vehicle $j$
else
  Confirm that vehicle $j$ is a dishonest vehicle
  Discard the interaction request information of the vehicle
end if
Upgrade the trust management information of vehicle $j$ in the LTD of vehicle $i$
Broadcast the trust management upgrade information of vehicle $i$
```

ALGORITHM 2: Hybrid trust calculation.

messages promptly. Due to the sensitive nature of the messages involved in the IoV, discarding safety messages can make a huge impact on the network. The attacker may use selfish behavior to manipulate the infected node so that it will not follow normal network protocol and provide necessary services for other nodes. For example, they will not forward data packets or spread route discovery requests. However, when the node is requested about the credibility of other nodes, it will not provide any false trust opinions.

4.1.2. Selective Misbehavior Attacks (SMAs). In this attack, malicious nodes provide false information to some nodes while providing normal information to other nodes. Attackers have inconsistent behavior patterns for different nodes, which will make the trust management between different nodes inconsistent and increase the difficulty of detection.

4.1.3. Time-Dependent Attacks (TDAs) [29]. Attackers use random patterns in the network to produce intelligent behavior. The attacker will initially act as a legitimate node in a short period to obtain the trust of vehicles in the network. The attacker can only start malicious behavior after gaining the trust of other vehicles and being a part of the legitimate network. In the attack mode, the attacker will share false messages and ratings with neighboring vehicles.

4.2. Simulation Setup. To facilitate the simulation, we used Veins [30], an open source platform widely used in vehicle network simulation. Veins is constructed by two mainstream simulators: traffic simulator SUMO [31] and discrete-time simulator OMNET++ [32]. Through the traffic control interface, events triggered by OMNET++ can deliver response instructions to SUMO to change vehicle paths and other information. We select part of the real road network in Zhengzhou City, Henan Province (as shown in Figure 5), as the simulated road network, with a topological area of 3 km × 3 km, and use SUMO to construct the initial road network (as shown in Figure 6). We randomly place 10 RSUs in the road network, and all vehicles are equipped with wireless communication standard protocol IEEE 802.11p. The system deploys one controller, and the infrastructure is connected to the controller through an Ethernet interface.

To ensure the reliability of the experimental results, 30 random experimental seeds have been carried out for each experimental scene and the average value has been taken. Table 2 provides details on the parameters used in the experimental environment. Since we believe that credibility is difficult to establish and easy to be destroyed, we set $\mu = 10\lambda$. To avoid the cold start problem [33], we set both $\mathrm{HT}_{\mathrm{thre}}$ and $\mathrm{OTV}_{\mathrm{ini}}$ to 0.5. The probabilities of malicious behaviors of the three malicious node attacks are all set to 0.5.

4.3. Evaluation Metrics. Since the weighted voting method has been widely used in many previous wireless network trust management schemes [15, 34], we use the weighted voting method as a baseline method when evaluating the performance of the HHTM scheme.

We utilize the following three parameters to evaluate the accuracy of the HHTM scheme: precision (P) and recall (R), which are widely used in machine learning and information retrieval to evaluate accuracy [35]. In this paper, we use both P and R to evaluate the accuracy of the proposed scheme for identifying dishonest nodes in VANET. F-score (F) is the weighted average of P and R values, used to reflect the overall accuracy of the trust management model. The parameters are defined as follows:

$$
\begin{aligned}
P &= \frac{\text{number of truly malicious nodes caught}}{\text{total number of dishonest nodes caught}},\\
R &= \frac{\text{number of truly malicious nodes caught}}{\text{total number of truly malicious nodes}},\\
F &= \frac{2 * P * R}{P + R}.
\end{aligned}
\tag{12}
$$

4.4. Result Analysis. As shown in Figure 7(a), the precision values of HHTM at different number of nodes are higher than those of the baseline method. As the node density continues to increase, its value exceeds by 90%. This is because when the total number of honest nodes increases, the trust evaluation node is more likely to receive real data from other nodes. Figures 7(b) and 7(c) show that the HHTM scheme is also superior to the baseline method in terms of recall value and F-score value. Similarly, when the node density is high, the value exceeds by 90%.

Figure 8 shows the changes of P and R values during SA. When the number of malicious nodes is small, the precision and recall of the two schemes are better. As the number of malicious nodes increases, the P and R values of the two schemes have both declined to a certain extent. It can be seen that the difference between the two schemes is not obvious. This is because SAs only maliciously discard or delay information and do not spread false trust opinions, so they are less destructive than other attacks.

Figure 9 shows the changes of P and R values during SMA. It can be seen that the P and R values of the baseline method are significantly lower than those of HHMT. When the number of malicious nodes reaches 40%, the P and R values of HHMT are 12.7% and 11% higher than those of the baseline method, respectively. This is because the baseline method relies on weighted voting, so the recognition of nodes with inconsistent trust management is reduced.

Figure 10 indicates the changes of P and R values during TDA. Because the attacker uses intelligent behavior to initiate attacks intermittently, the values of P and R of HHTM are lower than those of the above two attacks. However, opposed to the baseline method, HHTM still shows good response capabilities due to the adoption of a hierarchical trust management strategy. When the number of malicious nodes reaches 40%, the P and R values of HHTM are 13.8% and 12.2% higher than those of the baseline method, respectively.

For different levels of security incidents, the trust threshold requirements are also different, such as setting a higher threshold for the determination of road traffic accidents to ensure the reliability of the event. We compare the detection rate and F-score value of the HHTM scheme with EBT [36] and AATMS [37] when the threshold is different to confirm the performance difference between the different schemes.

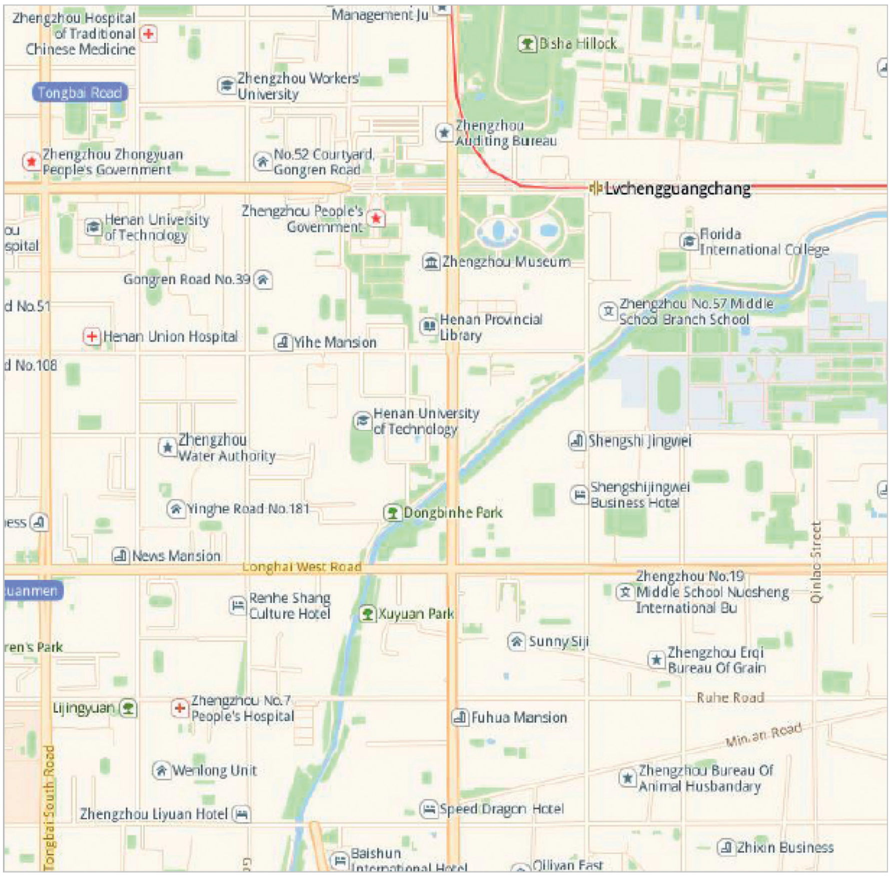

Zhengzhou Hospital of Traditional Chinese Medicine
Management Ju
Bisha Hillock
Tongbai Road
Zhengzhou Workers' University
Zhengzhou Auditing Bureau
Zhengzhou Zhongyuan People's Government
No.52 Courtyard, Gongren Road
Lvchengguangchang
Henan University of Technology
Zhengzhou People's Government
Florida International College
Zhengzhou Museum
Gongren Road No.39
Zhengzhou No.57 Middle School Branch School
Henan Provincial Library
Henan Union Hospital
Yihe Mansion
Henan University of Technology
Shengshi Jingwei
Zhengzhou Water Authority
Shengshijingwei Business Hotel
Yinghe Road No.181
Dongbinhe Park
News Mansion
Longhai West Road
Zhengzhou No.19 Middle School Nuosheng International Bu
Renhe Shang Culture Hotel
Xuyuan Park
Qinlao Street
Sunny Siji
Zhengzhou Erqi Bureau Of Grain
Ruhe Road
Lijingyuan
Zhengzhou No.7 People's Hospital
Fuhua Mansion
Min.an Road
Wenlong Unit
Zhengzhou Bureau Of Animal Husbandary
Tongbai South Road
Zhengzhou Liyuan Hotel
Speed Dragon Hotel
Baishun International Hotel
Zhixin Business

Figure 5: Extracted city map.

