Security and Communication Networks

Data Security and Privacy for Fog/ Edge Computing-Based IoT 2022

Lead Guest Editor: Jie Cui
Guest Editors: Antonio Liotta, Ke Gu, and Lu Liu

Data Security and Privacy for Fog/Edge Computing-Based IoT 2022

Security and Communication Networks

Data Security and Privacy for Fog/Edge Computing-Based IoT 2022

Lead Guest Editor: Jie Cui
Guest Editors: Antonio Liotta, Ke Gu, and Lu Liu

Contents

Hindawi
Security and Communication Networks
Volume 2023, Article ID 9143563, 19 pages
https://doi.org/10.1155/2023/9143563

Research Article

SAIFC: A Secure Authentication Scheme for IOV Based on Fog-Cloud Federation

Yashar Salami, Vahid Khajehvand, and Esmaeil Zeinali

Department of Computer and Information Technology Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran

Correspondence should be addressed to Vahid Khajehvand; vahidkhajehvand@gmail.com

Received 1 July 2022; Revised 19 October 2022; Accepted 27 January 2023; Published 24 April 2023

Academic Editor: Ke Gu

The Internet of Things allows vehicles to communicate with their surroundings and send various traffic and road conditions to other vehicles, making driving better. Data received from other vehicles sometimes need to be processed. In this state, the data should be sent to the Roadside unit for processing and fog if necessary. The source and destination must know each other's identities to send and receive information, and various attacks can threaten source and destination authentication. The paper presents secure authentication scheme based on fog-cloud for the Internet of Vehicles. Formal and informal security analyses verify that the SAIFC demonstrates resistance to famous attacks. The SAIFC is compared with another scheme regarding security features, computing, and communication costs. The SAIFC simulated with the NS3 tool and compared several routing protocols in packet loss, packet delivery, end-to-end delay, throughput, and MAC/PHY overhead.

1. Introduction

Today, intelligent transportation systems in many countries are increasingly expanding [1]. That has improved road safety, traffic monitoring [2], automatic driving [3], and passenger comfort, one of the main goals of intelligent transportation systems [4]. Furthermore, IOT is an emerging technology connected to the physical and digital worlds [5]. The Internet of Things has revolutionized the relationship between objects and humans. The IOV [6] is one of the most active research areas created by combining the Internet of Things and the Vanets. IOV has solved many traffic problems, thus leading to passenger safety and facilitating the driving experience [7, 8]. There are several ways to transfer data between vehicles, one of which is the HTTP protocol. Almost all devices connecting to the Internet use HTTP protocol [9]. HTTP can connect the vehicle to vehicle and vehicle to RSU. Figure 1 shows the vehicle communication with the HTTP protocol.

Before sending information, the sender and receiver must authenticate each other to trust the data's integrity.

Hasrouny et al. [10], in 2015, presented a group-based authentication from vehicle to vehicle. Their method did not support AKE in the fog based. In 2017, Yang et al. [11] proposed a AKE for the IOV environment. However, the protocol is based on ECC but cannot perform mutual authentication in the environment of fog based on HTTP. Protocol ensuring privacy and authentication for a vehicle to vehicle resource sharing was presented in 2017 by Benarous and Kadri [12]. Nevertheless, this method is vulnerable to rainbow attacks, and the fog environment does not support key exchange and mutual authentication. The design of authentication protocol for the automotive system is based on wireless sensor network by Mohit et al. [13] in 2017. This protocol supports fog and mutual authentication, but key exchange and ECC are not supported, and in terms of security, it is weak to RTA. In 2017, lightweight AKE for IOV was presented by Ying and Nayak [14]. This method is inefficient in terms of security, does not use the ECC method in AKE, and does not support the fog environment. An efficient anonymous authentication scheme for the IOV was presented by Liu et al. [15] in 2018. Although this method has used ECC, it is still weak against RTA. Structurally, it does not support fog environment and HTTP. In 2019, Lim and Tuladhar [16] presented a Lidar information-based dynamic V2V authentication. This scheme does not use

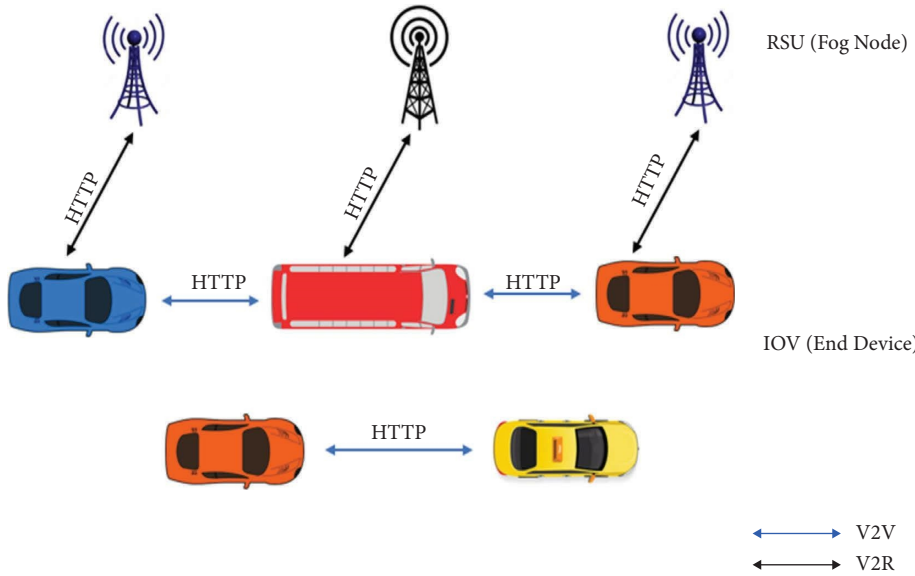


Figure 1: Communication through HTTP protocol.

ECC, is vulnerable to various attacks, and does not support AKE in the HTTP base fog environment.

A secure authentication protocol for the IOV presented by Chen et al. [17] in 2019. This method did not support AKE in the fog environment and was vulnerable to the RTA. Vasudev et al. [18] presented lightweight authentication protocol for IOV in 2020. Kalra and Sood [19] 2015 and Kumari et al. [20] 2017 presented an authentication scheme for IOT and cloud. The schemes are based on HTTP and support mutual authentication; however, they do not support key exchange in the fog and are vulnerable to RTA. In 2019, Wazid et al. [21] presented an AKM protocol in fog-based IOV. Although this method could support mutual authentication and key exchange in a fog environment, it is weak to RTA. Table1 shows a comparison of the related works to the SAIFC.

1.1. Our Contribution

(i) First, we examine the security problem of Kumari et al. Then, we use the AVISPA to perform a safety analysis.

(ii) We testbed the scheme Kumari et al. using the Arduino board.

(iii) We propose an HTTP-based authentication scheme for IOV-fog-based, which sends data for authentication via a cookie.

(iv) We have SAIFC a secure authentication for the IOV environment, which is resistant to various attacks.

(v) We have SAIFC informal and formal AVISPA tools for security analysis. We have also compared the SAIFC scheme with other protocols in terms of security features.

(vi) We compare the SAIFC scheme's computational and communication costs with other protocols.

(vii) We have implemented the SAIFC scheme with NS3 simulation to obtain the most appropriate routing protocol for the SAIFC scheme.

1.2. Structure of the Paper. The structure of this paper is as follows: Section 2 explains the problems of the scheme of Kumari et al. and then provides the security analysis and testbed results. Section 3 introduces the problem statement and network model. Section 4 presents the SAIFC scheme, and in Section 5, Security Analysis and Result are discussed. Performance analysis and comparison of security features do provide in Section 6. Section 7 presents the simulation of the SAIFC scheme with the NS3 tool and analysis results. Finally, Section 8 concludes this work.

2. The Security Problem of Kumari

This section discusses the RTA on the Kumari scheme.

2.1. RTA. RTA is a precalculated table that is used to break the hash. RTA is sometimes used to recover passwords or credit card numbers. This attack has tables of specified length and contains a limited number of components [22]. To study the working method, you can visit Rainbow Crack.

2.2. Notations. Table 2 shows the notations used in the paper.

2.3. RTA in the Registration Phase

Step 1: Edi merges the Idi and Pwi values in the registration phase, hashes them into *li*, and sends them to the CS. Attackers can break the hash value generated by Edi at this point using a rainbow attack. After this step, it can read the data sent by Edi to the CS or change this

Table 1: Comparison of related works with SAIFC scheme.

Related works	Fog based	RTA	Mutual authentication	Key exchange	ECC based	HTTP based
[10]	No	No	No	No	No	No
[11]	No	No	No	Yes	Yes	No
[12]	No	No	Yes	No	No	No
[13]	Yes	No	Yes	No	No	No
[14]	No	No	No	Yes	No	No
[15]	No	No	Yes	No	Yes	No
[16]	No	No	No	No	No	No
[17]	No	No	No	No	No	No
[18]	No	No	Yes	No	No	No
[19]	No	No	Yes	No	Yes	Yes
[20]	No	No	Yes	No	Yes	Yes
[21]	Yes	No	Yes	No	Yes	No
SAIFC	Yes	Yes	Yes	Yes	Yes	Yes

Table 2: Used notations in this paper.

Notations	Description
Vdi	Vehicle
Idi	Identity of Vdi
Pwi	Password of Vdi
CS	Cloud server
R	RSU (fog node)
F	Fog
Idcs	Identity of CS
IdR	Identity of R
XR	The secret key of R is based on ECC
Zp	Finite field group
p	Prime number of the order $>2^{160}$
r1, r2	Random numbers generated for ECC
rs	Random number generated by R
G	Generator point of a large order n
Ck	Cookie information
Et	Cookie expiration time
$h(\cdot)$	HF
$\oplus$	XOR
$\|\|$	Concatenation
ΔT	Expiration time
TV	Timestamp Vdi
TR	Timestamp R
TF	Timestamp fog

information and send it back to the CS. The steps are as follows:

Edi $\longrightarrow$ CS: $\{Ii\}$
Attacker $\longrightarrow$ CS: $\{Ii\}$
Rainbow Crack{Ii}
Resend {Ii} for CS

Step 2: The CS then performs a series of calculations to respond to Edi's registration and sends the hashed PIdi and Ck′ to Eddie to continue. Attackers can break the hash value generated by CS at this point using a rainbow attack. After this step, it can read the data sent by CS to the Edi or change it and send it back to the Edi. The steps are as follows. Figure 2 shows the stages of a RTA in the registration phase.

CS $\longrightarrow$ Edi: {PIdi, Ck′}
Attacker $\longrightarrow$ Edi: {PIdi, Ck′}
Rainbow Crack {PIdi, Ck′}
Resend {PIdi, Ck′} for Edi

2.4. RTA in the Login and Authentication Phase

Step 1: Edi performs a series of calculations in login and authentication, generates $P1$, $P2$, and PIdi data, and sends it to the cloud. Attackers can break the hash value generated by Edi at this point using a rainbow attack. After this step, it can read the data sent by Edi to the CS or change this information and send it back to the CS. The steps are as follows:

Step $R3$: Edi $\longrightarrow$ CS: $\{P1, P2, P\,\text{Idi}\}$
Step $A3$: Attacker $\longrightarrow$ CS: $\{P1, P2, P\,\text{Idi}\}$
Rainbow Crack $\{P2\}$
Resend $\{P1, P2, P\,\text{Idi}\}$ for CS

The attacker can listen to the sent messages in steps 2 and 3 because she has obtained the sent data in the previous step. Figure 3 shows the stages of an RTA in the login and authentication phase.

2.5. Security Analysis Kumari. Avispa is used to evaluate the security of Internet protocols [23]. Avispa will provide an HLPSL to define the security of protocols and display their security specifications. We analyzed the security of Kumari with Avispa, and the results show that it is vulnerable to a rainbow attack. In the first stage, attackers can break the hashed data sent from Edi by using a rainbow attack. Figure 4 shows the security weakness of Kumari.

2.6. Testbed. The tools we used to test the attack on the Kumari scheme were the Linux operating system and the board. We chose MD5 by default because the author did not mention the type of hash used in his work, and the MD5 library we used for the Arduino is available on GitHub. Table 3 shows information about the environment and the tools used.

Arduino is a hardware and processing platform designed as open source. This platform is based on a simple I/O and

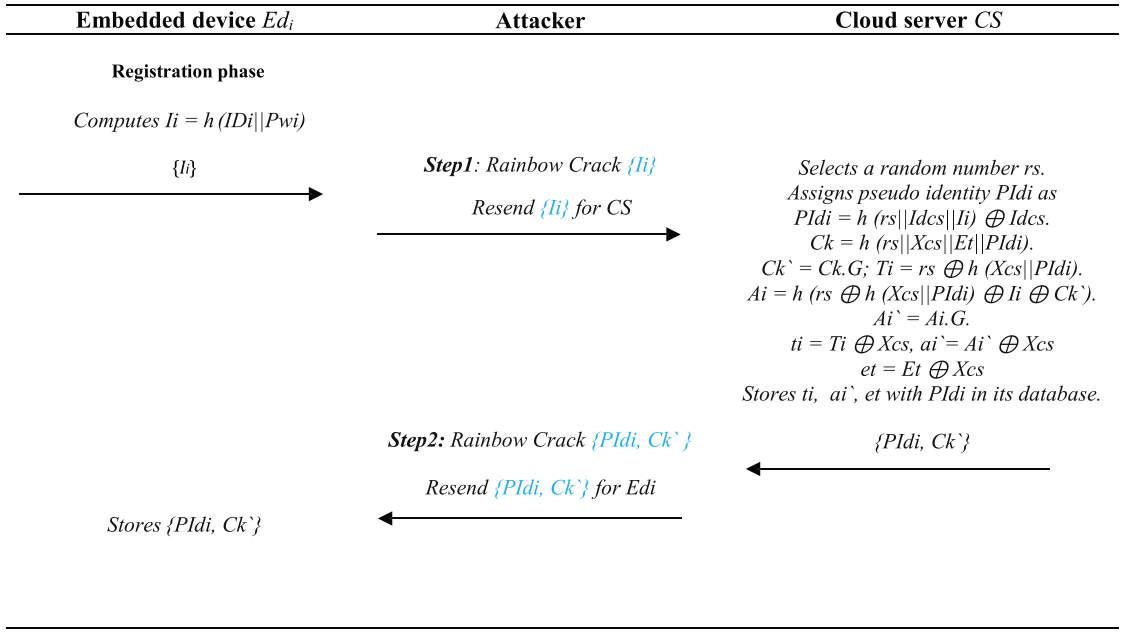


Figure 2: Stages of a RTA in the registration.

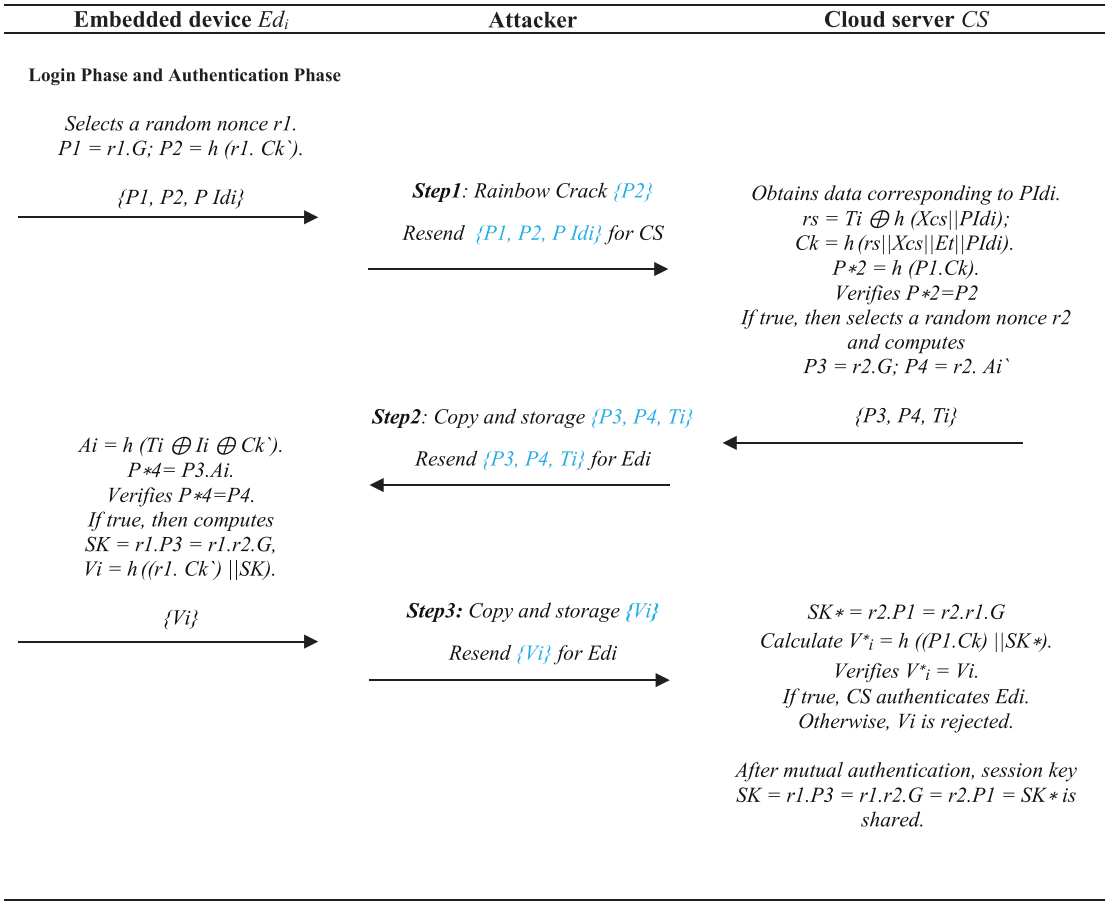


Figure 3: Stages of a RTA in the login and authentication.

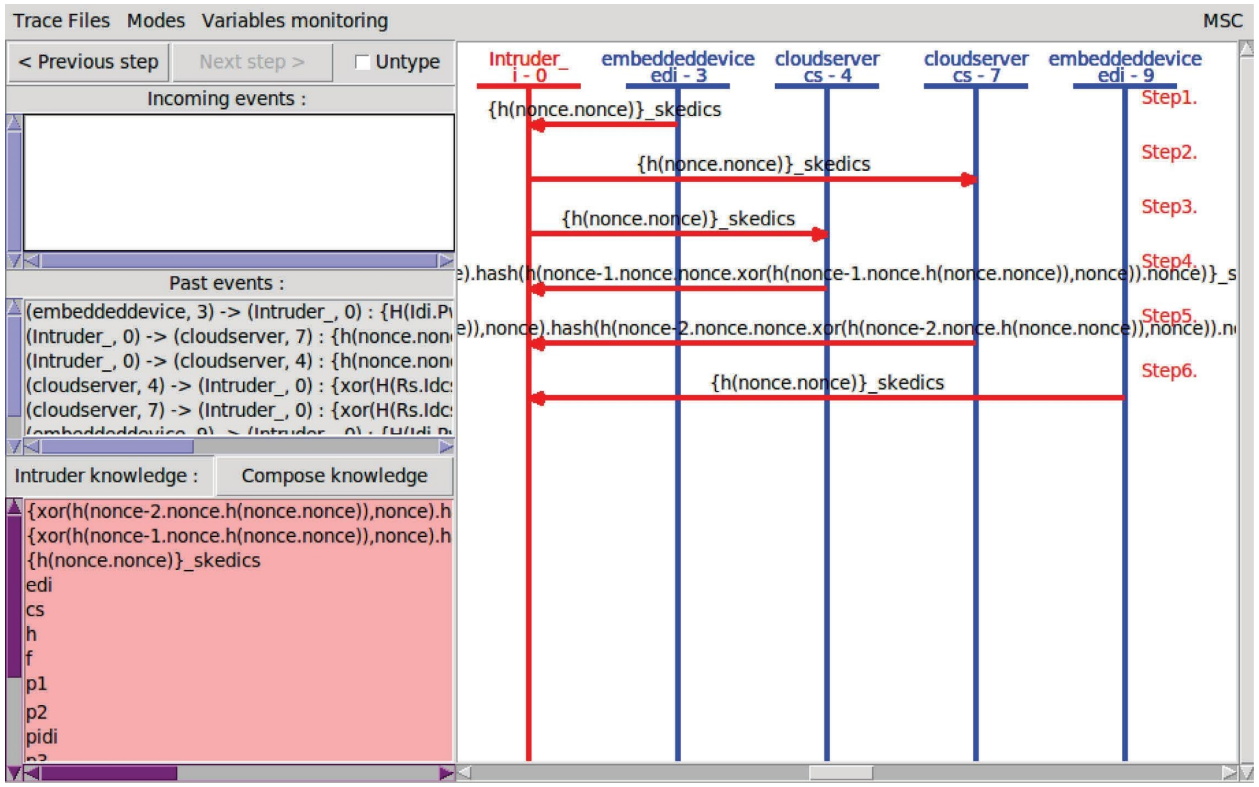


Figure 4: Vulnerability of Kumari et al. in the tool AVISPA.

Table 3: Environmental information and testbed tools.

Environment	Description
Operating system	Kali Linux ver: 2020
Boards	Arduino Uno R3
Programming languages	ANSI C
Type hash	MD5 Arduino libs
Attack tools	Rainbow table ver: 1.8

the designed processing/wiring language. Also, this platform is suitable for communication with external systems and software. Our work in the article used the Arduino board model Uno *R*3, microcontroller ATmega328, and input voltage, 7–12 with memory of 32 KB, and a speed clock of 16 MHz. Figure 5 shows the implementation of the Arduino boards.

2.6.1. Appointed Data. To attack the Kumari scheme, in the register phase, we need ID and Pwi, equal to $a1$ and 260 each. Table 4 shows the required data in the registration phase. Next, we will implement the given data in the Arduino board, and we will get each of the provided data from the serial port of the hash port. Finally, we will use these data to prove the authenticity of our attack. Figures 6 and 7 show the hash output in the Arduino board, and Figure 8 shows the MD5 source code used.

2.6.2. Testbed Result. At this stage, we will attack the registration phase of the Kumari scheme using a Rainbow. Table 4 shows the values and hashes of ID and Pwi. To

Figure 5: Implementation of the Kumari scheme of the Arduino board.

perform the attack, we assign the hash values of ID and Pwi to the Rainbow tables. The output results show that the Rainbow breaks the hashes given to the plain text quickly, which shows the weakness of the design against this attack. Figure 9 shows the results of an attack on an ID, and Figure 10 shows an attack on Pwi. The ciphertext/plain text and statistical results are marked with a red box in the image.

The results of the statistics of the RTA to reach the plain text ID are as follows: total time 0.58, time of chain traverse 0.58, hash and reduce the calculation of chain traverse 7216200, hash and reduce the calculation of alarm traverse 16391, number of alarm 11, the performance of chain traverse 12.46 million/s, and the performance of chain alarm 12.46 million/s. The results of the statistics of the RTA to reach the plain text Pwi are as follows: total time 0.58 s, time

Table 4: Data used in the testbed.

Notations	Values	HEX	MD5
ID	$a1$	6131	8a8bb7cd343aa2ad99b7d762030857a2
Pwi	260	323630	a4f23670e1833f3fdb077ca70bbd5d66

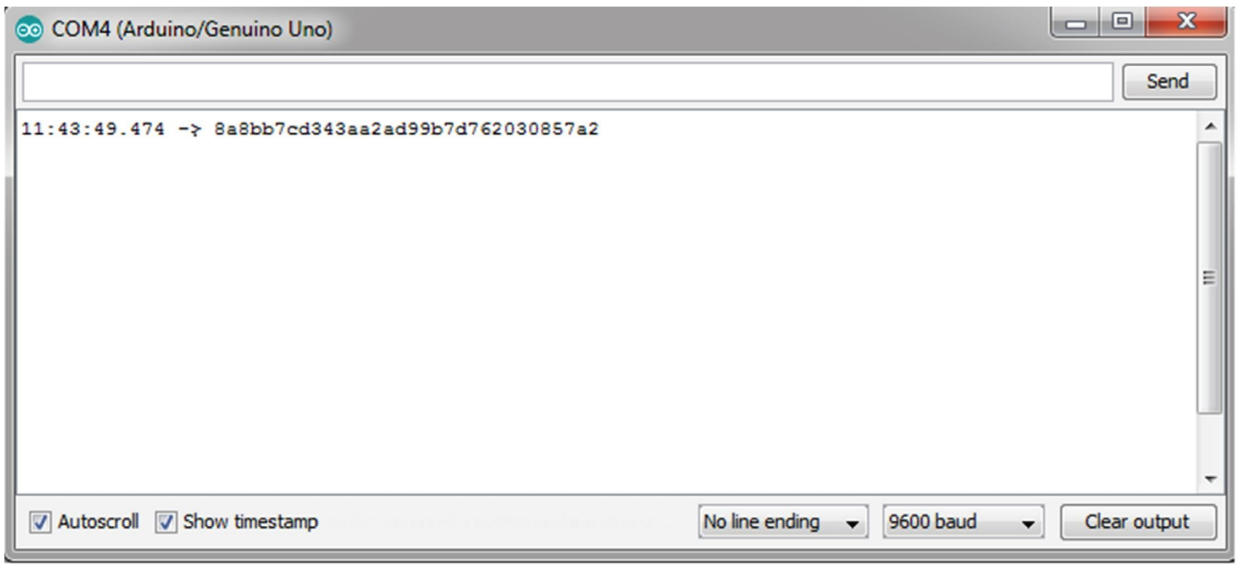


Figure 6: The results of ID hash.

Figure 7: The results of Pwi hash.

```
#include <MD5.h>
Void setup ()
{  //initialize serial
  Serial. Begin (9600);
  //give it a second
  Delay (1000);
  //generate the MD5 hash for our string
  Unsigned char* hash=MD5:: make hash ("260");
  //generate the digest (hex encoding) of our hash
  Char *md5str = MD5::make_digest (hash, 16);
  Free (hash);
  //print it on our serial monitor
  Serial.println (md5str);
  //Give the Memory back to the System if you run
the md5 Hash generation in a loop
  Free (md5str); }
Void loop () { }
```

Figure 8: MD5 source code.

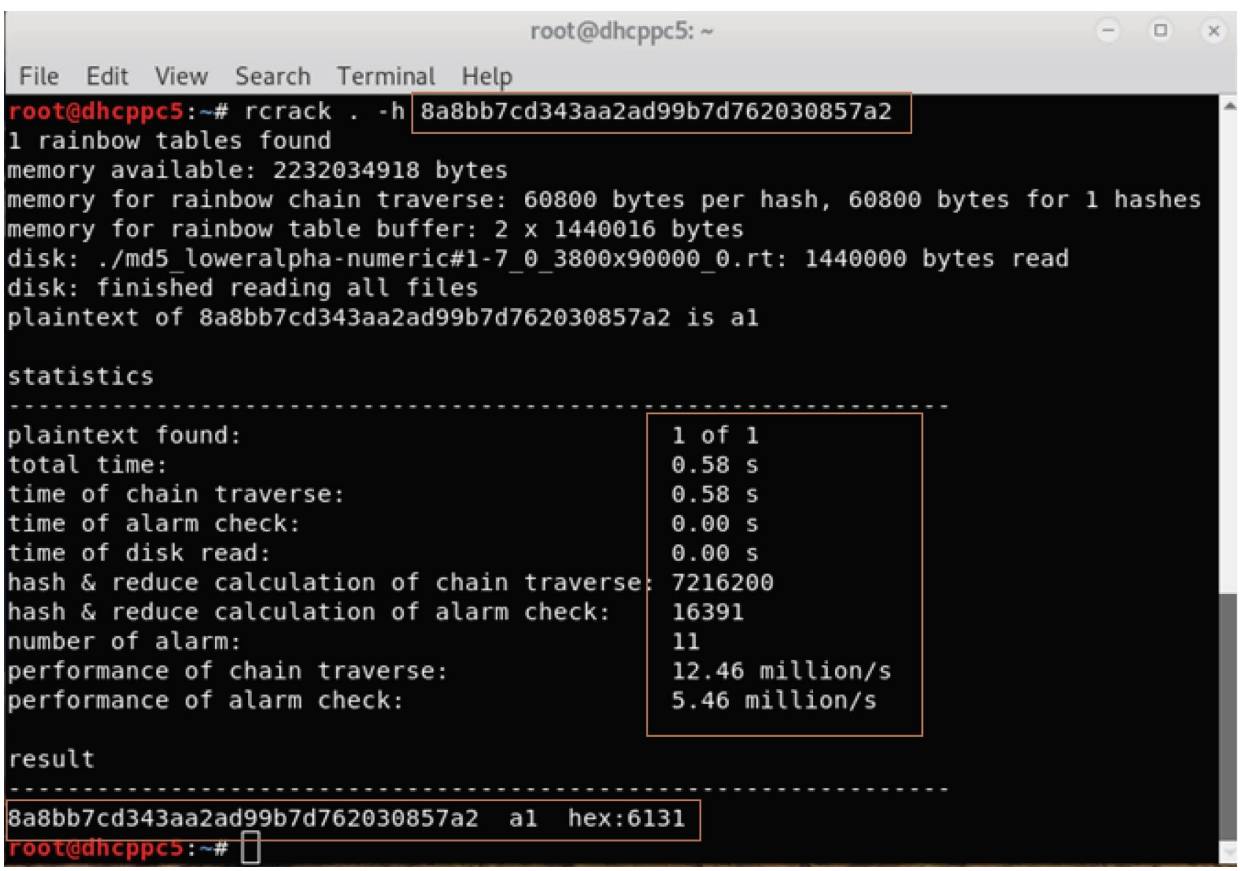


Figure 9: The results of the RTA on the ID hash.

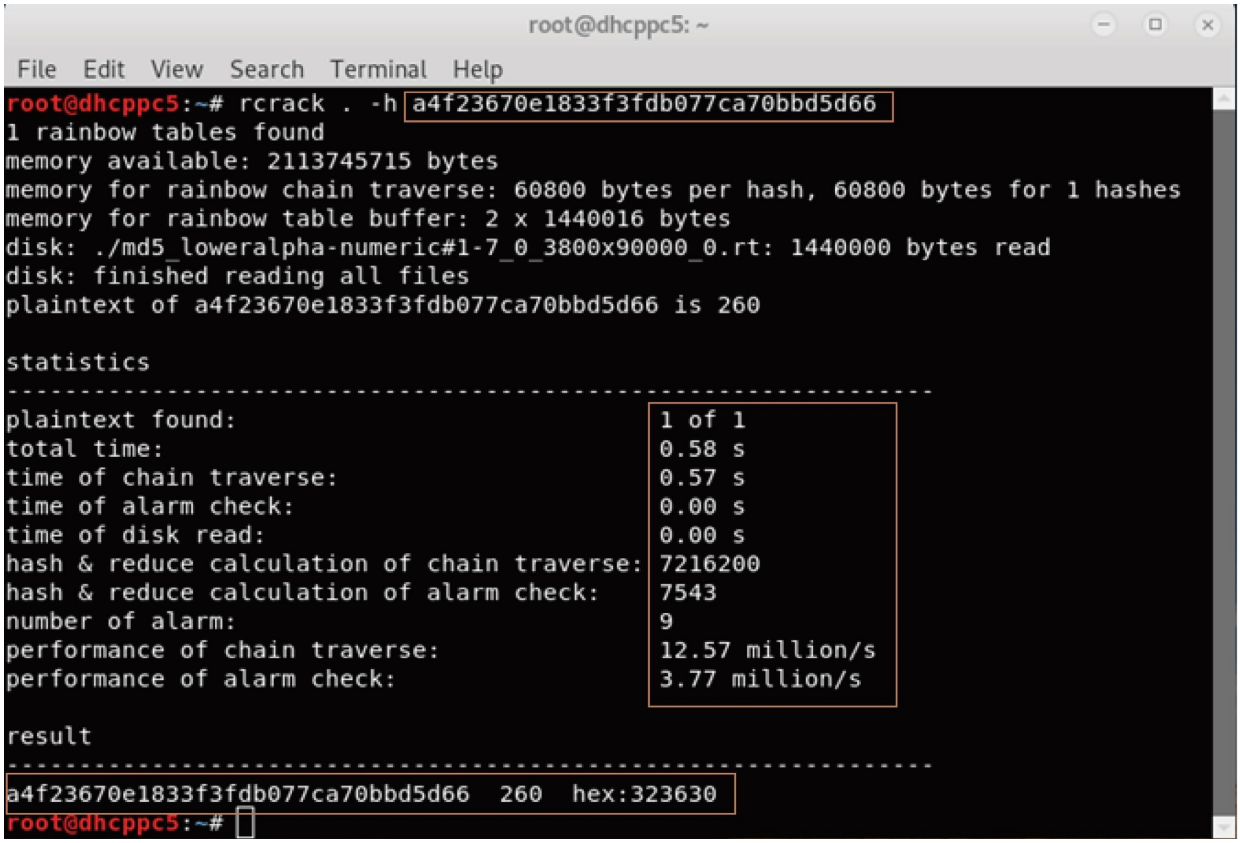


Figure 10: The results of the RTA on the Pwi hash.

of chain traverse 0.57 s, hash and reduce the calculation of chain traverse 7216200, hash and reduce the calculation of alarm traverse 7543, the number of alarm 9, the performance of chain traverse 12.57 million/s and, the performance of chain alarm 3.77 million/s.

3. Network Model

This section describes the network, assumption, and adversary models and reviews ECC.

The network model for fog computing-based IOV is shown in Figure 11. This network model has a variety of connections between different parties, such as "V2V, V2R, R2F, F2F, F2C." V2V: vehicles can communicate with other vehicles, and receive and send traffic information and other data. V2R: RSU can communicate with vehicles, exchange information, and be associated with other RSUs. R2F: sometimes, the received data require complex processing beyond the power of RSU, in which case the data do transmit to fog for processing. F2F: fog can communicate with other fog and support each other to process data [24]. F2C: when fog cannot do the necessary processing, they send the data to the cloud. The assumptions in this model are as follows:

(i) The time of all devices is synchronized

(ii) Cloud and fog and fog nodes know each other

(iii) Fog and clouds are resistant to various attacks that also do not leak any data

(iv) The transmission channel is not secure in the network

3.1. Problem Statement. The main challenge in the IOV environment is to ensure the source of the data sent. Authentication allows us to identify the source of the transmitted data and to be able to detect fake data. The Kumari scheme is a cookie-based authentication scheme for IOT environments. However, this scheme is vulnerable to rainbow attacks. This paper presents a SAIFC scheme based on the ECC, which uses the cookie in the HTTP protocol to send data. This scheme provides secure authentication between source and destination and resists active and passive attacks.

3.2. Adversary Model. The following are some of the attacks that can be dangerous in an IOV environment:

(i) Replay attack: an attack method where an intruder records a communication session and then broadcasts it again

(ii) Man-in-the-middle attack: MITM attack is when an intruder uses session data to forge connections or change data

(iii) Sybil attack: an attack in which the attacker can have multiple identities and deceive other vehicles

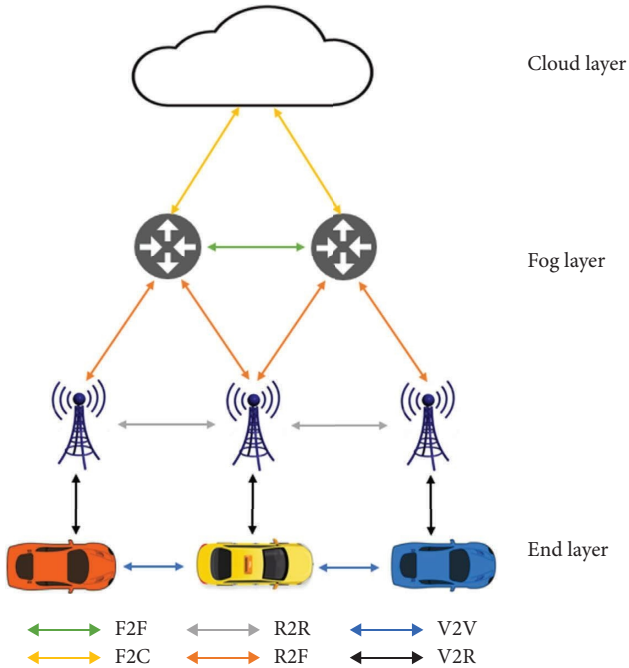


Figure 11: Network model of IOV and fog computing.

(iv) Impersonation attack: the intruder intends to forge the other person's identity in these attacks

(v) Brute force attack: the attacker uses all possible modes to break the encrypted text

(vi) Rainbow table: an attacker uses tables where the hash text output does save to break the hash text

Our SAIFC will be resistant to the attacks.

3.3. Review ECC. The ECC is a PKE method based on an algebraic structure of EC on finite fields. The use of EC in encryption was proposed independently by Neal et al. in 1985. The PKE is based on the difficulties in some math problems. Earlier, systems based on the public key were considered safe, assuming that finding two or more prime factors for a large integer was difficult. For EC-based algorithms, it is assumed that finding the DL from a random element of EC is impractical, given a publicly known base point. The size of the EC determines the difficulty of the problem. The main advantage of the ECC was a key with a smaller size, which means reduced storage. The EC is a flat curve composed of equation (1) for today's encryption purposes.

$$y^2 = x^3 + a\,x + b. \tag{1}$$

4. SAIFC

In this section, the different phases of the SAIFC scheme are described. Figure 12 shows the roadmap of the SAIFC.

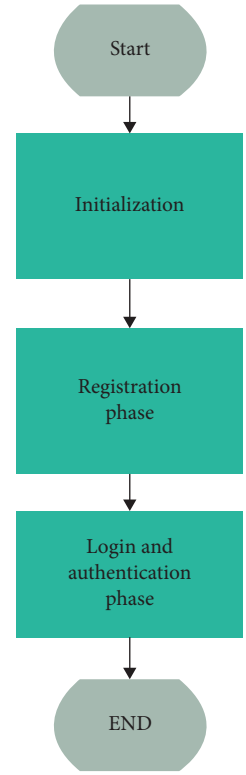


Figure 12: SAIFC roadmap.

4.1. Initialization. The first R chooses equation (1) on the EC of Zp. After R choose the element is f, $a, b \in Zp$ each in which a, b fulfill condition $y^2 = x^3 + ax + (b) \pmod p$. In the EC, G is the foundation point, with a prime order of $n(n > 2^{160})$. If the O is a point, then the equation $n.G = O$ is at infinity. XR are randomly selected as the secret keys of R.

4.2. Register Phase. The Register phase is as follows:

Step 1: Vdi to register in R, $Ii = h$ ($IdR||$Pwi), Tv $= h$ (TV) computes and sends Ii, Tv, and TV to R.

Step 2: When the registration request was received, R checks TV in the computes of TV$' = h$ (TV) and the result obtained with Tv, and if it is the same, it checks in terms of timestamp. If it is small from the expiration time, continue the steps. R produces a rs and computes PIdR $= h$ (rs||IdR||Ii) $\oplus$ IdR for Vdi and storage IdR. Then, R computes the Ck and other component. R storage ti, ai$'$, and et corresponding to PIdR of Vdi in its DB. The expiration time of the Et that corresponds to Ii of Vdi is storage by R himself. R timestamp calculated itself in the Tr $= h$ (TR) After, sends {PIdR, Ck$'$, Tr, TR} to Vdi through a communication channel.

Step 3: Vdi checks TR in the computes of TR$' = h$ (TR), and the result obtained with Tr; if it is the same, it checks in terms of timestamp. After receiving {PIdR, Ck$'$}, the Vdi stores PIdR and Ck$'$ in its memory. Ck $= h$

(rs||XR||Et||PIdR) can update its expiration time. Figure 13 shows the flowchart of the SAIFC registration phase.

4.3. Login and Authentication Phase

Step 1: For each entry, the Vdi selects a $r1$ and computes the ECC point $P1 = r1.G$; $P2 = h(r1.\mathrm{Ck}')$. Then, it stores $P1$ in its memory, and $\mathrm{Tv} = h$ (TV) computes and sends the login request {P1, P2, PIdR, Tv, TV} to R.

Step 2: Upon receiving the login request, R checks TV in the computes of $\mathrm{TV}' = h$ (TV) and the result is obtained with Tv, and if it is the same, it checks in terms of timestamp. If it is small from the expiration time, continue the steps: R data corresponding to the receives PIdR and computes $rs = \mathrm{Ti} \oplus h(\mathrm{XR} \| \mathrm{PIdR})$. Next, R computes the $\mathrm{Ck} = h$ (rs ||XR||Et ||PIdR) and $P * 2 = h(P1.\mathrm{Ck})$.

Step 3: R selects a random nonce $r2$, computes the ECC point $P3 = r2.G$; $P4 = r2.Ai'$ and $\mathrm{Tr} = h$ (TR) compute sends {$P3$, $P4$, Ti, Tr, TR} to Vdi.

Step 4: Upon receiving, Vdi checks TR in the computes of $\mathrm{TR}' = h$ (TR), and the result obtained with Tr, and if it is the same, it checks in terms of timestamp. The next step computes $Ai = h(\mathrm{Ti} \oplus Ii \oplus \mathrm{Ck}')$ and the ECC point $P * 4 = P3.Ai$. Then, it verifies $P * 4 = P4$ to authenticate R. If the verification holds, then Vdi authenticates R and continues the next step;

Step 5: Vdi computes the session key $\mathrm{SK} = r1.P3 = r1.r2.\mathrm{G}$ and $\mathrm{VR} = h((r1.\mathrm{Ck}')\|SK)$ and $\mathrm{Tv} = h$ (TV) sends to R.

Step 6: Upon receiving the login request, R checks TV in the computes of $\mathrm{TV}' = h$ (TV) and the result obtained with Tv, and if it is the same, it checks in terms of timestamp. If it is minor from the expiration time, continue the steps: R computes the session key $SK = r1 \cdot P3 = r1 \cdot r2 \cdot G = r2 \cdot P1 = SK*$. Then, R verifies $V * R = \mathrm{VR}$ to authenticate Vdi. If the verification holds, R authenticates Vdi; otherwise, VR is rejected. From then on, all the after messages transmitted between Vdi and R are XOR with the session key $SK = r1\,P3 = r1\,r2\,G = r2\,P1 = \mathrm{SK}*$. Figure 14 shows the flowchart of the login and authentication phase of the SAIFC.

Step 7: R calculates $\mathrm{Tr} = h$ (TR), $\mathrm{IDR}' = h$ (IDR) and sending to F. F checks TR in the computes of $\mathrm{TR}' = h$ (TR), and the result obtained with Tr, and if it is the same, it checks in terms of timestamp. In the next step, calculate $\mathrm{IdR}' = h$ (IDR), check whether $\mathrm{IDR}' = \mathrm{IDR}'$ that is true storage the IDR. In the next step, calculate $\mathrm{Tf} = h$ (TF) and {IDR′, IDR, Tf, TF} sending to CS.

Step 8: F checks Tf in the computes of $\mathrm{TF}' = h$ (TF), and the result obtained with Tf; if it is the same, it checks in terms of timestamp. In the next step, calculate $\mathrm{IDR}' = h$ (IDR), check whether $\mathrm{IDR}' = IDR'$ that is true storage of the IDR. Figure 15 shows the authentication steps.

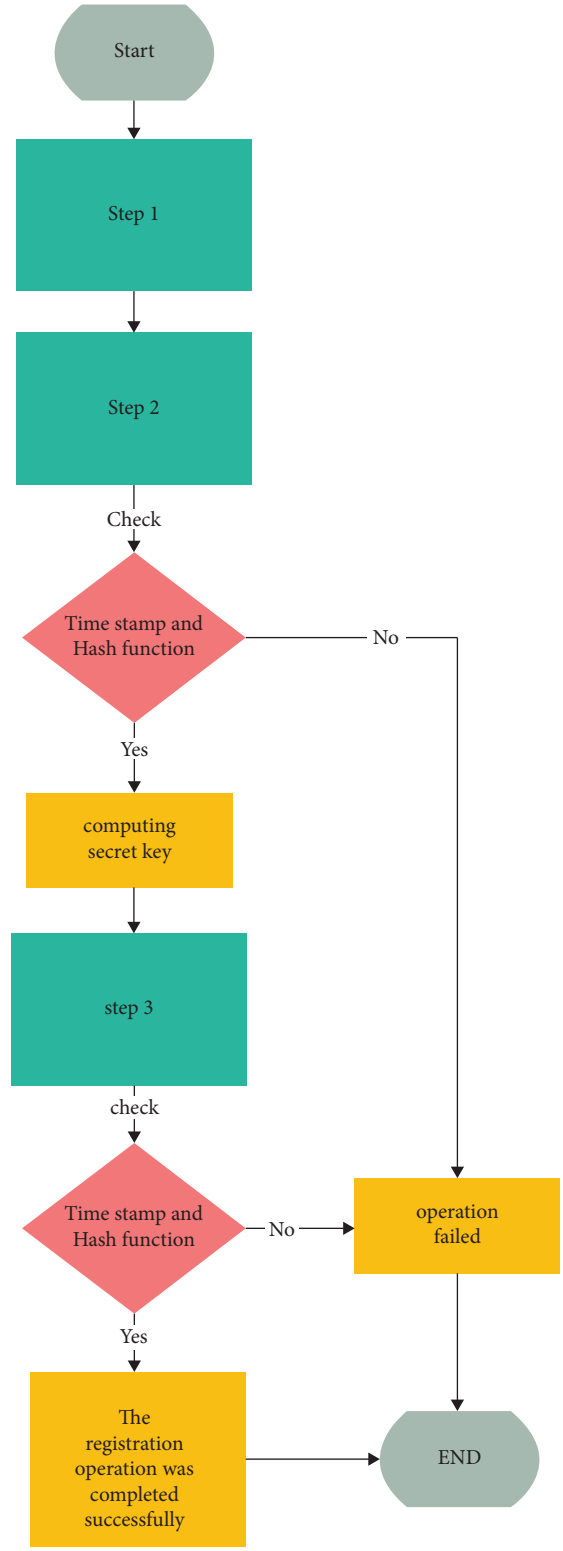


Figure 13: Flowchart of the registration phase.

5. SAIFC Security Analysis

In this section, security analyzes our SAIFC and discusses the results and analysis, followed by an informal security analysis.

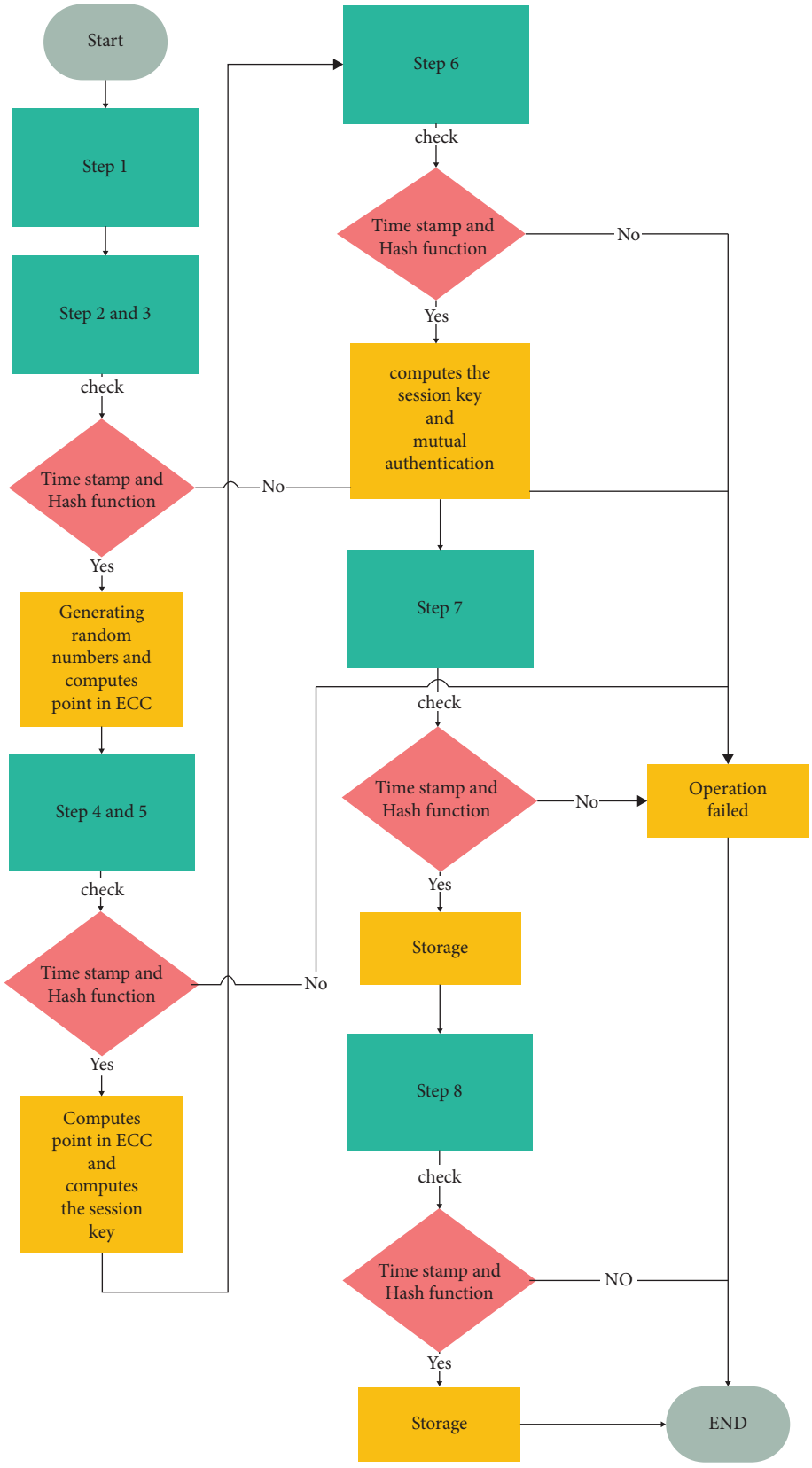


Figure 14: Flowchart of the login and authentication phase.

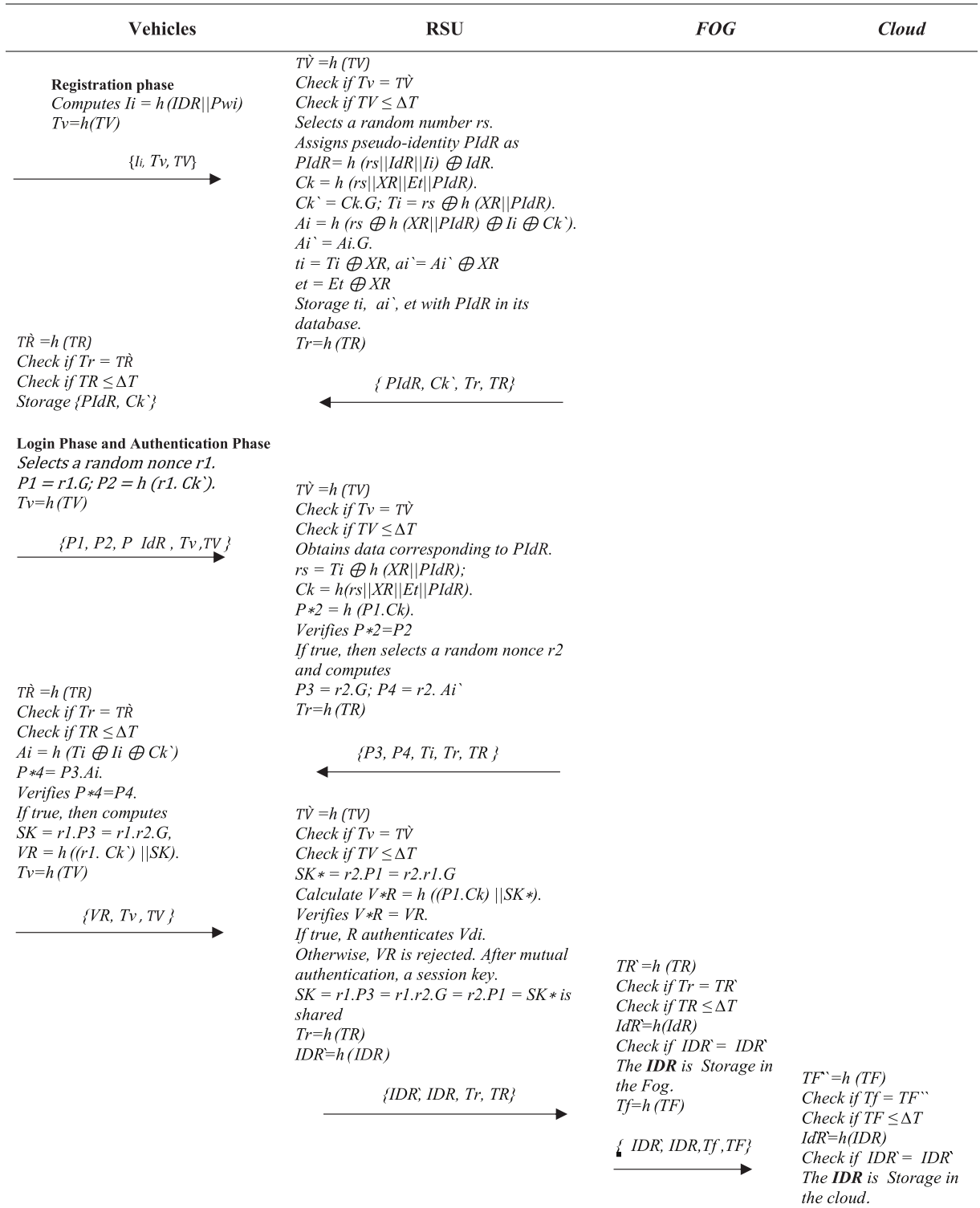


Figure 15: Authentication of the SAIFC scheme.

AVISPA is a formal verification tool for evaluating a safe protocol that combines several methods to model checking [21, 25, 26]. AVISPA is an HLPSL used to define the security of schemes and their security specifications [27, 28]. In HLPSL, an attacker always plays a legal role that is indicated by (i). The *D-Y* threat model [29] has been embedded. AVISPA uses four tools OFMC [30], CL-AtSe [31], SATMC [32], and TA4SP [33] to analyze security targets. Out of these

four tools, SATMC and TA4SP do not support xor operation; therefore, in the simulation, we have used other tools (OFMC and CL-AtSe) to test.

5.1. Analysis of AVISPA Results. Our SAIFC scheme, a replay, MITM, and other attacks are discussed in Section 4.2, with tools OFMC and CL-AtSe tested. The simulation results of OFMC and CL-AtSe are shown in Figures 16 and 17, respectively. The total number of visited nodes is 12, while the number of depth four plies with a search time of 0.44 seconds and CL-AtSe analyzed 0 states, and the translation time was 0.15 seconds. Thus, the overall results of the two tools show that the SAIFC scheme is safe.

5.2. Informal Security Analysis

(i) Replay attack: In the SAIFC scheme, the attacker can intercept the messages exchanged at the registration, login, and authentication phases and take legal registration or login in the future. In the SAIFC scheme, TV, TF, and TR parameters are used to prevent this attack, and before any processing, the receiver of the message first checks the time stamp, and if it is smaller than the expiration time, it performs processing; otherwise, the communication channel is closed. The reason for the SAIFC scheme is that it is resistant to replay attacks.

(ii) MITM: The SAIFC scheme can be an attacker placed between vehicle and RSU and intercept or modify the exchanged messages. In the SAIFC scheme to prevent this attack, mutual authentication is used, as shown in steps 3, 4, and 5. Also, in the SAIFC scheme, using a timestamp and HF in each message has made it resistant to a MITM.

(iii) Sybil attack: To prevent a Sybil attack in the SAIFC, in the registration phase, the vehicle first sends its password to RSU. RSU calculates parameters PIdR and Ck' based on XR and R1 due to the use of Pwi, PIdR, and Ck', and the SAIFC scheme is resistant to Sybil attack.

(iv) Impersonation attack: We have used mutual authentication in the SAIFC scheme to prevent this attack, which is discussed in steps 3, 4, and 5.

(v) Brute force attack: The attacker wants to check all possible situations until the answer is reached. Assuming this, the attacker can extract the parameters of $P1$, $P2$, $P3$, and $P4$ from the exchanged messages. He cannot attack because key XR is unknown to him, and he has no way to guess the random numbers $r1$ and $r2$. For this reason, the SAIFC schema is resistant to Brute force attacks.

(vi) RTA: If the attacker wants to break the HF of the messages sent between the communication parties, this process takes time. In the SAIFC scheme, a time stamp is used in each message, which makes the attack impossible because it takes time to break the HF. If the time stamp of the received message is greater than the expiration time, the message is considered invalid, and the communication channel is closed. Therefore, the SAIFC scheme against the RTA is resistant.

```
% OFMC
% Version of 2006/02/13
SUMMARY
  SAFE
DETAILS
  BOUNDED_NUMBER_OF_SESSIONS
PROTOCOL
  /home/span/span/testsuite/results/Auth_Fog. if
GOAL
  as_specified
BACKEND
  OFMC
COMMENTS
STATISTICS
  parseTime: 0.00 s
  searchTime: 0.44 s
  visitedNodes: 12 nodes
  depth: 4 plies
```

Figure 16: Results of the SAIFC in the OFMC.

```
SUMMARY
  SAFE
DETAILS
  BOUNDED_NUMBER_OF_SESSIONS
    TYPED_MODEL
PROTOCOL
  /home/span/span/testsuite/results/Auth_Fog.if
GOAL
  As Specified
BACKEND
  CL-AtSe

STATISTICS
  Analysed : 0 states
  Reachable: 0 states
  Translation: 0.15 seconds
  Computation: 0.00 seconds
```

Figure 17: Results of the SAIFC in the CL-ATS.

6. SAIFC Performance Analysis

The performance analysis of the SAIFC scheme and security features are compared in this section with protocols by Wazid [21], Liu [15], Liu [11], Kalra [19], Kumari [20], Vasudev [18], Ming Chen [17], Ying and Nayak [14] , Mohit [13] in this section.

Table 5: Comparison of the different schemes in terms of communication costs.

No	Schemes	HF	ECC	PKE	PKD	SKE	SKD	Total cost	Total cost (ms)
1	[21]	35 Thf	4Teccm	0TPKe	0TPKd	0TSKe	0TSKd	35 Thf + 4Teccm	8.9845
2	[11]	8 Thf	11Teccm	0TPKe	0TPKd	0TSKe	0TSKd	8 Thf + 11Teccm	24.5044
3	[15]	10 Thf	6Teccm	0TPKe	0TPKd	0TSKe	0TSKd	10 Thf + 6Teccm	13.5518
4	[19]	9 Thf	7Teccm	0TPKe	0TPKd	0TSKe	0TSKd	9 Thf + 7Teccm	15.8043
5	[20]	13 Thf	8Teccm	0TPKe	0TPKd	0TSKe	0TSKd	7 Thf + 8Teccm	17.8379
6	[18]	17 Thf	0Teccm	0TPKe	0TPKd	0TSKe	0TSKd	17 Thf	0.0391
7	[17]	17 Thf	0Teccm	0TPKe	0TPKd	0TSKe	0TSKd	17 Thf	0.0391
8	[14]	12 Thf	0Teccm	0TPKe	0TPKd	2TSKe	2TSKd	12 Thf + 2TSKe + 2TSKd	0.046
9	[13]	20 Thf	0Teccm	0TPKe	0TPKd	0TSKe	0TSKd	20 Thf	0.046
10	SAIFC	30 Thf	8Teccm	0TPKe	0TPKd	0TSKe	0TSKd	30 Thf + 8Teccm	17.877

6.1. Computational Cost. The computational costs of the SAIFC scheme and other schemes [21], [11], [15], [19], [20], [18], [17], [14], [13] are tabulated in Table 5.

For analysis, the following symbols are defined. Thf is the number execution of HF. Teccm is the number execution of an ECC point multiplication operation. TPKe is the number execution of PKE. TPKd is the execution number of PKD. TSKE is the number execution of SKE. TSKd is the number execution of SKD. The time required to calculate the XOR operation is small, and we do not consider this. We use the paper [30] evaluation results for different cryptographic.

Our observations show that protocols by Vasudev et al. [18] and Chen et al. [17], with 0.0391 ms, have a lower cost compared to Ying and Nayak [14] and Mohit et al. [13] protocols which cost 0.046 ms. Wazid et al. [21] and Liu et al. [15], and Kalra and Sood [19] protocols have costs of 8.9845 ms and 13.5518 ms, and 15.8043 ms, respectively. The cost of the SAIFC is slightly higher than the Kumari et al. [20] protocol, and Liu et al. [11] protocol has the highest computation cost.

6.2. Communication Cost. A comparative study of the communication costs and total bits of different schemes is presented in Table 6. The obtained results have been measured manually and with the E3C tool [34]. Our observations show that Liu et al. [15] protocol has the lowest communication cost. The next is Kalra and Sood [19] and Kumari et al. [20] protocols with communication costs of 3. Next, Chen et al. [17] and Ying and Nayak [14] protocols are the communication cost. The SAIFC scheme costs more than Wazid, [21], Liu [11], and Vasudev [18] protocols, and Mohit et al. [13] protocol has the highest communication cost. First, Mohit [13] protocol has the least bits, and then, the SAIFC scheme and Kalra and Sood [19] and Kumari et al. [20] protocol are 1760 bits. Next are Ying and Nayak [14] and Liu et al. [15] and Vasudev et al. [18] and Chen et al. [17], and Wazid et al. [21] protocols, 1952 bit and 2272 bit and 2496 bit and 3024 bit and 3392 bit, respectively. Finally, Liu et al. [11] protocol has the most bit.

6.3. Security Features Comparison. Our observations show that all protocols are resistant to the replay attack. However, it is vulnerable to Mohit [5] protocol and MITM and Kalra [11] protocol insider attack. All protocols are resistant to stolen-verifier attacks, impersonation attacks, Brute force attacks, and offline password-guessing attacks. However, Kalra and Sood [19] protocol is vulnerable to offline password-guessing attacks. Except for Kalra and Sood [19] protocol, everyone can support device anonymity, mutual authentication, session key agreement, and forward secrecy. In the SAIFC scheme, a timestamp is considered for sending each message, which is checked at the destination with expiration time. For this reason, the SAIFC scheme can be resistant to rainbow attacks. The SAIFC scheme is based on HTTP Protocol and can support fog, OFMC, and CL-ATSE used for security evaluation. Table 7 shows a comparison of security features.

Note: FV1: replay attack; FV2: MITM; FV3: insider attack; FV4: stolen-verifier attack; FV5: impersonation attack; FV6: Brute force attack; FV7: offline password guessing attack; FV8: device anonymity; FV9: mutual authentication; FV10: session key agreement; FV11: forward secrecy; FV12: confidentiality; FV13: RTA; FV14: OFMC; FV15: CL-ATSE; FV16: fog-based; FV17: HTTP-based.

7. Simulation Results and Analysis

A feasible demonstration of the SAIFC by the NS3 presents in this section.

7.1. Simulation Environment and Settings. Table 8 presents the parameters used in the NS3.

7.2. SAIFC Simulation Results. We simulated our SAIFC using the three routing protocols AODV, DSDV, and OLSR. The results of packet delivery show that DSDV protocol performed better, OLSR protocol performed moderately, and AODV protocol performed poorly. Figure 18 shows the packet delivery rate comparison. A comparison of throughput shows that DSDV protocol performed better and AODV protocol performed poorly. Figure 19 shows the throughput comparison. In packet loss, DSDV protocol is higher after AODV and OLSR protocol is placed, respectively. Figure 20 shows the packet loss comparison. OLSR protocol has less delay than DSDV and AODV. Figure 21 shows the end to end delay comparison. According

Table 6: Comparison of the different schemes in terms of communication cost and the number of bits.

No	Schemes	Number of messages	Total bits
1	[21]	6	3392
2	[11]	6	8992
3	[15]	2	2272
4	[19]	3	1760
5	[20]	3	1760
6	[18]	6	2496
7	[17]	4	3024
8	[14]	4	1952
9	[13]	9	1280
10	SAIFC	7	1760

Table 7: Comparison of the different schemes in terms of security features.

Security features	Schemes									SAIFC
	[21]	[11]	[15]	[19]	[20]	[18]	[17]	[14]	[13]	
FV1	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FV2	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
FV3	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
FV4	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FV5	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FV6	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FV7	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
FV8	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
FV9	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
FV10	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
FV11	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
FV12	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FV13	No	No	No	No	No	No	No	No	No	Yes
FV14	Yes	No	No	Yes	Yes	No	No	No	No	Yes
FV15	Yes	No	No	Yes	No	No	No	No	No	Yes
FV16	Yes	No	No	No	No	No	No	No	No	Yes

Table 8: Simulation parameters.

Parameters	Description
OS	Ubuntu-20.04.1
Hardware	Dell 5110, Core i5, 4 GB RAM
Tool	NS 3 2.29
No. of V	30
No. of R	10
No. of fog	5
Mobility of V	20 m/s (no pause)
Mobility model	Random
Environment area	300 $*$ 1500 M
Loss model	Two-ray ground loss
Transmit power	7.5 dBm
Routing protocol	AODV-DSDV-OLSR
MAC	IEEE 802.11
Wireless protocol	802.11 p
Communication range of R to V	145 M
Simulation scenario	Highway
Simulation time	300 seconds

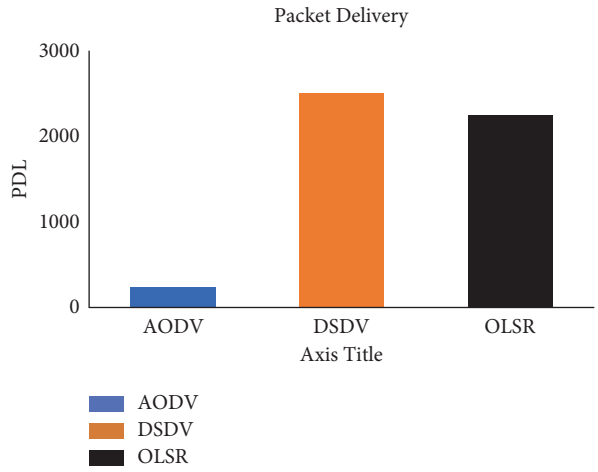


Figure 18: Comparison of packet delivery.

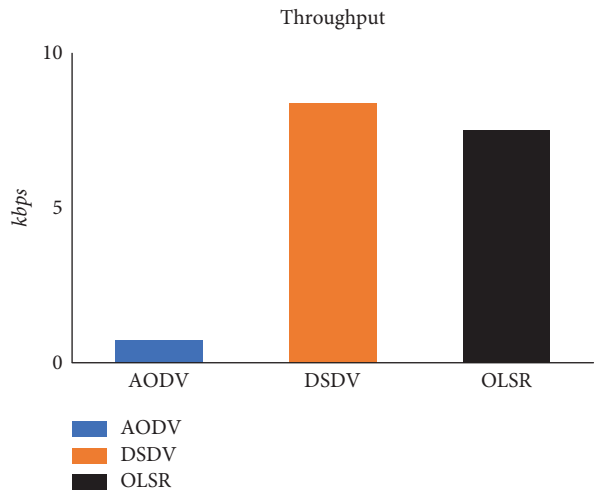


Figure 19: Comparison of throughput.

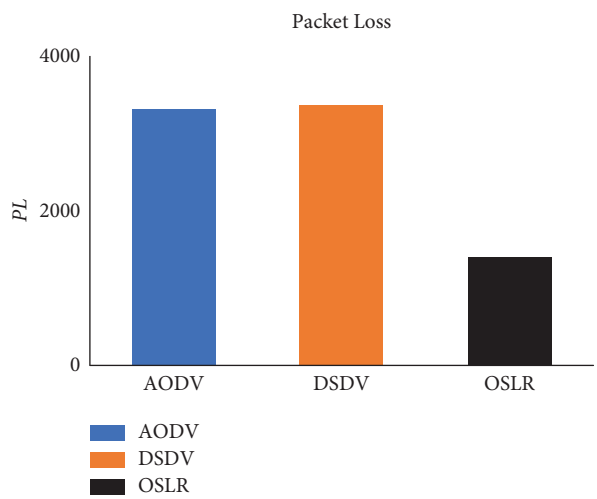


Figure 20: Comparison of packet loss.

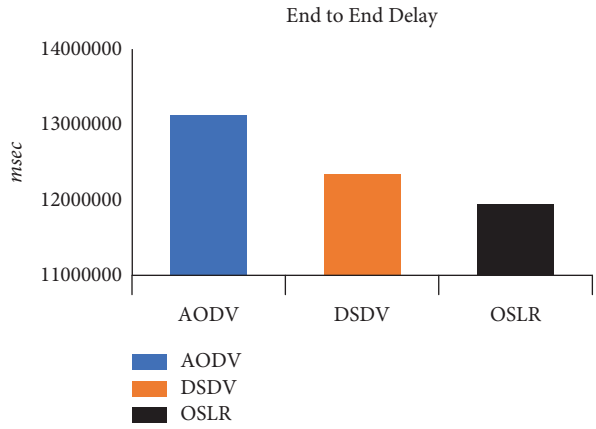


Figure 21: Comparison of end-to-end delay.

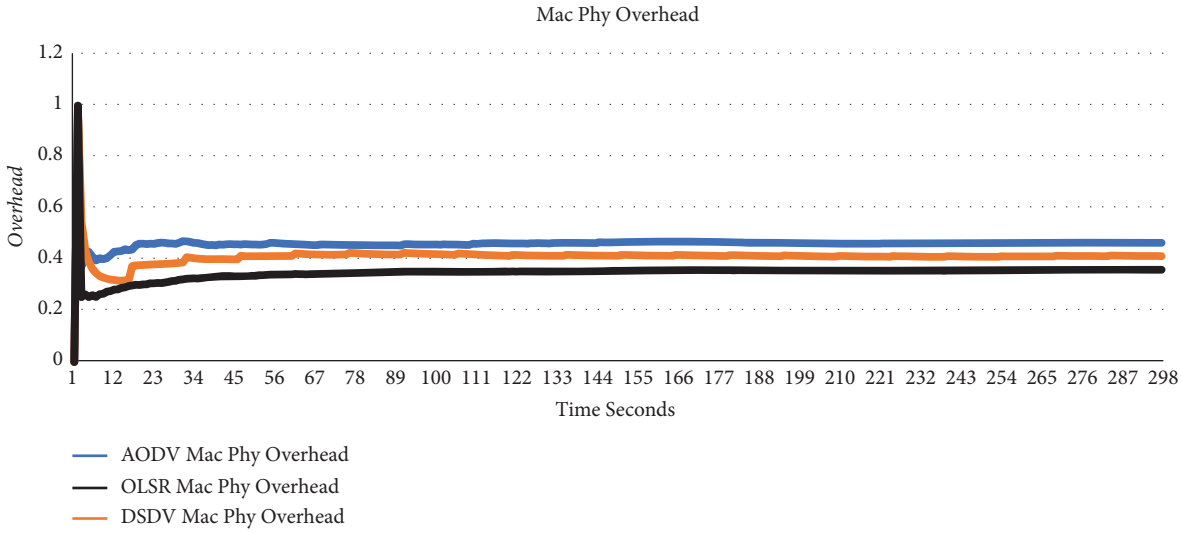


Figure 22: Comparison of MAC/PHY overhead.

to the results, it is impossible to say precisely, which routing protocol works best for the SAIFC scheme.

For this reason, we have measured the overhead of routing protocols. The results show that OLSR, DSDV, and AODV protocols have the lowest overhead and are suitable for the SAIFC OLSR protocol scheme. Figure 22 shows the overhead comparison.

8. Conclusion

We have dealt with an important emerging research topic to secure authentication between IOV and fog computing. We propose an HTTP-based secure mutual authentication scheme for IOV-fog-based, which sends data for authentication via a cookie. We used informal and AVISPA for the security analysis SAIFC scheme; the security analysis results show that the SAIFC resists famous attacks. The performance analysis of the SAIFC with other protocols about the number of bits, computation, and communication cost shows that the cost of computation and communication has increased in the SAIFC. We simulated the SAIFC scheme with the tool NS3 and then compared the AODV, DSDV, and OLSR routing protocols. Among the routing protocols compared, OLSR is more efficient. The results show that the SAIFC scheme works well on highways with the OLSR routing protocol and can be used in applications related to the exchange of information in fog-based environments. In future work, the SAIFC scheme can be developed based on the blockchain system, and a lightweight scheme can be reached by reducing communication and computing costs.

Acronyms

AVISPA: Automated validation of internet security protocols and applications
IOT: Internet of things
IOV: Internet of vehicles
RSU: Roadside unit
V2V: Vehicles to vehicles
V2R: Vehicles to roadside unit
R2F: Roadside unit to fog
F2F: Fog to fog

F2C: Fog to cloud
HLPSL: High-level protocol specification language
OFMC: On-the-fly model-checker
CL-ATSE: CL-based attack searcher
SATMC: SAT-based model-checker
TA4SP: Tree automata-based protocol analyser
AKE: Authentication and key exchange
AKM: Authenticated key management
ECC: Elliptic curve cryptography
EC: Elliptic curves
DL: Discrete logarithm
HF: Hash function
PKE: Public key encryption
PKD: Public key decryption
SKE: Symmetric key encryption
SKD: Symmetric key decryption
RTA: Rainbow table attack.

Data Availability

The data used to support this novel scheme are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] S. Sharma and B. Kaushik, "A survey on internet of vehicles: applications, security issues & solutions," *Vehicular Communications*, vol. 20, pp. 100–182, 2019.

[2] Z. Zhang, G. De Luca, B. Archambault, J. Chavez, and B. Rice, "Traffic dataset for dynamic routing algorithm in traffic simulation," *Journal of Artificial Intelligence and Technology*. vol. 2, 2022.

[3] M. Yang, "Research on vehicle automatic driving target perception technology based on improved MSRPN algorithm," *Journal of Computational and Cognitive Engineering*, vol. 1, no. 3, pp. 147–151, 2022.

[4] 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. 4662, no. c, p. 1, 2020.

[5] S. il Hahm, "Reliable real-time operating system for IoT devices," *IEEE Internet of Things Journal*, vol. 4662, no. c, pp. 1–11, 2020.

[6] L. Xu, H. Wang, and T. A. Gulliver, "Outage probability performance analysis and prediction for mobile IoV networks based on ICS-BP neural network," *IEEE Internet of Things Journal*, vol. 4662, no. c, p. 1, 2020.

[7] Y. Salami and V. Khajehvand, "SMAK-IOV: secure mutual authentication scheme and key exchange protocol in fog based IoV," *Journal of Computer and Robotics*, vol. 13, no. 1, pp. 11–20, 2020.

[8] N. Moustafa, B. Turnbull, and K. K. R. Choo, "An ensemble intrusion detection technique based on proposed statistical flow features for protecting network traffic of internet of things," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 4815–4830, 2019.

[9] Y. An, F. R. Yu, J. Li, J. Chen, and V. C. M. Leung, "Edge intelligence (EI)-Enabled HTTP anomaly detection framework for the internet of things (IoT)," *IEEE Internet of Things Journal*, vol. 4662, no. c, p. 1, 2020.

[10] H. Hasrouny, C. Bassil, A. E. Samhat, and A. Laouiti, "Group-based authentication in V2V communications," in *Proceedings of the Fifth International Conference on Digital Information and Communication Technology and its Applications (DICTAP)*, pp. 173–177, Beirut, Lebanon, April 2015.

[11] Y. Liu, Y. Wang, and G. Chang, "Efficient privacy-preserving dual authentication and key agreement scheme for secure V2V communications in an IoV paradigm," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 10, pp. 2740–2749, 2017.

[12] L. Benarous and B. Kadri, "Ensuring privacy and authentication for V2V resource sharing," in *Proceedings of the Seventh IEEE International Conference on Emerging Security Technologies 2017*, pp. 1–6, Canterbury, UK, September 2017.