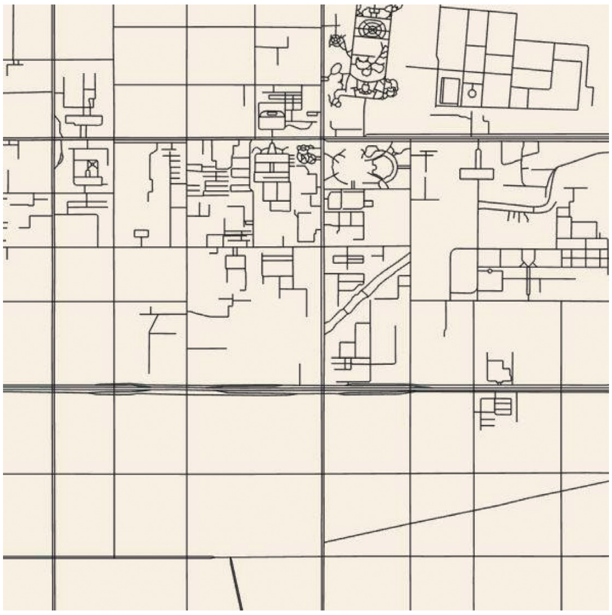
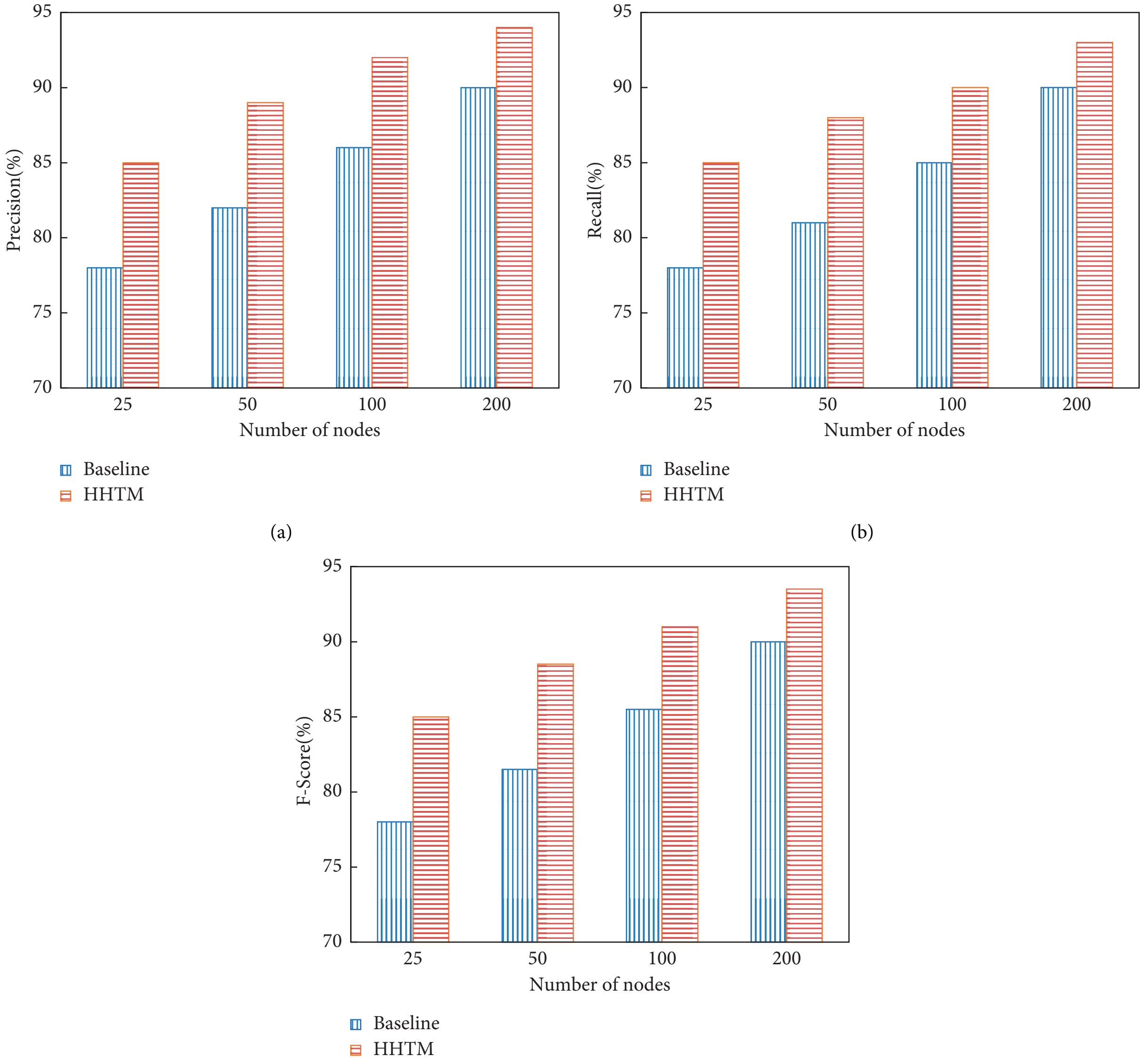


Figure 6: Initial road network model.

It can be seen from Figure 11 that the higher the trust threshold, the lower the detection rate. The proposed trust management scheme is superior to the comparison scheme in terms of detection rate. When the trust threshold is set to 0.9, the detection rate of HHTM is still above 20%. It can be seen from Figure 12 that with the increase of the trust

Table 2: Simulation parameters.

Parameter	Value
Simulation area (km × km)	3 × 3
Simulation time (sec)	800
Number of vehicles	25, 50, 100, 200
Location of vehicles	Random
Num. of attackers (%)	10, 20, 30, 40, 50
MAC protocol	IEEE 802.11p
ROV_{thre}	0.5
QOD_{thre}	0.5
HT_{thre}	0.5
OTV_{ini}	0.5
Reward factor λ	0.01
Penalty factor μ	0.1

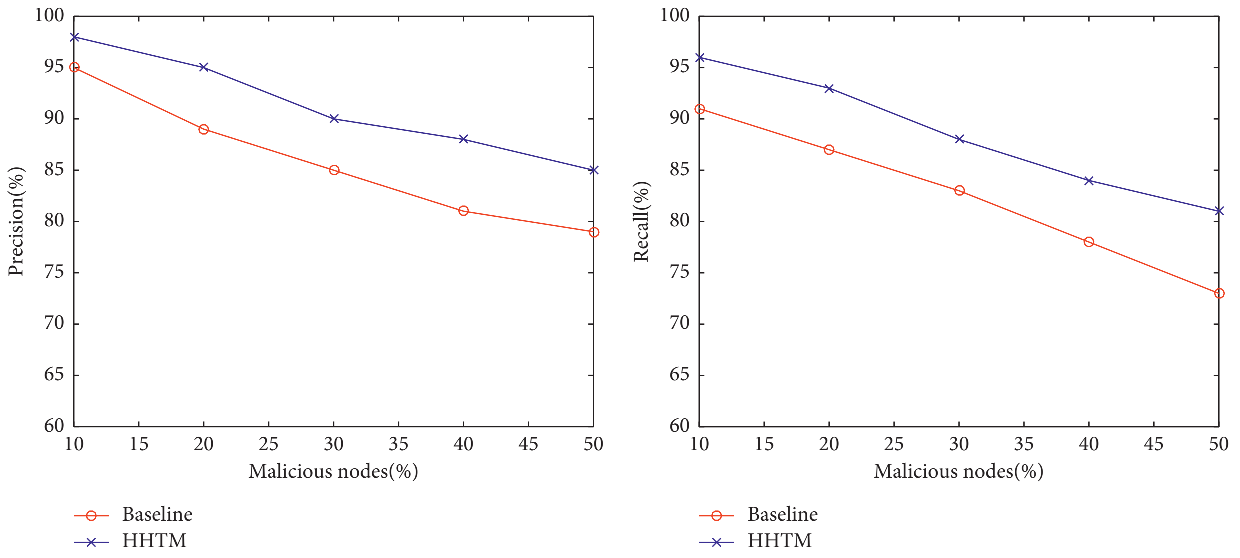


(a)

(b)

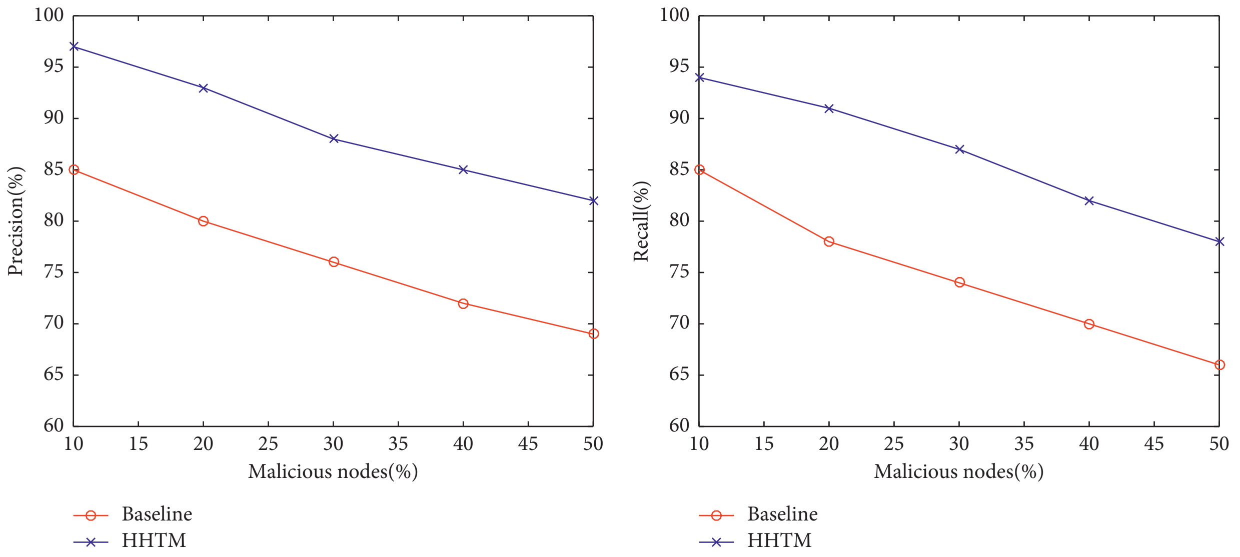
(c)

Figure 7: (a) Precision at different number of nodes. (b) Recall at different number of nodes. (c) *F*-score at different number of nodes.



(a) (b)

Figure 8: (a) *P* value during SA. (b) *R* value during SA.



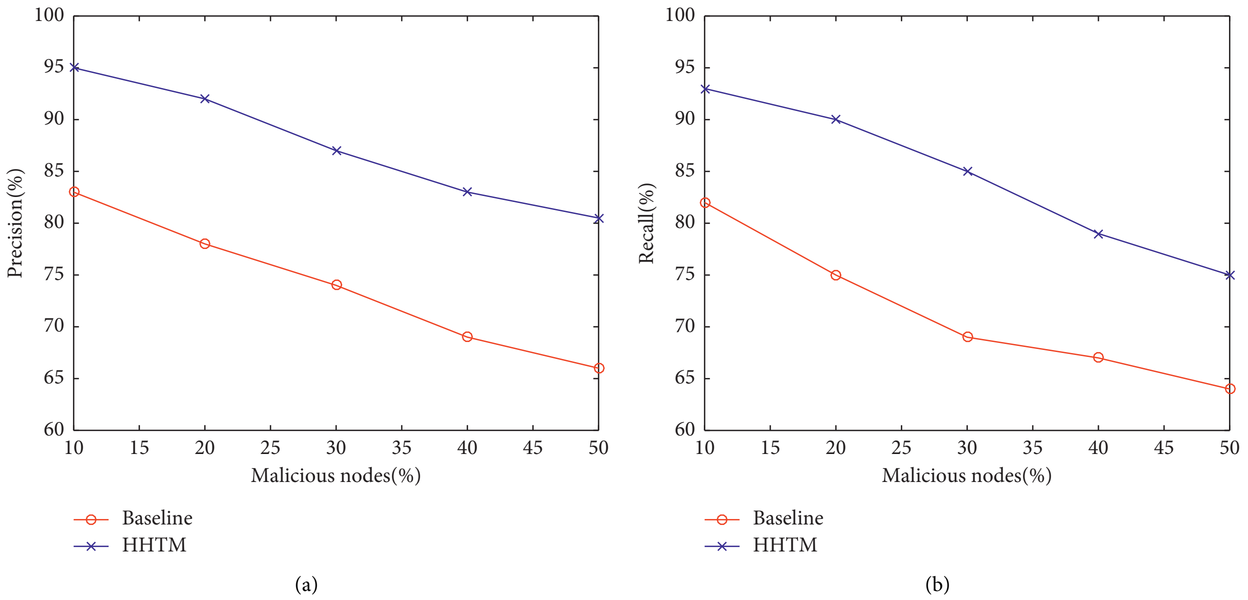
(a) (b)

Figure 9: (a) *P* value during SMA. (b) *R* value during SMA.

threshold, the *F*-score decreases continuously. When the trust threshold reaches 0.8, the *F*-score of all schemes decreases significantly. At the same time, the *F*-score of the proposed scheme is always better than that of the contrast schemes.