[13] P. Mohit, R. Amin, and G. P. Biswas, "Design of authentication protocol for wireless sensor network-based smart vehicular system," *Vehicular Communications*, vol. 9, pp. 64–71, 2017.

[14] B. Ying and A. Nayak, "Anonymous and lightweight authentication for secure vehicular networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 12, pp. 10626–10636, 2017.

[15] J. Liu, Q. Li, R. Sun, X. Du, and M. Guizani, "An efficient anonymous authentication scheme for internet of vehicles," *IEEE International Conference on Communications*, vol. 2018, Article ID 8422447, pp. 16, 2018.

[16] K. Lim and K. M. Tuladhar, "LIDAR: lidar information based dynamic V2V authentication for Roadside infrastructure-less vehicular networks," in *Proceedings of the 2019 16th IEEE Annual Consumer Communications & Networking Conference (CCNC)*, pp. 1–6, Las Vegas, NV, USA, January 2019.

[17] C. M. Chen, B. Xiang, Y. Liu, and K. H. Wang, "A secure authentication protocol for internet of vehicles," *IEEE Access*, vol. 7, no. c, pp. 12047–12057, 2019.

[18] H. Vasudev, V. Deshpande, D. Das, and S. K. Das, "A lightweight mutual authentication protocol for V2V communication in internet of vehicles," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 6, pp. 6709–6717, 2020.

[19] S. Kalra and S. K. Sood, "Secure authentication scheme for IoT and cloud servers," *Pervasive and Mobile Computing*, vol. 24, pp. 210–223, 2015.

[20] S. Kumari, M. Karuppiah, A. Kumar, D. Xiong, L. Fan, and N. Kumar, "A secure authentication scheme based on elliptic curve cryptography for IoT and cloud servers," *The Journal of Supercomputing*, vol. 74, pp. 6428–6453, 2017.

[21] M. Wazid, P. Bagga, A. K. Das, S. Shetty, J. J. P. C. Rodrigues, and Y. Park, "Akm-IoV: authenticated key management protocol in fog computing-based internet of vehicles deployment," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 8804–8817, 2019.

[22] E. R. Sykes and W. Skoczen, "An improved parallel implementation of RainbowCrack using MPI," *Journal of Computer Science*, vol. 5, no. 3, pp. 536–541, 2014.

[23] Y. Glouche, T. Genet, and E. Houssay, "SPAN--a Security Protocol ANimator for AVISPA--User Manual," *IRISA/ Université de Rennes*, vol. 1, p. 20, 2006.

[24] Y. Salami, Y. Ebazadeh, and V. Khajehvand, "Cost-effective secure key exchange scheme in Fog Federation," *Iran J. Comput. Sci.* vol. 4, no. 3, pp. 1–13, 2021.

[25] I. Konnov, *Handbook of Model Checking* Springer International Publishing AG, Cham, Switzerland, 2018.
[26] Y. Salami and V. Khajehvand, "LSKE: lightweight secure key exchange scheme in fog federation," *Complexity*, vol. 2021, Article ID 4667586, 2021.
[27] A. Gotsman, F. Massacci, and M. Pistore, "Towards an independent semantics and verification technology for the hlpsl," *Electronic Notes in Theoretical Computer Science*, vol. 135, pp. 59–77, 2005.
[28] M. Wazid, A. K. Das, N. Kumar, and A. V. Vasilakos, "Design of secure key management and user authentication scheme for fog computing services," *Future Generation Computer Systems*, vol. 91, pp. 475–492, 2019.
[29] D. Dolev and A. C. Yao, "On the security of public key protocols," *IEEE Transactions on Information Theory*, vol. 29, no. 2, pp. 198–208, 1983.
[30] D. Basin, S. Mödersheim, and L. Viganò, "An on-the-fly model-checker for security protocol analysis," *Computer Security -- ESORICS 2003*, vol. 2808, pp. 253–270, 2003.
[31] M. Turuani, "The CL-atse protocol analyser," in *Term Rewriting and Applications*, F. Pfenning, Ed., pp. 277–286, Springer, Berlin Germany, 2006.
[32] A. Biere and D. Kröning, "SAT-based model checking," in *Handbook of Model Checking*, E. M. Clarke, T. A. Henzinger, H. Veith, and R. Bloem, Eds., Springer International Publishing, Berlin Germany, pp. 277–303, 2018.
[33] L. Vigan, "Automated security protocol analysis with the avispa tool 1," *Electronic Notes in Theoretical Computer Science*, vol. 155, no. 1, pp. 61–86, 2006.
[34] Y. Salami, V. Khajehvand, and E. Zeinali, "E3c: a tool for evaluating communication and computation costs in authentication and key exchange protocol," 2022, https://arxiv.org/abs/2212.03308.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 1729413, 10 pages
https://doi.org/10.1155/2022/1729413

Research Article

Internet of Vehicles Information Processing Method with Vehicle-Mounted Cloud Grid as the Underlying Data Fusion Structure

Yibo Han,[1] **Xia Li**,[2] **Xiaocui Li**,[3] **Zhangbing Zhou**,[3] **and Jingshuo Li**[4]

[1]*Nanyang Big Data Research Institute, Nanyang Institute of Technology, Nanyang, Henan 473000, China*
[2]*School of Communication, Nanyang Institute of Technology, Nanyang, Henan 473000, China*
[3]*School of Information Engineering, China University of Geosciences (Beijing), Beijing 100083, China*
[4]*School of Electronic Information Engineering, Henan Polytechnic Institute, Nanyang, Henan 473000, China*

Correspondence should be addressed to Yibo Han; hanyibo@nyist.edu.cn

Received 13 July 2022; Revised 4 August 2022; Accepted 26 August 2022; Published 26 September 2022

Academic Editor: Ke Gu

With the development of electronic information network technology, large car networking systems can produce all kinds of data such as text, images, and videos, a large number of heterogeneous data, different features of heterogeneous data, and different data structures. In the Internet of vehicles, the beacon message generation strategy needs to be researched and designed on the premise of meeting the requirements of vehicle location accuracy and wireless communication performance. According to the Kalman filter differential prediction equation, the message generation model of Kalman filter beacon is established. In the deep learning research on the underlying data fusion algorithm, the most effective way to solve the problem of insufficient integration degree between the underlying data is to improve the data quality and ensure data sharing and reuse between multisource heterogeneous data. Therefore, the D-S evidence theory fusion model and rough set underlying model are proposed in the vehicle-mounted cloud network. Among them, the D-S evidence theory fusion model ensures the improvement of underlying data quality, forms effective rule combination, and reduces conflicts between rules through filtering evidence theory. The rough set underlying data fusion model optimizes the underlying data of each device by improving the rough set attribute reduction method of particle swarm optimization algorithm.

1. Introduction

The Internet of vehicles (IOV) is a combination of vehicles, mobile Internet, and Internet of things. It refers to the network interconnection between vehicles and people, vehicles and roads, and vehicles and supporting infrastructure through vehicle-mounted devices or related mobile devices and the use of communication technology, intelligent terminals, and vehicle navigation systems. Thus, the network system of intelligent supervision, vehicle scheduling, and other related functions can be effectively implemented for the whole ecosystem of people, cars, roads, and the environment [1]. Compared with the traditional vehicle management system, the data scale, data types, and real-time data acquisition have made substantial progress. On this basis, more professional data processing technology and the application of more abundant vehicle management services for the heterogeneous underlying data fusion of the Internet of vehicles are still needed.

Although the vehicle-mounted cloud network is also a wireless network, it has different characteristics from other wireless networks because it covers the road network: (1) Node mobility: the fast and frequent movement of vehicles not only leads to the dynamic change of V2X wireless communication frequent short links (short link connection) and link capacity but also makes the dynamic change of network topology, which makes it impossible to form a stable topology. (2) The vehicle position can be predicted. The moving trajectory of a vehicle is limited, and it always drives along the given direction of the road. Its driving speed

is affected by the moving state of the vehicle in front, and its position and moving direction and speed are predictable to a certain extent. (3) Local information acquisition: because the storage and computing functions of vehicle nodes are very limited, they cannot store a large amount of data or carry out complex application calculations. Therefore, in most cases, vehicle nodes only need to obtain local traffic information and realize complex applications by using the network. (4) Energy is basically unrestricted. The vehicle can carry and continuously supplement external energy to continuously power the on-board equipment, making the on-board equipment have strong performance. (5) GPS, vehicle sensor-assisted positioning: at present, many vehicles carry GPS and other on-board sensors for vehicle positioning, which can not only ensure that the vehicle has accurate global synchronization clock but also provide accurate position, speed, direction, and other state information for the vehicle. Through these devices, communication and interaction between vehicles can be well supported.

The big data scenario formed by the periodic dissemination of safety information by vehicle nodes in the Internet of vehicles is exactly in line with the extended application field of data fusion in the Internet of vehicles. The knowledge discovery of massive vehicle node information through the data fusion algorithm of universities can accurately locate the vehicle position information. Based on the vehicle random path prediction model, a moving vehicle location data update strategy based on BM-KFFPP and a beacon information generation strategy based on threshold were proposed on the premise of meeting the requirements of vehicle information accuracy and wireless communication performance. Aiming at the problem of vehicle position information loss in the process of beacon message transmission, a beacon lost data complement algorithm based on least square support vector machine was proposed, which was simulated and verified by example. The multisource heterogeneous underlying vehicle cloud network data fusion framework uses the ontology idea and D-S evidence theory method to perform feature fusion on the underlying data to reduce data redundancy, improve data accuracy, and optimize decision-making efficiency. Section 2 describes the related work, Section 3 describes the overall design of data fusion at the bottom of vehicle cloud network, Section 4 provides example verification, and Section 5 gives the conclusion.

2. Related Work

2.1. Internet of Vehicles. The Internet of vehicles is an information interaction network composed of vehicle speed, route, and location. It is a vehicle-network joint technology that is aimed towards safety, energy saving, environmental protection, and information communication [2]. The Internet of Vehicles realizes the collection of vehicle, traffic environment and road information through electronic equipment such as RFID, GPS, Beidou positioning, high-definition cameras, sensors, and image processing. According to certain communication protocols and standards, wireless communication or information exchange can be carried out among one person, one road, one environment, one network, and one infrastructure. The cloud computing center uses computer technology to process vehicle data information, so as to timely report the road conditions, calculate the best route of different vehicles, arrange the signal cycle, and achieve intelligent scheduling, monitoring, and management of vehicles, people, and roads. The Internet of vehicles is the inevitable result of human society entering the information age and automobile age and is the extension and application of Internet of things technology in the field of transportation [3]. The formation of Internet of vehicles industry can be regarded as the collection of vehicle-borne information service system and intelligent traffic system (ITS) [4]. The vehicle information service system is to point to by vehicle electronic equipment in a timely manner to understand the status of the vehicle driving and the information service system. The intelligent transportation system mainly refers to providing the traffic information system, vehicle management system, etc., in an important mode of the future intelligent car [5]. The underlying data fusion is a multidisciplinary computer technology, which has related applications in many fields. The United States is the fastest developing and earliest in the underlying data fusion technology. In the early stage, the underlying data fusion was mainly applied to the military field. As early as the 1970s, the United States funded the research on sonar signal understanding and fusion and developed a series of CISR (Command, Control, Communications, Computers, Intelligence, Surveillance, and Reconnaissance) and IW (Intelligence Weapon) systems [6, 7].

2.2. Underlying Data Fusion Technology. After years of research and development, multisensor underlying data fusion technology has achieved fruitful theoretical and application results. The progress of computer technology, communication technology, and data processing technology also provides new impetus for the development of multisensor data fusion technology. Among them, real-time processing technology plays an increasingly important role in the multisensor underlying data fusion system and is applied more widely. As the system structure becomes more complex and the data scale becomes larger, a real-time data processing strategy is needed to maintain the stable operation of the system and ensure the real-time utilization value of the collected data flow [8]. Due to the heterogeneity of data acquired by multiple sensors, distributed underlying data fusion will also bring into play its potential value. In order to meet the various needs of underlying data fusion, many underlying data fusion models have been proposed, which can be roughly divided into two categories: (1) functional fusion model, which is mainly constructed by the sequence of functions realized by the underlying data fusion in each node, and (2) data fusion model, which is mainly constructed through data extraction in the underlying data fusion [9]. Literature [10] is a fusion model of human-machine information interaction. This model takes into account the role of users in the terminal of the Internet of vehicles and mainly uses the role and reaction of people in

the fusion process. The classic fusion method is modified to enhance the independence of different fusion stages, which is beneficial for better coordination of different fusion models. However, the weakness of this fusion model is that the user participates too much in the process, resulting in the lack of entity abstraction. The fusion model in literature [11] is applied to the Internet of vehicles in the Internet of things: the whole realization process is the network connection between vehicles and sensors of the surrounding environment infrastructure equipment, and the key to this fusion is the real-time correlation of data [12]. This method is divided into two different systems: data filtering and underlying data fusion, to ensure the accuracy of the data. The adaptive exponential smoothing method is adopted to rapidly fuse the road segment travel time based on a fixed detector and floating car under different reliability, so that the road segment travel time can be obtained accurately and efficiently [13]. Considering the continuity of traffic status (travel time) of adjacent sections, a fuzzy regression model is proposed, and only a small amount of floating car data are needed to accurately predict the travel time of interflow interruption [14].

2.3. Dynamic Road Network Induction. The heterogeneous system data: floating car GPS data, coil detection data, and video detection data, are fused and matched to the GIS to achieve dynamic road network induction [15]. By filtering out the floating vehicle data which are greatly affected by signal control and mining the historical floating vehicle data, a road travel time estimation method with missing signal timing information is proposed. For coarse-grained floating vehicle data [16], the average absolute error of this estimation method is superior to that of the traditional direct and indirect methods, and a new method for real-time capacity of urban road network of Internet of vehicles based on immune theory is proposed. This method introduces the immune network theory into the vehicle self-organization network and obtains the traffic flow in the road network through the recognition of statistical antigen and antibody, which not only solves the real-time problem but also solves the problem that traditional methods need to establish an accurate mathematical model [17]. In view of the characteristics of traffic information in the environment of Internet of vehicles, the improved artificial neural network and improved support vector regression method are used to predict the road travel time [18].

The research on the underlying data fusion in the aspect of vehicle-mounted cloud network mainly includes the following: Literature [19] puts forward the ontological research on ontology fusion methods among various specialties of high-speed railway in China. Literature [20] uses ontology to solve the problem of disunity between developers and operators of railway information in Europe and realizes data sharing and interoperability. Literature [21] introduces the overall design scheme and key technologies of railway big data platform. Literature [22] mainly studies and analyzes the standardization of high-speed railway data and proposes the construction of standard indicators. Literature [23] is aimed at the problem of multisource heterogeneity of data information in intelligent maintenance decision of the high-speed railway signal system, where the ontology fusion algorithm was used to reduce the computational complexity and time complexity compared with the classical closure algorithm, and the running time of the algorithm was also much lower than that of the classical closure algorithm. Finally, a unified framework for heterogeneous underlying data fusion and intelligent decision-making of the high-speed railway signal system is proposed. The experimental verification and analysis show that the diagnostic accuracy of this framework is significantly improved and its applicability is good. Literature [24] takes the high-speed railway as the research object and adopts feedback D-S evidence theory to optimize the algorithm for problems such as insufficient integration degree caused by multichannel data transmission during train operation, which is verified feasible by experiments. Through the research and analysis of the working state of the track circuit, the characteristic attributes of the data are extracted in literature [25], and the fusion of the two kinds of monitoring data of the signal equipment and the microcomputer equipment is realized, which is of practical significance for the research on the fusion of the underlying data of the track circuit. Literature [26] takes the TDCS/CTC system as the main research object and proposes a system hierarchical information aggregation scheme for the problem of high real-time requirements of massive train control data. The scheme was successfully run in the railway general dispatching command center, proving its practicability. Literature [27] uses the Kalman underlying data fusion algorithm to conduct real-time modeling of random vibration interference with the second-order autoregressive model and builds the strapdown inertial testing method for tracking geometric parameters. It is verified that the test degree of the modified system is effectively improved. Through the analysis and study of data transmission characteristics of railway train control in literature [28], the importance of fusion technology to improve safety performance of train control is clearly identified. In this regard, the colored Petri net (CPN) algorithm is used for modeling, and the experimental results show that this method is effective and feasible. Studies have shown that the different semantic alignment of vehicle-mounted data records in cloud network systems can easily lead to the problem of data conflict and information islands. The multisource heterogeneous data fusion method can integrate structured, semi-structured, and unstructured data to ensure data consistency and provide data guarantee for intelligent decision-making in railway operation and maintenance systems.

3. Overall Design of Data Fusion at the Bottom of Vehicle Cloud Network

3.1. Architecture of Underlying Data Fusion. The data fusion system for heterogeneous data in the Internet of vehicles adopts the architecture of C/S (client and server) and B/S (browser and server) for mixed development. The client uses mobile devices with the Android operating system to obtain corresponding data. The heterogeneous data are then uploaded to the Tomcat server through wireless LAN 802.11

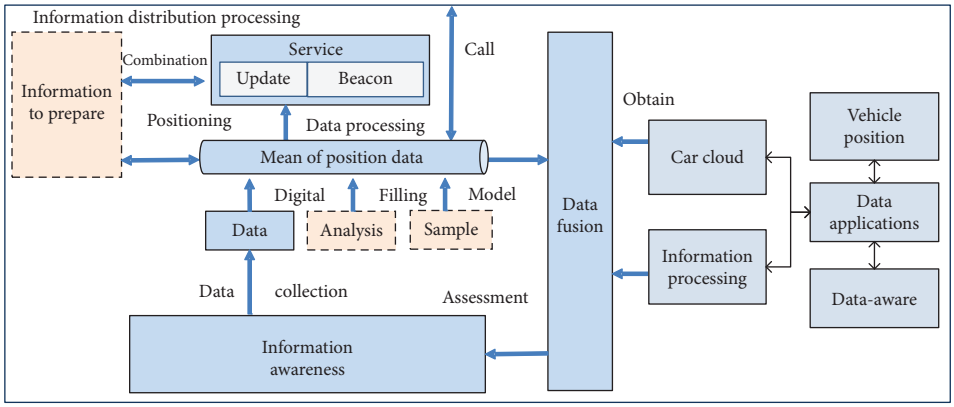


Figure 1: Overall architecture of the system.

(Wi-Fi) or GPRS traffic data, and the Tomcat server cleans the heterogeneous data and transfers them to the corresponding cloud storage platform in real time. The server side mainly refers to the cloud storage platform where data are stored, while the data fusion display side extracts data from the cloud storage platform through related services to achieve related fusion. Figure 1 shows the overall architecture of the system.

The data acquisition layer is composed of various terminal devices installed on the vehicle, mainly using mobile devices with the Android operating system, such as mobile phones, tablets, vehicle intelligent rearview mirrors, and other devices. By using different sensors such as GPS and camera of mobile devices, text, picture, and video data are collected according to the different running time of vehicles. At the same time, the collected data are uploaded to the Tomcat server in real time through 3G/4G or Wi-Fi, and then the data are saved to the corresponding cloud storage platform in real time after relevant service processing.

3.2. Underlying Data Storage and Fusion. The data storage layer uses the cloud storage mode and manages data of the entire system. The Hadoop distributed file system (HDFS) stores uploaded video data in the data acquisition layer, the distributed nonrelational database MongoDB stores image data, and the relational database MySQL stores text data. The classified storage method is adopted to improve the storage efficiency and later utilization efficiency of data. The data interaction between the data storage layer and the data acquisition layer is shown in Figure 2.

The data fusion display layer is mainly used to fuse the three heterogeneous data—text, picture, and video—stored in the data storage layer and display them to the web front end, so as to satisfy people's data utilization. This layer mainly develops corresponding web-side applications through the node.js platform and Java Web to realize the real-time vehicle position display, path tracking, and real-time upload data fusion and display on the browser side.

Packet routing data transmission is one of the key technologies for information transmission in the Internet of vehicles, and it is also the most common data transmission technology. It has important theoretical significance and practical application value and has a profound impact on the communication networking capability of the Internet of vehicles. In the environment of high mobility of vehicle nodes, topology changes greatly and links are unreliable. Networking routing is a complex task. The key point of routing design is to design and maintain routes from the source node to the destination node to ensure real-time, complete, and effective data transmission. The main challenges of routing protocol design include transmission delay from the source node to the destination node, routing cost and stability, and reliability and QoS. Performance indicators include average end-to-end latency, route cost, packet delivery rate, and throughput. Figure 3 shows the routing protocol diagram of packet transmission.

Routing protocols based on the topology structure use existing links to transmit data, which is inefficient, and the existing protocols have poor data transmission effect in high dynamic environments. For typical topology-based protocols such as DSDV (destination-sequenced distance-vector) routing, AODV (ad hoc on-demand distance vector) routing, and DSR (dynamic super-resolution) routing, NS2 is used to simulate urban scenarios. The simulation environment includes a 4 km section, 80 vehicle nodes, 90 min time, 6 MHz transmission rate, and 28.8 dbm power. The speed is 10 m/SEC, and the measurement analysis is simulated from the parameters such as throughput, packet loss rate, and data collision. Compared with the other two routing protocols AODV and DSDV, the DSR routing protocol has higher performance in traffic safety transmission in urban environments. Table 1 compares the transmission performance of typical routing protocols based on topology.

3.3. Bottom-Level Multiple Information Fusion of Internet of Vehicles Information Processing. More research on information fusion focuses on multisensor information fusion and comprehensive processing, so as to get more accurate and reliable conclusions. The multisensor data fusion is generally divided into three levels: data layer, feature layer, and decision layer.

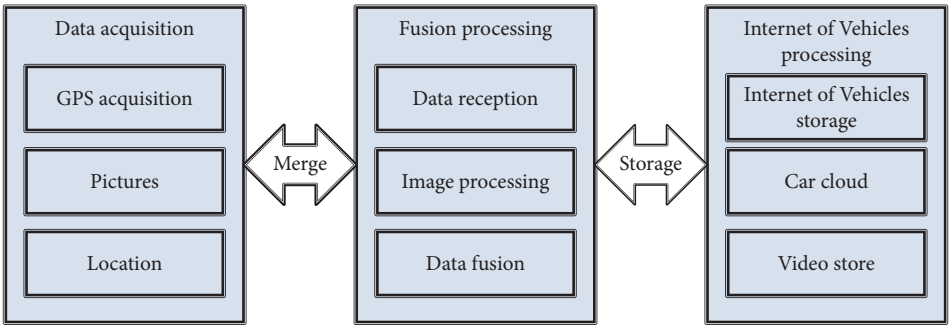


Figure 2: Data interaction diagram.

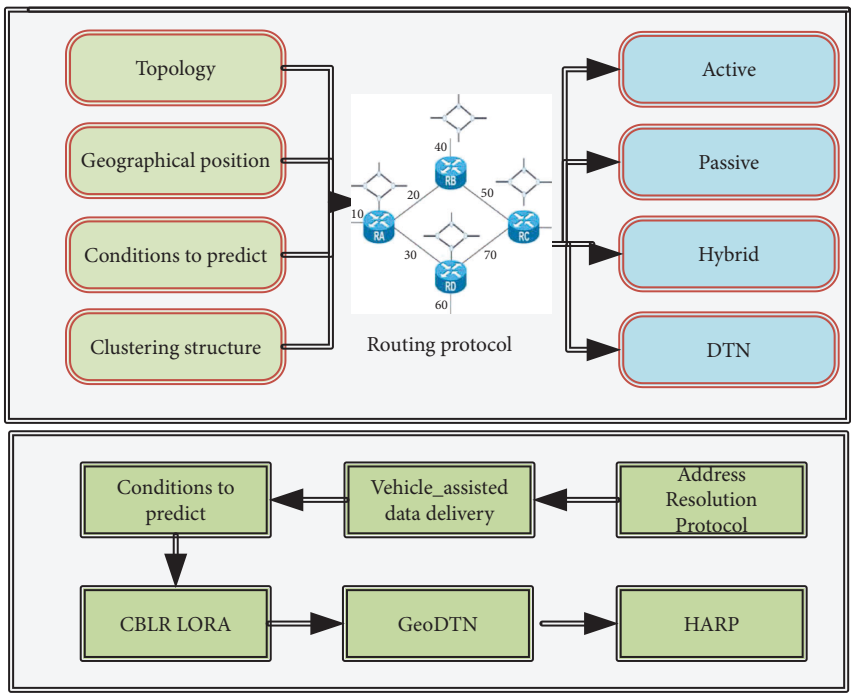


Figure 3: Routing protocol diagram of packet transmission.

Table 1: Comparison of typical routing protocols based on topology.

Parameter	Routing protocol type	Throughput	Packet loss rate	Data collision
DSDV	Active table drive	High	Low	Middle
AODV	Active equation	Middle	Middle	High
DSR	Passive equation	Low	High	Low

Data layer fusion requires that sensors observe the same physical quantity. If multiple sensors observe different physical quantities, data can only be fused at the feature layer or decision layer. Data layer fusion is the direct fusion of the observation data of the observation sensor. The original measurement data of each sensor are directly fused without analysis and processing. This fusion has the advantage of collecting as much field data as possible, providing raw concrete information for other fusion levels. However, it requires a large number of sensors, high processing cost, long processing time, and poor real-time performance.

Feature layer fusion is the fusion of the information collected by the underlying sensor. Firstly, the original sensor data are collected, and then the features of the original data are analyzed and extracted. Finally, the data are classified and collected according to the characteristic information.

Decision level fusion is a kind of high level fusion, that is, the optimal decision is made according to certain judgment criteria. In order to meet the needs of specific decision problems, this paper makes full use of all kinds of feature information of measurement objects extracted from feature level fusion and adopts appropriate fusion technology to realize it.

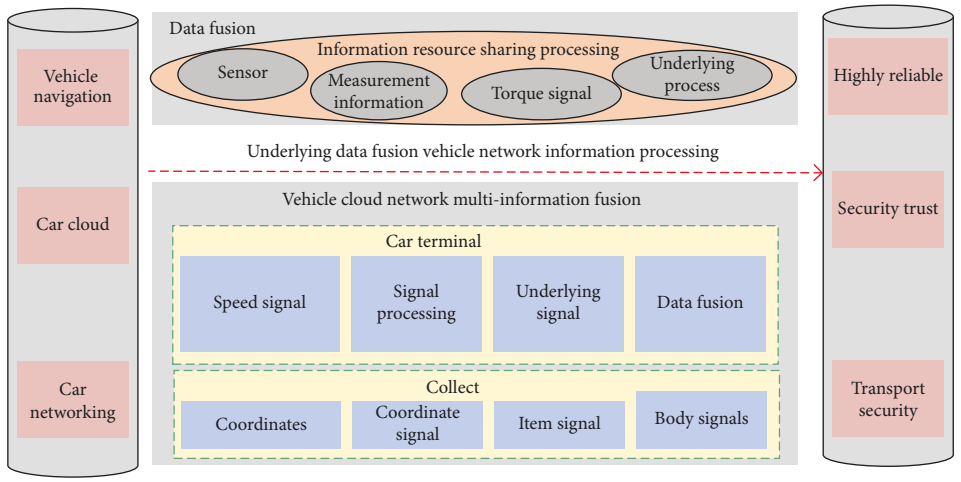


Figure 4: Vehicle-mounted terminal multiple information fusion model.

Assume that the probability model of the measured value of sensor J is described by the Gauss probability distribution function:

$$F(X) = e^{\left[(1/2\delta^2)(x-x_i)^2\right]}\left(\frac{1}{2\sqrt{\alpha\delta^2}}\right). \tag{1}$$

As shown in Figure 4, there are three relationships between sensor I and sensor J: (1) sensor I and sensor J are independent; (2) sensor I strongly supports sensor J, and sensor J weakly supports sensor J; and (3) sensor J strongly supports sensor I, and sensor I weakly supports sensor I. For two sensors, if the measurement information of the two sensors is inconsistent, there must be sensor measurement error, and the fusion of the two sensors has no practical significance. When the measured values of the two sensors support each other, the fusion results minimize the uncertainty and unreliability of the measured values.

The rules in fuzzy theory select the minimum or maximum value directly from the dataset u, which is not suitable for the decision value of the fusion sensor k. Therefore, the following algorithm is improved:

$$f\left(u_x, u_y\right) = \frac{\left[(k-1)^2\left(u_x u_y\right) + k\left(u_x + u_y\right)\right]}{\left[1 + k^2\right]}. \tag{2}$$

The channel load data sequence is used as the dependent variable reference data column, and the communication environment parameters and traffic flow data sequence are used as the independent variable comparison data column. According to the following calculation steps, the correlation degree of the main influencing factors of channel load in this section is determined:

Step 1 (dimensionless processing of data): before the association analysis, in view of the different dimensions of the original data, it is necessary to carry out the dimensionless processing of range standardization on the original data and transform the data of different time and space into comparable standardized data. For the processing of original data, the standardized transformation method is adopted.

Step 2 generate the difference data sequence. The difference in the standard data series after dimensionless processing is calculated, that is, the absolute value of the difference between the standardized dependent variable series and the independent variable series.

Step 3 calculate the correlation coefficient according to Deng's correlation degree calculation method:

$$\begin{aligned}\delta(j) = &\left(\min Z_i(t) - \min Z_j(t)\right) \\ &+ \rho\left(\max Z_i(t) - \max Z_j(t)\right).\end{aligned} \tag{3}$$

Step 4 rank the calculated correlation coefficients according to their sizes. The larger the value is, the stronger the correlation degree between the influencing factor and the channel load is at time T; otherwise, the correlation degree is weaker.

In order to maximize transmission, the delay forwarding protocol is distance-based forwarding in essence. The setting of forwarding rules tries to make each distributed message cover the "jump" range as long as possible, that is, the node furthest from the source node is selected for forwarding, and the single hop is transmitted to the subsequent node for the next hop relay. Sending the end node sends messages, and the nodes within its transmission range receive messages. When receiving messages, the receiver starts the setting of delay forwarding timer T, and its model is as follows:

$$T(x) = \triangle T\left(\max \frac{d_{ij}}{R_{IJ}}\right) + T_{\min}. \tag{4}$$

In message distribution, it is assumed that the vehicle node VS sends a new message, and all vehicle nodes that receive the message directly within the transmission range consider themselves potential relay and forwarding nodes (RNs). Nodes elected as RNs by policy will distribute

messages to other vehicles in the transmission range; if not elected, the message is discarded. If two RNs are close to each other, because the transmission coverage area between them is basically the same, the situation of conflicting interference may occur. In addition, the two elected RNSs (radio network subsystems) also need to be in the communication range to ensure successful reception of relay messages.

The appropriate transmission range depends on the transmission rate of the channel and the target bit error rate. Therefore, the protocol is usually designed to maximize the transmission range of a relay configuration that uses as few destination RNs as possible, as long as transmission is ensured.

Beaconless message distribution usually uses the following basic methods to suppress redundant spurious distribution:

(1) Message distribution based on probability: the vehicle node receiving the new message selects the node for message distribution based on probability P, which can be set to a fixed value or dynamically changed. According to the dynamic traffic density scenario, the probability P can be dynamically adjusted according to the number of receiving groups or the interest degree of neighboring vehicles in the message, which has a wider coverage and higher efficiency than the algorithm with fixed probability.

(2) Counter-based message distribution: the vehicle node receiving the new message sets the counter and waiting time. When the vehicle receives the same message within the waiting time, the counter will be accumulated, and at the end of the waiting time period, the counter value will be compared with the set value (which can be fixed or dynamically adjusted according to some rules). If the counter value is less than the threshold, the message is forwarded. Otherwise, the message is discarded.

(3) Message distribution based on delay forwarding timer: the vehicle node receiving new messages sets timer T to delay message distribution. When the value of T expires, the message is forwarded. However, if the same data are received during the delay forwarding timer period, the node is forbidden to forward messages, which is called the suppression rule. The message distribution rule based on delay timer T is as follows: the vehicle with the lowest timer value will distribute the message first, and the other adjacent vehicle nodes whose delay timer does not reach will cancel the message distribution. The setting of delay timer is usually related to distance. The node with a larger distance from the sending vehicle has a smaller T value of delay timer and can obtain the priority of forwarding.

4. Example Verification

The floating vehicle collects the number of vehicles, traffic flow density, speed, and other data within the vicinity of 500 meters once every 5 minutes, with a data scale of 300.

Table 2: Simulation parameters of urban road intersections.

Parameter	Value
Length of import (export)	1.5 km
Vehicle position sampling interval	6 s
Average traffic flow in one lane	1,660 vph
Number of incoming lanes	6
Number of exit lanes	6
Average headway	2.5 s
Average velocity	58 km/h
The signal cycle	65 s
Green letter than	2/5
Green time	36 s
Green interval	3.8 s

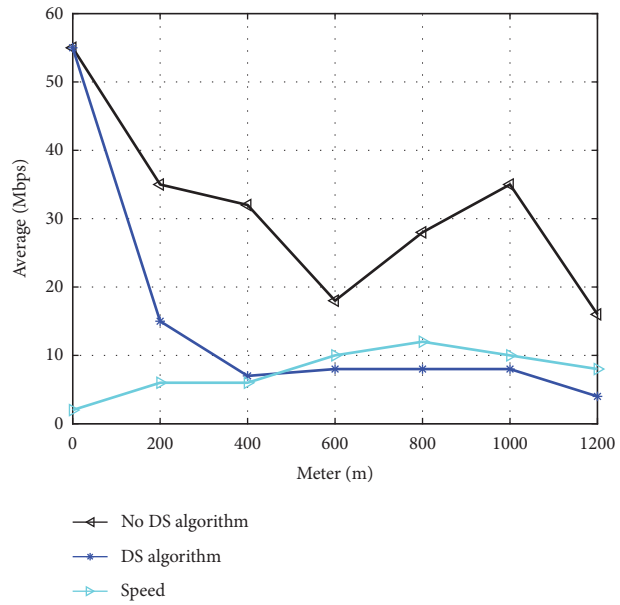


Figure 5: Load-distance relationship of the channel (exit channel).

The last 24 data are taken as test samples, and the channel load sequence is taken as the research object to conduct short-term channel load prediction experiments. According to the requirements of periodic beacon message generation rate and packet size in vehicle-mounted security applications, the size of beacon message is 800 bytes and the message generation rate is 15 bits/second, that is, the beacon message rate of each vehicle is 96 K. The initial communication distance is 250 meters, the initial carrier detection distance is 500 meters, the maximum communication distance is 500 meters, the maximum carrier detection distance is 1500 meters, the minimum allowable channel load threshold is 3 Mbps, the maximum allowable channel load threshold is 6 Mbps, the power adjustment step is 0.01, the channel load prediction period is 1 minute, the average velocity is 58 km/h, and the green interval is 3.8 s. The simulation parameter settings are shown in Table 2.

After using the DS algorithm, the average channel load generated by the beacon message near the intersection is shown in Figure 5.

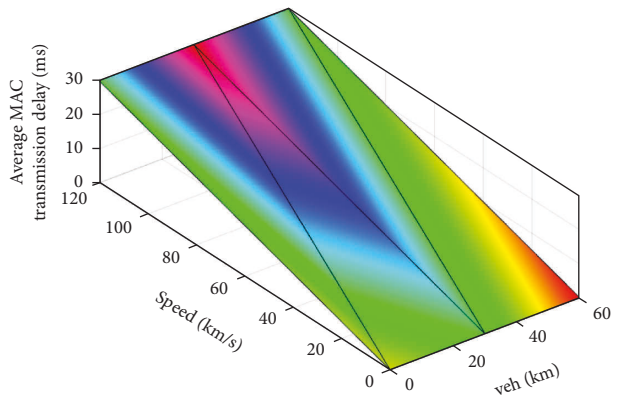


Figure 6: Average MAC layer transmission delay at different vehicle densities and speeds.

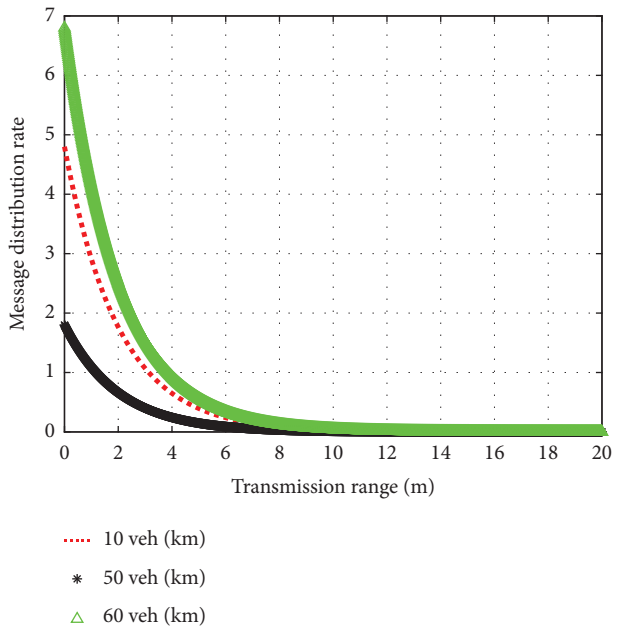


Figure 7: Relationship between the message distribution rate and the transmission range at different vehicle densities.

As shown in Figure 5, after a certain oscillation period, the channel load converges to the preset threshold range of 3–6 Mbps. The closer to the intersection, the greater the channel load value. When vehicles leave the intersection, the channel load value begins to decrease significantly. At 400 m away from the intersection, the channel load begins to increase again.

Figure 6 sets the average transmission delay and MAC layer redundant stray distribution probability of the actual experiment under different vehicle densities P and different vehicle speeds on the experimental road with 3D graphics and establishes a model to compare the simulation results. The linear surface in the figure represents the simulation results of the real road, and the model analysis results are represented by dashed lines. The average MAC layer transmission delay time of v (80, 100, and 120 km/h) is plotted.

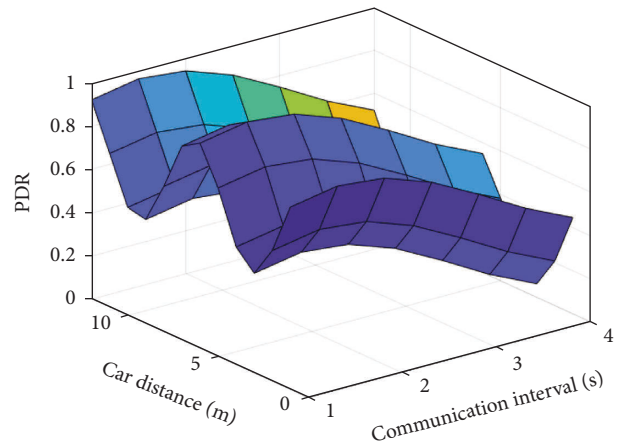


Figure 8: PDR without underlying data fusion algorithm (Car No. 7).

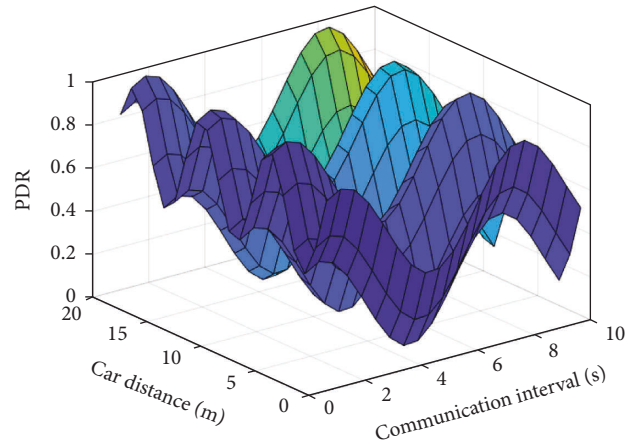


Figure 9: PDR using the underlying data fusion algorithm (Car No. 7).

Figure 7 shows the relationship between the message distribution rate and the transmission range at different vehicle densities. As can be seen from Figure 7, for values below 400 m, the probability of repeated forwarding is very low, the message distribution efficiency is high, and the influence of vehicle density can be ignored.

As for the relationship between the PDR and the communication distance between two vehicles, the expected PDR is set as 9, the minimum retransmission distance is 50 meters, the maximum retransmission times is 5, the retransmission interval is 50 meters, and the maximum retransmission distance is 500 meters. The American Denso V2X vehicle-mounted machine supporting DSRC protocol was selected as the workshop communication platform, and the underlying data fusion decision algorithm was set according to the above parameters to conduct data acquisition and analysis on the position message generated by the No. 7 and 8 vehicles in the fleet and the communication process with the neighboring vehicles. PDR value pairs without transfer mode algorithm and with the transfer mode algorithm are shown in Figures 8 and 9.

5. Conclusion

For real-time and heterogeneous data acquired in the data awareness layer, relevant databases or distributed file systems of different cloud storage models are adopted for different data formats. The classification and storage of heterogeneous data facilitates related operations in data fusion. A threshold-based message generation strategy for vehicle beacons is proposed. According to the relation of Kalman filter prediction time domain, the message generation model of Kalman filter prediction beacon is established. According to the measured value and preset threshold of channel load, a beacon message generation model and strategy with time interval adaptive adjustment are established. This distributed strategy can effectively reduce channel load, avoid channel congestion, and ensure the fairness of message generation and transmission of each node while meeting the requirements of location information accuracy required by the application of Internet of vehicles. In the case of the large-scale road network, it is necessary to reduce the burden of path calculation in the Internet of vehicles data center and optimize and improve the efficiency of path calculation. In addition, the method in this paper does not consider the special needs of customers, such as the shortest route, the lowest driving cost, the safest travel, and other personalized needs. The above problems are the key points to be improved and perfected in the next step.

Data Availability

Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

Consent

Informed consent was obtained from all individual participants included in the study references.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This work was supported by the 2019 Cross Science Research Project of Nanyang Institute of Technology (Grant no. 201913502) (Research on Intelligent Mining and Recommendation of Zhang Zhongjing Prescription Based on Deep Neural Network) and Henan Science and Technology Plan Project (Grant no. 222102210134) (Research on Key Technologies of Cloud Security Desktop Based on Kunpeng Architecture).

References

[1] F. E. Da Silva Barbosa, F. F. De Mendonça Júnior, and K. L. Dias, "A platform for cloudification of network and applications in the internet of vehicles," *Transactions on Emerging Telecommunications Technologies*, vol. 31, no. 2, pp. 67–78, 2020.

[2] D. B. Rawat, R. Alsabet, and C. Bajracharya, "On the performance of cognitive internet-of-vehicles with unlicensed user-mobility and licensed user-activity," *Computer Networks*, vol. 137, pp. 98–106, 2018.

[3] F. Sattar, F. Karray, M. Kamel, L. Nassar, and K. Golestan, "Recent advances on context-awareness and data/information fusion in ITS," *International Journal of Intelligent Transportation Systems Research*, vol. 14, no. 1, pp. 31–19, 2016.

[4] R. G. A. Congalton, "Comparison of methods for determining forest composition from high-spatial-resolution remotely sensed imagery," *Forests*, vol. 12, pp. 156–167, 2021.

[5] J. Du, R. Guo, G. Suo, and X. Zhang, "A multisource alarm information fusion processing method for network attack situation," *IOP Conference Series: Materials Science and Engineering*, vol. 466, pp. 012050–012478, 2018.

[6] Y. Li, H. Wang, W. Wang, S. Liu, and Y. Xiang, "Reducing the risk of rear-end collisions with infrastructure-to-vehicle (I2V) integration of variable speed limit control and adaptive cruise control system," *Traffic Injury Prevention*, vol. 17, no. 6, pp. 597–603, 2016.

[7] S. S. Monfort and J. M. Nolan, "Trends in aggressivity and driver risk for cars, SUVs, and pickups: vehicle incompatibility," *Traffic injury preventio*, vol. 20, no. 1, pp. 92–96, 2020.

[8] F. M. A. Henk, "Framework for automated acquisition and processing of as-built data with autonomous unmanned aerial vehicles," *Sensors*, vol. 19, no. 20, pp. 3224–3234, 2019.

[9] X. Xu, B. Shen, and S. Ding, "Service offloading with deep Q-network for digital twinning empowered internet of vehicles in edge computing," *IEEE Transactions on Industrial Informatics*, vol. 5, no. 9, pp. 789–802, 2020.

[10] C. Wu, X. Chen, T. Yoshinaga, Y. Ji, and Y. Zhang, "Integrating licensed and unlicensed spectrum in the internet of vehicles with mobile edge computing," *IEEE Network*, vol. 33, no. 4, pp. 48–53, 2019.

[11] R. Hao, H. Li, and Y. Dai, "Querying in internet of things with privacy preserving: challenges, solutions and opportunities," *IEEE Network*, vol. 3, no. 6, pp. 121–128, 2018.

[12] S. Du, "Functional classification of urban parks based on urban functional zone and crowd sourced geographical data," *ISPRS International Journal of Geo-Information*, vol. 10, pp. 763–778, 2021.

[13] M. A. Che, "Collaborative sensing system for farmland water conservancy project maintenance through integrating satellite, aerial, and ground observations," *Water*, vol. 13, pp. 278–298, 2021.

[14] W. Wang and M. Zhang, "Tensor deep learning model for heterogeneous data fusion in internet of things," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 4, no. 1, pp. 32–41, 2020.

[15] K. D. Returi and Y. Radhika, "An artificial neural networks model by using wavelet analysis for speaker recognition," *Advances in Intelligent Systems and Computing*, vol. 340, pp. 859–874, 2015.

[16] G. Sun, S. Sun, H. Yu, and M. Guizani, "Toward incentivizing fog-based privacy-preserving mobile crowdsensing in the internet of vehicles," *IEEE Internet of Things Journal*, vol. 7, no. 5, pp. 4128–4142, 2020.

[17] M. van der Kam and W. van Sark, "Smart charging of electric vehicles with photovoltaic power and vehicle-to-grid technology in a microgrid; a case study," *Applied Energy*, vol. 152, pp. 20–30, 2015.

[18] S. A. Elsagheer Mohamed and K. A. Alshalfan, "Intelligent traffic management system based on the internet of vehicles

(IoV)," *Journal of Advanced Transportation*, vol. 2021, no. 4, Article ID 4037533, 23 pages, 2021.

[19] D. Y. Kim, M. Jung, and S. Kim, "An internet of vehicles (IoV) access gateway design considering the efficiency of the in vehicle ethernet backbone," *Sensors*, vol. 21, no. 1, pp. 98–121, 2020.

[20] H. Teng, Y. Liu, A. Liu et al., "A novel code data dissemination scheme for Internet of things through mobile vehicle of smart cities," *Future Generation Computer Systems*, vol. 94, pp. 351–367, 2019.

[21] K. M. Alam, M. Saini, and A. El Saddik, "Toward social internet of vehicles: concept, architecture, and applications," *IEEE Access*, vol. 3, pp. 343–357, 2015.

[22] K. M. Alam, M. Saini, and A. E. Saddik, "Workload model based dynamic adaptation of social internet of vehicles," *Sensors*, vol. 15, no. 9, pp. 23262–23285, 2015.

[23] Y. Li, M. Wang, R. Zhu et al., "Intelligent augmented keyword search on spatial entities in real-life internet of vehicles," *Future Generation Computer Systems*, vol. 94, pp. 697–711, 2019.

[24] M. K. Priyan and G. U. Devi, "A survey on internet of vehicles: applications, technologies, challenges and opportunities," *International Journal of Advanced Intelligence Paradigms*, vol. 12, pp. 98–121, 2019.

[25] G. Pirelli, R. K. Otto, and A. Estoup, "Using internet and social media data as collateral sources of information in forensic evaluations," *Professional Psychology: Research and Practice*, vol. 47, no. 1, pp. 12–17, 2016.

[26] C. Im and D. Kim, "Real-time traffic information and road sign recognitions of circumstance on expressway for vehicles in CITS environments," *Journal of the Institute of Electronics and Information Engineers*, vol. 54, no. 1, pp. 55–69, 2017.

[27] C. Liu, J. Zhang, G. Li, S. Gao, and Q. Zeng, "A two-layered framework for the discovery of software behavior: a case study," *IEICE - Transactions on Info and Systems*, vol. 101, no. 8, pp. 2005–2014, 2018.

[28] D. Lin, F. Labeau, Y. Yao, A. V. Vasilakos, and Y. Tang, "Admission control over internet of vehicles attached with medical sensors for ubiquitous healthcare applications," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 4, pp. 1195–1204, 2016.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 3214913, 10 pages
https://doi.org/10.1155/2022/3214913

Research Article

A New Certificateless Signcryption Scheme for Securing Internet of Vehicles in the 5G Era

Beibei Cui,[1,2] **Lu Wei**,[2] **and Wei He**[3]

[1]*Department of Electronic Information, Huishang Vocational College, Hefei 230039, China*
[2]*School of Computer Science and Technology, Anhui University, Hefei 230039, China*
[3]*School of Mechanical and Automotive Engineering, Anhui Water Conservancy Technical College, Hefei 231603, China*

Correspondence should be addressed to Beibei Cui; cuibei3@163.com

Received 25 February 2022; Revised 13 July 2022; Accepted 23 July 2022; Published 19 September 2022

Academic Editor: Anwar Ghani

The application of digital signature technology to the Internet of vehicles (IoV) is affected by its network and communication environment. In the 5G era, the influx of a large number of intelligent devices into the mobile Internet requires a low transmission delay and power consumption as well as high-security requirements. To the best of our knowledge, a well-designed solution in which signcryption technology is used has not been proposed in the IoV research area. Motivated by the fact, a certificateless signcryption scheme based on the elliptic curve digital signature algorithm, in which pseudonym and timestamp mechanism are also considered, has been designed in this paper. We prove that the scheme proposed by us can be reduced to solving the difficulty of the computational Diffie–Hellman problem with a standard model, showing that the scheme meets requirements on both security and efficiency, which provides a comparative analysis with the state-of-the-art schemes in terms of security analysis, computational cost, and communication cost, demonstrating that the scheme proposed by us is suitable to be deployed in the IoV environment, which is of the characteristics of high-speed vehicle movement.

1. Introduction

The Internet of vehicles (IoV) has made significant progress in the 5G era. To meet the needs of research and application, IoV communication can be divided into vehicle to vehicle (V2V), vehicle to infrastructure (V2I), vehicle to pedestrian (V2P), and vehicle to network (V2N). We call them vehicle to everything (V2X). At present, data transmission of the IoV is realized with the help of the DSRC and cellular network, and then, data are stored in the cloud [1]. Among them, V2X communication is based on the 5G network [2], which has been widely used by global operators and automobile manufacturers. Security issues such as counterfeiting, manipulation, and forgery exist in all the IoV links [3]. Since they are critical aspects in solving the problems of information security and privacy protection, anonymous authentication has become a hotspot of research in recent years. Kamat et al. [4] proposed an identity-based and cryptography-based VANET security framework (IBC). Shamir[5] proposed the concept of an identity-based system. In 1984, a cryptosystem based on arbitrary strings could use conventional anonymity approach for the first time, which entails a third-party trustworthy institution storing the correspondence between all vehicles and anonymous certificates. According to the report, if the authority is not authorized, it may intentionally disclose personal information of the vehicle, forge, and tamper with the legal vehicle identification. Tzeng et al. [6] integrated the identity-based public-key cryptosystem into the Internet of vehicles to meet this challenge. The user's private key is generated by a third-party private key generator (PKG). What can be done if a third-party private key creation center is dishonest or malicious as public keys. For instance, Zhang et al. [7] recommended that fingerprint information be used for identity authentication. Cui et al. [8] adopted edge computing in VANETs to apply privacy protection. Raya and Hubaux[9] proposed that signature of any user can be forged, causing the problem of key escrow. As a result, Al-

Riyami and Paterson[10] presented the concept of a key generation center (KGC), pointing out that any effective key can be generated by the secret value of OBU and partial keys distributed by KGC. A certificateless signature system was presented by Liu et al. [11] in 2007. Keys are no longer solely determined by the CA, and the traditional signature method was broken. Shim [12] devised a novel certificateless signature system and assessed its security using computational Diffie–Hellman (CDH), and Yang et al. [13] considered that the scheme was vulnerable to malicious and passive KGC attacks. Thumbur et al. [14] suggested a certificateless signature technique without bilinear pairing in 2020, claiming that the scheme can be used in IoV with limited resources. Mei et al. [15] suggested a bilinear pairing-based certificateless signature aggregation approach with conditional privacy protection. Under the random oracle paradigm, the approach achieved complete aggregation and was proved to be safe. For V2V secure communication, Ali et al. [16] devised an identity-based message authentication technique without bilinear pairing. When vehicles request to register with the trusted authority (TA), the TA creates pseudonyms and keys for them to secure its anonymity during the communication process. Barbosa and Farshim[17] proposed the certificateless signcryption (CLSC) concept, which can transmit signing and encryption simultaneously. Processing time, broadband occupation, and key management can all benefit from signcryption, which was first proposed by Zheng [18]. Barbosa's method, however, has been shown to be vulnerable to malicious passive KGC assaults. For bilinear pairs, Barreto et al. [19] suggested a certificateless signcryption approach. Suzhen et al. [20] proposed a signcryption technique that includes a privacy protection feature in 2018. Vehicle keys and pseudonyms were generated by TA and PKG, respectively. The bilinear pairing operation was used in the same way in documents [20, 21], with low computational efficiency. Many researchers are now studying signcryption technology [22–25], but no systematic scheme is formed. Du et al. [26] put forward a certificateless signature scheme based on elliptic curve cryptosystems, but there is a replacement key attack. We improve Du et al.'s scheme, propose a new certificateless signcryption scheme based on an elliptic curve, and apply this scheme to the privacy protection of the IoV. We construct a new CLSC scheme to obtain a higher level of security. Our CLSC scheme proves its security of the scheme by using two different types of adversary selection message attacks. Compared with other existing schemes, this scheme avoids expensive bilinear pairing, is more cost-effective, and is suitable for rapidly changing IoV environment. The main contributions of this paper are as follows:

(i) To create pseudonyms, ECC cryptography is employed; the standard tamper-proof device (TPD) and password (PWD) are not used. Instead, the pseudonym is formed using the intermediate variables false identity and timestamp. Therefore, the hidden danger of password theft is avoided, and the system has a high level of privacy protection.

(ii) Combining certificateless and signcryption theory, anonymous is introduced into the scheme. Key generation is related to RSU, OBU, and KGC; the IBC algorithm is improved by two-way authentication among them. Thus, the security of the key is enhanced.

(iii) When compared to other related systems, the computational cost decreased. The scheme satisfies the security requirements of IND-CCA and EUF-CMA, giving the IoV system forward security, anonymity, traceability, and the capacity to prevent replay attacks.

2. Elliptic Curve

If q is a large prime, it satisfies $q \geq 2^{160}$, and Z_q includes all solutions in the finite domain F_q. Elliptic curve E: $y^2 = x^3 + ax + b \bmod q$, and $E(Z_q)$ denotes the set of pairs $(x, y) \in (Z_q \times Z_q)$, satisfying the above equation along with a special value O. That is, $E(Z_q) = \{(x, y) | x, y \in Z_q,\ y^2 = x^3 + ax + b \bmod q \cup O$. The elements $E(Z_q)$ are called the points on the elliptic curve E, where $4a^3 + 27b^2 \neq 0$, and O is called the point at infinity.

(i) Elliptic curve digital signature algorithm (ECDSA) [27]: it is an algorithm through which a random integer k is generated and calculates the point $P = kG$ as well as the number $r = x_p \bmod q$ is calculated, where x_p is the x coordinate of P. Finally, $s = k^{-1}(z + rd_A) \bmod q$ is calculated as a signature, and z is the hash truncation of message M.

(ii) Elliptic curve discrete logarithmic problem (ECDLP): there are two points M, N on the elliptic curve $E(a, b)$, and $M = k \cdot N (\forall k \in Z^*)$ is calculated, when the points M, N are known, the problem of solving the coefficient k is called an elliptic curve discrete logarithmic problem, and the coefficient k cannot be calculated in the polynomial time.

(iii) Elliptic curve Diffie–Hellman problem (ECDHP): the problem is that on inputs $a, b \in Z^*$, point G is taken as the base point in the finite field of elliptic curve $E(a, b)$ to have the given equation, $M = a \cdot G, N = b \cdot G, R = ab \cdot G$ when the values of M and N are known, solving the value of R is called an elliptic curve Diffie–Hellman problem, which cannot be effectively solved in the polynomial time.

2.1. System Overview. In our scheme, the IoV model consists of vehicles, roadside units, key generator centers, and trusted authorities. The specific division of labor is as follows:

Onboard unit (OBU): intelligent vehicles with OBU can exchange information and data with roadside units or other vehicles. Each vehicle periodically broadcasts information for safe driving. To ensure location privacy, each vehicle needs to use a pseudonym to replace its real identity to transmit information.

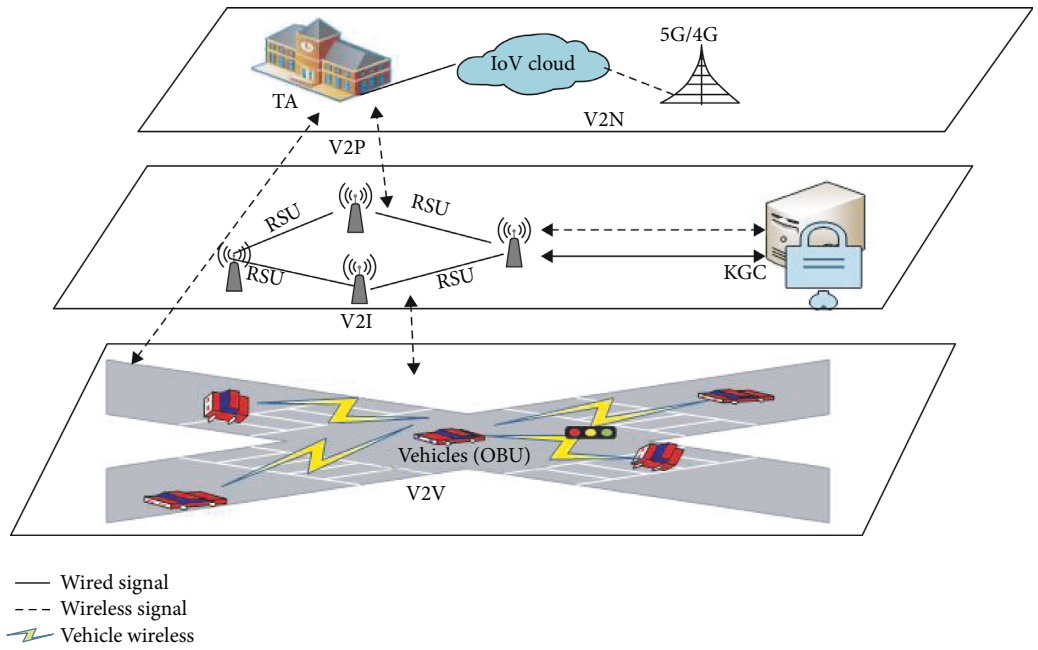


Figure 1: System structure diagram of the IoV.

RSUs (roadside units): RSUs are deployed alongside urban roads, which consist primarily of a wireless communication interface and a local data preprocessing unit. The roadside units are deployed by specific guidelines. As a result, the vehicle can access the roadside units. All the RSUs should be interconnected with the intelligent transportation information data center.

Trusted authority (TA): the TA is managed by the traffic management department and is mainly in charge of OBU identity registration and authentication. It is fully trusted in this scheme and is responsible for generating the false identity of the vehicle.

Key generation center (KGC): the KGC is in charge of communicating with TA to generate partial public/private keys for legitimate OBU and RSUs.

The model is shown in Figure 1.

2.2. Scheme. Our CLSC scheme is designed for IoV communication, eliminates the issue of key escrow, and makes use of a pseudonym mechanism to protect the real identities of both parties to the communication, so ensuring the privacy of the identity and vehicle traceability.

First, in order to eliminate the impact of replacing the public key, the system master key is added to the pseudonym generation formula to make it more difficult for attackers to forge signatures, and make the s impossible to bypass. It can be seen that in the Du et al.' scheme [26], part private key SK_i was calculated by the system master key. The malicious signer cannot calculate the value of the system master key and SK_i through technical means, but the public key of the certificateless signature scheme is not authenticated between the signer and the verifier. The malicious signer forges the signature by forging the secret value and bypassing the unknown system master key. Therefore, there is a key replacement attack. So, in our scheme, signcryption algorithm is introduced to ensure the confidentiality of transmission and improve transmission efficiency. Finally, the security of the scheme is proved in the standard model. The meaning of relevant symbols is shown in Table 1. The flowchart of the algorithm is shown in Figure 2. The algorithm steps are provided.

2.3. Algorithm. There are five participants in the improved certificateless signcryption scheme algorithm: KGC, TA, RSU, the sender of vehicle (V_A), and the receiver of vehicle (V_B). OBU and RSU conduct two-way authentication through TA [28]. We divide the entire scheme into six algorithms, which are listed as follows.

2.3.1. Initialization. The KGC chooses five collision-resistant Hash functions:

$$\begin{aligned} H_0&: \{0,1\}^* \longrightarrow Z_q^*, \\ H_1&: \{0,1\}^* \times G \longrightarrow Z_q^*, \\ H_2&: \{0,1\}^* \times G \times G \longrightarrow Z_q^*, \\ H_3&: \{0,1\}^* \times Z_q^* \times G \times G \longrightarrow Z_q^*, \\ H_4&: \{0,1\}^* \times G \times G \longrightarrow Z_q^*. \end{aligned} \tag{1}$$

The KGC secret saves system master key s and encrypted transmits s to TA, and TA saves (s, RID_i) and generates system public key $P_{\mathrm{pub}} = sP$. The common parameter is $pp = (q, G, P, P_{\mathrm{pub}}, H_0, H_1, H_2, H_3, H_4)$.

Table 1: Parameter description table.

Parameter	Implications
G	Additive cyclic group of order q
P	Generator of group G
s	System master key
Z_q^*	$Z_q^* = \{x: 0 < x < q, \gcd(x, q) = 1\}$
H_0, H_1, H_2, H_3, H_4	Five safe hash functions
P_j, K_j, k_j	The identity of roadside unit j, public key Y_j, and private key y_j
S_i	Partial private key
r_i	KGC generate the secret value to generate public/private keys
x_i	Secret value of the vehicle
ξ_i	Secret value for the RSU
PK_i, SK_i	Public key and private key for a vehicle
RID_i	List of true vehicle identities
F_i	False identity of a vehicle
FID_i	Pseudonym of a vehicle
T_i	Current timestamp of a vehicle
δ	Ciphertext between two vehicles
Y, Y^*	Encryption key and decryption key
V_A, V_B	Vehicle of data sender and vehicle of data receiver
$\mathbb{A}_I, \mathbb{A}_{II}$	Type-I and type-II adversaries

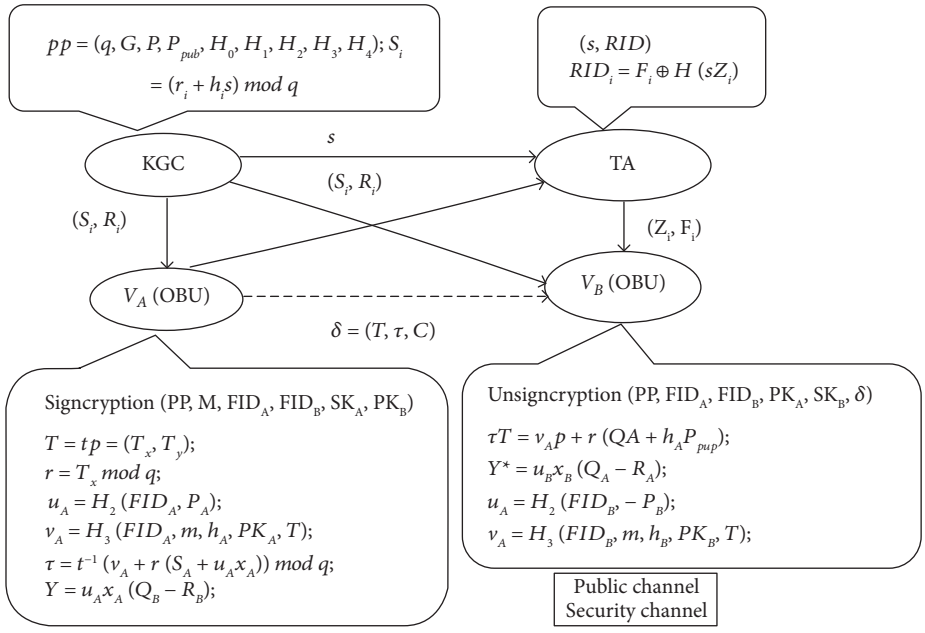


Figure 2: The CLSC of our scheme.

2.3.2. Registration. OBU executes the algorithm, randomly selects $z_i \in Z_i^*$, calculates the negotiation key [29] $Z_i = z_i P$, generates false identity $F_i = \mathrm{RID}_i \oplus H_0(z_i P_{\mathrm{pub}})$, and then sends (Z_i, F_i) to TA. The algorithm is executed by TA, and TA receives the message (Z_i, F_i) from OBU. TA calculates $\mathrm{RID}_i = F_i \oplus H_0(sZ_i)$ and queries whether the vehicle identity list containing RID_i. If not, the algorithm is terminated by TA, and the OBU is determined to be illegal. RSU sets identity as P_j, randomly selects $k_i \in Z_q^*$ as its private key, RSU calculates negotiation key $K_i = k_i P$ and public key $K_j = k_i P_{pub}$, and sends (P_j, K_i) to TA, and TA calculates $\mathbb{K}_j = sK_i$ and forwards $(P_j, \mathbb{K}_j)$ to the legitimate OBU.

2.3.3. Pseudonym Generation. The trusted organization no longer issues the public-key certificates (PKI) to vehicles but generates pseudonyms for them. In this scheme, the generation of a pseudonym consists of three parameters, including false identity of its own, RSU identity information, and timestamp, rather than the device password information. When the vehicle enters the area responsible for RSU, OBU receives K_j from the RSU broadcast. When OBU receives multiple RSU broadcast signals at the same time in a critical environment, it can only record the strongest RSU broadcast information and discard relatively weak RSU broadcast information. The OBU checks the RSU's public

key; if $K_j \notin (P_j, \mathbb{K}_j)$, the RSU will be illegal, and the algorithm will not be executed. Otherwise, OBU obtains the current timestamp T_i and the public key K_j of the current RSU, then selects the secret value $\xi_i \in Z_q^*$ for the RSU. The OBU calculates $\mathrm{FID}_{i1} = F_i \oplus H_0(K_j \xi_i \| T_i)$, $\mathrm{FID}_{i2} = P_j Q$, and sets the pseudonym of the vehicle $\mathrm{FID}_i = (\mathrm{FID}_{i1}, \mathrm{FID}_{i2}, T_i)$.

Through the above operations, TA indirectly judges the legitimacy of RSU. OBU generates the pseudonym through legal RSU, false identity of the vehicle, and the timestamp.

2.3.4. Key Generation

(i) Secret value: OBU chooses a random $x_i \in Z_q^*$ as its secret value. When the pseudonym is updated, the secret value should also be changed randomly, to maintain forward safety [30].

(ii) Partial private/public key: KGC inputs the pseudonym of the vehicle FID_i and the parameter value PP, KGC chooses $r_i \in Z_q^*$ randomly and calculates partial public key $R_i = r_i P$ and partial private key $S_i = (r_i + h_i s) \bmod q$, which is $h_i = H_1(\mathrm{FID}_i, R_i)$. KGC sends (S_i, R_i) to OBU via secure channel.

(iii) Public key extract: OBU calculates $P_i = x_i P$, $u_i = H_2(\mathrm{FID}_i, P_i)$, $Q_i = R_i + u_i P_i$ then generates the public key, which is $PK_i = (R_i, Q_i)$.

(iv) Private key extract: OBU checks whether the $S_i P = R_i + h_i P_{\mathrm{pub}}$ is established. If so, it will be accepted. If not, it will be rejected. The private key is generated as $SK_i = (S_i, x_i)$. Proof of correctness is as follows: $S_i P = (r_i + h_i s)P = R_i + h_i P_{\mathrm{pub}}$.

2.3.5. Signcryption. V_A is the sender of OBU, while V_B is the receiver of OBU, and V_A takes message M, $\mathrm{FID_A}$, FID_B, PP, SK_A, and PK_B as input; generates a random integer t; and produces signcryptext δ. The signcryption generation process is based on ECDSA, and the specific calculation process is as follows:

(i) $T = tP = (T_x, T_y)$, T_x, T_y are the x coordinate value and y coordinate value of point T.

(ii) $\tau = t^{-1}(v_A + r(S_A + u_A x_A)) \bmod q$, where $r = T_x \bmod q$ can be considered as an important parameter for verifying signatures, and there has three hash functions.

$$\begin{aligned} h_A &= H_1(\mathrm{FID}_A, R_A), \\ u_A &= H_2(\mathrm{FID}_A, P_A), \\ v_A &= H_3(\mathrm{FID}_A, m, h_A, PK_A, T). \end{aligned} \tag{2}$$

Hash functions h_1, h_2, and h_3 are used to protect the pseudonym FID_A, message m, and public key PK_A.

$$C = M \oplus w, \tag{3}$$

C is signcryptext, which is generated by M XOR W.

$$w = H_4(\mathrm{FID}_A, \mathrm{FID}_B, Y), \tag{4}$$

$$Y = u_A x_A (Q_B - R_B), \tag{5}$$

V_A sends $\delta = (T, \tau, C)$ to V_B.

2.3.6. Unsigncryption. V_B takes δ, FID_A, FID_B, PP, SK_B, and PK_A as input and returns massage M, if $\tau T = v_A P + r(Q_A + h_A P_{\mathrm{pub}})$ is hold. V_B performs the following steps:

$$w^* = H_4(\mathrm{FID}_A, \mathrm{FID}_B, Y^*), \tag{6}$$

$$\begin{aligned} Y^* &= u_B x_B (Q_A - R_A), \\ u_B &= H_2(\mathrm{FID}_B, P_B). \end{aligned} \tag{7}$$

V_B executes the algorithm $M = C \oplus w^*$ to decrypt the signcryption.

3. Correctness

Only if the following two equations are true, respectively, the scheme meets the correctness.

(i) Public verifiability. The message is signed by V_A, if the verification signature is valid, V_B receives the message. Otherwise, if the signature is invalid, V_B rejects the message.

$$\begin{aligned} \tau T &= t^{-1}(v_A + r(S_A + u_A x_A)) tp \bmod q \\ &= (v_A + r(r_A + h_A s + u_A x_A))P \\ &= v_A P + r(R_A + u_A P_A + h_A P_{\mathrm{pub}}) \\ &= v_A P + r(Q_A + h_A P_{\mathrm{pub}}). \end{aligned} \tag{8}$$

(ii) Consistency of encryption and decryption. If $Y^* = Y$ is true, $w^* = w$ must be true, and $M = C \oplus w^* = M \oplus w \oplus w^*$ must be established.

$$\begin{aligned} Y &= u_A x_A (Q_B - R_B) \\ &= u_A x_A u_B x_B P, \end{aligned} \tag{9}$$

$$\begin{aligned} Y^* &= u_B x_B (Q_A - R_A) \\ &= u_B x_B u_A x_A P. \end{aligned} \tag{10}$$

Both Y and Y^* are deduced from the public key generation algorithm $Q_A = R_A + u_A P_A$, $Q_B = R_B + u_B P_B$, $P_A = x_A P$, and $P_B = x_B P$. From the formulas (4), (6), (9), and (10), it is deduced that the equation $w^* = w$ holds.

$$\begin{aligned} M &= C \oplus w^* \\ &= M \oplus w \oplus w^* \\ &= M \oplus w \oplus w. \end{aligned} \tag{11}$$

Thus, the message M can be restored.

4. Security Proof

Two types of adversaries are considered to prove the security of our scheme [31]. These requirements on security are described via some games between adversaries ($\mathbb{A}_I$ or $\mathbb{A}_{II}$) and a challenger $\mathbb{C}$. Adversaries can be divided into two cases: one is that the adversary $\mathbb{A}_I$ is a malicious who does not know the system master key s, but can replace the public key of any user; the second type of adversary $\mathbb{A}_{II}$ is a malicious KGC attacker, who knows the master key s but cannot replace any public key. In our CLSC scheme, the adversaries may access the following oracles:

(i) H_{PK}: FID_i is entered as an identifier, and a public key PK_i matching FID_i will be returned.

(ii) H_d: FID_i is entered as an identifier, and a partial private key S_i will be returned.

(iii) $H_{\mathrm{Replace}.PK}$: FID_i is entered as an identifier, a new public key PK_i' that can be used will replace the original public key PK_i.

(iv) H_{SK}: FID_i is entered as an identifier, a private key SK_i matching FID_i will be returned, when the public key is not replaced.

(v) $H_{\mathrm{Signcrypt}}$: When there is a message M, identity of a sender is FID_A, and identity of a receiver is FID_B as input, and an available signcryption δ on M will be returned.

(vi) $H_{\mathrm{Unsigncrypt}}$: When a signcryption δ, identity of a sender is FID_A, and identity of a receiver is FID_B as input, the message M will be restored, when δ is available.

$\mathbb{A}_I$ can access all the above oracles, while $\mathbb{A}_{II}$ can access all of them except $H_{\mathrm{Replace}.PK}$ and H_d, because $\mathbb{A}_{II}$ owns the system master key s, $\mathbb{A}_{II}$ can forge partial private key γ; meanwhile, $\mathbb{A}_I$ and $\mathbb{A}_{II}$ can suppose $H_I = \{H_{PK}, H_d, H_{\mathrm{Replace}.PK}, H_{SK}, H_{\mathrm{Signcrypt}}, H_{\mathrm{Unsigncrypt}}\}$ and $H_{II} = \{H_{PK}, H_{SK}, H_{\mathrm{Signcrypt}}, H_{\mathrm{Unsigncrypt}}\}$, respectively. We prove our CLSC scheme from two aspects: confidentiality and unforgeability.

4.1. Confidentiality. This property is considered as the indistinguishability under chosen-ciphertext attack (IND-CCA). In this section, the security proof is proved by some games between adversaries ($\mathbb{A}_I$ or $\mathbb{A}_{II}$) and a challenger $\mathbb{C}$.

Game 1: The game interactions between an adversary $\mathbb{A}$ and a challenger $\mathbb{C}$ are as follows:

(i) Setup: $\mathbb{C}$ inputs a security parameter λ, a common parameter pp and α are generated, of which α is kept as a secret.

(ii) Phase 1 queries: $\mathbb{A}_I$ sends bounded queries in polynomial time to the oracles H_I, and the $\mathbb{C}$ responds to the queries passing through these oracle models.

(iii) Challenge: $\mathbb{A}_I$ sends two equal length messages m_0 and m_1 to challenger $\mathbb{C}$ with FID_A^* and FID_B^* as identifiers. A bit $\gamma \in \{0, 1\}$ is randomly selected by $\mathbb{C}$, through which $\mathrm{Signcrption}(PP, M, \mathrm{FID}_A^*, \mathrm{FID}_B^*, SK_A^*, PK_B^*)$ is implemented by $\mathbb{C}$ and δ is sent to $\mathbb{A}_I$.

(iv) Phase 2 queries: $\mathbb{A}_I$ sends bounded queries in polynomial time to the oracle H_I, and the $\mathbb{C}$ responds to the queries passing through these oracle models.

(v) Guess: $\mathbb{A}_I$ outputs a guess of γ, which is γ^*.

It is said that $\mathbb{A}_I$ wins game 1, if $\gamma^* = \gamma$ and the following conditions are established:

(1) SK_A^* cannot be extracted by $\mathbb{A}_I$ at any point

(2) S_A^* cannot be extracted by $\mathbb{A}_I$, if $\mathbb{A}_I$ has replaced PK_A^* with PK_A' before accepting the challenge

(3) In phase 2 queries, $\mathbb{A}_I$ is unable to perform unsigncryption query on δ^* under FID_A^* or FID_B^*, and signcryption FID_B^*, PK_A^*, or PK_B^* has been replaced after the challenge is issued.