Figure 13 shows the impact of the delay on the trust management scheme under the three types of attacks. It can be observed that in the SA, the end-to-end delay of the three schemes is not much different. In the SMA, with the increase of malicious nodes, the delays of the three schemes are comparable. When the number of malicious nodes is 50%, the delay of HHTM is reduced by 38.6% and 25.2% compared with EBC and AATMS, respectively. In the TDA, when the number of malicious nodes exceeds 30%, the delay of EBC increases significantly. It shows that this scheme has no advantage in dealing with TDAs. In the three attack modes, the delay of HHTM is better than that of the comparison schemes.

In summary, compared with other solutions, HHTM has achieved better results in resisting the attacks of the three models and can better deal with a higher proportion of malicious nodes.



(a)

(b)

Figure 10: (a) *P* value during TDA. (b) *R* value during TDA.

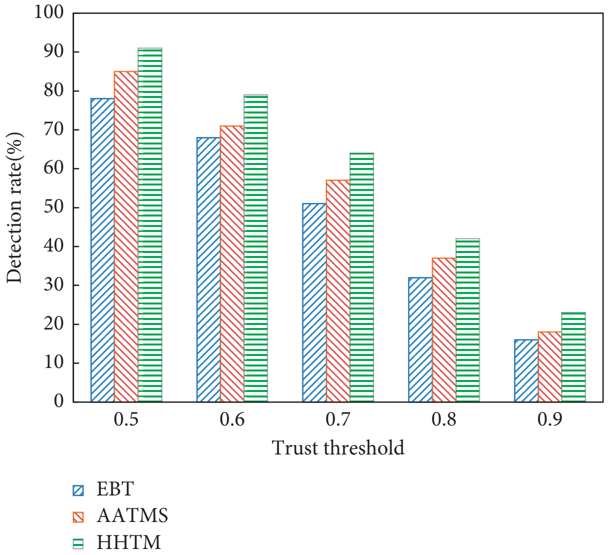


Figure 11: The impact of trust threshold on detection rate.

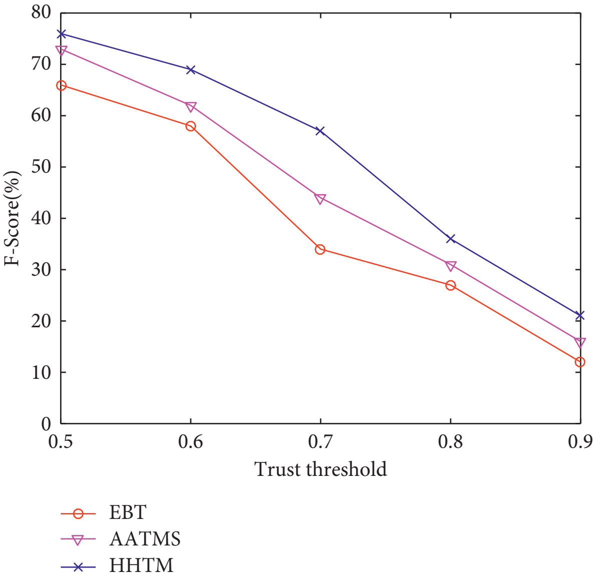


Figure 12: The impact of trust threshold on *F*-score.

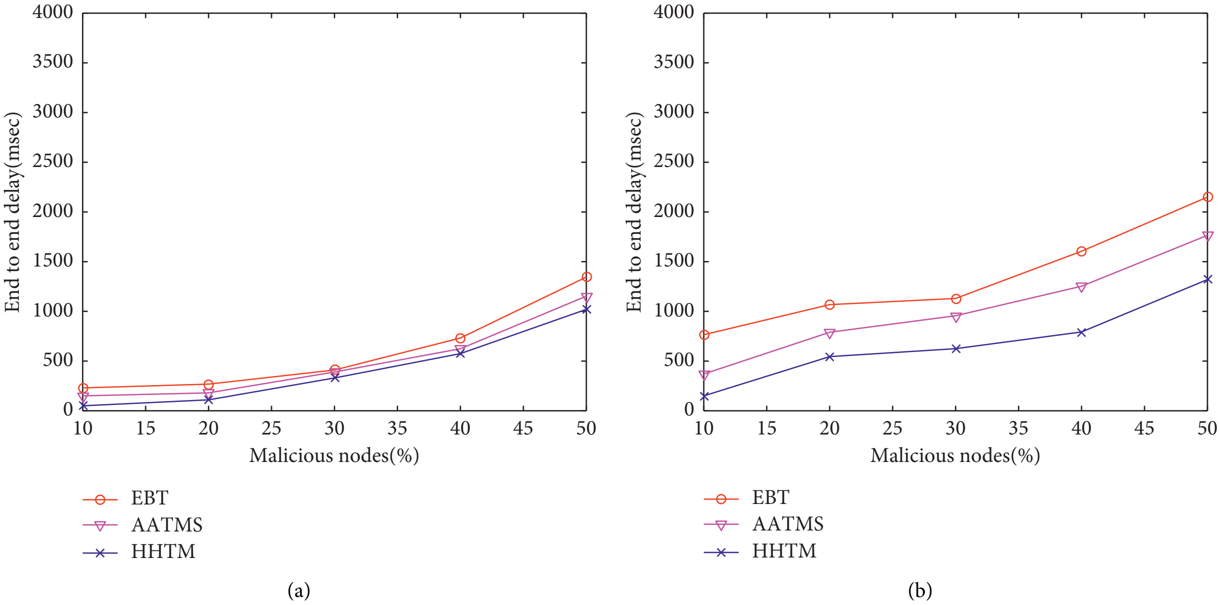
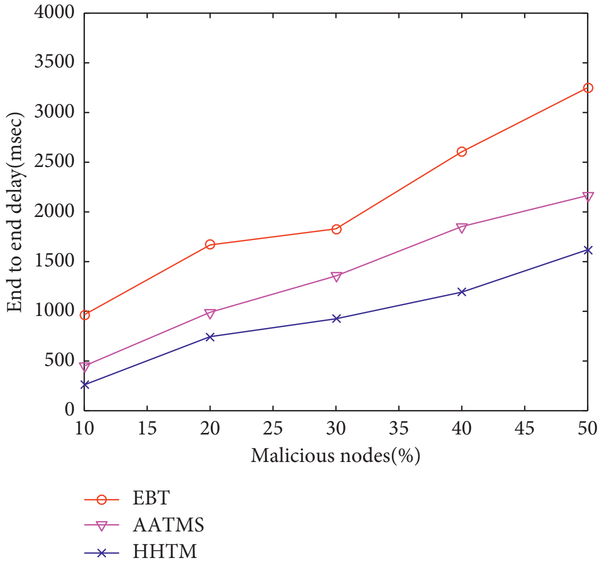


Figure 13: Continued.



(c)

Figure 13: (a) End-to-end delay during SA. (b) End-to-end delay during SMA. (c) End-to-end delay during TDA.

5. Conclusions

In VANET, a safe and attack-free environment is essential for the transmission of trusted messages between the vehicle and the infrastructure. However, because VANET involves a variety of different application environments, it is a very challenging task to ensure the trust foundation in each environment when an attacker penetrates the network and pollutes the network with fake information. A robust TM architecture should be established to achieve vehicle and message verification.

In this article, with the help of SDVN's fast flow forwarding mechanism, a trust management scheme named HHTM is proposed to evaluate the credibility of vehicles and traffic data in VANET. In the HHTM scheme, the trustworthiness of the EvN is modeled and evaluated as two independent indicators, namely, the trust between vehicles and the trust between nodes and infrastructure. Among them, it focuses on the use of the inter-vehicular trust to evaluate whether the received node data are credible and to what extent. On the other hand, we use the node-infrastructure trust to strengthen the trust of vehicles sending data in VANET. Extensive experiments are carried out to verify the robustness of the proposed trust management scheme. The experiment results show that compared with the comparative schemes, the proposed HHTM scheme can accurately assess the credibility of nodes and data in VANET and deal with various malicious attacks.

Based on the realization of the trust management of the Internet of Vehicles, in order to strengthen the data transmission security of the Internet of Vehicles, future work should be aimed at establishing security mechanism for vehicular data sharing. At the same time, the fine-grained access control of the Internet of Vehicles is also a direction worth studying.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (nos. 61802429, 61872382, and 61521003) and the National Key R&D Program of China (nos. 2018YFB0804002, 2019YFB1802505, 2019YFB1802501, 2019YFB1802502, and 2020YFB1804803).

References

[1] N. A. Jagadeesan and B. Krishnamachari, "Software-defined networking paradigms in wireless networks: a survey," *ACM Computing Surveys*, vol. 47, no. 2, pp. 1–11, 2015.

[2] C. Jiacheng, Z. Haibo, Z. Ning, Y. Peng, G. Lin, and S. Xuemin, "Software defined Internet of vehicles: architecture, challenges and solutions," *Journal of Communications and Information Networks*, vol. 1, no. 1, pp. 14–26, 2016.

[3] C. Y. Yeun, "Security protocol model for ubiquitous networks," US patent, 2006.

[4] J. Wang, Y. Zhang, Y. Wang, and X. Gu, "RPRep: a robust and privacy-preserving reputation management scheme for pseudonym enabled VANETs," *International Journal of Distributed Sensor Networks*, vol. 2016, Article ID 6138251, 15 pages, 2016.

[5] J. Grover, M. S. Gaur, and V. Laxmi, "Trust establishment techniques in VANET," *Wireless Networks and Security*, Springer, Berlin, Germany, pp. 273–301, 2013.

[6] F. Li and Y. Wang, "Routing in vehicular Ad Hoc networks: a survey," *IEEE Vehicular Technology Magazine*, vol. 2, no. 2, pp. 12–22, 2007.

[7] S. Park, B. Aslam, and C. C. Zou, "Long-term reputation system for vehicular networking based on vehicle's daily commute routine," in *Proceedings of the 2011 IEEE Consumer Communications and Networking Conference (CCNC'11)*, pp. 436–441, Las Vegas, NV, USA, January 2011.

[8] X. Li, J. Liu, X. Li, and W. Sun, "RGTE: a reputation-based global trust establishment in VANETs," in *Proceedings of the 5th IEEE International Conference on Intelligent Networking and Collaborative Systems (INCoS'13)*, pp. 210–214, IEEE, Xi'an, China, September 2013.

[9] Z. Huang, S. Ruj, M. A. Cavenaghi, M. Stojmenovic, and A. Nayak, "A social network approach to trust management in VANETs," *Peer-to-Peer Networking and Applications*, vol. 7, no. 3, pp. 229–242, 2014.

[10] W. Bamberger, J. Schlittenlacher, and K. Diepold, "A trust model for intervehicular communication based on belief theory," in *Proceedings of the 2nd IEEE International Conference on Social Computing (SocialCom'10)*, pp. 73–80, IEEE, Minneapolis, MN, USA, August 2010.

[11] X. Hong, D. Huang, M. Gerla, and Z. Cao, "SAT: situation-aware trust architecture for vehicular networks," in *Proceedings of the 3rd International Workshop on Mobility in the Evolving Internet Architecture, MobiArch'08*, pp. 31–36, Seattle, WA, USA, August 2008.

[12] F. Ahmad, V. N. L. Franqueira, and A. Adnane, "TEAM: a trust evaluation and management framework in context-enabled vehicular ad-hoc networks," *IEEE Access*, vol. 6, pp. 28643–28660, 2018.

[13] Z. Lu, G. Qu, and Z. Liu, "A survey on recent advances in vehicular network security, trust, and privacy," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 2, pp. 760–776, 2019.

[14] F. Ahmad, A. Adnane, F. Kurugollu, and R. Hussain, "A comparative analysis of trust models for safety applications in IoT-enabled vehicular networks," in *Proceedings of the 2019 Wireless Days (WD)*, Manchester, UK, 2019.

[15] M. Raya, P. Papadimitratos, V. D. Gligor, and J. P. Hubaux, "On data-centric trust establishment in ephemeral ad hoc networks," in *Proceedings of the IEEE 27th Conference on Computer Communications (INFOCOM)*, April 2008.

[16] T. Gazdar, A. Belghith, and H. Abutair, "An enhanced distributed trust computing protocol for VANETs," *IEEE Access*, vol. 6, pp. 380–392, 2018.

[17] A. Wu, J. Ma, and S. Zhang, "RATE: a RSU-aided scheme for data-centric trust establishment in VANETs," in *Proceedings of the 7th International Conference on Wireless Communications, Networking and Mobile Computing*, September 2011.

[18] S. Gurung, D. Lin, A. Squicciarini, and E. Bertino, "Information-oriented trustworthiness evaluation in vehicular ad-hoc networks," *Network and System Security*, Springer, Berlin, Germany, pp. 94–108, 2013.

[19] U. Khan, S. Agrawal, and S. Silakari, "Detection of malicious nodes (DMN) in vehicular ad-hoc networks," in *Proceedings of the 2014 International Conference on Information and Communication Technologies (ICICT)*, pp. 965–972, Elsevier, Chengdu, China, December 2014.

[20] N. Yang, "A similarity based trust and reputation management framework for VANETs," *International Journal of Future Generation Communication and Networking*, vol. 6, no. 2, pp. 25–34, 2013.

[21] F. G. Mármol and G. M. Pérez, "TRIP, a trust and reputation infrastructure-based proposal for vehicular ad hoc networks," *Journal of Network and Computer Applications*, vol. 35, no. 3, pp. 934–941, 2012.

[22] S. Ahmed and K. Tepe, "Using logistic trust for event learning and misbehaviour detection," in *Proceedings of the IEEE 84th Vehicular Technology Conference (VTC-Fall)*, September 2016.

[23] Y.-M. Chen and Y.-C. Wei, "A beacon-based trust management system for enhancing user centric location privacy in VANETs," *Journal of Communications and Networks*, vol. 15, no. 2, pp. 153–163, 2013.

[24] R. Shrestha and S. Y. Nam, "Trustworthy event-information dissemination in vehicular ad hoc networks," *Mobile Information Systems*, vol. 2017, Article ID 9050787, 16 pages, 2017.

[25] A. Alioua, S.-M. Senouci, S. Moussaoui, H. Sedjelmaci, and A. Boualouache, "Software-Defined heterogeneous vehicular networks: taxonomy and architecture," in *Proceedings of the 2017 Global Information Infrastructure and Networking Symposium (GIIS)*, Saint Pierre, France, 2017.

[26] C. Piao, J. Zhao, and J. Feng, "Research on entropy-based collaborative filtering algorithm," in *Proceedings of the 2007 IEEE ICEBE*, Hong Kong, China, October 2007.

[27] J.-M. Chen, T.-T. Li, and J. Panneerselvam, "TMEC: a trust management based on evidence combination on attack-resistant and collaborative Internet of vehicles," *IEEE Access*, vol. 7, pp. 148913–148922, 2019.

[28] A. Singhal and I. Google, "Modern information retrieval: a brief overview," *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, vol. 24, no. 4, pp. 35–43, 2001.

[29] H. Xia, S.-s. Zhang, Y. Li, Z.-k. Pan, X. Peng, and X.-z. Cheng, "An attack-resistant trust inference model for securing routing in vehicular ad hoc networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 7, pp. 7108–7120, 2019.

[30] Veins, "Vehicles in network simulation, the open source vehicular simulation framework," 2018, http://veins.car2x.org.

[31] M. Behrisch, L. Bieker, J. Erdmann et al., "SUMO-simulation of urban mobility: an overview," in *Proceedings of the 3rd International Conference on Advances in System Simulation*, Barcelona, Spain, 2011.

[32] OMNET, "OMNET++: discrete event simulator," 2018, https://omnetpp.org/.

[33] L. H. Son, "Dealing with the new user cold-start problem in recommender systems: a comparative review," *Information Systems*, vol. 58, pp. 87–104, 2016.

[34] I.-R. Chen, F. Bao, M. Chang, and J.-H. Cho, "Dynamic trust management for delay tolerant networks and its application to secure routing," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 5, pp. 1200–1210, 2014.

[35] J. Davis and M. Goadrich, "The relationship between precision-recall and ROC curves," in *Proceedings of the ACM 23rd International Conference on Machine Learning*, Pittsburgh, PA, USA, 2006.

[36] U. F. Minhas, J. Zhang, T. Tran, and R. Cohen, "A multifaceted approach to modeling agent trust for effective communication in the application of mobile Ad Hoc vehicular networks," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 41, no. 3, pp. 407–420, 2011.

[37] J. Zhang, K. Zheng, D. Zhang, and B. Yan, "AATMS: an anti-attack trust management scheme in VANET," *IEEE Access*, vol. 8, pp. 21077–21090, 2020.

Hindawi
Mobile Information Systems
Volume 2021, Article ID 9160516, 11 pages
https://doi.org/10.1155/2021/9160516

Research Article

An Efficient Routing Approach to Maximize the Lifetime of IoT-Based Wireless Sensor Networks in 5G and Beyond

C. Jothikumar,[1] **Kadiyala Ramana**,[2] **V. Deeban Chakravarthy**,[1] **Saurabh Singh**,[3] **and In-Ho Ra**[4]

[1]*School of Computing, SRM Institute of Science and Technology, Chennai, India*
[2]*Department of Artificial Intelligence & Data Science, Annamacharya Institute of Technology and Sciences, Rajampet, India*
[3]*Department of Industrial and System Engineering, Dongguk University, Seoul 04620, Republic of Korea*
[4]*School of Computer,Information and Communication Engineering, Kunsan National University, Gunsan 54150, Republic of Korea*

Correspondence should be addressed to In-Ho Ra; ihra@kunsan.ac.kr

Received 13 May 2021; Revised 14 June 2021; Accepted 4 July 2021; Published 13 July 2021

Academic Editor: Zengpeng Li

The Internet of Things grew rapidly, and many services, applications, sensor-embedded electronic devices, and related protocols were created and are still being developed. The Internet of Things (IoT) allows physically existing things to see, hear, think, and perform a significant task by allowing them to interact with one another and exchange valuable knowledge when making decisions and caring out their vital tasks. The fifth-generation (5G) communications require that the Internet of Things (IoT) is aided greatly by wireless sensor networks, which serve as a permanent layer for it. A wireless sensor network comprises a collection of sensor nodes to monitor and transmit data to the destination known as the sink. The sink (or base station) is the endpoint of data transmission in every round. The major concerns of IoT-based WSNs are improving the network lifetime and energy efficiency. In the proposed system, Optimal Cluster-Based Routing (Optimal-CBR), the energy efficiency, and network lifetime are improved using a hierarchical routing approach for applications on the IoT in the 5G environment and beyond. The Optimal-CBR protocol uses the *k*-means algorithm for clustering the nodes and the multihop approach for chain routing. The clustering phase is invoked until two-thirds of the nodes are dead and then the chaining phase is invoked for the rest of the data transmission. The nodes are clustered using the basic *k*-means algorithm during the cluster phase and the highest energy of the node nearest to the centroid is selected as the cluster head (CH). The CH collects the packets from its members and forwards them to the base station (BS). During the chaining phase, since two-thirds of the nodes are dead and the residual energy is insufficient for clustering, the remaining nodes perform multihop routing to create chaining until the data are transmitted to the BS. This enriches the energy efficiency and the network lifespan, as found in both the theoretical and simulation analyses.

1. Introduction

The Internet of Things has increased its adoption, but these are only a few of the endless fields in which it can be applied, providing infinite uses. A lot of things will be connected through the Internet of Things (IoT) and allow for automatic (human-to-machine) machine (M2M) communication. Providing all the hardware and software for the Internet of Things (IoT) with eyes and auditory and optical sensors (WSNs), the subject becomes important due to state-of-the-art applications and cutting-edge technologies.

WSN is a cornerstone of IoT, and all depend on it. WSN's main function is in the promotion and growth of IoT is allowing lower resource and life-changing services. It links tens of thousands of sensors using wireless technology. Advancement in sensors technology makes smaller, more intelligent appliances feasible for low-cost and large-scale applications. The sensor nodes are usually made up of various numbers of WSN. WSN can be used in various industrial applications for humidity, temperature, pressure, light, and movement control, as well as in agriculture, logistics, and military, and for transport and communications.

Regardless, however, existing telecommunication technologies have not yet kept up with the increasing demands of the digital age. Better performance, higher bandwidth, lower latency, and less power consumption for the Fifth Generation (5G) are required. It provides a better, more sophisticated device and a more dependable technology. Also, with all the promises of 5G, the network infrastructure coverage is a limiting factor. The 5G network system uses a millimeter band that affects the continuity of coverage. Despite increased data throughput, the 5G network has lower service availability. More subsidiary repeaters would be needed to propagate the waves in heavily populated areas, such as a megalopolis, to maintain stable data speeds. Thus, a large base station and antenna network of deployment are needed to adequately cover a 5G coverage area. Building the network and stations would not be cost-effective.

Wireless sensor networks (WSNs) fill in as a scaffold between the physical and virtual universes. These are exceptionally scattered networks of little, lightweight sensor nodes outfitted with batteries that are answerable for detecting and communicating information to the Internet. WSN is basic in giving the most challenging solution and most alluring regions for an assortment of use regions, including military observation, torrent identification, patient health monitoring, disaster surveillance and emergency management, environment checking, and mechanical computerization. Memory, processors, detecting components, batteries, and transceivers contain the sensor nodes. Accordingly, the organization of sensor nodes is deployed in the observation region, creating a huge measure of data that should be communicated to the BS. Notwithstanding, on the grounds that sensor nodes are so little, they have a few limitations as far as memory, transmission capacity, data processing, and battery life [1].

It can be contended that energy management ought to be the essential thought when designing an effective WSN. Regularly, when WSNs are used for remote area observing, a lifetime of the sensor network guarantees the efficiency of the system and reliability on data transmission. Perceiving the components that add to the energy utilization needed to support all tasks within the WSN, an enormous extent of energy is regularly utilized for data communication. By minimizing the total number of jumps and the gaps, energy utilization can be decreased. At last, as expected, the WSN's life expectancy will be adequately broadened.

Due to the immense utility potential of sensors in diverse systems, perhaps environmental monitoring, industrial automation, healthcare, target tracking, and localization, the popularity of research in wireless sensor networks (WSN) is increasing day by day. Large numbers of sensor nodes that are compact and stocked with less power are the primary components of WSNs. Sensor nodes sense, process, and transfer the observed data of the surroundings to the destination, thus making it easier to monitor the hard environments which are inconvenient to monitor otherwise. A typical WSN comprises nodes ranging from a few hundred to several thousand [2, 3]. Dynamic network topology, power constraints, heterogeneous nature of nodes, limited preloaded energy, mobility of nodes, and adaptability during node failures are the major characteristics of WSNs. The system without any routing approach disseminates the incoming packets to every link in the network through its neighbors. The transmission of the packet is guaranteed from the source to the destination since each node recognizes the data of every other node through its neighbors [4, 5]. The system does not require complex routing techniques. The main issues of flooding refer to the blindness of the resource.