Game 2: The game interactions between an adversary $\mathbb{A}$ and a challenger $\mathbb{C}$: the challenge steps are the same as those of game 1.

(i) Setup: $\mathbb{C}$ inputs a security parameter λ, and a common parameter pp and α are generated. $\mathbb{C}$ sends parameter pp and α to $\mathbb{A}_{II}$.

(ii) Phase 1 queries: $\mathbb{A}_{II}$ sends bounded queries in polynomial time to the oracle H_{II}, and $\mathbb{C}$ responds to the queries passing through these oracle models.

(iii) Challenge: $\mathbb{A}_{II}$ sends two equal length messages m_0 and m_1 to challenger $\mathbb{C}$ with FID_A^* and FID_B^* as identifiers. A bit $\gamma \in \{0, 1\}$ is randomly selected by $\mathbb{C}$, through which $\mathrm{Signcryption}(PP, M, \mathrm{FID}_A^*, \mathrm{FID}_B^*, SK_A^*, PK_B^*)$ is implemented, and then, δ is sent to $\mathbb{A}_{II}$.

(iv) Phase 2 queries: $\mathbb{A}_{II}$ sends bounded queries in polynomial time to the oracle H_{II}, and $\mathbb{C}$ responds to the queries passing through these oracle models.

(v) Guess: $\mathbb{A}_{II}$ outputs a guess γ^* of γ.

It is said that $\mathbb{A}_{II}$ wins game 2 if $\gamma^* = \gamma$ and the following conditions are hold:

(1) $\mathbb{A}_{II}$ cannot extract SK_A^* at any point. Because the secret value x_i cannot be obtained by $\mathbb{A}_{II}$, $\mathbb{A}_{II}$ solves x_i as ECDLP problem.

(2) In phase 2 queries, $\mathbb{A}_{II}$ is unable to perform an unsigncryption query on δ^* under FID_A^* or FID_B^*.

If the probability $\mathrm{Adv}(\mathbb{A}) = 2 * |\Pr[\mathbb{A} - 1/2]|$ is negligible, we say that the scheme is IND-CCA safe. We know that $\mathbb{A}_I$ can access to all of the oracles, while $\mathbb{A}_{II}$ can access to all of them except $H_{\mathrm{Replace}.PK}$ and H_d.

$\mathbb{A}_I$ sends bounded queries in polynomial time to the oracle H_I making a signcryption query $H_{\mathrm{Signcrypt}}$ but cannot win δ under FID_A^* or FID_B^*. The key generation process is $Q_A^* - R_A^* = u_A^* x_A^* P$, $Q_B^* - R_B^* = u_B^* x_B^* P$, and $Y = u_B^* x_B^* u_A^* x_A^* P$. It is still difficult to solve Y, which is an ECDHP problem.

$\mathbb{A}_{II}$ sends bounded queries in polynomial time to the oracle H_{II}, making a public key query H_{PK}, but H_{II} cannot

be used to obtain x_i^*; thus, $\mathbb{A}_{II}$ cannot obtain PK_i, and solving x_i^* is an ECDLP problem.

The probability for $\mathbb{A}_I$ and $\mathbb{A}_{II}$ to win game 1 and game 2 is negligible.

4.2. Unforgeability. This property is considered as the existential unforgeability against the chosen message attack (EUF-CMA). In this section, the security proof is proved through some games between adversaries ($\mathbb{A}_I$or$\mathbb{A}_{II}$) and a challenger $\mathbb{C}$.

Game 3: The game interactions between an adversary $\mathbb{A}$ and a challenger $\mathbb{C}$ are as follows:

(i) Setup: $\mathbb{C}$ inputs a security parameter λ, a common parameter pp and α are generated, and α is kept as a secret.

(ii) Phase 1 queries: $\mathbb{A}_I$ sends bounded queries in polynomial time to the oracle H_I, and $\mathbb{C}$ responds to the queries passing through these oracle models.

(iii) Forgery: $\mathbb{A}_I$ forges the message M^* and signcryption $\delta^* = (T^*, \tau^*, C^*)$ from the sender V_A^* to the receiver V_B^*.

If the decryption output is M^* and the following conditions are met, it is said that $\mathbb{A}_I$ wins game 3.

(1) $\mathbb{A}_I$ cannot extract SK_A^* at any point

(2) $\mathbb{A}_I$ cannot extract SK_i^* for any pseudonym FID_i, if PK_i^* has been replaced

(3) $\mathbb{A}_I$ cannot extract x_A^*

(4) $\mathbb{A}_I$ cannot make a signcryption query on M^* under FID_A^* or FID_B^*

Game 4: The game interactions between an adversary $\mathbb{A}$ and a challenger $\mathbb{C}$: the challenge steps are the same those of as game 3.

(i) Setup: $\mathbb{C}$ inputs a security parameter λ, and a common parameter pp and α are generated. $\mathbb{C}$ sends parameter pp and α to $\mathbb{A}_{II}$.

(ii) Queries: $\mathbb{A}_{II}$ sends bounded queries in polynomial time to the oracle H_{II}, and the $\mathbb{C}$ responds to the queries passing through these oracle models.

(iii) Forgery: $\mathbb{A}_{II}$ creates a forged message m^* or signcryption $\delta^* = (T^*, \tau^*, C^*)$ from the sender V_A^* to the receiver V_B^*.

If the decryption output is M^* and the following conditions are met, it is said that $\mathbb{A}_{II}$ wins game 4.

(1) $\mathbb{A}_{II}$ cannot extract SK_A^* at any point

(2) $\mathbb{A}_{II}$ cannot make a signcryption query on M^* under FID_A^* or FID_B^*

If it is negligible $\mathbb{A}_I$ or $\mathbb{A}_{II}$ to win game 3 and game 4 $(\mathrm{AdvSig}_{\epsilon,A}^{\mathrm{CMA}}(k) \le \mathrm{negl}(k)$, we say that the scheme is EUF-CMA safe. Note that $\mathbb{A}_I$ has access to all of the mentioned oracles, while $\mathbb{A}_{II}$ has access to all of them except $H_{\mathrm{Replace}.PK}$ and H_d.

$\mathbb{A}_I$ executes public key replacement queries from $H_{\mathrm{Replace}.PK}$, which can replace the public key with $PK_A' = (R_A, Q_A')$, $PK_B' = (R_B, Q_B')$, signcryption queries from $H_{\mathrm{Signcrypt}}$, and unsigncryption queries from $H_{\mathrm{Unsigncryption}}$; $\mathbb{A}_I$ randomly selects $t^* \in Z_q^*$, $x_A^* \in Z_q^*$, and $x_B^* \in Z_q^*$, which is used to $T^* = t^*P = (T_x, T_y)$, $r^* = T_x \bmod q$, $v_A^* = H_3(\mathrm{FID}_A^*, m, h_A^*, PK_A', T)$, $Q_A' = x_A^*P - h_A^*P_{\mathrm{pub}}$, and $Q_B' = x_B^*P - h_B^*P_{\mathrm{pub}}$, which are forged, so as to signcrypt the message m^*. Then, signcryption $\delta^* = (T^*, \tau^*, C^*)$ is forged, V_B receives δ^*, and feasibility verification is conducted:

$$\begin{aligned}\tau^*T^* &= t_A^{*-1}(v_A^* + r^*x_A^*)t_A^*P \\ &= (v_A^* + r^*x_A^*)P \\ &= v^*P + r^*\left(Q_A' \quad + h_A^*P_{\mathrm{pub}}\right).\end{aligned} \tag{12}$$

If it is only a signature algorithm without signcryption, the adversary can still forge a signature and pass the authentication by signing before encryption or encrypting before signature, which is the same as Du et al. [26].

$$\begin{aligned}Y' &= u_A^*x_A^*(Q_B' - R_B) \\ &= u_Ax_A\left(x_B^*P - h_B^*P_{\mathrm{pub}} - R_B\right),\end{aligned} \tag{13}$$

$$\begin{aligned}Y^* &= u_B^*x_B^*(Q_A' - R_A) \\ &= u_Bx_B\left(x_A^*P - h_A^*P_{\mathrm{pub}} - R_A\right).\end{aligned} \tag{14}$$

According to formulas (13) and (14), it is known that $Y^* \neq Y'$, so $w^* \neq w'$, the adversary $\mathbb{A}_I$ cannot pass the encryption consistency verification. Public key replacement fails. $\mathbb{A}_{II}$ cannot execute query partial private key from H_d; thus, γ is forged to replace x_A^*, and $t' \in Z_q^*$ is selected to forge $\delta^* = (T^*, \tau^*, C^*)$, where $T^* = t'P$, $\tau^* = t'^{-1}(v_A + r(S_A + u_A'\gamma)) \bmod q$, and $C^* = m^* \oplus w$, in which $P_A' = \gamma P$ and $u_A' = H_2(\mathrm{FID}_A, P_A')$, and V_B gets δ^*; then, a feasibility verification is done.

$$\begin{aligned}\tau^*T^* &= \left(t'^{-1}(v + r(S_A + u_A'\gamma))\right)t'P \bmod q \\ &= (v' + r(r_A + h_As + u_A'\gamma)) \\ &= P \\ &= v'P + r\left(R_A + h_AP_{\mathrm{pub}} + u_A'P_A\right),.\end{aligned} \tag{15}$$

$\mathbb{A}_{II}$ cannot replace any public key. It is known that $Q_A \neq R_A + h_AP_A$; thus, $\tau^*T^* \neq vP + r(Q_A + h_AP_{\mathrm{pub}})$. The output will be INVALID, and V_B discards the ciphertext.

The probability of $\mathbb{A}_I$ and $\mathbb{A}_{II}$ to win game 3 and game 4 is negligible.

5. Security Analysis

5.1. Forward Security. If the system master key s is omitted, it is calculated due to the difficulty of ECDLP, it is still difficult to calculates r_i and x_i, and (PK_i, SK_i) remains unknown. Therefore, it is guaranteed that the past signcryption information will not be disclosed, because of the randomness of r_i and x_i. When the system master key is omitted, the new values will immediately replace it. The key

Table 2: Run time of the different encryption operations.

Symbol	Operation	Parameter	Runtime (ms)
T_{em}	Elliptic curve point multiplication	$x \cdot P(P \in G, x \in z_q^*)$	0.341
T_{in}	Inverse mode	$t^{-1} \bmod q(t \in z_q^*, q \in z_q^*)$	0.029
T_{ea}	Elliptic curve point plus	$P + Q(P \in G, Q \in G)$	0.002
T_{bp}	Time required for the bilinear pairing	$e(\overline{S}, \overline{T})(\overline{S} \in G_1, \overline{T} \in G_1)$	4.669
T_{pm}	Pairing multiplication operation	$\overline{x} \cdot \overline{P}(\overline{x} \in z_q^*, \overline{P} \in G)$	0.788
T_{pa}	Pairing addition	$\overline{S} + \overline{T}(\overline{S} \in G_1, \overline{S} \in G_1)$	0.002
T_{mtp}	MapToPoint hash function	H_1: $0, 1^* \longrightarrow G_1$	0.145
T_e	Modular exponentiation	$g^* \bmod n$	1.915

update is realized, and these actions further confirm the security of the communication [32].

5.2. Traceability. The ciphertext should contain relevant information about the vehicle identity. In the scheme, TA can be used to calculate $\mathrm{RID}_i = F_i \oplus H_0(sZ_i)$ using the system master key s, which queries whether RID_i is listed in the vehicle identity. It seems that only the trusted authority TA can track the vehicle according to the relevant information. In addition, the IoV requires an extremely high real-time nature. The ciphertext contains timestamp information, which can also prevent replay attacks. Because ciphertext $C = M \oplus w$; $w = H_4(\mathrm{FID}_A, \mathrm{FID}_B, Y)$, here we can use the pseudonym of the vehicle $\mathrm{FID}_i = (\mathrm{FID}_{i1}, \mathrm{FID}_{i2}, T_i)$, making the ciphertext contains timestamp information.

5.3. Anonymous. Pseudonyms are used in V2V and V2I communications to protect the true identity of the vehicle. The pseudonym of the vehicle consists of three parts: $\mathrm{FID}_i = (\mathrm{FID}_{i1}, \mathrm{FID}_{i2}, T_i)$, where FID_{i1} is generated by the false identity F_i of the vehicle, $\mathrm{FID}_{i1} = F_i \oplus H_0(K_j \xi_i \| T_i)$, $F_i = \mathrm{RID}_i \oplus H_0(z_i P_{\mathrm{pub}})$, $\mathrm{FID}_{i2} = P_j$, and T_i is the timestamp to ensure the anonymity of the vehicle. It is necessary to protect the identity information RID_i of the vehicle when the pseudonym information is disclosed. According to the irreversibility of a hash function and the difficulty of ECDLP, the attacker cannot calculate z_i, ξ_i, or k_i in polynomial time, so the RID_i of the vehicle cannot be obtained. In addition, vehicles carry different pseudonyms in different RSU communication ranges and timestamps; that is, the pseudonym information of the vehicle changes with position and time, which makes the generation process of a pseudonym a one-way trapdoor function.

5.4. Unforgeable. The unforgeability of the CLSC scheme is proven in the unforgeability section using a (existential unforgeability against selected message attacks, EUF-CMA) security model. The signature ciphertext forged by an attacker does not satisfy the encryption consistency or convey the attacker's intentions.

6. Performance Evaluation

Computational cost, communication cost, and safety analysis are analyzed in this section compared with other relevant schemes [33–38]. The schemes selected for comparison are certificateless signcryption, which can be applied to the IoV.

The computational cost mainly depends on the amount of signcryption and unsigncryption algorithms, which can be measured based on the number of execution times of statistical elliptic curve scalar multiplication, elliptic curve scalar addition, bilinear pairing, and mapping to point operation. The computational cost of XOR operation on Z_q^* is too small to make comparison. The operation results are listed in Table 2. The experimental system environment is as follows:

CPU: Intel core i7-6700@3.40 GHz; RAM: 8 GB;

OS: Ubuntu 16.04;

Library: MIRACL, a public C++cryptographic library; [https://github.com/miracl/MIRACL/archive/master.zip].

Under the same operating environment, our scheme costs 1.397 ms, Kasyoka et al.'s scheme [33] costs 1.705 ms, Karati et al.'s scheme [34] costs 2.424 ms based no pairing, Karati et al.'s scheme [35] costs 18.913 ms based on bilinear pairing, He et al.'s [36] scheme costs 2.05 ms, and Seo et al.'s [38] scheme costs 3.41 ms. Compared with the other schemes [33–36, 38], our scheme in this paper decreases by 18.06%, 42.37%, 92.61%, 31.85%, and 59.03%, respectively.

Communication cost is measured by the length of a single ciphertext. In the bilinear pairing operation scheme, the length of $|G_1|$ is 1024 bit, and that of $|G_2|$ is the same. To provide the security schemes of the same level for a scheme based on the elliptic curve, q is the prime number and the length of $|Z_q^*|$ is 160 bit. The additive cyclic group with q order generation for point P on a nonsingular elliptic curve is G, and the length of $|G|$ is 320 bit.

The superiority of this scheme is illustrated by comparing the computation and communication overhead of a single ciphertext, which is statistically analyzed in Table 3.

In the comparative analysis of communication cost, the length of a single ciphertext is used as the unit of comparison, which is 640 bit in our scheme, slightly higher than that of Kasyoka et al.'s [33] and Seo et al.'s [38] and is lower than that of Karati et al.'s [35] and He et al.'s [36] bilinear pairing scheme, the same as no pairing scheme of Karati et al. [34].

Table 3: Performance comparison of different signcryption schemes.

Scheme	Calculate cost			Communication cost	
	Signcryption	Unsigncryption	Runtime	Signcryptext	Length
[33]	$2T_{em}$	$3T_{em}$	1.705 s	$3\|Z_q^*\|$	480 bit
[34]	$3T_{em} + 2T_{ea} + T_{in}$	$4T_{em} + 2T_{ea}$	2.424 s	$2\|Z_q^*\| + \|G\|$	640 bit
[35]	$3T_e$	$2T_e + 2T_{bp}$	18.913 s	$4\|G_1\| + \|Z_q^*\|$	4256 bit
[36]	$3T_{em}$	$3T_{em} + 2T_{ea}$	2.05 s	$3\|G\| + \|Z_q^*\|$	1120 bit
[38]	$3T_{em}$	$7T_{em}$	3.41 s	$3\|Z_q^*\|$	480 bit
Our CLSC	$T_{in} + T_{em}$	$2T_{ea} + 3T_{em}$	1.397 s	$2\|Z_q^*\| + \|G\|$	640 bit

Table 4: Safety comparison.

Scheme	Confidentiality	Unforgeability	Forward security	Anonymous
[33]	False	True	False	False
[34]	False	False	False	False
[35]	True	False	False	False
[36]	False	True	False	True
[37]	False	True	False	False
[38]	True	True	False	False
Our CLSC	True	True	True	True

Our CLSC scheme is designed according to a certificateless signcryption model and relies on ECDSA, which depends on the difficulty of pseudonyms generation. In this section, the security of the algorithm is compared and with that of similar schemes and is then analyzed. The result is in Table 4.

7. Conclusion

In this paper, we construct a reliable certificateless signcryption scheme without bilinearity, where a pseudonym mechanism is also designed to protect the privacy of vehicles. We use certificateless signcryption technology to implement the scheme, which can secure vehicular communication with a low computation overhead. Performance analysis demonstrates that the scheme proposed by us can be used to reduce computational and communication cost compared with other related schemes. Security proofs and analyses show that the scheme proposed by us can be used to avoid replacement public key attacks, and ensure the satisfaction of the security of IND-CCA as well as EUF-CMA. Other requirements on security including perfect forward secrecy, anonymity, traceability, and resistance of replay attacks can also be ensured.

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 there are no conflicts of interest.

Acknowledgments

This work was supported by the funding project for Top Talent Cultivation in Colleges and Universities in Anhui Province (gxgnfx2020178) and the Natural Science Research Project of Colleges and Universities in Anhui Province (KJ2018A0944).

References

[1] J. Cui, X. Zhang, H. Zhong, J. Zhang, and L. Liu, "Extensible conditional privacy protection authentication scheme for secure vehicular networks in a multi-cloud environment," *IEEE Transactions on Information Forensics and Security*, vol. 15, no. 1, pp. 1654–1667, 2020.

[2] W. Qi, B. Landfeldt, Q. Song, L. Guo, and A. Jamalipour, "Traffic differentiated clustering routing in DSRC and C-V2X hybrid vehicular networks," *IEEE Transactions on Vehicular Technology*, vol. 69, 2020.

[3] C. Song, M. Y. Zhang, W. P. Peng, Z. Z. Liu, Z. P. Jia, and X. X. Yan, "Research on anonymous authentication scheme in VANET," *Journal of Chinese Computer Systems*, vol. 39, no. 5, pp. 899–903, 2018.

[4] P. Kamat, A. Baliga, and W. Trappe, "An Identity-Based Security Framework for VANETs," in *Proceedings of the International Workshop on Vehicular Ad Hoc Networks*, January 2006.

[5] A. Shamir, "Identity-based cryptosystems and signature schemes," in *Proceedings of the CRYPTO 84 on Advances in Proceedings of CRYPTO 84 on Advances in Cryptology*, Santa Barabara, California, USA, 1985.

[6] S. F. Tzeng, S. J. Horng, T. Li, X. Wang, P. H. Huang, and M. K. Khan, "Enhancing security and privacy for identity-based batch verification scheme in VANETs," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 4, pp. 3235–3248, 2017.

[7] J. Zhang, J. Cui, H. Zhong, Z. Chen, and L. Liu, "Pa-Crt: Chinese remainder theorem based conditional privacy-preserving authentication scheme in vehicular ad-hoc networks," *IEEE Transactions on Dependable and Secure Computing*, vol. 18, 2019.

[8] J. Cui, L. Wei, H. Zhong, J. Zhang, Y. Xu, and L. Liu, "Edge computing in VANETs-an efficient and privacy-preserving

cooperative downloading scheme," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 6, pp. 1191–1204, 2020.

[9] M. Raya and J. P. Hubaux, "Securing vehicular ad hoc networks," *Journal of Computer Security*, vol. 15, no. 1, pp. 39–68, 2007.

[10] S. Al-Riyami and K. G. Paterson, "Certificateless Public Key Cryptography," in *Proceedings of the International Conference on the Theory and Application of Cryptology and Information Security*, pp. 452–473, Taipei, Taiwan, 2003.

[11] J. K. Liu, M. H. Au, and W. Susilo, "Self-generated-certificate public key cryptography and certificateless signature/encryption scheme in the standard model," *Proceedings of the 2nd ACM Symposium on Information, Computer and Communications Security*, pp. 273–283, Springer, Berlin, Germany, 2007.

[12] K. A. Shim, "A new certificateless signature scheme provably secure in the standard model," *IEEE Systems Journal*, vol. 13, no. 2, pp. 1421–1430, 2019.

[13] W. Yang, S. Wang, W. Wu, and Y. Mu, "Top-level secure certificateless signature against malicious-but-passive KGC," *IEEE Access*, vol. 7, Article ID 112870, 2019.

[14] G. Thumbur, G. S. Rao, P. V. Reddy, N. B. Gayathri, and D. V. R. K. Reddy, "Efficient pairing-free certificateless signature scheme for secure communication in resource-constrained devices," *IEEE Communications Letters*, vol. 24, no. 8, pp. 1641–1645, 2020.

[15] Q. Mei, X. Hu, J. H. Chen, M. Yang, S. Kumari, and M. K. Khan, "Efficient certificateless aggregate signature with conditional privacy preservation in IoV," *IEEE Systems Journal*, vol. 15, pp. 1–12, 2020.

[16] I. Ali, T. Lawrence, and F. G. Li, "An efficient identity-based signature scheme without bilinear pairing for vehicle-to-vehicle communication in VANETs," *Journal of Systems Architecture*, vol. 103, Article ID 101692, 2020.

[17] M. Barbosa and P. Farshim, "Certificateless signcryption," in *Proceedings of the ACM Symposium on Information, Computer and Communications Security-ASIACCS*, pp. 369–372, Tokyo, Japan, March 2008.

[18] Y. Zheng, "Digital Signcryption or How to Achieve Cost(-signature & Encryption)<<cost (Signature)+cost (encryption)," in *Proceedings of the Annual International Cryptology Conference*, pp. 165–179, Berlin, Germany, May 1997.

[19] P. S. L. M. Barreto, A. M. Deusajute, E. De, S. Cruz, and R. R. D. Silva, "Toward Efficient Certiicateless Signcryption from (And without) Bilinear Pairings [EB/OL]," 2008, https://pdfs.semanticscholar.org/c42d/307a94023543067d9668b2fc9442d443070a.pdf.

[20] C. A. O. Suzhen, X. Lang, X. Liu, and F. Wang, "New heterogeneous signcryption scheme under 5G network," *Netinfo Security*, vol. 18, no. 11, pp. 33–39, 2018.

[21] F. G. Li, M. Shirase, and T. Takagi, "Certificateless hybrid signcryption," *Mathematical and Computer Modelling*, vol. 57, no. 3-4, pp. 324–343, 2013.

[22] Z. Liu, Y. Hu, X. Zhang, and H. Ma, "Certificateless signcryption scheme in the standard model," *Information Sciences*, vol. 180, no. 3, pp. 452–464, 2010.

[23] C. Zhou, W. Zhou, and X. Dong, "Provable certificateless generalized signcryption scheme," *Designs, Codes and Cryptography*, vol. 71, no. 2, pp. 331–346, 2014.

[24] M. Luo, M. Tu, and J. Xu, "A security communication model based on certificateless online/offline signcryption for Internet of Things," *Security and Communication Networks*, vol. 7, no. 10, pp. 1560–1569, 2013.

[25] H. F. Yu and B. Yang, "Provably secure certificateless hybrid signcryption," *Chinese Journal of Computers*, vol. 38, no. 4, pp. 804–813, 2015, in Chinese with English abstract.

[26] H. Du, Q. Wen, S. Zhang, and M. Gao, "A new provably secure certificateless signature scheme for Internet of Things," *Ad Hoc Networks*, vol. 100, Article ID 102074, 2020.

[27] D. Johnson, A. Menezes, and S. Vanstone, "The elliptic curve digital signature algorithm (ECDSA)," *International Journal of Information Security*, vol. 1, no. 1, pp. 36–63, 2001.

[28] L. Wei, J. Cui, H. Zhong, I. Bolodurina, and L. Liu, "A lightweight and conditional privacy-preserving authenticated key agreement scheme with multi-TA model for fog-based VANETs," *IEEE Transactions on Dependable and Secure Computing*, p. 1, 2021.

[29] J. Zhang, H. Zhong, J. Cui, Y. Xu, and L. Liu, "SMAKA: secure many-to-many authentication and key agreement scheme for vehicular networks," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 1810–1824, 2021.

[30] G. Itkis and L. Reyzin, "Forward-secure signatures with optimal signing and verifying," *Advances in Cryptology - CRYPTO 2001*, Springer, vol. 2139, pp. 332–354, Santa Barbara, CA, USA, 2001.

[31] R. Parvin, S. Willy, and D. Mohammad, "Efficient certificateless signcryption in the standard model: revisiting Luo and wan's scheme from wireless personal communications," *The Computer Journal*, vol. 62, no. 8, 2018.

[32] L. Wei, J. Cui, Y. Xu, J. Cheng, and H. Zhong, "Secure and lightweight conditional privacy-preserving authentication for securing traffic emergency messages in VANETs," *IEEE Transactions on Information Forensics and Security*, vol. 16, 2020.

[33] P. Kasyoka, M. Kimwele, and S. M. Angolo, "Cryptanalysis of a pairing-free certificateless signcryption scheme," *ICT Express*, vol. 7, no. 2, pp. 200–204, 2021.

[34] A. Karati, C. I. Fan, and J. J. Huang, "An efficient pairing-free certificateless signcryption without secure channel communication during secret key issuance ☆," *Procedia Computer Science*, vol. 171, pp. 110–119, 2020.

[35] A. Karati, C. I. Fan, and R. H. Hsu, "Provably secure and generalized signcryption with public verifiability for secure data transmission between resource-constrained IoT devices," *IEEE Internet of Things Journal*, vol. 6, no. 6, Article ID 10431, 2019.

[36] D. He, S. Zeadally, B. Xu, and X. Huang, "An efficient identity-based conditional privacy-preserving authentication scheme for vehicular ad hoc networks," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 12, pp. 2681–2691, 2015.

[37] X. Jia, D. He, Q. Liu, and K. K. R. Choo, "An efficient provably-secure certificateless signature scheme for Internet-of-Things deployment," *Ad Hoc Networks*, vol. 71, pp. 78–87, 2018.

[38] S. H. Seo, J. Won, and E. Bertino, "pCLSC-TKEM: a pairing-free certificateless signcryption-tag key encapsulation mechanism for a privacy-preserving IoT," *Transactions on Data Privacy*, vol. 9, no. 2, pp. 101–130, 2016.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 4645585, 10 pages
https://doi.org/10.1155/2022/4645585

Research Article

Blockchain-Based Electronic Medical Records System with Smart Contract and Consensus Algorithm in Cloud Environment

Sanjeev Kumar Dwivedi,[1] Ruhul Amin,[1] Jegatha Deborah Lazarus,[2] and Vijayakumar Pandi[2]

[1]*Department of Computer Science & Engineering, DR SPM International Institute of Information Technology (IIIT-NR), Atal Nagar-Nava Raipur, Chattisgarh, India*
[2]*Department of Computer Science & Engineering, University College of Engineering Tindivanam, Tindivanam, India*

Correspondence should be addressed to Jegatha Deborah Lazarus; blessedjeny@gmail.com

Received 6 May 2022; Accepted 29 July 2022; Published 15 September 2022

Academic Editor: Jie Cui

The blockchain is a peer-to-peer distributed ledger technology that works on the precept of "write-once-read-only." In a blockchain, pieces of information are arranged in the form of blocks, and these blocks are linked together using the hash value of previous blocks. The blocks in a blockchain mechanism are appended only, which means that once information is stored in a block and it cannot be changed; no one tampers the block's content. The traditional electronic medical records (EMRs) based system stores the patients' information in a local database or server, which provides centralization of information, and traditional EMRs are more centric on the health providers. So, security and sharing of patients' information are difficult tasks in the traditional EMR system. The blockchain mechanism has the potential to resolve these existing problems. Due to the appended-only-ledger principle and decentralization of blocks between the network participants, blockchain technology is suited to the EMR system. In this article, first, we discuss all the existing EMR systems and discuss their drawbacks. Keeping all the drawbacks in our mind, we propose a blockchain-based medical record system that utilizes clouding technology for storage purposes. Furthermore, we have designed a smart contract and consensus algorithm for our proposed EMR. Our system only uses a permissioned blockchain model so that only verified and authenticated users can generate their data and participate in the data-sharing system.

1. Introduction

In the recent epoch, patients' medical information is growing rapidly due to the collaboration of information technology (wearable Internet of things devices, e.g., wearable sensors) and the healthcare system. The patients' medical information is important because it provides significant help for medical researchers as well as service providers to turn up with the proper result, which will help to diagnose patients [1]. Medical researchers and service providers often want to share patients' data. So securely storing the patient data is a crucial task [2]. The traditional electronic medical record-based (EMR) system does this task. The EMR-based system provides real-time patient records, ease of access to these records, improved accuracy, sharing of patient data between different researchers, and safety and security to the patient data compared to the paper-based system [3]. In the paper-based system, maintaining and storing patient data is a difficult task (such as huge numbers of rooms are needed to store patients' records, etc.), and the safety of records is also not guaranteed [4, 5]. Malicious users are easily able to do harmful activity in these records. To rectify these problems, the electronic medical record-based system is used in comparison to the paper-based system. But still, in the traditional EMRs-based system, many pitfalls are present.

1.1. Shortcoming of Existing Electronic Medical Record (EMR)-Based Systems. Currently, electronic medical records (EMRs)-based systems are widely popular because they can manage the huge volume of patients' data and provide easy

access to these data. But although, there are several drawbacks present in the existing EMRs-based system.

(i) The EMRs-based system stores data related to a patient in a local database or server, which provides centralization to the patients' data [6]. If users or service providers want to access the patients' data, they directly access it without the patient's intervention.

(ii) The traditional EMRs-based systems are more centric on health providers. The health providers (i.e., hospitals, authorities, etc.) share the patient's data without the patient's knowledge. Therefore, they can manipulate the data. As a result, the originality and integrity of data are at high risk.

(iii) The records of the patients are not secure and safe in the traditional EMRs-based system. Malicious users (or) attackers enter the EMR system due to the lack of security and privacy mechanism [7, 8] present in these systems, and then they tamper (or access) the patient's data.

(iv) The sharing of patients' data in the traditional EMRs-based system becomes problematic because different health providers use different encryption methods and schemas [9]. (even if the patient has agreed to share the data with service providers).

(v) Currently, IoT-based smart devices [4, 10, 11] (i.e., wearable sensors) are also generating the patient's data. Generally, cloud servers (such as storage) are used to store patient's data [12, 13]. But, this mechanism demands more cost and time in maintenance. Therefore, the system's overall efficiency has degraded [14, 15].

(vi) Due to the health providers' centric approach, they can modify the patients' data. So updating the medical records in the EMR system is also a big challenge.

In summary, the current EMR system has several pitfalls, such as centralized storage and inadequate access control mechanism. Therefore, a decentralized technique is highly required to store the patient's data, and at the same time, it should provide privacy and security, proper access control mechanisms [16], and authenticity for the patients' data [7, 17, 18]. The blockchain mechanism can solve the problems mentioned earlier. Due to its inherent characteristics, it is well suited to healthcare applications [4, 5, 14, 15].

1.2. Motivation behind the Proposed Work. Many pitfalls exist in the traditional electronic medical records (EMRs) system, which motivated us to propose a new framework for the EMR system using blockchain technology. A few of them are discussed here.

(i) In the traditional EMR system, patients' data is stored in the central database, and in a centralization system, the "single-point-failure" problem exists. This existing problem motivates us to design a new system in which patients' data is stored in a decentralized way such that if any node fails, then also we will be able to retrieve the patient's data. The blockchain mechanism is well suited for the abovementioned problem.

(ii) In the traditional EMR system, the security and privacy of patients' data in vulnerable. The present system does not provide a sufficient solution for these problems. With the integration of blockchain technology in the current EMR system, patient data security and privacy can be achieved.

(iii) The traditional EMR systems are more centric on health providers. They share the patients' data without the knowledge of patients. To rectify this problem, a suitable system is needed where patients are the central authority for sharing their medical data with other providers.

(iv) In the traditional EMR system, health providers cannot share the patient's data, even if patients concur to share their data. The reason behind this problem is that health providers use different schemas to store the data in their local databases.

So, a proper mechanism is needed which can resolve all these problems. Blockchain technology has the potential to resolve all these problems.

1.3. Major Contributions. This article presents the following contributions:

(i) We have rigorously performed the literature review for the blockchain-based electronic medical health record system and then we have also discussed the shortcomings of the existing system.

(ii) A proposed framework for an electronic medical health record system with blockchain technology has been proposed by considering all aspects of the EMR system.

(iii) We have designed a smart contract algorithm using a finite state machine for the proposed EMR system.

(iv) We have also designed a consensus algorithm for the proposed EMR system.

(v) Finally, we have given some future research challenges with security concerns.

1.4. Organization of the Article. The rest of the article is organized as follows: the literature review for the EMR system is presented in Section 2. Section 3 deals with the proposed architecture with smart contract and consensus algorithm, followed by concluding remarks for the article in Section 4.

2. Related Works

Many authors attempted to solve the problems of the traditional EMRs based system by using the blockchain

mechanism [4, 5, 19, 20]. Some authors also used the smart contracts mechanism to solve it. Uddin et al. [4] proposed "A Patient Agent (PA) Based Remote Patient Monitoring (RPM) Architecture." Every patient has its own patient agent in their architecture, which is stored on the patient local server (PLS). In their architecture, PA selects one node as a miner among the available nodes, and the miner's work is to generate a hash value of the current block. Xia et al. [5] proposed a blockchain-based data-sharing scheme. The framework proposed by Xia et al. addresses the problem associated with sensitive data stored in the cloud environment. The authors suggest the patient-centric solution in [12] for health data sharing system, using a private blockchain. Azaria et al. [19] proposed a MedRec: a decentralized record management system to handle electronic medical records using blockchain technology. The problem with the proposed architecture is the security of the individual database. They did not address this problem, and the key management problem remains unsolved in the proposed architecture. Chen et al. [20] proposed a new business process for medical information sharing based on a blockchain mechanism. The proposed approach is "patient-centric," where patients are the central authority for viewing and sharing their medical records. The limitation of this method is the smart contract mechanism.

Yang and Li [7] proposed a blockchain-based architecture for EHRs systems, using a new incentive mechanism to create a new block. The architecture works on top of the existing database, which healthcare provider maintains. Griggs et al. . [13] proposed blockchain-based smart contracts to secure remote patient monitoring. The proposed framework uses a private (permissioned) blockchain. Al Omar et al. [8] proposed a permissioned blockchain-based healthcare data management system to attain privacy and security. The proposed solution is a "Patient-Centric" approach in which the patient is the sole authority to keep the data on a blockchain. Dubovitskaya et al. [9] proposed a blockchain-based healthcare data management framework, especially for electronic medical records (EMRs) systems. They provide a secure and trustable framework for sharing in the EMR system. Novikov et al. [21] presented a decentralized blockchain-based infrastructure to store patients' electronic medical records (EMRs) in a healthcare system.

Table 1 compares the existing blockchain-based medical records system concerning blockchain taxonomy (i.e., smart contracts, consensus algorithm, authentication, key management, and 51% attacks, etc.). In Table 1, two abbreviations are used: ND and PB. In this specific column, PB means the type of blockchain is permissioned blockchain, and ND means that the authors did not discuss the type of blockchain. The solution provided by the authors is applicable for both permissioned and permissionless blockchains. ND means that the respective authors did not discuss the corresponding blockchain taxonomy in other columns. Table 2 compares the existing approach with its advantage and disadvantage.

3. Proposed Architectures

3.1. System Overview. Our proposed architecture comprises the following components (or entities): a central authority and a management system, known as CAMS, a list of the service provider (e.g., doctors, insurance companies, and research organizations, etc.), the user (generally, a patient), a pool of Data Lake, hash generators, and a cloud server. These entities are connected using a decentralized peer-to-peer architecture. The architecture of the proposed system is shown in Figure 1.

3.2. Role and Responsibilities of the Involved Entities

(i) Central Authority and Management System (CAMS): the CAMS is responsible for generating a pair of keys and issuing the same keys using the cryptographic mechanism to the user and service provider. The proposed framework utilizes the permissioned blockchain system. If any new user or service provider wishes to join the system, firstly, they take permission from CAMS. Here permission means that the CAMS authenticate them because the new user or service provider may be a malicious user. CAMS is also responsible for managing the entire system. CAMS has a list of users and service providers who are already present in the network. If any new user or service provider joins, then after the process of key generation and authentication, CAMS updates the list, which tells that currently how many users and service providers exist in the system.

(ii) User: the user is generally a patient who wants services from service providers. All users have a copy of smart contracts which tells about the agreement or set of protocols—the copy of smart contracts issued by the CAMS. If any user wishes for services from service providers, this request is checked by smart contracts. If smart contracts are executed correctly, then, only the user can take the services from service providers; otherwise, not. The service providers are doctors, hospital authorities, insurance claim companies, and medical researchers, etc.

(iii) Service providers: service providers provide their services to the user according to the need of the user. Service providers are doctors, insurance companies, laboratory offices, and scientific researchers, etc. The user consults with a doctor for specific medical treatment. The doctor gives suggestions accordingly. Doctors often suggest a specific medical test according to the user's problem. For this, the user goes to the laboratory office (inside or outside the hospital) to perform the test. If the laboratory office gives the result immediately to the patient, then the patient shows these results to the doctor and gets suggestions for some medicine, if required.