For potential large-scale networks, lightweight, low-cost, and ultimately expendable sensor nodes are needed. Also, to extend the lifetime of an individual sensor node and also that of the network, each node needs to use as little power as possible because of power limitations. The node lifetime is the duration during which data can be received, transmitted, or forwarded to others by a node. Life and energy use are also critical concerns for WSNs. Routing algorithms can make intelligent decisions with a reliable lifetime estimate that can help save resources and extend the lifetime of the node.

The amalgamation of data from various sources is termed data aggregation. The sensor nodes can generate homogeneous data packets from several nodes. Aggregation of these packets reduces the number of transmissions. Partial or complete execution of the above functions can be done in each sensor node. In comparison with communication, computations consume minimum energy, thus saving an ample amount of energy. Aggregation of data is an efficient way to attain a significant saving of energy and can further lead to the traffic optimization of various routing protocols. In general, in several network architectures, highly dominant and specific nodes are allocated for aggregation and computation tasks. Much research has been done to study various algorithms and protocols to decrease the total energy consumed by the sensor network. With sensible designing of routing protocols and application layers of the operating system concerning energy conservation, the lifespan of a sensor network can be greatly increased. The algorithms and protocols should also consider the hardware, and further, they should have the ability to utilize the distinct features of transceivers and microprocessors such that the energy consumed by a sensor node gets reduced. This capability facilitates a customized solution for various sorts of sensor node designing. Various sensor networks utilize distinct sensor nodes, which further leads to collective algorithms in the field of WSNs.

Optimization of energy consumption is considered as the key objective in the study of WSN system architecture, due to the limited energy supply of each node. Clustering of nodes is done to reduce the energy consumption of the network in WSNs, by utilizing the energy efficiently and thereby improving the network lifespan. k-means clustering is one of the numerous clustering algorithms that can enhance cluster formation in WSNs. Though the k-means algorithm enhances the cluster formation, there are drawbacks due to the random selection of the initial centroid and results in the formation of an unbalanced cluster [6]. The selection of the initial centroid is enhanced in our proposed method, and the residual energy of the nodes is considered for balancing the clusters in the cluster head selection, thus resolving the creation of unbalanced clusters in the network. The optimal chain is generated based on the threshold energy of the nodes in the network and thus allows for the maximum use of the nodes that extend the lifespan of the network.

2. Related Works

In this segment, algorithms of hierarchical routing protocols are briefly discussed, and five are used for comparative purposes with the proposed system.

The protocol Low-Energy Adaptive Clustering Hierarchy (LEACH) chooses the CHs at arbitrary and sets up the cluster hierarchy. The benefit in the system refers to a centralized approach named LEACH-C (LEACH-Centralized). The starting assigns of the head of the cluster are done dependent on arbitrary likelihood and the head publicizes to the neighbor hubs to connect as a member node [7]. The member node exchanges the information with the accessible vitality to the cluster head. By utilizing time-division multiplexing access (TDMA), the nodes in the network remain in sleep mode after transmission. Here, LEACH-C is a single-hop data transmission framework, the CHs send the melded information specifically to the sink directly with the nodes' vitality level. The node having vitality over the threshold is considered as the head for the current round, and this data is broadcast to the complete arrange. The head for the cluster is rendered for the span of that round when the hub identity matches the broadcast identity of the same hub. LEACH-C gives a break-even with a dispersion of vitality utilization between the accessible hubs. The protocol is used only for the shortest distance and it is not scalable.

A chain-based routing approach refers to a Chain-Based Hierarchical Routing Protocol (CHIRON) [8, 9] which aims to alleviate the various flaws associated with data propagation delay. In the beginning phase, the network is segregated into different fan-shaped segments. Further, the control message from the BS is delivered to every single node, and each node decides the group to which it belongs. In the second phase, the far-away node present in the BS is triggered to form a group chain within the individual group. With the aid of the greedy algorithm, the closest neighborhood node is linked to the node, which later turns into a new node that initiates in the succeeding linking step. In the third phase, as per the highest level of residual energy associated with the group nodes, the election to choose the leader node is carried out. The node located far away from the base station is initially considered to be the head of the community chain. Later, as the group chain leader, the node identified with the highest residual energy is nominated. In the fourth phase, originally, data is transmitted to the group chain head in each group through the chain. Also, together, the chain heads dispatch their collected data in a leader-by-leader transmission manner to the base station. This mechanism outputs the low-energy dissipation, and clustering overhead will occur in this environment.

The simplest clustering algorithm named k-means clustering belongs to the unsupervised clustering technique. In this clustering algorithm, the given set of nodes is partitioned into k-clusters by calculating the centroid mean value [1, 10]. Initially, the k points are randomly selected to be the centroid of k-clusters, respectively, and the nodes closest to each point are grouped to form the clusters. For the further rounds, the centroid of each cluster is calculated and the nodes closest to it are grouped to form new k-clusters. This process continues until there is no change in the clusters. The sequence of the k-means algorithm is as follows:

(1) Select the k points randomly to be centroids, where k is the number of clusters

(2) Assign each node to its closest centroid using Euclidean distance to form clusters

(3) Compute the new centroid of every cluster

(4) Repeat the second and third steps until there is no variation in the centroid of every cluster

The system supports increasing the lifetime of the network. The performance gain is maximum only when the region of the sensor is reduced.

The system employs a hierarchical clustering algorithm, namely, Energy-Aware Unequal Clustering using the Fuzzy (EAUCF) approach to extend the network lifetime [11]. In this system, cluster head selection depends on the residual energy and also the distances from the BS of the nodes. Tentative cluster heads will be elected in the network according to the random probability approach. The Fuzzy Inference System (FIS) will calculate the competition radius of the tentative cluster heads. The tentative cluster heads will collect information about residual energy from the nearby tentative cluster heads within the competition radius. If more numbers of tentative cluster heads are available within the competition radius, the lower energy nodes will be discarded from the cluster head election. The selection of the CH will not consider the node proximity which rises in the communication cost that perhaps minimizes the network's lifespan. The system employs the balancing of energy among the clusters but has a delay in data transmission.

The Hierarchical Power-Aware Routing (HPAR) communication protocol splits the network into various zones [12]. Every single zone, which is deemed to be an entity, is a collection of location-specific sensor nodes. Therefore, the foremost step performed by HPAR is formatting all the clustered entities (zones). The following step involves the communication scheme to finalize the way information can be hierarchically directed across various zones such that the life expectancy of the network is improved. The second step can be accomplished by routing a message through a path with the highest energy over the leftover paths with the lowest energy. The respective path is known as the max-min path, and the routing scheme offers an approximation algorithm known as the max-min ZPmin algorithm, which first identifies a path associated with the lowest energy consumption with the aid of the Dijkstra algorithm. Later, it identifies another path that can raise the network's residual energy. Further, the HPAR protocol is involved in optimizing the two solution principles. The prime benefit of this communication protocol is its attention toward the node's transmission energy and the lowest energy present in the path. Moreover, it utilizes the zones to monitor a majority of the sensor nodes. In this system, the overhead is associated with the network while estimating the energy.

The system decreases the distance between the transit of received data but does not support a larger network.

The Position-Based Aggregator Node Election Protocol (PANEL), one of the grid-dependent hierarchical algorithms in WSN, utilizes the data of the node's topographical location to ascertain the node's aggregators [13, 14]. Fulfilling the requirements of both synchronous and asynchronous systems is by far the best distinguishing attribute of PANEL. In this algorithm, the network is separated into various topographical clusters. Based on the lower-left bend location of the cluster, nodes determine a reference point in every cluster. The node adjoining the reference point is selected as the cluster head. Data, in this algorithm, can be transmitted in two ways. They are the intracluster and intercluster transmission. If the data is provided to the aggregator of a distinct cluster, then it is called intracluster transmission which has the benefit of transferring the data among the cluster in the course of aggregator election. When the information is transferred between the BSs and far-away clusters, then it is termed as intercluster transmission. The system reduces the energy consumption by employing the optimal path for data transit and the balancing of load between the nodes is less.

In addition, the authors have given and analyzed an overview of the study in the energy efficiency problems and possible solutions for 5G broadband wireless access networks. Some more 5G-related work has been introduced in [15–25]. It generates an energy-saving challenge that combines storage and data transmission costs. Furthermore, strategies for 5G network resource distribution are studied to increase energy efficiency.

3. Proposed Model

The system reduces the dissipation of the individual nodes' energy, thereby maximizing the network's lifetime. For the selection of the cluster head, the system uses the enhanced k-means algorithm and converts it to the chain route when the threshold value is greater than the energy of the nodes in the network [26–31].

3.1. Network Model. The hubs are set arbitrarily in a locale to screen the environment ceaselessly. The total number of nodes is represented as N, where $N=\{n_1, n_2, \ldots, n_n\}$. The assumption has been carried out to design the Optimal-CBR protocol:

(1) There should be a fixed base station above the sensor field

(2) The nodes are stationary in the sensing region

(3) Initially, all the nodes have an energy level

(4) The cluster head performs data aggregation before forwarding the data to the next level

Figure 1 represents the topology representation of the Optimal-CBR when two-thirds of the nodes are exhausted and the remaining energy of those nodes is insufficient to form a cluster for the clustering phase. Hence, the remaining node forwards the unit of data to the BS through chain formation. The communication is based on the close ideal way approach.

3.2. Energy Model. The data transceiver energy includes the energy of the device circuitry and the volume of data transmission and reception. The vitality is required for transmission circuitry and the information parcels are transmitted. Essentially, indeed the energy is required for getting bargains with the same variables. The vitality required to transmit a unit of information is

$$\begin{aligned} E_{\mathrm{Tr}}(p,d) &= E_{\mathrm{ele}}(p) + E_{\mathrm{amp}}(p,d), \\ \Longrightarrow E_{\mathrm{Tr}}(p,d) &= \left\{pE_{\mathrm{ele}} + p\varepsilon_{\mathrm{fs}}d^2, \quad d < d_0\right\}, \\ \Longrightarrow E_{\mathrm{Tr}}(p,d) &= \left\{pE_{\mathrm{ele}} + p\varepsilon_{\mathrm{mp}}d^4, \quad d \geq d_0\right\}. \end{aligned} \tag{1}$$

The vitality required to get a unit of information is

$$E_{\mathrm{Rr}}(p) = E_{\mathrm{ele}}(p) = pE_{\mathrm{ele}}, \tag{2}$$

where $E_{\mathrm{ele}}(p)$ is the vitality misfortune per bit of transceiver circuitry. Based on the transmission extend, free space (d^2) or multipath (d^4) propagation is used. $E_{\mathrm{amp}}(p,d)$ is the enhancement vitality with distance, d.