Table 1: The comparison of existing approaches with their advantages and disadvantages.

Researcher	Blockchain	Type smart contract mechanism	Consensus algorithm	Authentication and key management	Scalability	Mining incentive	Blockchain specific vulnerability
Uddin [4]	ND	No	No	No	ND	Yes	ND
Xia [5]	PB	No	No	Yes	Yes	No	ND
Liang [12]	PB	No	No	No	No	No	ND
Azaria [19]	ND	Yes	Yes	No	ND	Yes	ND
Chen [20]	PB	No	Yes	Yes	Yes	No	ND
Yang [7]	ND	Yes	No	No	Yes	Yes	ND
Griggs [13]	PB	Yes	Yes	No	ND	ND	ND
Al Omar [8]	PB	Yes	No	Auth.	ND	ND	ND
Dubovitskaya [9]	PB	No	Yes	Yes	No	No	ND
Novikov [21]	ND	Yes	No	Auth.	ND	ND	ND

ND: not discussed; PB: permissioned blockchain; Auth.: authentication; No: not present or did not discuss the required algorithm or mechanism; and yes: provided required algorithm or mechanism.

Table 2: The comparison of existing approaches with their advantages and disadvantages.

Reference	The idea of the article	Advantages	Disadvantages
[4]	A patient agent (PA) based remote patient monitoring (RPM) architecture.	The PA selects miners based on the available CPU resources and previous performance of miners. So by doing this time is minimized.	The smart contract mechanism and consensus mechanism are not discussed in this article. This architecture is also vulnerable to denial of service attacks and ransom cyber-attack.
[20]	A new business process for medical information sharing based on a blockchain mechanism.	The proposed approach is "patient-centric" where the patient has all the authority for viewing and sharing his/her medical records.	The authors do not investigate and analyze the smart contracts mechanism under the permissioned blockchain.
[7]	A blockchain-based architecture for EHR systems using a new incentive mechanism for the creation of any new block in a blockchain-based system has been proposed.	The proposed architecture uses a smart contract mechanism for agreement between patient and provider.	The proposed architecture is a "provider-centric" approach.
[13]	A blockchain-based smart contract for secure remote patient monitoring has been discussed.	The proposed method uses smart contracts and the PBFT consensus mechanism.	Key management and authentication and blockchain-based specific vulnerability (51% attacks) part are not discussed by the author.
[21]	A decentralized blockchain-based infrastructure for storing the electronic medical records (EMR) of the patients in a health care system.	This scheme uses a patient-centric model with the support of smart contracts to access patient data.	The consensus mechanism and blockchain-related vulnerability are not discussed by the author.
[5]	A blockchain-based data sharing scheme.	The proposed architecture is scalable to any number of nodes.	The communication and authentication protocols are not discussed.
[12]	A solution for health data sharing using a private blockchain has been discussed.	The method proposed is a patient-centric approach.	The authors do not explore the underlying smart contract and consensus mechanism.
[8]	A permissioned blockchain-based healthcare data management system to attain privacy and security for healthcare data has been proposed.	The proposed solution is a "patient-centric" approach. Smart contracts are helpful for interaction with the blockchain.	They assume that the user is having a key and password. How these keys are generated, they did not discuss.
[9]	A blockchain-based healthcare data management especially for electronic medical records (EMR) has been proposed. The proposed framework consists of the membership service, local database, cloud server, chain code, and the user (either patient or doctor) and nodes with his own ledger. Practical byzantine fault tolerance (PBFT) for the consensus mechanism has been used.	This scheme supports access control and data availability with desired security.	The issue with the framework is scalability. Blockchain-based specific vulnerability (51% attack) is not discussed by the author.
[19]	MedRec: a decentralized record management system, to handle electronic medical record systems by using blockchain technology has been proposed.	In this architecture, they used smart contracts.	The problem with the proposed architecture is the security of an individual database.

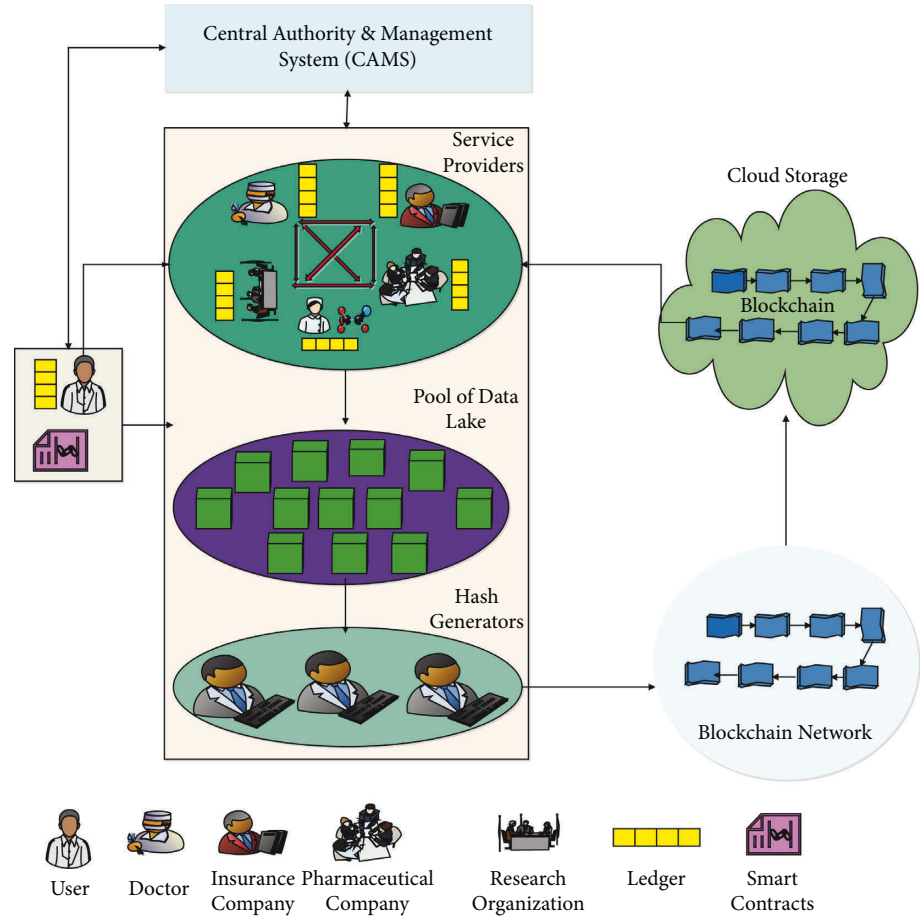


Figure 1: Proposed architecture.

Sometimes laboratory office directly gives the result to the doctor. Many users also take insurance plans like health insurance plans and term insurance plans according to their needs; therefore, insurance companies have also come into the picture as service providers. Pharmaceutical companies interact with doctors and insurance companies to brand and sell their medicine. Scientific researchers interact with different service providers (doctors, pharmaceutical companies, and users) for their research. Due to all these activities, a huge volume of data is generated. The EMRs system without a blockchain cannot handle this (as discussed in section 1). But by using the blockchain with EMRs system, trust in a patient's data increases and transparency of the entire system

(iv) Pool of data lake: the pool of Data Lake contains the bunch of data that users and service providers generate. Service provider gives their services (a patient consults a doctor and provides description of health records and insurance records) to the user; and all these huge amounts of data are kept inside the pool of Data Lake. The pool of Data Lake is a container (or database) used only to store the generated data.

(v) Hash generators: the hash generators generate the hash value of the current block. The hash generators module takes the data from a pool of Data Lake and converts it into the size of a specified block. First, they validate the block. After the validation process, they keep the block inside the blockchain system. CAMS specifies the size of the block. In the proposed system, more than one hash generators exist. The CAMS module picks the suitable hash generator, depending on the existing performance of hash generators. So at any point in time, only one or two online hash generators are available.

(vi) Cloud server: the cloud server only stores the blocks in the blockchain network. Our proposed architecture uses the cloud server instead of a local database because the volume of data is high. As per the discussion in section 2, these blocks are connected by using the hash value of previous blocks, so tampering with the data in any block is impossible. If scientific researchers want to use the patient's data for research purposes, they can use it with the user's permission only. Without the user's permission, scientific researchers, as well as service providers, are not able to take and share the user's data. Since the user data are stored in the blockchain system, using the cryptographic mechanism, it provides security for tamper-proof and immutable of user data.

3.3. Algorithm for the Proposed System. The algorithm for the proposed system is as follows:

(i) *Step 1*. CAMS authenticates the user who wants services from the service provider.

(ii) *Step 2*. If the user's authentication is successful, then the user is granted to take services from providers; otherwise, the error message is generated: authentication is not successful.

(iii) *Step 3*. If a new user wants to join the system, the new user may join after CAMS approval. The CAMS generates the pair of keys, and steps 1 and 2 are repeated.

(iv) *Step 4*. The user consults with service providers, and the data generated by them are kept in a pool of Data Lake.

(v) *Step 5*. The hash generators collect the data from the Data Lake pool, verify it, convert it into a block, and add it to a blockchain by using the previous block's hash value. This step is repeated after a while.

(vi) *Step 6*. Steps 4 and 5 are repeated for every user who wants services from the provider.

(vii) *Step 7*. Finally, the blockchain is stored in a cloud server.

3.4. Smart Contract Algorithm of the Proposed System. As per discussion, in Figure 2 in section 2.6, smart contracts are based on the state machine model, and the state machine model is always a deterministic state machine model. Smart contracts are define the set of rules, which are written in the form of the program (or scripts). This set of scripts are stored on all the nodes of the blockchain system. In turn, the blockchain nodes execute these scripts to perform certain activities or transactions in the network [22, 23]. By using the same concept, we also propose a deterministic state machine model for the proposed system since a deterministic state machine model is represented as a directed graph, and a directed graph consists of a set of vertices and a set of edges. In the deterministic state machine model, these sets of vertices are referred to as a set of states, and a set of edges is referred to as a transition from one state to another state or in the same state. The advantage of showing the smart contracts using the state machine model is that it is very easy for the developer to write the code by seeing the flow of the state machine. Moreover, it triggers the events to achieve the necessary behavior, which suits the EMR system. Furthermore, disclosing the smart contract code to external parties is not required. They can predict the system behavior and write the code with add-on requirements. Our proposed state machine consists of 4 states: labeled as state 0, state 1, state 2, and state 3. The user is represented as state 0; CAMS is represented as state 1; service providers are represented as state 2. State 3 is called a dead state. The proposed state machine model is shown in Figure 2.

In the state machine model, certain actions are defined: Authentication, No Action, Violation, and Permission. This set of actions is used to transition from one state to another. For example, if the machine is in state 0 and the action is Authentication, then the machine automatically moves from state 0 to state 1. If the machine is in state 1 and the action is Violation, then the machine automatically moves from state 1 to state 2. If the machine is in state 2 and the action is No Action, then the state does not change and so on (see Algorithm 1).

The solidity programming language can be used to implement the proposed smart contract for the EMR system. It is a statically-typed programming language influenced by other languages such as JavaScript, C++, and Python. Moreover, this language is designed to run on the Ethereum Virtual Machine (EVM). The Remix IDE with the solidity version 0.5.10 is used to execute the suggested smart contract. Remix IDE provides a convenient platform to deploy smart contracts. It provides three different environments (JavaScript VM, Injected Web3, and Web3 Provider) to execute and deploy smart contracts [24]. Furthermore, the execution cost plays a crucial role when smart contracts are executed in any of these three environments. Execution cost determines the total cost (in terms of "gas") required to execute the defined computational operations.

3.5. Consensus Mechanism for Proposed System. The consensus algorithm is the set of rules to reach a common viewpoint or agreement. The consensus algorithm is designed so that, after executing the block, all the nodes (or majority of nodes) in a network agree that the block is valid and can be included in the blockchain network. Once the agreement is done, no node can change the decision. For the proposed system, we also designed a consensus mechanism for the verification and validation of a new block [25]. In our proposed architecture, a new block is verified and validated with the help of hash generators. The main work of hash generators is to validate the block (whether the correct user sends the block or data or not because it may be possible that malicious users send the data in a pool of data lake, so validation is needed) and after the validation of new block, generate the hash value of new block, and finally add them in a blockchain.

In the proposed architecture, more than one hash generator exists but only one hash generator is responsible for validating and generating the hash value of the new block. The work of other hash generators is to validate the new block. The selection of hash generators is based on the previous performance of the hash generators or on the stake or wealth deployed in the network because the service providers also act as hash generators. The reason behind this is if the service providers act as a hash generator, based on their wealth deployed in a system, then the chance of malicious activity is very less because in that case, if they are performing a malicious activity, they are damaging their own wealth. In the proposed system, two categories of hash generators exist. In the first category, only one hash generator exists which is responsible for both validating as well as generating the hash value of the new block, and in the second category, remaining hash generators exist which are responsible for validating the new block only (see Algorithm 2).

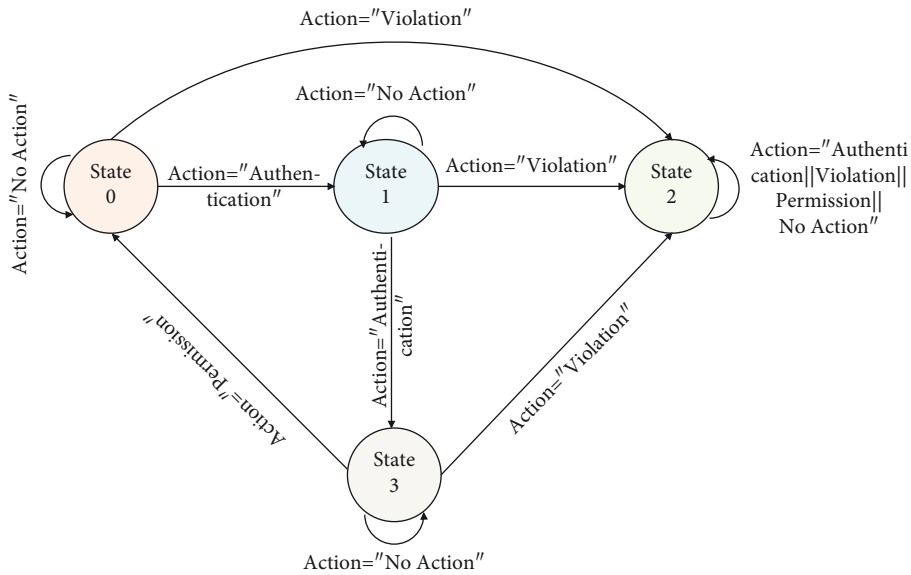


Figure 2: Smart contract using a finite state machine.

1 **Require:** *Actions such as Authentication, Violation, Permission, No Action*
2 **Ensure:** *Messages such as Authentication successful, Error and abort, Permission granted, No Action required*
3 FOR STATE 0 AND STATE 1:***Action*: Authentication, violation, No action**
4 **If** (*f(Action)* = = *Authentication*) **then**
5 f(Message) = authentication is successful;
6 move: S (1)
← S(0);
7 **Else if** (*f(Action)* = = *No Action*) **then**
8 f(Message) = No action required;
9 move: S (0)
← S(0);
10 **Else if** (*f(Action)* = = *Violation*) **then**
11 f(Message) = error and abort;;
12 move: S (3)
← S(0);
13 **End if**
14 FOR STATE 2:***Action*: Permission, violation, No action.**
15 **If** (*f(Action)* = = *Permission*) **then**
16 f(Message) = take permission;
17 move: S (1)
← S(2);
18 **Else if** (*f(Action)* = = *No Action*) **then**
19 f(Message) = No action required;
20 move: S (2)
← S(2);
21 **Else if** (*f(Action)* = = *Violation*) **then**
22 f(Message) = error and abort;
23 move: S (3)
← S(2);
24 **End if**
25 FOR STATE 3:***Action*: Permission, authentication, violation, No action**
26 **If** (*f(Action)* = = *Permission/Authentication/Violation/No Action*) **then**
27 f(Message) = error and abort;
28 move: S (3)
← S(3);
29 **End if**

Algorithm 1: Smart contracts as state machine model.

Require: *Authentication value, Validation value, Genesis block*
Ensure: *New block, Blockchain length, Nonce-value* SP[n] = list of service providers; HG[] = hash generators; Per[] = performance for hash generators; BCl = block chain length;
A_v = authenticating value provided by hash generators; V_v = validating value provided by hash generators; BC: Blockchain; B0: Genesis block;
$A_v = 0$;
$V_v = 0$;
$BCl = 1$;
$Per = 0$;
For ($i = 0$ to hash_generator − 1) **do**
Execute Per[HG[i]];
End for
Per ← Per[HG[0]];
For ($i = 1$ to hash_generator) **do**
If ($Per[HG[i]]\>; Per$) **then**
Per ← Per[HG[i]];
End if
End for
Display: Selected hash_generator)
For ($i = 0$ to hash_generator − 1) **do**
Check authentication of selected hash_generator;
If (Authentication == true) **then**
$A_v = A_v + 1$;
End if
End for
If ($A_v \geq \lceil HG[m]/2 \rceil$) **then**
Display: Authentication is successful;
End if
Create a new block by using the selected hash_generator
For ($i = 0$ to hash_generator − 1) **do**
Check validation of new_block by all hash_generators;
If (Validation == true) **then**
$V_v = V_v + 1$;
End if
End for
If ($V_v \geq \lfloor HG[m]/2 + 1 \rfloor$) **then**
Display: Validation is successful;
End if
While (TRUE) **do**
Calculate the Proof_Hash value for new_block;
If (Proof_Hash == Target_Value) **then**
$BCl = BCl + 1$;
End if
Change Nonce_value;
End while

Algorithm 2: A consensus algorithm for proposed architecture

3.5.1. Consensus Mechanism of the Proposed System

(1) In the first phase, after the selection of the hash generator, the remaining hash generators first authenticate this selected hash generator. If [ceiling (N/2)] number of hash generators authenticates this selected hash generator (assume that in a system "*N*" number of hash generators exist, excluding the selected one), then authentication is successful and it proceeds further; otherwise, the system aborts it with a message: authentication not successful; error message.

(2) In the second phase, the selected hash generator picks the data from the pool of Data Lake, converts it into a block, and sends the new block to other hash generators for validation. All the hash generators, including the selected one, validate the new block. If [floor (*N*/2) +1] number of hash generators, validate the new block (assume that in a system "*N*" number of hash generators exist, excluding selected one, +1 is used for selected hash generator) then validation of a new block is successful and it proceeds further, otherwise, the system aborts it with a message: validation not successful; error message.

(3) In the third phase, the selected hash generator generates the hash value of the new block and is added to the blockchain system.

4. Conclusion

In this article, we present the blockchain-based novel approach for electronic medical records (EMRs) systems. The proposed blockchain-based system provides several advantages compared to traditional electronic medical records-based systems. Traditional EMRs systems are more centric on healthcare providers. Whereas the proposed BMRS approach is centric on the patient only, which means that if the healthcare providers want to access the patient's data, they can access and share the patient's data with the patient's permission, which is an advantage over the traditional EMRs system. In the proposed architecture, hash generators are responsible for the maintenance of the blockchain system, including the creation of a new block, validation of a new block, and finally, adding the block to the blockchain network. The service providers also act as hash generators. The proposed system considers both smart contracts mechanism as well as a consensus mechanism. The smart contracts mechanism is based on the state machine modal. Hash generators use the consensus algorithm to authenticate the healthcare providers and to validate new blocks.

In future work, our research team will try to incorporate the incentive mechanism with its mathematical model and provide a solution for the mitigation of various attacks, such as routing attacks and phishing attacks that increase the security of the EMR system.

Data Availability

The datasets generated or analyzed during the current study are not publicly available because the data are strictly confidential because the manuscript is based on patient records that are maintained electronically. The authors understand that these data are to be maintained confidentially and hence they are not provided in the manuscript or elsewhere.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] A. Saha, R. Amin, S. Kunal, S. Vollala, and S. K. Dwivedi, "Review on "Blockchain technology based medical healthcare system with privacy issues"," *Security and Privacy*, vol. 2, no. 5, p. e83, 2019.

[2] J. Zhang, N. Xue, and X. Huang, "A secure system for pervasive social network-based healthcare," *IEEE Access*, vol. 4, pp. 9239–9250, 2016.

[3] P. T. S. Liu, "Medical Record System Using Blockchain, Big Data and Tokenization," in *Information and Communications Security. ICICS 2016*, K. Y. Lam, C. H. Chi, and S. Qing, Eds., Springer, Berlin, Germany, pp. 254–261, 2016.

[4] M. A. Uddin, A. Stranieri, I. Gondal, and V. Balasubramanian, "A patient Agent to manage blockchains for remote patient monitoring," *Studies in Health Technology and Informatics*, vol. 254, pp. 105–115, 2018.

[5] Q. Xia, E. Sifah, A. Smahi, S. Amofa, and X. Zhang, "BBDS: blockchain-based data sharing for electronic medical records in cloud environments," *Information*, vol. 8, no. 2, p. 44, 2017.

[6] D. Ivan, "Moving toward a blockchain-based method for the secure storage of patient records," in *Proceedings of the 2016 ONC/NIST Use of Blockchain for Healthcare and Research Workshop*, Gaithersburg, ML, USA, August 2016.

[7] G. Yang and C. Li, "A design of blockchain-based architecture for the security of electronic health record (EHR) systems," in *Proceedings of the 2018 IEEE International Conference on Cloud Computing Technology and Science (CloudCom)*, pp. 261–265, Nicosia Cyprus, December 2018.

[8] A. Al Omar, M. S. Rahman, A. Basu, and S. Kiyomoto, "Medibchain: A Blockchain Based Privacy Preserving Platform for Healthcare Data," in *Security, Privacy, and Anonymity in Computation, Communication, and Storage. SpaCCS 2017. Lecture Notes in Computer Science*, G. Wang, M. Atiquzzaman, Z. Yan, and K. K. Choo, Eds., Springer, Berlin, Germany, pp. 534–543, 2017.

[9] A. Dubovitskaya, Z. Xu, S. Ryu, M. Schumacher, and F. Wang, "Secure and Trustable Electronic Medical Records Sharing Using Blockchain," in *Proceedings of the 2017. AMIA Annu Symp Proc American Medical Informatics Association*, Washington, DC, USA, November 2017.

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

[11] E. Karafiloski and A. Mishev, "Blockchain solutions for big data challenges: a literature review," in *Proceedings of the IEEE EUROCON 2017 -17th International Conference on Smart Technologies*, pp. 763–768, Ohrid, Macedonia, July 2017.

[12] X. Liang, J. Zhao, S. Shetty, J. Liu, and D. Li, "Integrating blockchain for data sharing and collaboration in mobile healthcare applications," in *Proceedings of the 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Montreal, QC, Canada, October 2017.

[13] K. N. Griggs, O. Ossipova, C. P. Kohlios, A. N. Baccarini, E. A. Howson, and T. Hayajneh, "Healthcare blockchain system using smart contracts for secure automated remote patient monitoring," *Journal of Medical Systems*, vol. 42, no. 7, p. 130, 2018.

[14] Z. Alhadhrami, S. Alghfeli, M. Alghfeli, J. A. Abedlla, and K. Shuaib, "Introducing blockchains for healthcare," in *Proceedings of the 2017 International Conference on Electrical and Computing Technologies and Applications (ICECTA)*, Ras Al Khaimah, UAE, November 2017.

[15] T. Bocek, B. B. Rodrigues, T. Strasser, and B. Stiller, "Blockchains everywhere-a use-case of blockchains in the pharma supply-chain," in *Proceedings of the 2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*, May 2017.

[16] J. P. Dias, L. Reis, H. S. Ferreira, and Â Martins, "Blockchain for Access Control in E-Health Scenarios," 2018, https://arxiv.org/abs/1805.12267.

[17] D. Ding, M. Conti, and A. Solanas, "A smart health application and its related privacy issues," in *Proceedings of the 2016 Smart City Security and Privacy Workshop (SCSP-W)*, April 2016.

[18] X. Li, P. Jiang, T. Chen, X. Luo, and Q. Wen, "A survey on the security of blockchain systems," *Future Generation Computer Systems*, vol. 107, pp. 841–853, 2020.

[19] A. Azaria, A. Ekblaw, T. Vieira, and A. Lippman, "Medrec: using blockchain for medical data access and permission

management," in *Proceedings of the 2016 2nd International Conference on Open and Big Data (OBD)*, August 2016.

[20] J. Chen, X. Ma, M. Du, and Z. Wang, "A blockchain application for medical information sharing," in *Proceedings of the 2018 IEEE International Symposium on Innovation and Entrepreneurship (TEMS-ISIE)*, April 2018.

[21] S. P. Novikov, O. D. Kazakov, N. A. Kulagina, and N. Y. Azarenko, "Blockchain and smart contracts in a decentralized health infrastructure," in *Proceedings of the 2018 IEEE International Conference "Quality Management, Transport and Information Security, Information Technologies" (IT&QM&IS)*, September 2018.

[22] S. K. Dwivedi, R. Amin, and S. Vollala, "Blockchain based secured information sharing protocol in supply chain management system with key distribution mechanism," *Journal of Information Security and Applications*, vol. 54, Article ID 102554, 2020.

[23] S. K. Dwivedi, R. Amin, and S. Vollala, "Blockchain-based secured IPFS-enable event storage technique with authentication protocol in VANET," *IEEE/CAA Journal of Automatica Sinica*, vol. 8, no. 12, pp. 1913–1922, 2021.

[24] S. K. Dwivedi, M. S. Obaidat, R. Amin, and S. Vollala, "Decentralized management of online user reviews with immutability using IPFS and Ethereum blockchain," in *Proceedings of the 2022 International Mobile and Embedded Technology Conference (MECON)*, March 2022.

[25] S. K. Dwivedi, P. Roy, C. Karda, S. Agrawal, and R. Amin, "Blockchain-Based Internet of Things and Industrial IoT: A Comprehensive Survey," *Security And Communication Networks*, vol. 2021, Article ID 7142048, 21 pages, 2021.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 2332769, 14 pages
https://doi.org/10.1155/2022/2332769

Research Article

Telematics Collaborative Resource Allocation Algorithm Based on Cloud Sidecar

Zheng Zhang,[1] **Yanling Shao**,[1] **Xing Liu**,[2] **and Yibo Han**[3]

[1]*School of Computer and Software, Nanyang Institute of Technology, Nanyang, Henan 473000, China*
[2]*Frontier Information Technology Research Institute, Zhongyuan University of Technology, Zhengzhou, Henan 450000, China*
[3]*Nanyang Institute of Big Data Research, Nanyang Institute of Technology, Nanyang, Henan 473000, China*

Correspondence should be addressed to Zheng Zhang; zhangzheng@nyist.edu.cn

Received 13 July 2022; Revised 4 August 2022; Accepted 25 August 2022; Published 8 September 2022

Academic Editor: Ke Gu

This paper provides an in-depth study and analysis of a distributed allocation algorithm for collaborative resources for cloud-edge-vehicle-based Telematics. The approach starts from the emerging application of urban environmental monitoring based on vehicular networking, with an integrated design of data sensing detection and transmission, and collaborative monitoring of vehicle swarm intelligence based on urban air quality collection to avoid redundancy of information and communication overload. A hybrid routing method with minimal delay for reliable data transmission is proposed. The power adjustment algorithm divides the channel into 3 states. When the CBR is less than 0.5, the channel is in an idle state, and when the CBR is greater than 0.5 and less than 0.8, the channel is in an active state. The algorithm designs redundancy strategies based on coding mechanisms to improve the reliability of data transmission, combines coding mechanisms with routing design, incorporates routing switching ideas, and performs probability-based routing decisions to minimize the delay. In straight-line road sections, a fuzzy logic prediction-based vehicle adaptive connectivity clustering routing algorithm is proposed to reduce the communication overhead during vehicle collaboration and ensure high network connectivity; at intersections, a probability-based minimum delay routing decision algorithm is proposed to reduce the information transmission delay. Experiments show that the proposed method effectively improves the efficiency of data-aware collection and transmission, and increases the reliability of transmission. With the explosive growth of video services, the problem of intelligent transmission of DASH-based video streams has become another research hotspot in mobile edge networks. Based on the edge container cloud architecture of vehicular networking, the resource constraints of microservices when deployed in the edge cloud platform were analyzed, and a multi-objective optimization model for microservice resource scheduling was established with the comprehensive performance objectives of shortest microservice invocation distance, highest resource utilization of physical machine clusters, and ensuring load balancing as much as possible.

1. Introduction

Before the emergence of container technology, microservices were mainly deployed on virtual machines directly on bare metal, which was difficult to operate and maintain. Container technology, on the other hand, is a lightweight virtualization technology that provides resource scheduling and isolation for microservices at the container engine layer, reducing concerns about inconsistencies across platforms between development, testing, and production environments [1]. When deploying specific functional modules in different locations, how to utilize limited resources in a more balanced and efficient manner and further improve application service quality and user experience is a challenge in the initialization process of mobile IoT slices. And because containers reduce the hardware system virtualization layer, they can use the hardware resources of the actual physical host directly and therefore make fuller use of hardware resources. In summary, lightweight, and fast start/stop containers are well suited for edge workloads and are a good vehicle for Telematics microservices. However, as the granularity of containers is smaller, the number of

containers that can be started by one physical machine is also larger, and the resource management for containers is more complex [2]. Common container scheduling tools such as Kubernetes and Docker Swarm Kit only provide some simple resource scheduling policies, which cannot fully utilize the performance of physical machines, so it is crucial to design a reasonable resource scheduling policy for microservice containers [3]. Vehicle networking refers to a system network that carries out wireless communication and information exchange among vehicles, roads, pedestrians, and the Internet on the basis of in-vehicle network, inter-vehicle network, and in-vehicle mobile Internet according to agreed communication protocols and data exchange standards. Traditional vehicle networking refers to the electronic label loaded on the vehicle through the wireless radio frequency and other recognition technology, effectively realized in the information network platform to the vehicle attributes and static, dynamic information extraction, and utilization, according to the different function demands to the vehicle operation status to provide effective supervision and comprehensive service system. With the rapid development of the Telematics industry and technology, the traditional definition can no longer cover all its contents.

The rapid development and widespread use of the Internet of Everything have led to a shift in the role of edge devices, from the role of a single consumer of data to the role of a consumer and producer of data, with edge devices becoming more intelligent and capable of autonomous deep learning, predictive analysis, and intelligent data processing of data at the edge of the network. Big data processing is slowly entering the era of edge computing with the Internet of Everything at its core from the cloud computing era [4]. Cloud computing relies on powerful resource provisioning in data centers to centrally process big data, compared to edge computing which relies on numerous edge devices at the edge of the network to process massive amounts of data, reduce the occupation of network resources, enhance real-time communication capabilities, and complete data processing and execution services with extremely low latency [5]. With the growth of the Internet of Everything, latency-sensitive and compute-intensive application services are increasing, and cloud computing solutions cannot meet the latency requirements of these application services. For example, autonomous driving, self-driving cars generate 4 TB of data per day, which is demanding in terms of computational latency. The accuracy of multiedge collaborative mobile IoT slicing can reach 85%, while the other two comparison models are 72% and 65%, respectively. WAN transmission brings latency uncertainty, and autonomous driving data needs to be processed at the edge of the network to ensure low latency, but the computational resources on the edge side of the network are far inferior to cloud computing, and the data are processed through deep neural networks under resource-constrained conditions [6].

With the continuous improvement of relevant standards and the increasing number of smart vehicles, it is foreseeable that more vehicles will be connected to the network through relevant protocols in the future. Along with the increasing number of vehicles, road hazards have become an issue that is faced in the development of Telematics. This makes it increasingly important to study the transmission strategy of vehicle safety services. In the process of vehicle communication based on IEEE 802.11P and LTE-V protocols, channel congestion, channel interference, shadow fading, and intelligent computational processing are the main factors affecting the performance of vehicle communication. It is important to study how to schedule the computational and communication resources in vehicular networking to improve the communication performance of vehicle safety services.

To sum up, the continuous development of the Internet of Vehicles business and the continuous increase in the number of connected vehicles have brought great challenges to the existing Internet of Vehicles solutions. In order to improve driving safety and travel efficiency, the problem of limited computing power of a single vehicle can be solved by offloading tasks to the MEC server for execution. On the one hand, different IoV services have different requirements for latency, bandwidth, and computing power. How to manage and allocate communication and computing resources to meet the needs of various services is a key issue in-vehicle edge computing networks. On the other hand, due to the distributed deployment of MEC servers, the communication and computing resources on edge nodes are relatively limited. However, the unloading requests of vehicles are usually random and sudden, and an unreasonable resource allocation scheme will cause problems such as increased delay, unstable network services, and poor service quality. Therefore, it is of great significance to study communication and computing resource allocation methods for task offloading in-vehicle edge computing networks.

This paper proposes a distributed end-edge collaboration algorithm for the edge network of intelligent networked vehicles. According to the characteristics of high reliability and low delay content transmission of the Internet of Vehicles, a limited block length mechanism is introduced. At the same time, the compression coding power consumption of the vehicle video information source is introduced, and the vehicle energy consumption model is established. According to the video quality requirements of the vehicle video information source, by adjusting the video coding rate, the information source transmission rate, and the selection of vehicle multipath routing, a fully distributed optimization algorithm is proposed to improve the utilization of network resources and ensure a single Equity in energy consumption of vehicles. This paper proposes a distributed edge-end collaborative algorithm based on the subgradient algorithm, which realizes the resource allocation strategy by adjusting the video coding rate, the information source transmission rate, and the vehicle multipath routing decision. The farther the terminal is from the communication node, the greater the data transmission delay. Therefore, this paper uses dynamic communication nodes to solve the delay and energy consumption problems faced by mobile terminals. The algorithm can be deployed and executed in each ICV and only needs to exchange a small amount of information with its neighbouring nodes.

2. Related Works

As transferring large amounts of data to the cloud not only takes up limited backhaul bandwidth resources, it also generates large transmission latency and poses security risks of data leakage. In response to these problems, edge computing was born [7]. Edge computing is a service that deploys data processing and storage capabilities from the cloud as close to the endpoint as possible, storing, and analyzing data at the edge of the network, solving many of the challenges that exist when transferring data to the cloud center. As smart chips continue to develop, the processing power of terminals is gradually increasing, so that some simple data processing can be done locally at the terminal [8]. Of course, edge intelligence and terminal intelligence also have problems that need to be solved, such as the uneven distribution of edge devices and the uneven data storage and processing capabilities; not only is the energy consumption of terminal devices relatively high during processing, the terminal itself has limited endurance, and the life span of the terminal is also a problem that cannot be ignored when processing large amounts of data [9].

A heuristic algorithm based on three scenarios is proposed for the task scheduling problem of edge servers in multiserver multiuser mobile edge computing systems. Experimental results show that the algorithm can significantly reduce the average task execution delay. An efficient lightweight offloading scheme is proposed for the multi-user edge system [10]. The results show that this offloading scheme can effectively reduce the execution time of end-to-end tasks and improve the resource utilization of the edge server [11]. The battery size of end devices is typically very limited due to device size, etc. [12].

Abreha et al. proposed an analytical framework that models downlink traffic in a drive-through vehicle networking scenario via a multidimensional Markov process. It can be speculated that the computing load brought by the number of tasks at this time is not high for the edge servers in the network. So, the effect is not obvious. As the number of tasks increases, the task completion rate varies greatly. When the number of tasks is 60, the lowest task completion rate is 79.1% and the highest is 94.8%. There is a 15.7% gap between the lowest and highest. The arrival of packets in the RSU buffer is constructed as a Poisson process, and the transit time is exponentially distributed [13]. Considering the state space explosion problem associated with multidimensional Markov processes, this paper uses an iterative per-duration technique to compute the stationary distribution of Markov chains [14]. Sar-dianos et al. studied the hybrid data dissemination problem, i.e., optimally determining the time and destination of data transmission vehicles, and whether the vehicles obtain the required data directly from the edge of nearby vehicles [15]. The authors proposed a new data propagation algorithm, called the hybrid data propagation offline algorithm, which prioritizes finding the most beneficial vehicle-to-vehicle broadcast, and then selected the feasible vehicle-to-base station propagation method [16]. Shakir et al. studied how to deploy drop box optimally by considering the trade-off between delivery delay and drop box deployment cost [17]. To address this issue, first, provide a theoretical framework to accurately estimate delivery delay; then, based on the dimension based on the idea of enlargement and dynamic programming [18]. In terms of content uploading, Guan et al. proposed to deploy dedicated access points (APs) at bus stops to facilitate video uploading to study the video uploading problem of mobile buses and proposed a water injection placement algorithm that aims to balance the distribution [19]. The aggregate bandwidth of each bus is analyzed by establishing a queuing model to analyze the upload delay of video content, and a machine learning model is further used to incorporate the impact of bus routes into the queuing model. Based on the different application conditions, it is a great challenge to meet the requirements of low delay, huge amount of calculation, high efficiency, high reliability, and meticulous precision. For example, each car has different requirements for communication; there are self-driving cars and ordinary vehicles, which need to be treated differently.

In an edge computing environment, there are two aspects of energy consumption by the end device when performing task offloading [20]. One is the computational energy consumed when the task is computed locally, and the other is the transmission energy consumed when the task is uploaded to the edge server and the results are received back from the edge server. Therefore, offloading strategies can be designed to reduce energy consumption by means such as adjusting CPU frequency and offloading intensive tasks to the server [21].

The joint proposes a layered, modular edge computing architecture that runs on the cloud, fog, and edge devices and provides containerized services and microservices. The proposed architecture has three main layers: the sensing layer, the intermediary layer, and the enterprise layer. The perception layer is the underlying layer that performs sensing and operations (edge computing); the intermediary layer represents intermediate devices and operations (gateways, fog computing); and the upper layer, called the enterprise layer, represents the cloud and operations such as long-term global storage. The proposed architecture ensures that the data are collected and analyzed in the most efficient and logical place between the source and the cloud, balancing the load and pushing the computation and intelligence to the appropriate layer. It is also necessary to allocate the microservice containers with call dependencies to the same physical host, so that the cross-server calls of the microservice container are as few as possible, to reduce the response time of the service. For the container scheduling problem under the microservice architecture, a container scheduling strategy based on an improved particle swarm algorithm that effectively reduces network calls to fast physical hosts in container-based microservice clusters is proposed, considering the invocation relationship conditions between microservices.