3.3. Operation of the Optimal-CBR Algorithm. The Optimal-CBR approach uses the k-means algorithm to form clusters and chooses the CH for each cluster based on the Euclidean distance and nodes energy. The hard threshold broadcasted by the CH to the respective cluster members is the attribute value above which the node is permitted to transmit the data to the CH. Once two-thirds of the nodes are dead and the residual energy of the remaining nodes is insufficient for clustering, the nodes use the greedy approach to form a chain-like multihop routing until the BS is reached [32–42]. The Optimal-CBR algorithm is divided into two phases, namely,

(1) Clustering phase

(2) Chaining phase

3.3.1. Clustering Phase. In this phase, the Optimal-CBR implements the k-means algorithm in which the network nodes are arranged into k-*clusters*. Initially, the k nodes are arbitrarily chosen as CHs in the network. The remaining node figures the closest CH using the Euclidean distance forming k-clusters in the initial round. For the further rounds, the centroid of each cluster is calculated. The centroid of the i^{th} node, C_i, is given as

$$C_i = \left(\frac{1}{|P|} \sum_{i \in P} x_i, \frac{1}{|P|} \sum_{i \in P} y_i \right), \tag{3}$$

where "p" represents the cluster member, and x and y represent the coordinates of the nodes.

The framework not only uses the distance between the centroid-based nodes but also considers the energy of the nodes for the network's collection of CHs. Here, E_i is the nodes' remaining energy. The selection of CHs, therefore, is the maximum of the remaining energy and the minimum distance between the nodes and is given as follows:

$$\mathrm{CH}_{\mathrm{select}} = \max_i \{ \min_i \{ f(C_i, E_i) \} \}. \tag{4}$$

There is a cluster sample with a randomly selected initial CH and a centroid calculated using the centroid formula. The centroid is a virtual node at the cluster core. The node adjacent to the centroid is selected as the tentative CH. An ID is allocated to every node considering its distance from the centroid. The nodes closer to the centroid have a smaller ID than that of the nodes away from the centroid. The node comprising the next ID number is chosen as the CH if its energy is greater than the threshold. If the value is lesser than the threshold, the current CH sends the energy of the cluster member to the base station before it quits the session. BS looks at another node for the selection of CH based on the energy. If none of the nodes is above the energy of the threshold, the system forms a chain for data forwarding. The threshold energy, E_{TD}, is calculated as

$$E_{\mathrm{TD}} = lE_{\mathrm{el}} \left(\frac{N}{K} - 1 \right) + lE_{\mathrm{DA}} \left(\frac{N}{K} \right) + lE_{\mathrm{el}} + l\varepsilon_{\mathrm{fs}} d^2, \tag{5}$$

where K is the number of clusters in the sensor region. The chosen CH broadcasts the signal, and the nodes closer to it will transmit data to the respective CHs, according to the time slot provided by the CH.

The suggested schema employs "multihop" data transfer between the nodes and the sink. On each round, the sensed data and remaining energy of every node are collected and transferred to the sink via the CHs. The entire utilization of vitality for the single cluster incorporates the utilization of vitality of the member node and the vitality utilization of the CH within the given cluster. The overall vitality devoured by the clusters is calculated as

$$E_{\mathrm{clust}} = (m_n - 1) E_{\mathrm{CM}} + E_{\mathrm{CH}}, \tag{6}$$

where m_n is the number of member nodes, E_{CH} is the energy of the cluster head, and E_{CM} is the energy of the cluster member.

Since "N" speaks to the node count and "M" demonstrates the sensor locale, the choice of optimal clusters within the framework is given utilizing the condition

$$K_{\mathrm{opt}} = \frac{M\sqrt{N}}{\sqrt{2\pi}} \sqrt{\frac{\varepsilon_{\mathrm{fs}}}{\varepsilon_{\mathrm{mp}}} . \frac{1}{d_{\mathrm{BS}}^2}}. \tag{7}$$

The vitality dissemination of the organization is the item of the utilization of vitality of one cluster and the total number of clusters, K_{opt}, displayed within the detecting locale. The overall vitality utilization by the arrange is given as

$$E_{\mathrm{network}} = K_{\mathrm{opt}} \times E_{\mathrm{clust}}. \tag{8}$$

The assumption is being taken care of for the even distribution of hubs within the sensor locale. The difference from the nodes to the CH is given as

$$\begin{aligned} E[d] &= \int\int \sqrt{(x - x_{\mathrm{CH}})^2 + (y - y_{\mathrm{CH}})^2} \rho(x, y) \mathrm{d}x \mathrm{d}y, \\ E[d] &= \int\int \sqrt{x^2 + y^2} \rho(x, y) \mathrm{d}x \mathrm{d}y, \\ &= \int\int r^2 \rho(r, \theta) \mathrm{d}r d\theta. \end{aligned} \tag{9}$$

Here, ρ (r, θ) gets to be steady, when the sensors are conveyed consistently within the environment. The selection of optimal route is done using

$$O_R = \max\{ E_i(\mathrm{CH}) \}, \tag{10}$$

where O_R is the optimal route in a network and E_i is the residual energy of the node.

3.3.2. Chaining Phase. When a node's residual energy is lower (i.e., almost drained), a beacon message about the network is sent by the BS to the alive nodes. If "s" denotes the absolute number of dead nodes, then the total network nodes that are alive is given by $m = N - s$. When "s" is more than two-thirds (i.e., two-thirds of the nodes are dead), the BS computes the route to form a chain. In this phase, the BS creates the path through multihop chain routing, using the greedy approach. The nodes transmit the information to the BS through the chain path. The selection of the best path is done using

$$B_{\mathrm{path}} = \{ n_i, \quad i \in m \}. \tag{11}$$

Data aggregating is done by the nodes in a chain path:

$$f(m_0, \ldots m_n) = r \sum_{i=0}^{n} m_i + c. \tag{12}$$

Here, considering a chain with node "m_0" being a source, and the member $i = 0, 1, \ldots, n$, the aggregation function of the node before data transmission is given as $f(m_0, \ldots m_n)$, where c corresponds to the overhead of the aggregation, while $r < 1$ is the compression ratio.

The process of chaining is dynamic based on the greedy approach concerning the residual energy of the live nodes in the network. An overview of the Optimal-CBR protocol is given in Figure 2.

4. Statistical Analysis

In this segment, the proposed model is compared with a few existing clustering algorithms like LEACH-C, CHIRON, and k-means. Table 1 displays the parameters of the simulation. The performance of the proposed protocol is compared with that of

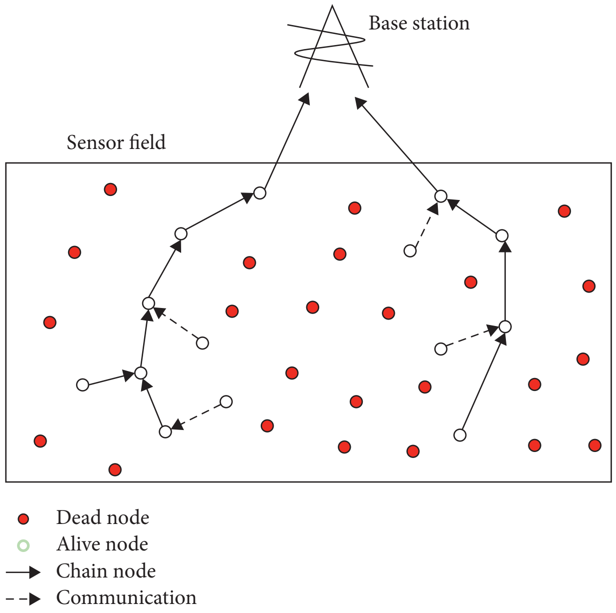


Figure 1: Topology of Optimal-CBR.

the existing technique, and the following metrics are considered:

(1) Average energy with several rounds
(2) Existence of living nodes in each round
(3) First node dead comparison for validation
(4) Energy dissipation of the CH
(5) Packet delivery ratio
(6) End-to-end delay analogy

The residual (leftover) energy of all the nodes assessed within the proposed conspire. The sum of the network's leftover vitality of Optimal-CBR is higher when compared to that of the existing framework like LEACH-C, CHIRON, and the *k*-means protocol. Figure 3 shows the entirety of the leftover vitality amid each circular and the advancement accomplished by Optimal-CBR. The improvements achieved by Optimal-CBR over *k*-means, CHIRON, and LEACH-C are 19.24%, 34.57%, and 71.46% of energy, respectively.

The Optimal-CBR model has a higher count of alive nodes per iteration when compared with the LEACH-C, CHIRON, and *k*-means protocols. The count of the living nodes in each iteration is represented in Figure 4. After completion of 1200 thousand rounds, Optimal-CBR, CHIRON, *k*-means, and LEACH-C have 258, 173, 142, and 76 alive nodes, respectively. In *k*-means, CHIRON, and LEACH-C, the energy value of the nodes is less than the threshold; the nodes communicate to the base station directly, and this leads to more energy depletion. But in the Optimal-CBR protocol, the system forms a chain route through the greedy approach to transmitting the data to the base station, which consumes less energy and increases the live nodes in the network comparatively.

As the iteration progresses, the count of the alive nodes is monitored to assess the efficiency in due consideration of the network's lifespan. As presented in Figure 5, the current model has a higher communication round than that of LEACH-C, CHIRON, and *k*-means. The rounds of communication for the first node dead for Optimal-CBR are 578 which is greater than the other algorithms because of the user-defined data threshold present on every node. This has reduced the energy used by the CH to aggregate the data, which, in turn, has reduced the number of times reclustering has to be done.

Figure 6 illustrates the dissipation of energy of the CHs with each algorithm. In comparison with the other protocols, the Optimal-CBR model is far more effective as it has the lowest consumption of energy with an approximate value of 0.12 J. Also, its curve is more refined than the others concerning each iteration, the reason being the short and balanced distances between the sensor nodes and the CHs.

The fraction of the number of packets collected by the sink to that transferred by the sensor nodes is termed as a packet delivery ratio. The higher the ratio, the larger the "number" of packets delivered or accumulated in the sink. From Figure 7, it is apparent that Optimal-CBR has a better packet delivery ratio than the *k*-means, CHIRON, and LEACH-C protocols. The chances of failure during the data delivery process are very high in the LEACH-C protocol, because of the direct transmission mode leading to the higher energy utilization of nodes. As a result, more time is drained in the process of data transfer, and a significant volume of packets is stalled.

The occurrence of a delay in the Optimal-CBR protocol is lower than that in LEACH-C, K-Mean, and CHIRON. Figure 8 compares the protocols, Optimal-CBR, LEACH-C,

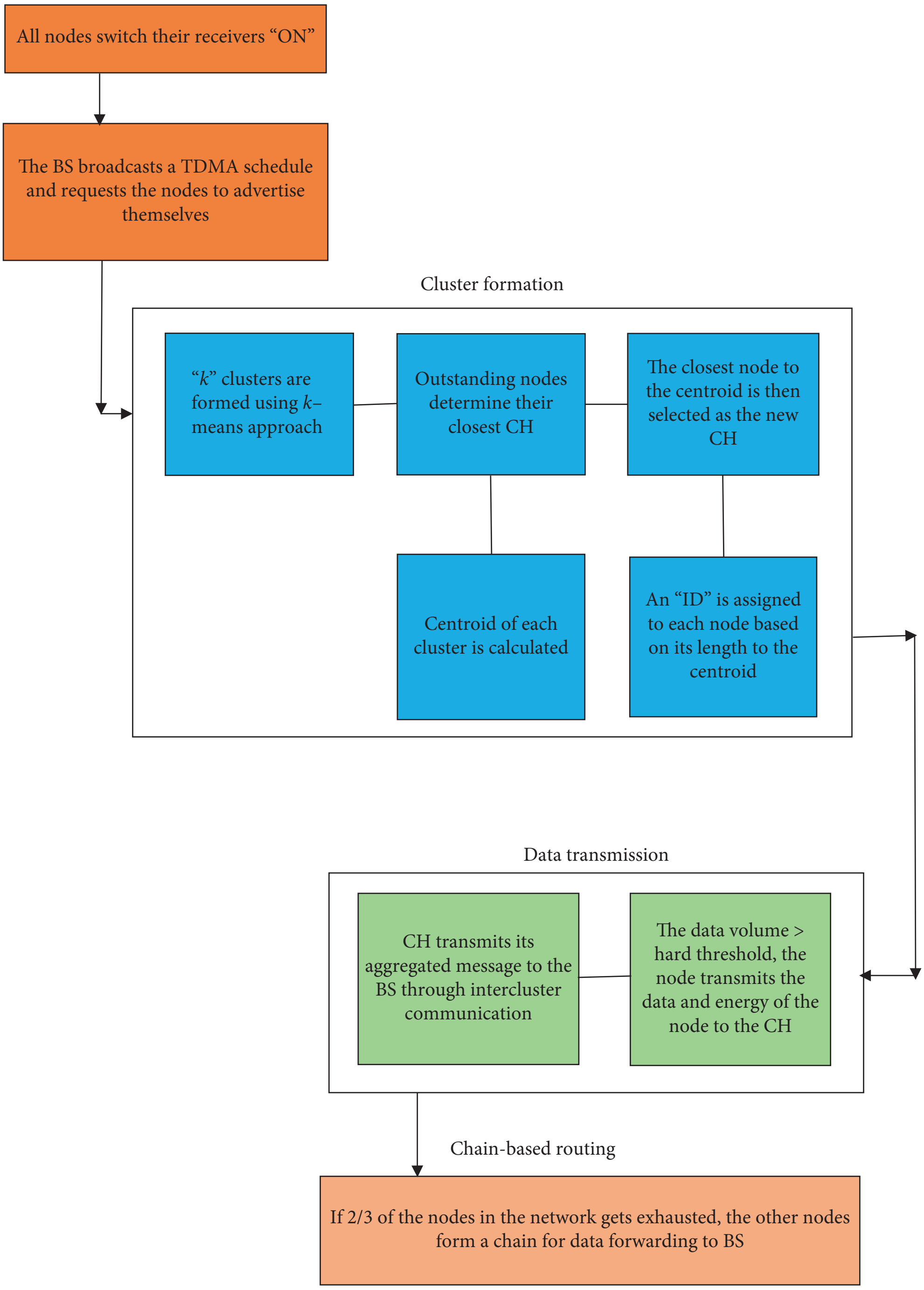

All nodes switch their receivers "ON"
The BS broadcasts a TDMA schedule and requests the nodes to advertise themselves
Cluster formation
"k" clusters are formed using k–means approach
Outstanding nodes determine their closest CH
The closest node to the centroid is then selected as the new CH
Centroid of each cluster is calculated
An "ID" is assigned to each node based on its length to the centroid
Data transmission
CH transmits its aggregated message to the BS through intercluster communication
The data volume > hard threshold, the node transmits the data and energy of the node to the CH
Chain-based routing
If 2/3 of the nodes in the network gets exhausted, the other nodes form a chain for data forwarding to BS

Figure 2: Overview of the Optimal-CBR protocol.

Table 1: Simulation parameters.

No.	Parameters	Specification
1	Size of the network	$500 \times 500\,\mathrm{m}^2$
2	Sensor count	500
3	The initial power of each node	1 J
4	Location of the base station	$X = 250\,\mathrm{m}$ $Y = 550\,\mathrm{m}$
5	Data packet size	300 bytes
6	Transmitter circuitry dissipation	50 nJ/bit
7	Energy for data aggregation, E_{DA}	5 nJ/bit/signal
8	Break parameter	Two third of nodes dead
9	No. of rounds taken for simulation	1200 rounds
10	Minimum threshold energy	10^{-4} J

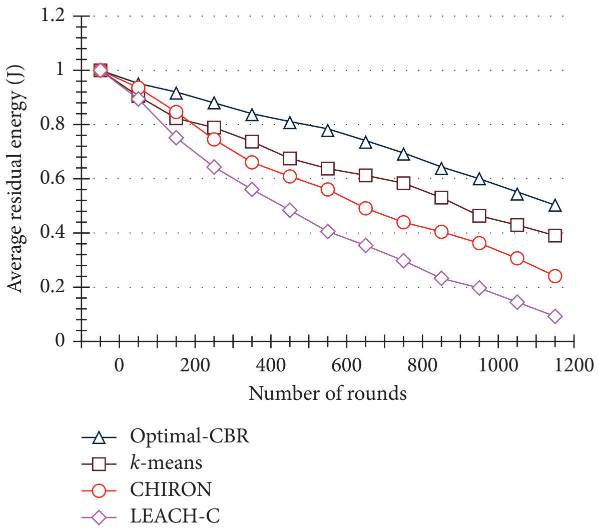


Figure 3: Average residual energy by the number of rounds.

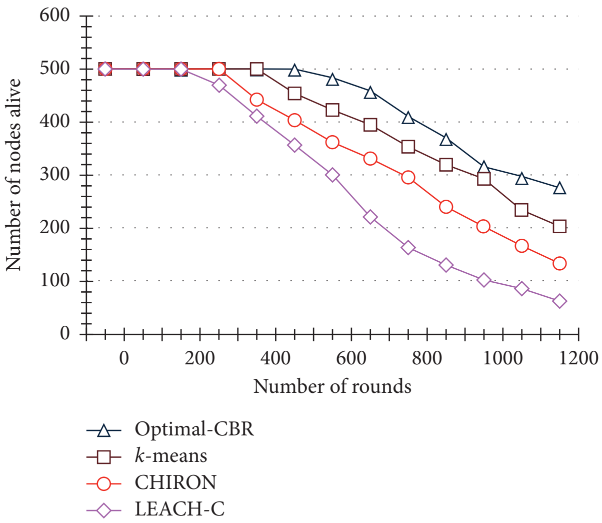


Figure 4: Existence of nodes under each round.

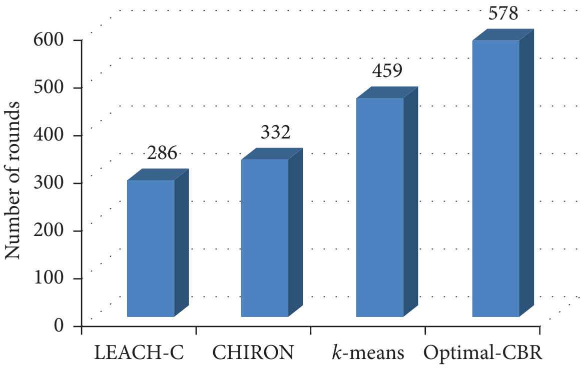


Figure 5: Comparison of the first dead node for validation.

k-means, and CHIRON in terms of delay in network data transfer. The LEACH-C protocol has a higher delay when compared to *k*-means, CHIRON, and Optimal-CBR protocols. Because of the direct transmission of data, the nodes utilize a large amount of energy in the LEACH-C protocol. In LEACH-C, a particular CH is chosen, and that

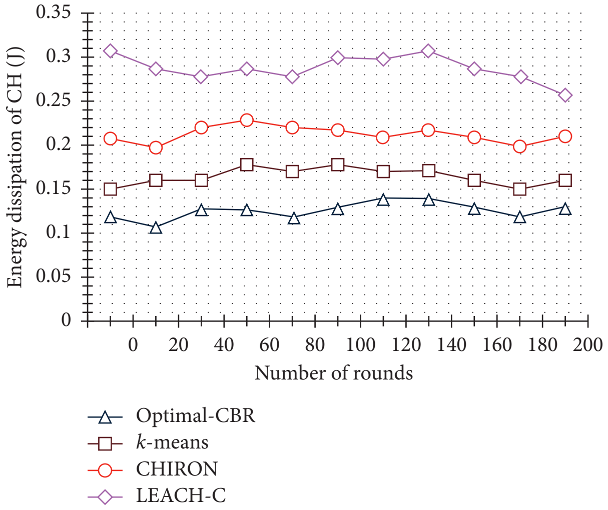


Figure 6: Energy dissipation of the cluster head.

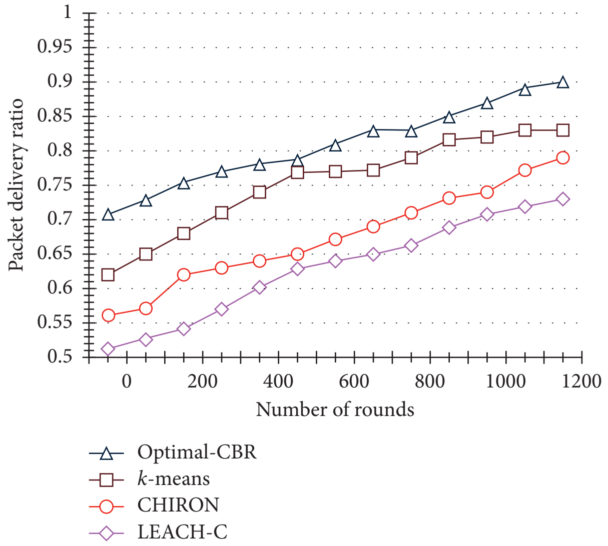


Figure 7: Packet delivery ratio.

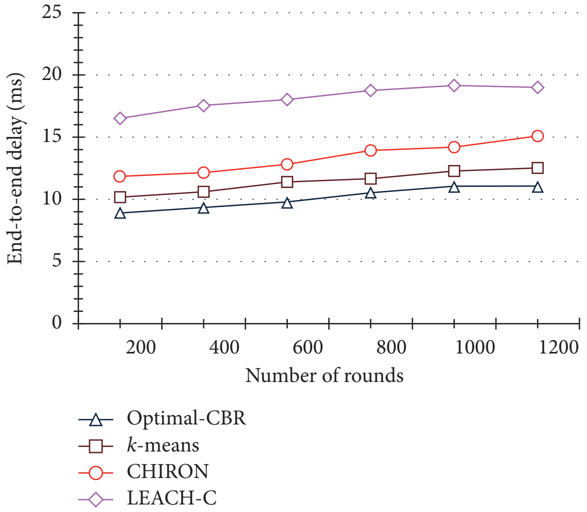


Figure 8: Comparison of end-to-end delay.

Table 2: Summary of the analysis.

Protocol	Average residual energy of the nodes at 1200 rounds	Number of live nodes at 1000 rounds	Energy dissipation of the CH at 200 rounds (J)	Packet delivery ration	End-to-end delay (ms)
Optimal-CBR	0.49	278	0.13	0.9	11
K-means	0.38	203	0.16	0.83	12.51
CHIRON	0.23	132	0.21	0.79	15.1
LEACH-C	0.09	63	0.26	0.73	19.1

cluster head is permitted to carry out data collection and data delivery. The delay occurs due to the number of data transmissions to the sink and reclustering. In *k*-means and CHIRON, to continue the operation, the election of the CH based on neighbor node proximity, multihop data transmission, and reclustering will occur, and this process generates a delay in the sensing field. The summary of the comparison is given in Table 2.