3. Analysis of the Distributed Allocation Algorithm for Collaborative Resource Allocation in the Cloud-Edge-Vehicle Side of Telematics

3.1. Collaborative Resource Design for Cloud-Edge-Vehicle Telematics. The user terminal in Telematics is a vehicle, and

because vehicles travel fast on the road with many vehicles and complex road conditions, smart driving vehicles under Telematics have high quality of service (QoS) requirements for various applications [22]. For this reason, when performing microservice deployment under Telematics, various aspects are considered including the dynamic characteristics of the vehicle (e.g., whether it is moving or not), the type of big data problems (e.g., speed, accuracy), and commutationally complex data analysis. The deployment architecture of microservices in the Telematics system is derived from the previous section on Telematics system architecture: cloud-side-end microservice layered deployment architecture, as shown in Figure 1.

This architecture consists of three parts: the vehicle and road test terminal, the edge cloud platform, and the central cloud. Unlike traditional cloud computing centers, the edge cloud layer deploys the cloud infrastructure near the service road section, which is not as powerful as traditional large cloud data centers but is closer to the specific service area and can effectively improve the quality of service (QoS) of Telematics applications. Vehicle and road information sensing through sensing technology, the On-Board Unit (OBU) enables vehicle-to-vehicle (V2V), vehicle-to-road, and vehicle-to-cloud communications. Microservices with functions such as onboard data fusion calculations, location positioning, road condition sensing, periodic or event data sending and receiving, and supporting autonomous driving fusion decisions are therefore deployed to the onboard and roadside terminals. The number of users accessing the Internet of Vehicles service will suddenly increase, resulting in the overload of some specific microservice resources. When the QoS of the application is reduced, the specific microservice container instance will be dynamically added.

The edge cloud platform of Telematics is a data processing center, through its strong computing capacity, it can realize the processing of massive real-time data of Telematics; it is an application software deployment platform, providing traffic-based cloud services to multiple users, to realize the interoperability of resources of different traffic systems; it is a resource management platform, through virtualization technology, realizing the unified management and elastic expansion of computing resources, thus increasing system stability, reducing costs, and saving energy consumption.

At the vehicle end, the vehicle unit and sensors are mainly used to collect the vehicle and environmental information; at the tube end, the wireless communication network is mainly responsible for the return transmission of information collected by the vehicle unit and roadside units, and the distribution of control information; at the edge of the vehicle network, the cloud end, using its strong computing and data processing capabilities, will process the collected data and information and calculate the integrated output for the application services required by the user. Accordingly, it is important to consider having enough CPU processing power available, memory, disk space available, and bandwidth resources to meet the hardware requirements when deploying microservices [16]. In the Telematics Edge Cloud, container-based microservice resource scheduling means that according to the resource requirements of the microservice, the cloud computing center allocates the corresponding container resources to it, and then the scheduling system deploys it to the physical machine to run. The operation logic of microservice scheduling is divided into four steps: (1) Grab the task in the task executor by annotation and report it to the task registration center. (2) The task orchestration center obtains data from the task registration center to schedule and save it into persistent storage. (3) The task scheduling center obtains scheduling information from persistent storage. (4) The task scheduling center accesses the task executor according to the scheduling logic. It has high precision and robustness, and is more suitable for solving problems such as missing data. The essence is that the scheduling system schedules the set of containers for deploying microservices to run on the set of physical machines according to the scheduling policy. The containers can be configured together with the resources required by the microservice programs, making full use of the hardware resources of the Telematics Edge Cloud platform, and making the physical machine clusters in the Edge Cloud as load balanced as possible while meeting the resource requirements of the microservices.

In addition, microservices are generally responsible for a single service function. When providing services to Telematics users, multiple microservices are often required to work together to meet user requirements, so there is a dependency relationship between microservices to invoke and be invoked. Each microservice is deployed into a container, and there is a one-to-one correspondence between the microservice and the container, thus creating a call dependency between each container. Therefore, in the scheduling and resource dispatching of containers, it is not only necessary to consider maximizing server resource utilization and load balancing between servers but also to allocate microservice containers with invocation dependencies to the same physical host as far as possible, so that the cross-server invocations of microservice containers are as few as possible to reduce the response time of the service, as shown in Figure 2.

The initial deployment of microservices means that the application estimates the number of resources it needs based on the actual number of users it serves daily while ensuring that there is a certain number of resources left over; then it applies for resources to the edge cloud based on its estimated resource characteristics; finally, the Telematics Edge Cloud platform schedules and deploys it to a designated physical host to run according to the microservice resource scheduling policy.

The initial deployment of microservices means that the application estimates the number of resources it needs based on the actual number of users it serves daily while ensuring that there is a certain number of resources left over; then it applies for resources to the edge cloud based on its estimated resource characteristics; finally, the Telematics Edge Cloud platform schedules and deploys it to a designated physical host to run according to the microservice resource scheduling policy. The transmission time of each batch is

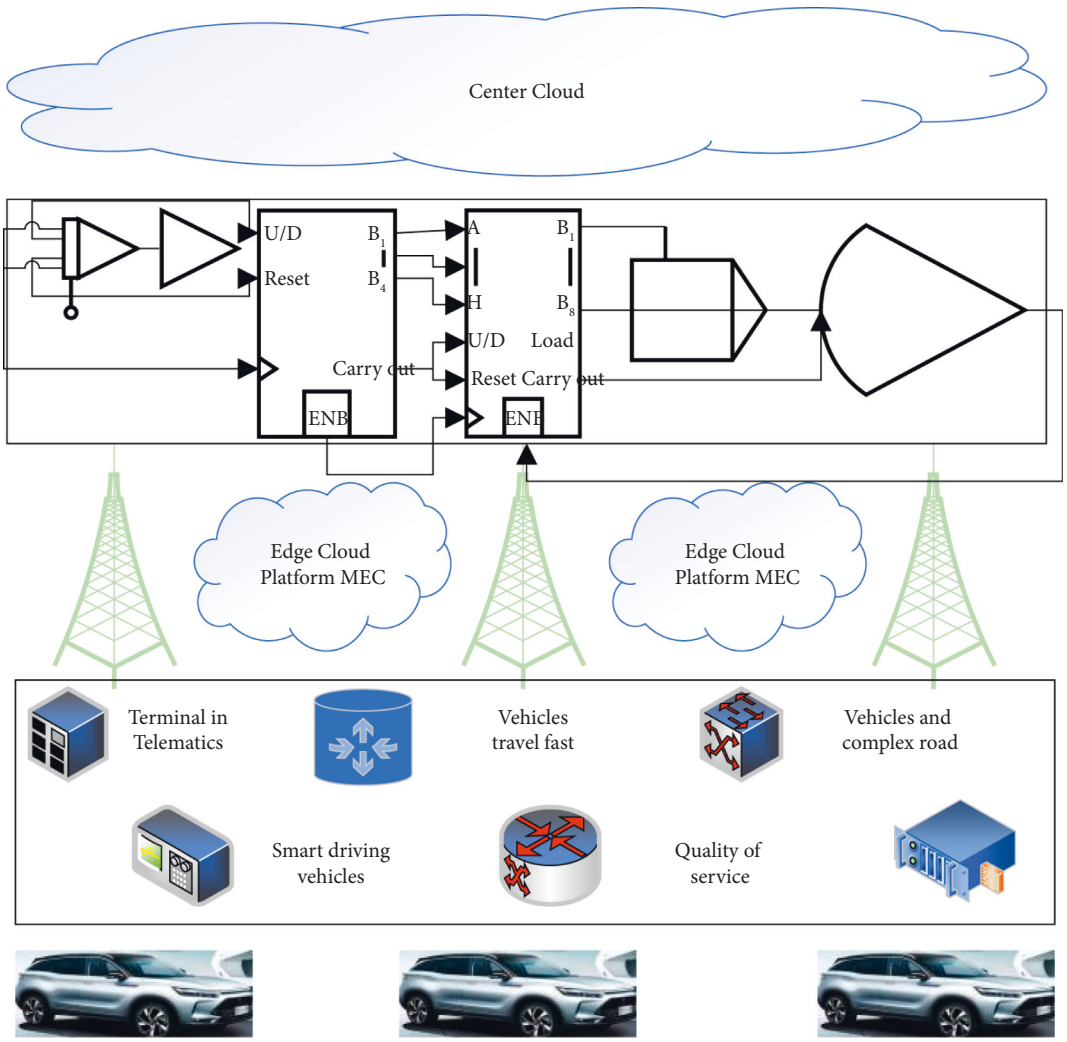


Figure 1: The layered deployment architecture of Telematics microservices.

minimized and the cost function is minimized. At the same time, the above algorithm also considers the upper limit of the bandwidth of each task, which can maximize the use of bandwidth resources. The initial deployment of microservices can ensure the number of resources required for their daily operation and meet their QoS requirements. The dynamic scaling of microservices refers to the dynamic addition of specific microservice container instances when unexpected conditions are encountered, such as a sudden increase in the number of vehicles in the service area of the Telematics Edge Cloud Platform, which results in the overloading of some specific microservice resources and a reduction in the QoS of the application, and then the Edge Cloud Platform deploys the added microservice instances to the specified physical hosts to run following the microservice resource scheduling policy to meet their resource requirements.

Whether it is the initial deployment of microservices or dynamic expansion, the edge cloud microservice resource scheduling policy is required to reasonably schedule microservice containers to deploy and run on the specified physical hosts. In the following, the microservice resource scheduling problem on the Telematics edge cloud platform is modeled according to its characteristics [23]. In the microservice resource scheduling problem on the edge cloud, it is crucial to make the most efficient use of resources in the resource scheduling process due to the limited computing resources compared to traditional cloud computing centers and the high user requirements for latency. In addition, as there are dependencies between microservices, the dependency between containers is also an important factor in the resource scheduling process, thus ensuring the responsiveness of the edge cloud to tasks, improving the utilization of cloud resources, and reducing the energy consumption of the edge cloud center.

3.2. Distributed Allocation Algorithm Design. In the system proposed in this paper, each buyer has a computationally intensive service to perform, but due to its computing resources and capacity constraints needs to migrate part of the service to a suitable service vehicle in the vicinity, and pay the final chosen service vehicle a certain amount of money. Each buyer is represented by a 7-tuple as follows:

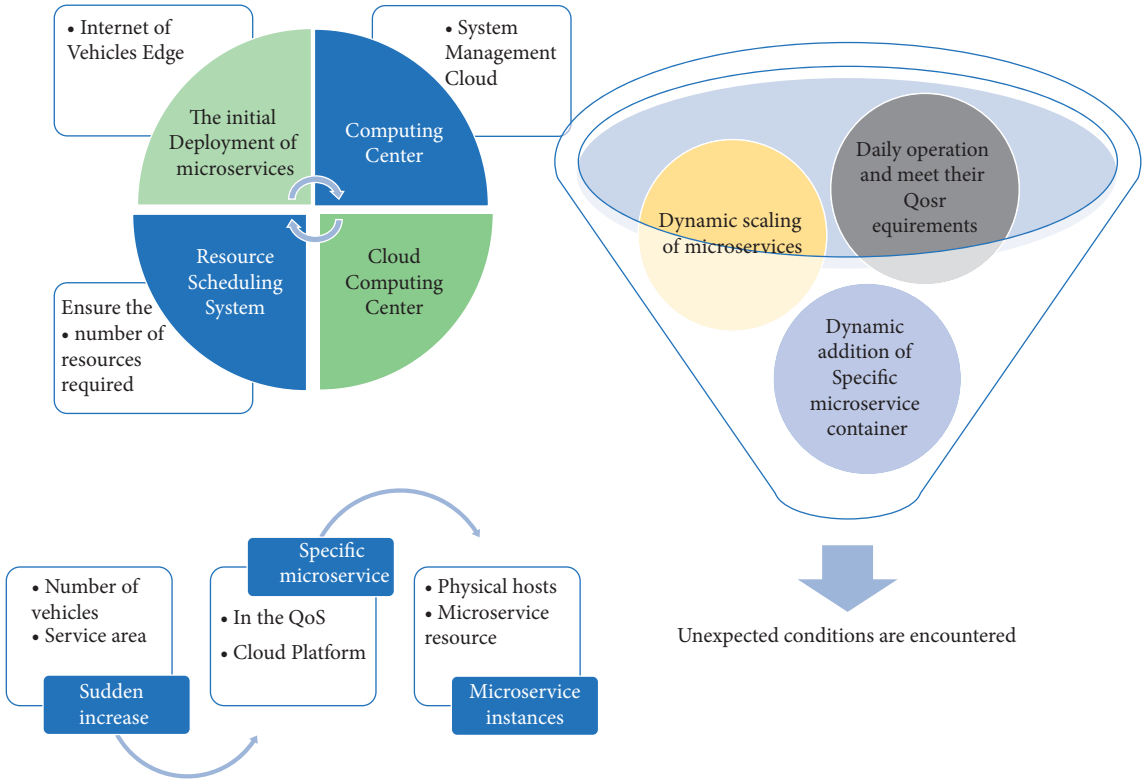


Figure 2: Schematic of cloud microservice resource scheduling in the telematics edge.

$$SR = \{\Gamma(t), f^T, v, \theta, \lambda^2, R\}, \tag{1}$$

where $\Gamma(t) = \{x, y\}$ denotes the geographic coordinates of the buyer at the time t, f^T is the buyer's local computing power (CPU cycle/s), v and θ are the buyer's speed (km/h) and direction of travel, respectively. d and $\lambda \in [0, 1]$ denote the data volume (bits) of the buyer's computationally intensive business, i.e., and the business migration rate, respectively. The data volume size (bit) of the migrated part of the business; R is the communication radius of the vehicle. Therefore, when scheduling resources for microservice containers, one of the optimization goals is to occupy the least number of physical hosts on the premise of meeting their needs, so that the resources of physical hosts can be effectively utilized.

The vehicle cloud consists of a set of computing services providing vehicles within the buyer's one-hop V2V communication range, where each member has more computing power and free resources compared to the buyer, represented by the following 7-tuple:

$$SP = \{\Gamma(t), f^S, v, \theta, \lambda^2, R^2\}, \tag{2}$$

$$P_{tot} = \sum_{k=1}^{K} \sum_{m=1}^{M} \sum_{l=1}^{L} \left(\xi P_{k,m,l}^2 - P_c\right). \tag{3}$$

The interference limit in this section considers the interference between the RSU sender pair and the receiver side of the vehicle node V, and the interference between the sender side of the K -relay forwarding and the receiver side of the vehicle node V. (3) is the interference between the K sender and the receiver V in different regions, where k, m, and$G$ denote the channel gain between the k-th relay and the m-th node on the l-th subcarrier.

$$I_{SBS} = \sum_{m=1}^{M} \sum_{l=1}^{L} P'_{k,m,l} V_{|l-m|} G^2_{k,m,l}. \tag{4}$$

In the actual process of vehicular network traffic flow data acquisition, it is often accompanied by loss communication such as sensor failure or transmission distortion, which inevitably results in the occurrence of missing, lost, or abnormal data, and may even lead to a high percentage of data loss, resulting in unreliable transmitted data [18]. Previous studies have shown that the higher-order tensor can tap higher-level data correlation, make full use of data dimensional information, improve the accuracy of data recovery, have higher accuracy and robustness, and is more suitable for solving problems such as missing data.

An analytical model of mobile IoT slices was established, abstracting different slices as different layers in a multilayer graph. The RSU can timely broadcast the vehicle density information and the priority information of the road condition warning message to the roadside cluster head vehicles in a timely manner. The aim is to solve the problems of how to deploy slices efficiently and flexibly, dynamically, and controllably allocate resources and optimize performance within and between slices according to the needs of different applications, and how to achieve highly reliable and low

latency edge computing slices in the application scenario where IoT terminals are constantly on the move, to maximize resource utilization and optimize the performance of mobile IoT slicing services, as shown in Figure 3.

In the subsequent study, we found that using only the trained neural network to predict the test set could not achieve better results, probably because the network decision space was too large and the training set could not cover all the decided cases. To solve the above problem, a genetic algorithm was used to perform a range search after the neural network decision to obtain a better decision result.

To solve (5), the task bandwidth allocation algorithm is designed because the core objective is to minimize the transmission completion time of each batch, i.e., to minimize the transmission time consumed by the last task to finish transmission within each batch, so with a certain bandwidth of the base station, the bandwidth is first allocated in equal proportion to the data size of the task, and if the bandwidth allocated to user i exceeds its bandwidth limit Bi, then if the bandwidth allocated to user i exceeds its bandwidth limit Bi, then the user's upper bandwidth limit is allocated, and then the remaining tasks are reallocated proportionally according to the above process [19]. In this way, the tasks within a batch can be transferred as simultaneously as possible, minimizing the transfer time of each batch and thus the cost function, while the algorithm also considers the bandwidth limit of each task, allowing for maximum utilization of bandwidth resources. Once the task transfer is complete, computational processing can begin, using a simple single-core processor to process incoming tasks serially, following the first-come, first-served principle, until all tasks are finally processed.

$$W_{ij} = WR_{ij}^2 RV_{ij}^2. \tag{5}$$

Since the adjacency matrix of each node of the graph random wandering model is a Markov matrix, the wandering probability of each node is a specific value in the adjacency matrix. Assuming that there are M components to be deployed, the matrix will reach a new state after M steps of wandering. The idea of routing switching is integrated, and the probability-based routing decision is made with the goal of minimizing the delay. Then, according to the final state of the matrix, the specific deployment probability of each node can be obtained, and the system needs to allocate more available resources to the node with the highest probability.

In the process of slicing resource management, this paper proposes to achieve this through a multimodel collaborative learning scheme, i.e., using Generative Adversarial Networks (GAN) and Deep Reinforcement Learning (DRL) to address resource demand prediction within slices and dynamic resource allocation between multiple slices, respectively, i.e., through multimodel collaboration to perform dynamic resource allocation for different slices to achieve efficient use of limited resources while providing slicing services for more applications.

$$\begin{aligned}\max \min V(d,G) = {} & E_{x\sim P_{\text{data}(x)}}[\log D(x) \\ & - E_{Z\sim P_Z(Z)}[1+\log D(g)]].\end{aligned} \tag{6}$$

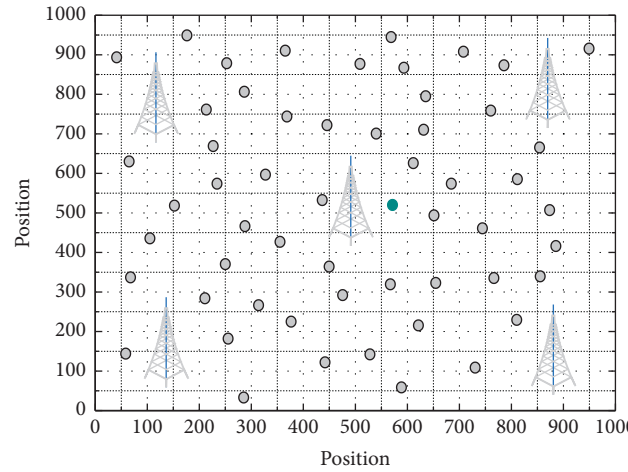


Figure 3: Network environment.

When a user sends a service request, a chain of invocations will be formed in the server to handle the user's demand. All microservices in the chain of invocations collaborate to meet the user's needs. Therefore, it is necessary to invoke the containers deployed in each physical host. The invocation of containers in the same physical host consumes significantly less time than the invocation across physical hosts. Therefore, in the process of container deployment and resource scheduling, the number of container calls across physical hosts should be minimized, so that the time for container calls across physical hosts can be reduced and the network resources wasted.

Therefore, when scheduling resources for microservice containers, one of the optimization goals is to minimize the number of physical hosts occupied while satisfying their requirements so that the resources of the physical hosts can be used effectively. In this paper, we use the defined parameter Z to denote the total number of physical hosts occupied by the deployment of microservices-equipped containers, while the combined resource utilization of the physical host population activated for the deployment of microservices in the entire edge cloud is expressed by the parameter U, as shown in the defined formula in (7).

$$Z = \sum_{i=1}^{n} P_i, \tag{7}$$

$$U = \left(\frac{\sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{l=1}^{s} p_{ij}^2 \times k_{ij}^2 \times r_{j,l}}{\sum_{m=1}^{M} \sum_{l=1}^{L} P'_{k,m,l} V_{|l-m|} G_{k,m,l}^2} \right). \tag{8}$$

The value of takes the range (0, 1), and the fewer physical hosts occupied by the same number of microservices, the higher the resource utilization of the physical machine cluster. The RSU, a network node located in the middle of a roadside or intersection, can communicate with the traffic management center and roadside vehicles to obtain timely information on the global traffic situation [24]. The centralized control mechanism allows the RSU to play an important regulatory role in congestion control and wireless channel control of in-vehicle communication. In the proposed algorithm, the RSU

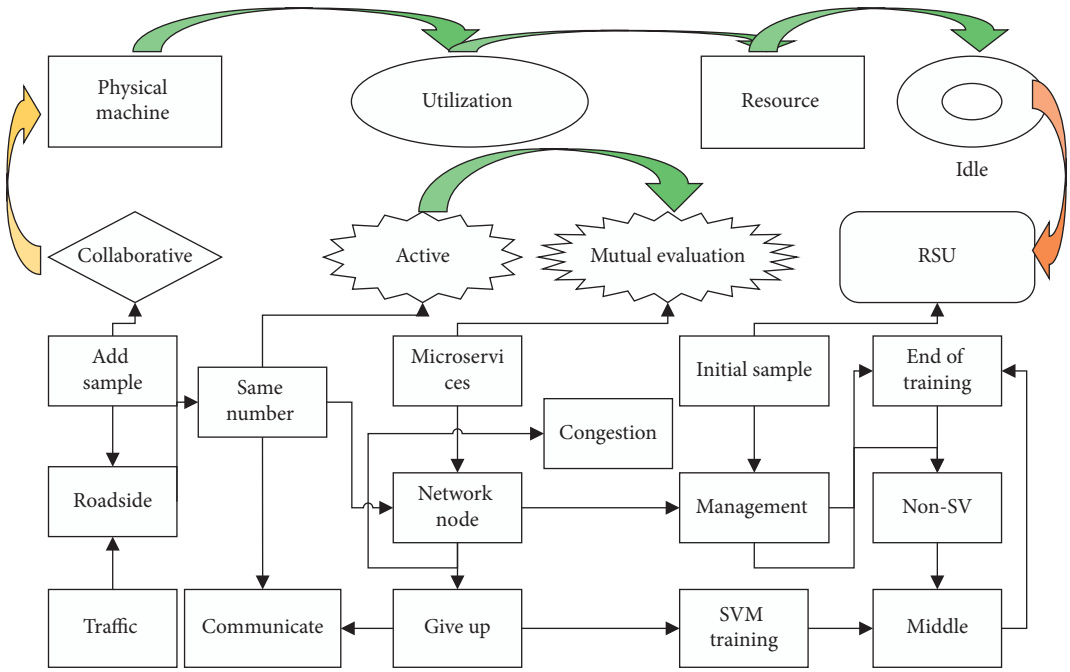


Figure 4: Channel state division diagram.

can broadcast accurate information on the density of vehicles and the priority of roadside alert messages to the roadside cluster headers on time by reusing the same communication resources. By combining the vehicle density information and the priority of the alert message, the vehicle node can calculate the information interference factor.

According to the CBR value of the cluster head vehicle, the power adjustment algorithm classifies the channel into three states, when the CBR is less than 0.5, the channel is in an idle state when the CBR is greater than 0.5 and less than 0.8, the channel is an inactive state, and when the CBR is greater than 0.8, the channel is in a congested state. The channel state transfer diagram is shown in Figure 4.

Resource management for containers is even more complicated. Commonly used container orchestration tools such as Kubernetes and Docker SwarmKit only provide some simple resource scheduling strategies and cannot fully utilize the performance of physical machines. Therefore, it is very important to design a reasonable resource scheduling strategy for microservice containers. When the channel is congested, by appropriately reducing the vehicle signal power, the conflict of vehicles competing for channel resources can be reduced, thus improving the efficiency of message delivery. When the channel is idle, the vehicle signal power is increased to improve the delivery success rate and the coverage of the alert message. When the channel state is active, the power of the vehicles in the cluster is not adjusted.

4. Analysis of Results

4.1. Performance Results of the Cloud-Edge-Vehicle-Side IoT Collaborative Resource Platform. When an IoT application starts running, the slicing system will perform business identification and related resource requirement analysis and use this as the basis for scalable deployment of functional modules. During the deployment process, depending on the type of service and demand, certain functional modules may be deployed at the core, at the edge, or even at the bearer network between the terminal and the core or other locations, enabling the flexible and scalable deployment of each functional module. These virtualized platforms for deploying different functional modules have different available resources, and the links connecting them have different transmission performances. Therefore, it is a challenge for the initialization process of mobile IoT slices to make more balanced and efficient use of limited resources when deploying specific functional modules at various locations and to further improve the quality of service and user experience of applications.

The latency in this experiment refers to the entire time that the data are sent from the terminal to the distributed computing platform until the calculation results are returned to the terminal. The goal of the multiedge collaborative mobile IoT slicing is to reduce the latency while maximizing recognition accuracy, and the experimental results confirm that the architecture achieves this goal. In contrast, edge computing relies on numerous edge devices at the edge of the network to process massive data, reduce the occupation of network resources, enhance real-time communication capabilities, and complete data processing and service execution with extremely low latency. There are three main reasons for this: firstly, the constant switching of communication nodes as the terminal moves; secondly, the multiedge collaborative slicing model that uses distributed deep learning from multiple edge nodes; and thirdly, the efficient task allocation and optimal transmission path selection method implemented through graph neural networks, as shown in Figure 5.

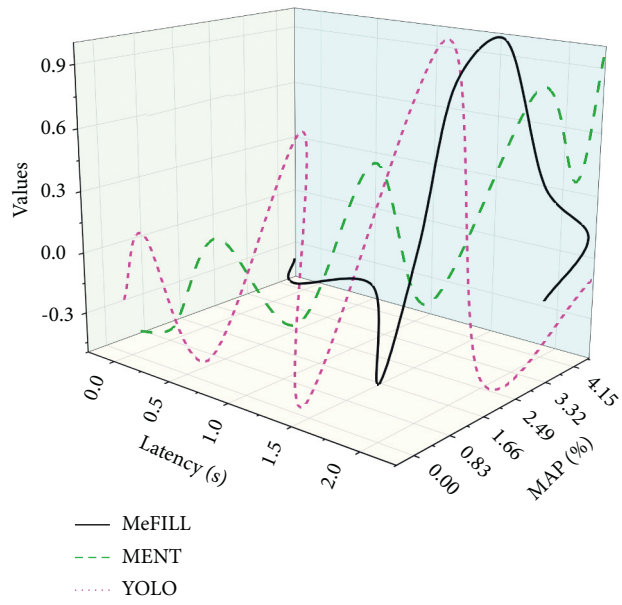


Figure 5: Comparison between time delay and accuracy.

In Figure 5, the comparison between latency and average correctness for the three architectures mentioned above is depicted. As can be seen from the figure, if higher accuracy is required, a larger latency is used. In the case of object recognition experiments, higher accuracy is used when the latency requirement is first guaranteed. For example, the requirement that the time delay must be less than 1.5 seconds enables the multiedge collaborative mobile IoT slice to achieve 85% accuracy, compared to 72% and 65% for the other two comparison models, respectively, thus showing that the multiedge collaborative mobile IoT slice has higher accuracy for the same time delay requirement, and similarly, the multiedge collaborative mobile IoT slice has lower latency for the same accuracy requirement.

As shown in Figure 6, both the delay and energy consumption in dynamic communication node mode is essentially constant, but both the delay and energy consumption in fixed communication node mode gradually increase as the terminal moves away from the communication node. This is because when the terminal is further away from the communication node, the terminal needs to increase its transmitting power to transmit data to a greater distance so that the communication node can receive it properly, which inevitably results in higher energy consumption. Similarly, the further away the terminal is from the communication node, the greater the data transmission delay. Therefore, this paper uses dynamic communication nodes to solve the problems of latency and energy consumption faced by mobile terminals. In some application scenarios of the Internet of Vehicles, such as autonomous driving, the latency requirement even needs to be lower than 10 ms. This makes the research on the transmission strategy of IoV security services more important.

In this network environment, using the CB-SIC solution (combined CB and SIC technology), all task nodes gain a total of 232-time slices to transmit data to the base station if they transmit with incremental CB power. When transmitting with fixed CB transmission power, a total of 219-time slices of data are transmitted to the base station. With the CB-only scheme (using only CB technology), only one task node in a time slice can transmit data to the base station. The interference avoidance scheme is like the CB-only scheme in this respect. The SIC-only scheme (using only SIC technology) has better data throughput than the CB-only and interference avoidance schemes but is still not comparable to the CB-SIC scheme.

By adjusting the number of tasks and idle nodes, 24 different network environments were obtained. In these network environments, the experimental results of the CB-SIC scheme, the CB-only scheme, the SIC-only scheme, and the interference avoidance scheme are compared. On-Board Units (OBUs) enable vehicle-to-vehicle (V2V), vehicle-to-road, and vehicle-to-cloud communications. The amount of data transmitted by each node in all-time slices is then calculated based on the transmission rate of the nodes, and the average data throughput in each network environment are shown in Figure 7.

As can be seen from the figure, the throughput obtained by the CB-SIC scheme is significantly improved compared to the other three schemes. Regardless of the number of task nodes and idle nodes, the CB-SIC solution consistently achieves more than twice the data throughput of the CB-only solution. The throughput can be further increased by increasing the CB transmission power, and the CB and SIC technologies increase the data throughput of the entire wireless network.

The task nodes need to transmit data directly to the base station by using CB technology. The base station, in turn, uses SIC technology to receive multiple signals simultaneously. Therefore, the container-based microservice architecture is adopted, and the application software functions are disassembled into microservices with smaller granularity for deployment, to achieve high reliability, flexibility, and high performance under limited resource conditions. In addition, the application scenarios of the edge cloud of the Internet of Vehicles are fixed. In this chapter, the structure of the network system is first given and a mathematical model is developed by analyzing it. The CB technique also generates a power gain, which further increases the number of signals that can be received and decoded at the same time by the base station using the SIC technique. In these ways, the amount of data transmitted to the base station in a fixed period is maximized. Simulation results show that the CB-SIC scheme, which combines CB and SIC technology, can significantly increase throughput compared to the CB-only scheme, the SIC-only scheme, and the interference avoidance scheme.

4.2. Performance of the Cooperative Resource Distributed Allocation Algorithm. In both the CB-SIC PRO and CB-SIC FIFO schemes, a suitability threshold is used to select an edge server for each task. Therefore, changes in the experimental results were observed by increasing the fitness thresholds in both schemes. To further demonstrate the

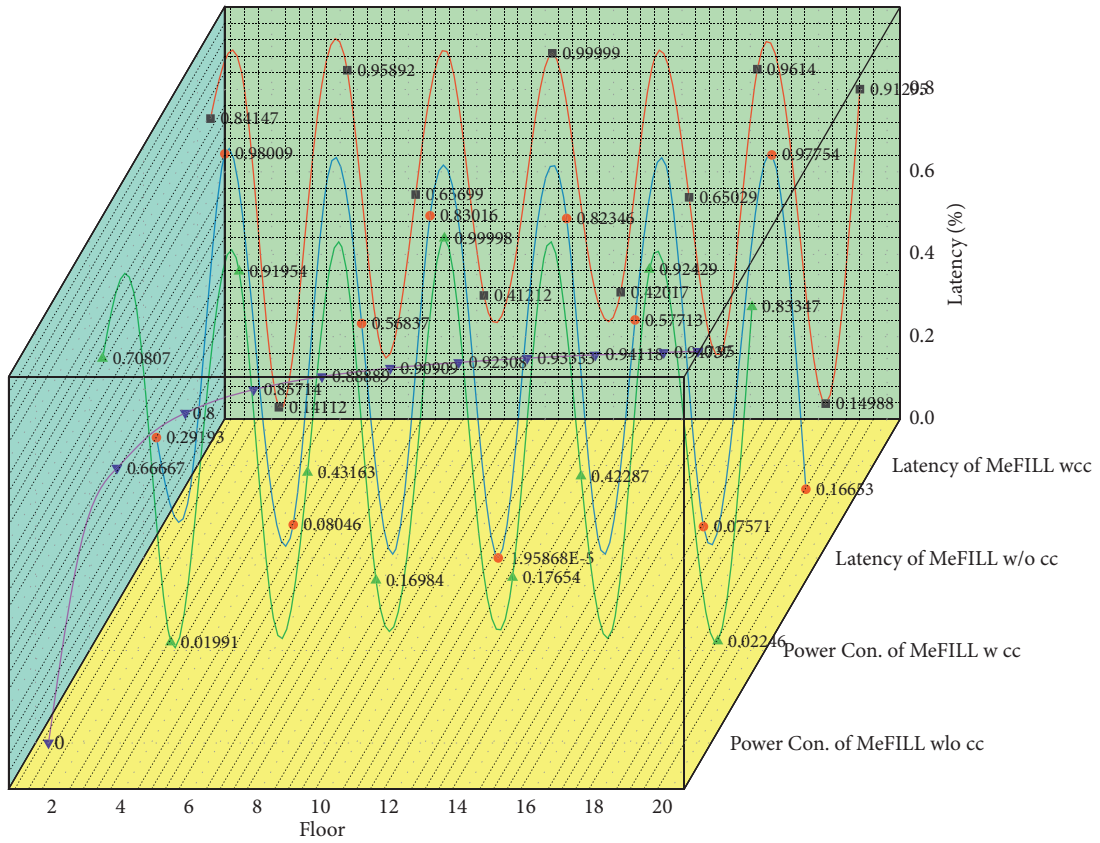


Figure 6: Comparison of terminal latency and energy consumption.

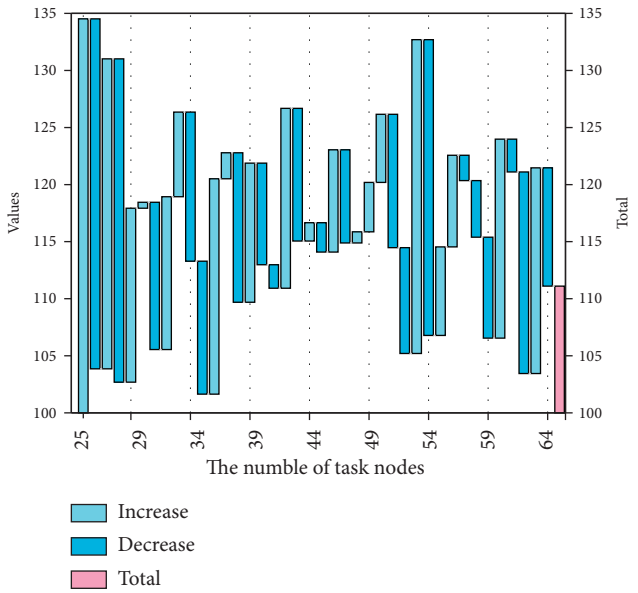


Figure 7: Schematic comparison of data throughput in different network environments.

effect of threshold α on task completion rate and task completion latency in different network environments, the number of randomly generated tasks in the network was set to 40, 50, 60, and 70.

From the function in the optimization objective, it can be analyzed that the closer the user is to the base station, the lower the bandwidth cost to achieve the same code rate. Therefore, the algorithm adopts the shortest distance access principle, which can greatly reduce the bandwidth consumption of the base station. The step-by-step training of the DNN network relies on its self-learning ability to finally obtain a better training model, and then make decisions on the incoming tasks, and output the best access strategy selected, as shown in Table 1.

It can be seen from Table 1 that the MBRA algorithm has nearly 50% of the data errors within 20%, followed by nearly 98% of the data errors within 40%, which is significantly higher than the other three heuristic algorithms, and the overall effect is better. It can also be seen from Table 1 that the average accuracy of the MBRA algorithm in the entire training process is 84.13%, the total time consumption for processing 20,000 pieces of data is 1471 s, and the time consumption for a single task decision is only 74 ms. The decision-making time of the heuristic algorithm is short, but the overall error is relatively large, so it cannot achieve a good decision-making effect.

The main purpose of the cooperative network is to make use of the high mobility and processing ability of the intelligent terminals carried by users, to make the intelligent terminals act as relays randomly, and to establish the communication between the sensor nodes and the infrastructure in a cooperative way, so that the data can be

Table 1: Time complexity and solution average error rate of MBRA and comparison algorithms.