5. Conclusion

The Optimal Cluster-Based Routing (Optimal-CBR) protocol for organizing Internet of Things-based Wireless Sensor Networks employs the *k*-means algorithm to construct clusters. When the residual energy of the CH is comparatively lesser than the threshold energy, a chaining phase is used to create a routing path. The cluster head is selected considering the Euclidean distance and the node's residual energy. The outcomes of the simulation graphs signify that Optimal-CBR has lower vitality scattering within the CH, and the sum of remaining vitality devoured is moo as compared with the *k*-means, CHIRON, and LEACH-C protocols, thus prolonging the node's lifespan. Hence, the current schema incorporates uniform energy distribution in all the network nodes and maximizes the transmission rounds that perhaps reduces the consumption of the energy under the 5G environment. Thus, the proposed system enhances the efficiency of the sensor nodes by minimizing the energy consumption of the nodes, which prolongs the lifetime of the network. The system does not focus on the security aspect when sharing the data through a multihop routing approach. The limitation has been overcome by employing the security mechanism in future work.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request (ihra@kunsan.ac.kr).

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korean Government (MSIT) under Project 2021R1A2C2014333.

References

[1] A. Ray and D. De, "Energy efficient clustering protocol based on *K*-means (EECPK-means)-midpoint algorithm for enhanced network lifetime in wireless sensor network," *IET Wireless Sensor Systems*, vol. 6, no. 6, pp. 181–191, 2016.

[2] I. F. Akyildiz, W. Weilian Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Communications Magazine*, vol. 40, no. 8, pp. 102–114, 2002.

[3] C. Jothikumar and R. Venkatraman, "A review of hierarchical routing protocol for wireless sensor network," *Indian Journal of Science and Technology*, vol. 9, no. 32, pp. 1–10, 2016.

[4] H. Lim and C. Kim, "Flooding in wireless ad hoc networks," *Computer Communications*, vol. 24, no. 3-4, pp. 353–363, 2001.

[5] V. S. Patel and C. R. Parekh, "Survey on sensor protocol for information via negotiation (spin) protocol," *IJRET: International Journal of Research in Engineering and Technology*, vol. 3, no. 3, pp. 208–211, 2014.

[6] G. Y. Park, H. Kim, H. W. Jeong, and H. Y. Youn, "A novel cluster head selection method based on *K*-means algorithm for energy efficient wireless sensor network," in *Proceedings of the 2013 27th International Conference on Advanced Information Networking and Applications workshops*, pp. 910–915, Barcelona, Spain, March 2013.

[7] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, pp. 660–670, 2002.

[8] K. H. Chen, J. M. Huang, and C. C. Hsiao, "CHIRON: an energy-efficient chain-based hierarchical routing protocol in wireless sensor networks," in *Proceedings of the 2009 Wireless Telecommunications Symposium*, pp. 1–5, Prague, Czech Republic, April 2009.

[9] S. Chatterjee and M. Singh, "A centralized energy-efficient routing protocol for wireless sensor networks," *International Journal of Advanced Networking and Applications*, vol. 3, no. 5, p. 12, 2012.

[10] P. Sasikumar and S. Khara, "*K*-means clustering in wireless sensor networks," in *Proceedings of the 2012 Fourth International Conference on Computational Intelligence and Communication Networks*, pp. 140–144, Mathura, India, November 2012.

[11] H. Bagci and A. Yazici, "An energy aware fuzzy approach to unequal clustering in wireless sensor networks," *Applied Soft Computing*, vol. 13, no. 4, pp. 1741–1749, 2013.

[12] Q. Li, A. A. Javed, and D. Rus, "Hierarchical power-aware routing in sensor networks," in *Proceedings of the DIMACS Workshop on Pervasive Networking*, pp. 102–122, New Brunswick, NJ, USA, May 2001.

[13] L. Buttyan and P. Schaffer, "Panel: position-based aggregator node election in wireless sensor networks," in *Proceedings of the IEEE International Conference on Mobile Adhoc and Sensor Systems*, pp. 1–9, Pisa, Italy, October 2007.

[14] L. Buttyan and P. Schaffer, "Position-based aggregator node election in wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2010, pp. 1–15, Article ID 679205, 2010.

[15] X. Zhao, W. Cheng, H. Zhu, C. Ge, G. Zhou, and Z. Fu, "A high gain, noise cancelling 2515–4900 MHz CMOS LNA for China mobile 5G communication application," *Computers, Materials & Continua*, vol. 64, no. 2, pp. 1139–1151, 2020.

[16] D. Ali Sehrai, F. Muhammad, S. Hassan Kiani, Z. Haq Abbas, M. Tufail, and S. Kim, "Gain-enhanced metamaterial based antenna for 5G communication standards," *Computers, Materials & Continua*, vol. 64, no. 3, pp. 1587–1599, 2020.

[17] D.-Y. Kim and S. Kim, "Network-aided intelligent traffic steering in 5G mobile networks," *Computers, Materials & Continua*, vol. 65, no. 1, pp. 243–261, 2020.

[18] L. Jianzhong, M. Dexiang, D. Dong, I. Khan, and P. Uthansakul, "Proportional fairness-based power allocation algorithm for downlink noma 5G wireless networks," *Computers, Materials & Continua*, vol. 65, no. 2, pp. 1571–1590, 2020.

[19] J. Khan, D. Ali Sehrai, M. Ahmad Khan et al., "Design and performance comparison of rotated y-shaped antenna using different metamaterial surfaces for 5G mobile devices," *Computers, Materials & Continua*, vol. 60, no. 2, pp. 409–420, 2019.

[20] A. Badshah, A. Ghani, M. Ahsan Qureshi, and S. Shamshirband, "Smart security framework for educational institutions using internet of things (IoT)," *Computers, Materials & Continua*, vol. 61, no. 1, pp. 81–101, 2019.

[21] S. K. Kim, M. Köppen, A. K. Bashir, and Y. Jin, "Advanced ICT and IOT technologies for the fourth industrial revolution," *Intelligent Automation & Soft Computing*, vol. 26, no. 1, pp. 83–85, 2020.

[22] C. Zhao, T. Wang, and A. Yang, "A heterogeneous virtual machines resource allocation scheme in slices architecture of 5G edge datacenter," *Computers, Materials & Continua*, vol. 61, no. 1, pp. 423–437, 2019.

[23] B. Jo, M. Jalil Piran, D. Lee, and D. Young Suh, "Efficient computation offloading in mobile cloud computing for video streaming over 5G," *Computers, Materials & Continua*, vol. 61, no. 2, pp. 439–463, 2019.

[24] S. K. Kim, M. Köppen, A. K. Bashir, and Y. Jin, "Advanced ICT and IOT technologies for the fourth industrial revolution," *Intelligent Automation & Soft Computing*, vol. 26, no. 1, pp. 83–85, 2020.

[25] Z. Liu, X. Qiu, S. Zhang, S. Deng, and G. Liu, "Service scheduling based on edge computing for power distribution IoT," *Computers, Materials & Continua*, vol. 62, no. 3, pp. 1351–1364, 2020.

[26] Y. Wu, S. Fahmy, and N. B. Shroff, "Energy efficient sleep/wake scheduling for multi-hop sensor networks: non-convexity and approximation algorithm," in *Proceedings of the IEEE INFOCOM 2007-26th IEEE International Conference on Computer Communications*, pp. 1568–1576, Anchorage, AK, USA, May 2007.

[27] C. J. Venkataraman, "A threshold sensitive dynamic cluster head for energy optimization in wireless sensor networks," *International Journal of Pure and Applied Mathematics*, vol. 115, no. 6, pp. 617–622, 2017.

[28] S. Lindsey, C. Raghavendra, and K. M. Sivalingam, "Data gathering algorithms in sensor networks using energy metrics," *IEEE Transactions on Parallel and Distributed Systems*, vol. 13, no. 9, pp. 924–935, 2002.

[29] Z. Wang, H. Ding, B. Li, L. Bao, and Z. Yang, "An energy efficient routing protocol based on improved artificial bee colony algorithm for wireless sensor networks," *IEEE Access*, vol. 8, pp. 133577–133596, 2020.

[30] C. Jothikumar, R. Venkataraman, T. Sai Raj, J. Selvin Paul Peter, and T. Y. J. Nagamalleswari, "EUCOR: an efficient unequal clustering and optimal routing in wireless sensor networks for energy conservation," *Journal of Intelligent & Fuzzy Systems*, vol. 40, no. 5, pp. 9187–9195, 2021.

[31] J. Qadir, U. Ullah, B. Sainz-De-Abajo, B. G. Zapirain, G. Marques, and I. de la Torre Diez, "Energy-aware and reliability-based localization-free cooperative acoustic wireless sensor networks," *IEEE Access*, vol. 8, pp. 121366–121384, 2020.

[32] S. Lindsey and C. S. Raghavendra, "PEGASIS: power-efficient gathering in sensor information systems," in *Proceedings of the IEEE Aerospace Conference*, Big Sky, MT, USA, March 2002.

[33] J. J. Lotf, M. N. Bonab, and S. Khorsandi, "A novel cluster-based routing protocol with extending lifetime for wireless sensor networks," in *Proceedings of the 2008 5th IFIP International Conference on Wireless and Optical Communications Networks (WOCN'08)*, pp. 1–5, Surabaya, Indonesia, May 2008.

[34] W. R. Heinzelman, J. Kulik, and H. Balakrishnan, "Adaptive protocols for information dissemination in wireless sensor networks," in *Proceedings of the 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking*, pp. 174–185, Seattle, WA, USA, August 1999.

[35] M. C. Thein and T. Thein, "An energy efficient cluster-head selection for wireless sensor networks," in *Proceedings of the 2010 International Conference on Intelligent Systems, Modelling and Simulation*, pp. 287–291, Liverpool, UK, January 2010.

[36] S. Rani, J. Malhotra, and R. Talwar, "Energy efficient chain based cooperative routing protocol for WSN," *Applied Soft Computing*, vol. 35, pp. 386–397, 2015.

[37] P. Yarde, S. Srivastava, and K. Garg, "Adaptive immune-inspired energy-efficient and high coverage cross-layer routing protocol for wireless sensor networks," *IET Communications*, vol. 14, no. 15, pp. 2592–2600, 2020.

[38] C. Xu, Z. Xiong, G. Zhao, and S. Yu, "An energy-efficient region source routing protocol for lifetime maximization in WSN," *IEEE Access*, vol. 7, pp. 135277–135289, 2019.

[39] M. Adil, R. Khan, J. Ali, B.-H. Roh, Q. T. H. Ta, and M. A. Almaiah, "An energy proficient load balancing routing scheme for wireless sensor networks to maximize their lifespan in an operational environment," *IEEE Access*, vol. 8, pp. 163209–163224, 2020.

[40] A. R. Basha, "Energy efficient aggregation technique-based realisable secure aware routing protocol for wireless sensor network," *IET Wireless Sensor Systems*, vol. 10, no. 4, pp. 166–174, 2020.

[41] M. H. Abidi, H. Alkhalefah, K. Moiduddin et al., "Optimal 5G network slicing using machine learning and deep learning concepts," *Computer Standards & Interfaces*, vol. 76, p. 103518, 2021.

[42] C. Iwendi, P. K. Maddikunta, T. R. Gadekallu, K. Lakshmanna, A. K. Bashir, and M. J. Piran, "A metaheuristic optimization approach for energy efficiency in the IoT networks," *Software: Practice and Experience*, pp. 1–14, 2020.