Algorithm	Solving time (s/data)	Average error rate (%)
Base station access decision algorithm MBRA	0.074	15.87
Random access RAS	0.00452	54.19
Access to NDAS at the closest distance	0.0029	29.4
Equal distribution of access to EDS	0.00481	53.97
Brute-force algorithm VA	5.262	0

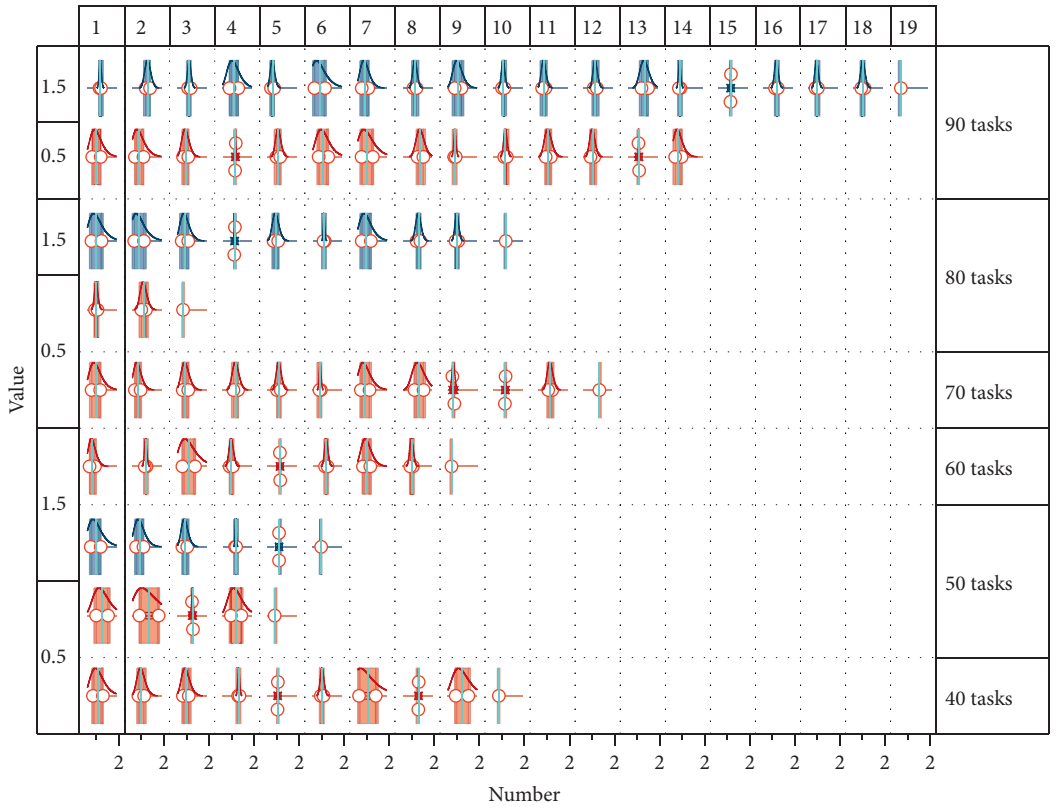


Figure 8: Diagram showing the effect of threshold α on the task completion rate of the CB-SIC PRO solution.

efficiently aggregated to the core network. Therefore, in the vehicle networking scenario, these ICVs need to consume more energy to relay the data sent by other sensor nodes. The algorithm in this paper considers the edge-end collaboration mechanism, and the ICVs closer to the base station assist the ICVs farther away from the base station to transmit data by means of data relay, thus resulting in more energy consumption.

Under the edge-end coordination mechanism of the algorithm in this paper, some ICVs that are far away from the base station do not communicate directly with the base station but use other ICVs to transmit data through multipath and multihop routing. Therefore, this algorithm can also reduce the number of links for vehicle-to-base station direct communication, thereby saving communication bandwidth resources.

Figure 8 illustrates the effect of progressively increasing suitability thresholds on the task completion rates. When the number of tasks is 40, the task completion rate increases slowly at first, and then gradually starts to decrease once it reaches 100%, and finally remains constant. When $\alpha = 0$, the task completion rate is 98.2%. When $\alpha = 14$, the task completion rate reaches 100%. At $\alpha = 22$, this starts to decrease and eventually stays at 98.2%. The task completion rate of the CB-SIC PRO solution remains at a high level for a task count of 40, and it can be assumed that the computational load from the task count at this point is not high for the edge servers in the network. Therefore, the impact is not significant. As the number of tasks increases, the task completion rate varies considerably. At a task count of 60, the task completion rate ranged from a low of 79.1% to a high of 94.8%. There is a 15.7% difference between the minimum and maximum.

The experimental results in Figure 9 show that the degree of load imbalance in the data center increases as the size of the microservice containers to be deployed increases. Then, the scheduling system deploys it to the physical machine to run. Its essence is that the scheduling system schedules the container set for deploying microservices to run on the physical machine set according to the scheduling policy, and

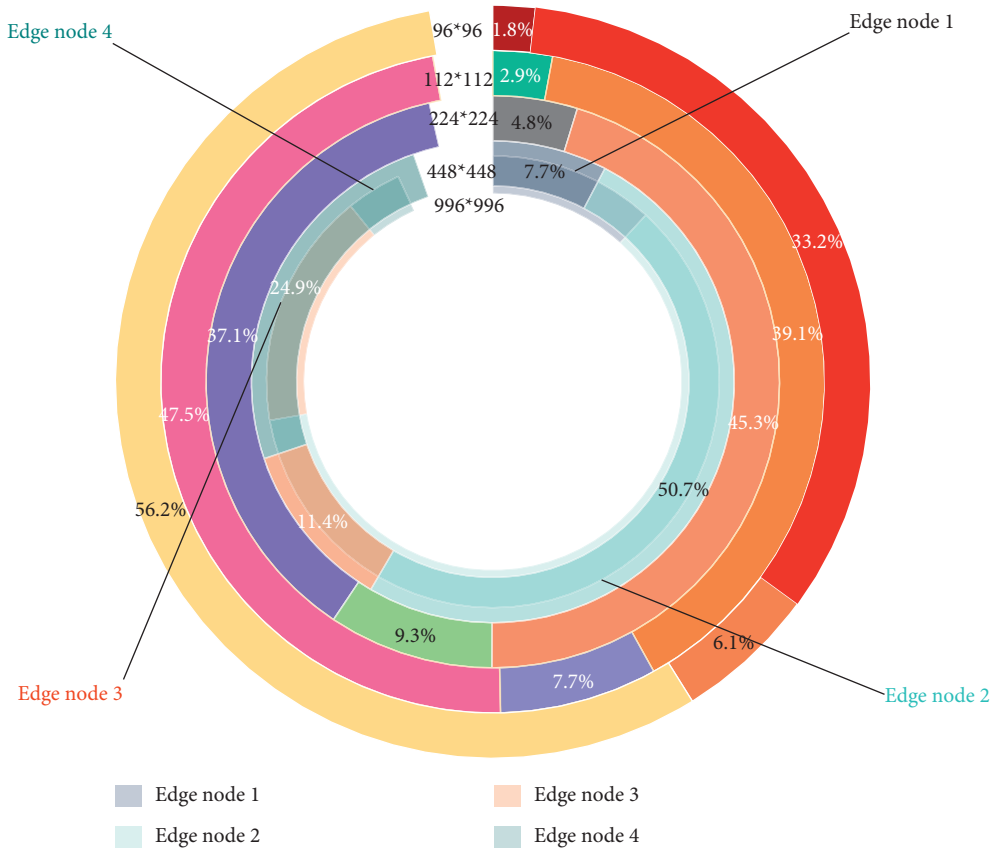


Figure 9: VGG16 model and the effect of feature map size on edge clustering.

the container can be configured with the resources required by the microservice program. For the same experimental conditions, the standard deviation of the data center load imbalance of the MFGA algorithm is better than the other three algorithms and stays in the lower range, achieving good load balancing.

From the experimental results in the previous section, communication time consumption has a more obvious impact on the overall time consumption, and the amount of data determines the communication time consumption. Here, we investigate the impact of the size of the input feature map of the deep network model on the computation time and acceleration ratio of the edge clusters, with five VM VMs-1 used as edge devices, one as the edge gateway and the remaining four as edge nodes, the number of data frames is 1, and the network bandwidth is set to 1000 Mbps.

Under the mechanism that the ICV communicates directly with the base station, the total energy consumption of video compression and communication varies greatly for each ICV, that is, the total energy consumption of video compression and communication of ICVs closer to the base station is higher than that of the ICV. This is because the ICVs that are closer to the base station have better signal-to-noise ratios of transmission channels and can support higher data transmission rates with less communication energy consumption.

One limitation of the value iteration method is that it requires a finite and minimal number of states, which makes solving the system of equations almost impossible when, as in this paper, the state space is growing exponentially. The policy iteration algorithm includes a process of policy estimation, which requires scanning all states several times, a complexity that seriously affects the efficiency of the policy iteration algorithm. Both value iteration and policy iteration require a known state transfer probability to compute the optimal policy, which is difficult to implement in real-world usage scenarios.

5. Conclusion

The Internet of Vehicles, as an application of the Internet of Things in the field of intelligent transportation, is an important development direction for future driving technology. In an IoT environment, it is impossible to realize autonomous driving without the collaborative cooperation of cloud, edge, and end. As a data processing center and application software deployment platform close to the end service terminal in the Telematics system, the edge cloud platform will carry most Telematics applications. As it is deployed at the edge of the network, it can greatly shorten the response time of Telematics applications, reduce bandwidth costs, and improve service quality. The edge cloud platform has relatively limited resources compared to traditional cloud data centers, and its resource scheduling algorithm directly affects the performance of Telematics applications that are not on it. Therefore, it is important how

to make limited use of edge cloud hardware resources in the Telematics edge cloud platform and ensure the reliability and high performance of Telematics applications. In this strategy, firstly, the clustering algorithm is used to cluster the vehicles on the road, secondly, it is introduced how to evaluate the channel busy status of the clustered vehicles using the RSU of the roadside node, then the vehicles are classified into different channels states according to different channel busy degrees, and the corresponding power adjustment strategies are carried out by the vehicles in different channel states to improve the communication performance of the alarm messages transmitted on the road. Finally, the performance of the proposed algorithm is effectively verified in simulation experiments.

Data Availability

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

Consent

Informed consent was obtained from all individual participants included in the study references.

Consent

The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgments

This work was supported by 2019 Cross Science Research Project of Nanyang Institute of Technology, Grant No. 201913502, Research on Intelligent Mining and Recommendation of Zhang Zhongjing Prescription Based on Deep Neural Network, and Henan Science and Technology Plan Project, Grant No. 222102210134, Research on Key Technologies of Cloud Security Desktop Based on Kunpeng Architecture.

References

[1] Y. Ding, M. Jin, S. Li, and D. Feng, "Smart logistics based on the internet of things technology: an overview," *International Journal of Logistics Research and Applications*, vol. 24, no. 4, pp. 323–345, 2021.

[2] K. Kardaras, G. I. Lambrou, and D. Koutsouris, "Telematics healthcare through digital terrestrial television networks: applications and perspectives," *International Journal of Sensors, Wireless Communications & Control*, vol. 11, no. 5, pp. 560–576, 2021.

[3] H. Tavolinejad, M. R. Malekpour, N. Rezaei et al., "Evaluation of the effect of fixed speed cameras on speeding behavior among Iranian taxi drivers through telematics monitoring," *Traffic Injury Prevention*, vol. 22, no. 7, pp. 559–563, 2021.

[4] D. Ivanov, C. S. Tang, A. Dolgui, D. Battini, and A. Das, "Researchers' perspectives on Industry 4.0: multi-disciplinary analysis and opportunities for operations management," *International Journal of Production Research*, vol. 59, no. 7, pp. 2055–2078, 2021.

[5] O. R. Sánchez, C. A. Collazos Ordonez, M. A. Redondo, and I. Ibert Bittencourt Santana Pinto, "Homogeneous group formation in collaborative learning scenarios: an approach based on personality traits and genetic algorithms," *IEEE Transactions on Learning Technologies*, vol. 14, no. 4, pp. 486–499, 2021.

[6] S. Jung, J. Kim, M. Levorato, C. Cordeiro, and J. H. Kim, "Infrastructure-assisted on-driving experience sharing for millimeter-wave connected vehicles," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 8, pp. 7307–7321, 2021.

[7] K. Yue, Y. Zhang, Y. Chen et al., "A survey of decentralizing applications via blockchain: the 5g and beyond perspective," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 4, pp. 2191–2217, 2021.

[8] C. H. Lai and J. S. Fu, "Exploring the linkage between offline collaboration networks and online representational network diversity on social media," *Communication Monographs*, vol. 88, no. 1, pp. 88–110, 2021.

[9] Q. Yu, M. Wang, H. Zhou, J. Ni, J. Chen, and S. Cespedes, "Guest editorial special issue on cybertwin-driven 6G: architectures, methods, and applications," *IEEE Internet of Things Journal*, vol. 8, no. 22, pp. 16191–16194, 2021.

[10] G. Fodor, J. Vinogradova, P. Hammarberg et al., "5G new radio for automotive, rail, and air transport," *IEEE Communications Magazine*, vol. 59, no. 7, pp. 22–28, 2021.

[11] Z. H. Ali and H. A. Ali, "Towards sustainable smart IoT applications architectural elements and design: opportunities, challenges, and open directions," *The Journal of Supercomputing*, vol. 77, no. 6, pp. 5668–5725, 2021.

[12] C. Na, D. Lee, J. Hwang, and C. Lee, "Strategic groups emerged by selecting R&D collaboration partners and firms' efficiency," *Asian Journal of Technology Innovation*, vol. 29, no. 1, pp. 109–133, 2021.

[13] H. G. Abreha, C. J. Bernardos, A. D. L. Oliva, L. Cominardi, and A. Azcorra, "Monitoring in fog computing: state-of-the-art and research challenges," *International Journal of Ad Hoc and Ubiquitous Computing*, vol. 36, no. 2, pp. 114–130, 2021.

[14] K. Yang, C. Hu, Y. Qin, Y. Huang, and X. Tang, "Potential and challenges to improve vehicle energy efficiency via V2X: literature review," *International Journal of Vehicle Performance*, vol. 7, no. 3/4, pp. 244–265, 2021.

[15] C. Sardianos, I. Varlamis, C. Chronis et al., "The emergence of explainability of intelligent systems: delivering explainable and personalized recommendations for energy efficiency," *International Journal of Intelligent Systems*, vol. 36, no. 2, pp. 656–680, 2021.

[16] C. Englund, E. E. Aksoy, F. Alonso-Fernandez, M. D. Cooney, S. Pashami, and B. Astrand, "AI perspectives in Smart Cities and Communities to enable road vehicle automation and smart traffic control," *Smart Cities*, vol. 4, no. 2, pp. 783–802, 2021.

[17] A. Rajesh and S. Shaffath Hussain Shakir, "Investigations on scheduling algorithms in LTE-advanced networks with carrier aggregation," *International Journal of Advanced Intelligence Paradigms*, vol. 18, no. 2, pp. 1–62, 2021.

[18] B. Bhushan, C. Sahoo, P. Sinha, and A. Khamparia, "Unification of Blockchain and Internet of Things (BIoT): requirements, working model, challenges and future directions," *Wireless Networks*, vol. 27, no. 1, pp. 55–90, 2021.

[19] S. Guan, J. Wang, C. Jiang, R. Duan, Y. Ren, and T. Q. S. Quek, "MagicNet: the maritime giant cellular network," *IEEE Communications Magazine*, vol. 59, no. 3, pp. 117–123, 2021.

[20] Y. Zhang, K. C. S. Lee, and D. Adams, "Visualizing research in educational technology leadership using CiteSpace," *Int. Online J. Educ. Leadership*, vol. 5, pp. 61–77, 2021.

[21] J. Moeyersons, S. Kerkhove, T. Wauters, F. De Turck, and B. Volckaert, "Towards cloud-based unobtrusive monitoring in remote multi-vendor environments. Software: practice and Experience," *Software: Practice and Experience*, vol. 52, no. 2, pp. 427–442, 2022.

[22] B. Chan, "Sidecar learning vs LibWizard: a comparison of two split-screen tutorial platforms," *Journal of Web Librarianship*, vol. 15, no. 2, pp. 90–103, 2021.

[23] C. Simon, M. Maliosz, M. Máté, D. Balla, and K. Torma, "Sidecar based resource estimation method for virtualized environments," *INFOCOMMUNICATIONS JOURNAL*, vol. 12, no. 2, pp. 4–11, 2020.

[24] P. Jamshidi, C. Pahl, N. C. Mendonça, J. Lewis, and S. Tilkov, "Microservices: the journey so far and challenges ahead," *IEEE Software*, vol. 35, no. 3, pp. 24–35, 2018.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 8758156, 15 pages
https://doi.org/10.1155/2022/8758156

Research Article

Trust-Based Certificateless Privacy-Preserving Authentication in Internet of Vehicles

Chang Yu [1] **and Kezhong Lu** [2]

[1]*College of Finance and Economics, Anhui Technical College of Industry and Economy, Hefei230000, Anhui, China*
[2]*School of Big Data and Artificial Intelligence, Chizhou University, Chizhou 247100, China*

Correspondence should be addressed to Kezhong Lu; luck@czu.edu.cn

Received 6 March 2022; Revised 16 April 2022; Accepted 21 May 2022; Published 18 August 2022

Academic Editor: Jie Cui

Owing to the security requirements of Internet of vehicles (IOV), it is necessary to design a secure privacy-preserving scheme for communication. Traditional privacy-preserving schemes have two deficiencies. One is the high cost of computation and communication. Another is the inability to prevent the spread of malicious or modified messages. Motivated by those facts, we proposed a trust-based authentication scheme for certificateless privacy-preserving of IOV, based on the advantages of the short key, fast speed, and high security performance of elliptic curve cryptography (ECC). We proposed a method to replace the revocation list by authenticating trust to prevent broadcasting fake and altered massages. Our scheme can encrypt the message sent by the node while adopting a certificateless authentication method to complete the anonymous authentication function, which protects the privacy of the node information and effectively reduces the system storage load. In addition, aggregate signatures can effectively reduce computational and communication overhead. It is proven theoretically that the proposed scheme can satisfy correctness, anonymity, confidentiality of messages, and unforgeability of signatures. Therefore, this scheme is more suitable for the deployment and application of physical IOV.

1. Introduction

Internet of vehicles (IOV) are applications of mobile ad hoc networks (MANETs) and wireless sensor networks in the field of intelligent transportation to implement the communication between intelligent vehicles and increase the safety and efficiency of road traffic. The key features that distinguish IOV from other MANETs are vehicle density, self-organization, multihop, rapid change of network topology, limited network capacity, no power and storage constraints, predictable node mobility patterns owing to fixed roads and lanes, and a large number of nodes in urban traffic [1]. A typical IOV architecture usually includes three components: service center, road side unit (RSU), and a vehicle node that configures the onboard unit (OBU), where the OBU is mounted on the vehicle to provide wireless communication capabilities. The RSU is used to provide a wireless and radio-covered vehicle interface [2]. As networks become more common, there is a growing need for vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications [3], and the communication between V2V and V2I is realized by dedicated short range communication (DSRC) [4] systems. Most importantly, the IOV is a promising technology for providing effective traffic management solutions, navigation-based services, infotainment, and vehicle safety.

Privacy and security issues in IOV have attracted a significant amount of attention. Since IOV supports emergent real-time applications and processes vital message, relevant schemes should meet security requirements such as privacy, confidentiality, integrity, and nonrepudiation to provide secure communication to attackers and malicious nodes [5]. All Kinds of security attacks such as denial of service (DOS), Sybil attack, illusion attack, and wormhole attack will affect the privacy of the vehicle and possibly lead to traffic congestion, misinformation dissemination, positioning and identity leakage, disguise or forgery of data, and intrusion of private information. Therefore, data security

and privacy-preserving issues in the IOV environment have become the focus of attention [2, 6].

A number of asymmetric cryptography-based security authentication schemes have been proposed to prevent such attacks. Anonymous authentication is one of the basic methods used for preserving privacy. The typical anonymous authentication mechanisms in IOV include pseudonyms, random silence, group signatures, ring signatures, blind signatures, and smart cards. In recent years, scholars have proposed a variety of anonymous schemes for IOV security authentication, such as digital signature scheme [7] and group signature scheme [8] based on public key infrastructure (PKI). However, these traditional anonymous authentication schemes have the following disadvantages. (1) The computational and communication costs of message authentication are large. In the case of high traffic density, there will be more delay, and a large number of messages will get lost. (2) Requirement for vehicle to store a large number of certificates and dependence on the revocation list to achieve vehicles revocation. It results in a large storage overhead of the system. Therefore, improving the efficiency of anonymous authentication based on ensuring security is also one of the principal challenges facing IOV [9].

Except for efficiency issues, authentication mechanism also has a major limitation, as it only ensures that the messages are transmitted from a legitimate sender, and does not prevent legitimate senders from maliciously spreading false or modified information to other vehicles. False or altered messages can reduce traffic efficiency and, at worst, threaten people's lives. The question to be considered is how a vehicle decides whether to believe a message sent by a dependable vehicle. In order to prevent the above problems from causing improper behavior of the vehicle, misconduct detection mechanisms [10] and reputation systems [11] have been put forward. Trust vehicles can be distinguished from untrusted vehicles by building trust relationships and detecting malicious behavior in IOV, thereby preventing the vehicle from being misdirected by other malicious vehicles. Therefore, trust is essential to protect IOV. Anonymous authentication trust is becoming a compelling method of preserving privacy in IOV. Nevertheless, there is a lack of research on this topic, especially for the IOV system. Trust management of IOV [12] has been studied and attempted.

In this study, we propose a certificateless anonymous authentication scheme based on the trust of the IOV. In our scheme, the trust value is combined with the traditional encryption scheme for preserving privacy. Only if the vehicle generating the message has a certain trust value, the message is thought to be reliable. The proposed method can not only ensure the effective communication of vehicles in the vehicle network but also make sure the vehicles receive information that is reliable. The basic principle of the scheme is to allow a trusted authority (TA) or authorized parties (AP) to announce the latest aggregate list of integrated node trust (INT) and verify the node trust without certificates. In our proposed scheme, a TA updates the trust value of each vehicle, stores the values in the trust value table using hashing techniques, and then broadcasts the trust value table. Thus, all vehicles can obtain the trust value of the adjacent vehicle by querying the trust value table to strengthen security. Depending upon the location of the trust value in the INT aggregation list, the receiving node can verify the sender's message anonymously and without a certificate, and aggregate signatures can effectively reduce the computational costs and communication overhead. Furthermore, multiple APs may flexibly coordinated to achieve trust authentication while supporting aggregation signature verification. The method can provide fast, anonymous authentication and preserve privacy, and can ensure the reliability of the message of V2V communication.

1.1. Our Contributions. The main contributions of the proposed scheme are summarized:

(i) We propose a scheme to guarantee the security of communication and the reliability of messages in IOV by combining trust with traditional privacy-preserving encryption scheme. We demonstrated that the proposed method was secure, and evaluated the performance by analyzing the proposed scheme.

(ii) We propose a method to replace the revocation list by authenticating trust, and our scheme does not involve PKI certificates, thus reducing the storage burden of the system vehicles. It also does not involve complex bilinear pairing operations, which effectively improves authentication efficiency.

1.2. Organization. The rest of this article is arranged as following: Section 2 describes the related work of the proposed scheme. Section 3 introduces preliminaries and background information. In Section 4, we described the proposed scheme in detail. Section 5 gives a proof of the security in the random oracle model under ECDLP. Security analysis and performance evaluation are described in detail in Section 6. Finally, Section 7 summarizes the future work of this paper.

2. Related Work

In the last several years, scholars have done a lot of research on the preserving privacy and data security of nodes in IOV.

2.1. Anonymous Authentication. Many anonymous authentication schemes have been proposed for IOV, which can be divided into five categories based on the encryption mechanism employed: public key infrastructure (PKI), certificateless signature, symmetric cryptography, identity-based signature, and group signature.

To realize preserving privacy and security in IOV, in 2007, Raya and Hubaux [13] used anonymous certificates to hide the identity of users and a PKI-based scheme is proposed. Raya advises to store huge amounts of public/private keys and corresponding certificates in each vehicle, and the vehicle randomly selects the certificate to sign the message. The privacy of the vehicle is protected by regular replacement of keys and certificates. In 2008, Lu et al. [2] proposed an efficient conditional privacy preservation (ECPP)

protocol based on bilinear mapping. The main limitation of ECPP is the large latency of RSU in generating pseudonym. In 2012, Shim proposed an identity-based signature scheme [14], which stores the master key in the vehicle's tamper-proof device. The vehicle can use the system master key to generate pseudo-names and other information. In 2013, Horng et al. proposed a scheme [15] to use RSU to generate different pseudo-names for vehicles to generate a distinctive anonymous authentication scheme, avoiding the use of a great deal of public and private key pairs by using pseudonym communication. However, guaranteeing the security of the RSU is also a problem. Shao et al. [16] through the use of the new group signature scheme proposed new IOV authentication protocol. However, it can cause random tracking, which reduces user privacy. In 2018, Li et al. [17] proposed an anonymous conditional privacy-preserving authentication scheme based on pseudoidentity method. Each OBU should prestore pseudoidentity in order to maintain their identity privacy. Liu et al. [18] designed a distributed MAC layer antiattack pseudonym scheme. In 2019, Liu et al. [19] designed an anonymous authentication scheme based on group signature, where area TA provided anonymous authentication services. Boualouache et al. [20] proposed an effective pseudonym changing and management framework. This approach can keep the message integrity, and the sender's privacy, but it also has some disadvantages. When the vehicle's private key has been revoked, the system needs to be updated regularly for vehicle certificate; it may take time. Key distribution, management, and storage are challenges. To solve these problems, Du et al. [21] designed a certificateless signature scheme combined with certificateless public key cryptography. Zhong et al. [22] presented a full aggregation authentication scheme for VANETs, which achieved conditional privacy protection by using pseudonyms. In 2020, Bayat et al. [23] proposed a new security and privacy protection scheme based on RSU. In this scheme, the TA stored the master key in the temper-proof device of the RSU, and the verifier used the public key of the RSU instead of the system to check whether the signature is valid. Therefore, vehicles cannot check the signatures of other vehicles on the road from other RSUs. However, bilinear pairing and map-to-point operations are used in the scheme, which results in high computational overhead. Verma et al. [24] proposed the pairing-free certificate-based aggregated signature scheme. Xu et al. [25] proposed a certificateless signature scheme based on the CDH assumption. However, the scheme utilized the expensive map-to-point hash function, which also increased computational and communication overhead. To reduce computational and bandwidth costs, Mei et al. [26] proposed a conditional privacy certificateless signature scheme, which achieved full aggregation. But the scheme is also based on bilinear pairing. To further reduce the overhead of the vehicle, Chen et al. [27] designed a certificateless aggregated signature scheme without the expensive map-to-point hash function and bilinear pairing operations. Ali et al. [28] proposed a certificateless short signature-based conditional privacy-preserving authentication scheme based on ECC, which supported the batch signature verification method.

Table 1: Properties of related IOV schemes.

Scheme	Crypto.primitive	Comp.&comm.cost
Raya and Hubaux [13]	PKI	High
Lu et al. [2]	Group signature	Medium
Shim [14]	ID based	Low
Shao et al. [16]	Group signature	Medium

Table 1 provides the nature of the above scheme for the sake of clarity.

However, only anonymity is not sufficient to prevent an attacker from illegally tracking, even if the broadcast message remains completely anonymous [29]. In addition, traditional public key infrastructure (PKI) guarantees user identity authentication in IOV; however, PKI cannot distinguish untrustworthy information from authorized users. Therefore, a trust evaluation is necessary to guarantee the trustworthiness of information by distinguishing malicious users from networks.

2.2. Trust. The issue of trust stems from the field of security and social psychology. In the past decade, the concept of trust has been suggested to introduce information and communication technology (ICT). There is little research about trust management of IOV during the preceding years. In 2014, MC Chuang and Lee [30] proposed a lightweight authentication scheme for distributed trust extension, called trust extension authentication mechanism, applicable to the vehicle network, with good anonymity and security. In fact, they are designed to further enhance the performance of the authentication process by using the concept of passing trust relationships. Nevertheless, because of the selfish and malicious nodes, the security of mobile ad hoc networks has been greatly reduced. Then, Sugumar et al. [31] proposed a trust-based authentication protocol for cluster-based IOV in 2016. The vehicles are clustered and the trust level of each node is estimated. Inspired by the estimated trust, the cluster head is selected. Because the CRL check requires time, the group signature-based scheme has a long computing delay. In 2018, Yan et al. [32] proposed a scheme to anonymously verify the trust of pervasive social networking (PSN) nodes in a semi-distributed way. It was emphasized that trust plays an important role in maintaining pervasive social networking. It can be seen that anonymous authentication of trust is emerging as a novel way to ensure privacy. In 2020, Liang et al. [33] proposed a reputation scheme based on implicit generalized mixed transition distribution model, which can evaluate the credibility of neighbor vehicles. Begriche et al. [34] proposed a vehicle-mounted network reputation system node based on Bayesian statistical filter that would establish a profile based on the behavior of its neighbors. However, there are only two categories of vehicle states. In the same year, Awan et al. [35] proposed a centralized trust-based clustering mechanism, using multiple parameters to select reliable cluster head and a backup cluster head, thus improving network security. In addition, the method selects a backup cluster head to achieve stable clusters. However, the scheme relies on the RSU. Alnasser

et al. [36] proposed a recommendation-based trust model. The trust of this model comes from two methods: direct trust and indirect trust, but the trust value is calculated in the way of weighted sum, which cannot resist collusion attacks. Chen et al. [37] proposed a decentralized trust management system based on blockchain. The trust model only allows trusted nodes to participate in the verification and consensus process, and a trusted execution environment is applied to protect the trust evaluation process and an incentive model for incentivizing more participation and punishing malicious behavior. Gao [38] proposed a trust management scheme. In the scheme, the trust of nodes includes direct trust and recommendation trust. Direct trust is computed dynamically through history and Bayesian inference. Recommendation trust takes into account the trust and reputation of other nodes and their reputation. Ahmad et al. [39] proposed a hybrid trust management scheme called NOTRINO, which calculates the trust value of nodes at the transport layer and calculates the trust value of data at the application layer.

Unlike all the previous work, this paper combines IOV application scenarios based on the research trust-based [32] and encryption scheme [40], a certificateless anonymous authentication scheme suitable for preserving privacy is proposed for IOV.

3. Preliminaries

In this section, we will briefly cover the mathematical foundations, system model, security and authentication requirements.

3.1. Mathematical Foundations. This subsection describes some of the basics associated with anonymous authentication protocols, namely, elliptic curve cryptography (ECC) and mathematical assumptions.

3.1.1. Elliptic Curve Cryptography (ECC). After elliptic curve cryptography was proposed by Koblitz [41] and Miller [42] in 1986, respectively, ECC began to be commonly used in security-related fields such as encryption and protocols. In the following sections, we briefly introduce elliptic curve cryptography, which is extensively used to design many encryption and security schemes because of its availability in computing and communication costs. In the case that the safety strength provided is the same as that of the discrete logarithm system, the parameters required by ECC are far less than those of the discrete log-based system [43]. The elliptic curve can be characterized by the set of solutions of a two element equation.

If the group $\mathbb{G}$ is a finite cyclic group on the elliptic curve E, its order is p and the generator is P. Let p be a prime number greater than 3, and the elliptic curve $y^2 = x^3 + ax + b$ on Z_p consist of a group of solutions $(x, y) \in Z_p \times Z_p$ based on congruence $y^2 \equiv x^3 + ax + b \bmod p$ and an exceptional point o called infinite point, where $a, b \in Z_p$ comprises two constants satisfying $4a^3 + 27b^2 \neq 0 \bmod p$. In addition, $\mathbb{G}$ has two rules of operation:

(1) Addition ($\pm$): let $P, Q \in \mathbb{G}$, if $P \neq Q$, $R = P + Q$, then R is the point where the line crosses P and Q and E; if $P = Q$, $R = P + Q$, then R is the intersection of the tangent of E and $P(Q)$; if $P = -Q$, there is $P + Q = P - P = \mathcal{O}$.

(2) Scalar multiplication ($\cdot$): let $P \in \mathbb{G}$, $m \in Z_q^*$, and P have a scalar multiplication of $m \cdot P = P + P + \cdots + P$ (m times in total).

3.1.2. Difficult Problem. Let $\mathbb{G}$ be a finite cyclic group with large prime q on an elliptic curve and P be a generator. To demonstrate the security of our scheme, two difficult problems are defined. The mathematical difficulties of participating in the proposed scheme are shown.

Definition 1. Elliptic curve discrete logarithm problem (ECDLP): random point $P, Q \in \mathbb{G}$ on E are presented, and $Q = xP$, output $x \in Z_q^*$.

Definition 2. Computation of Diffie–Hellman problem (CDHP): given $P, aP, bP \in \mathbb{G}$, where $a, b \in Z_q^*$, calculate abP.

If the algorithm of the ECDLP or the CDHP on the group $\mathbb{G}$ cannot be solved by a nonnegligible probability ε within the time τ, then the ECDLP or the CDHP is difficult in the group $\mathbb{G}$.

3.2. System Model. We describe the system model of the proposed anonymous authentication scheme in Figure 1. The trusted authority (TA) has adequacy functions and is trusted to provide identity management and trust management. What is more, TA or IOV nodes that are more stable and dependable than other vehicle nodes (for example, wi-fi access points and base stations) can act as authorizers (AP). AP uses adequate information about nodes to estimate the trust value of the node. In order to achieve instant communication, the nodes interact with each other. Because message integrity and privacy are important, it is necessary to verify node trust anonymously for reliable communication and preserving privacy. TA is used by vehicle nodes to manage the correspondence among real identity, pseudonym, key and trust in the cloud to save computing and storage costs. When the TA is inaccessible, the IOV node can use some of the IOV nodes as APs to correspond to each other.

(1) Trusted authority (TA): it is based on the assumption that TA is fully trusted and has sufficient computing and storage capacity. Through a secure channel, entities (vehicles and RSU) must register with the TA using some personal credentials that uniquely identify the entity. TA is responsible for the registration of fixed RSU on the roadside and mobile OBU installed on vehicles and can reveal the true OBU identity of secure messages.

(2) Road side units (RSU): suppose the RSUs are widely deployed on the road and can be viewed as the router

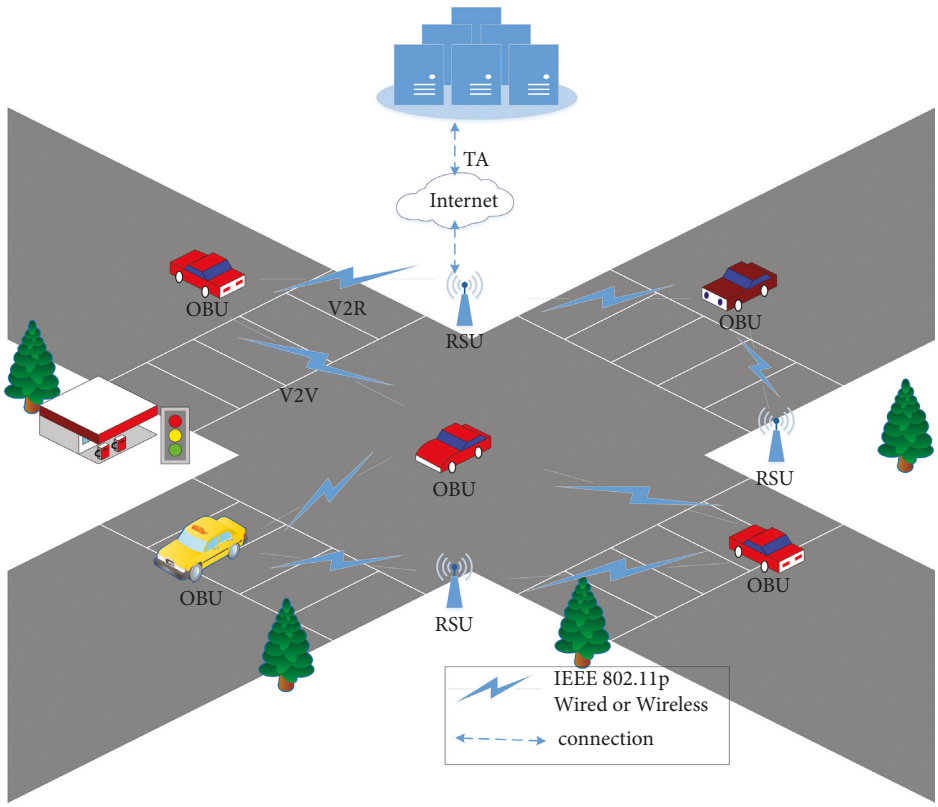


Figure 1: The system model of IOV.

between the TA and vehicle nodes. RSU are not entirely credible, so they have to be supervised by TA.

(3) Vehicles (OBU): each vehicle is equipped with OBU which has a shorter communication range and less computing power than RSU. With the built-in OBU and DSRC protocols, each vehicle can communicate with neighboring vehicles, RSU and TA. The real identity of the vehicle and some secret information about the operation are stored in the OBU.

3.3. Security Requirements. Because messages are transported in an open access environment, security and privacy issues related to IOV must be considered. For anonymous authentication on trust in IOV, the following safety requirements must be met [6]:

3.3.1. Authentication. This requirement consists of vehicle authentication and message integrity. Vehicle authentication allows the receiver to verify the authenticity of the sender, and the message integrity ensures that the message is not changed during the transmission.

3.3.2. Anonymity. The system proposed in this scheme is shown in Figure 1. No entity other than TA can know any information about the real identity of the vehicle, that is, only TA can reveal the real identity of the participating vehicle.

3.3.3. Traceability. This function is used to identify malicious vehicles that may transmit false messages. Vehicles and RSU have no way to know the real sender of the received message, but TA can recover the true identity of the sender in case of an accident, which is called conditional traceability in IOV.

3.3.4. Unlinkability. The user's unlinkability means that the attacker could not judge whether any two messages are from the same vehicle.

3.3.5. Replaying Resistance. Malicious vehicles cannot collect and send messages that have been received by the recipient.

3.4. Authentication Requirements. In order to ensure the safety of IOV communication, the following authentication requirements must be met:

(1) The computational and communication overhead of digital signatures must be low

(2) Authentication should be robust and extensible

(3) The process of reauthentication and revocation should be provided

4. The Proposed Scheme

In this section, we describe a trust-based authentication scheme proposed in this paper, which can authenticate node trust and verify node signature by anonymous method,

which is suitable for secure V2V communication in IOV. Specifically, after system settings and node registration, authorized parties (AP) issue aggregated lists of INT values and INT hash (in short, aggregated lists) to each IOV node. On the basis of INT, nodes generate their one-time key pairs to sign their messages. Based on previous research on trust in IOV, we can assume that the trust of a node is a specific value, such as context-aware trust generation [12].

The scheme is divided into seven phases: system initialization, node registration, issue trust value, aggregate list, one-off key pair generation, signature generation, and verification. The symbols used in the proposed scheme are given in Table 2. Detailed procedures for the proposed scheme are as follows:

4.1. System Initialization. In this subsection, TA generates system parameters and loads them to the vehicle node. The system initialization of the scheme is the responsibility of TA, which consists of two parts, namely, key generation center (KGC) and tracing authorization (TRA), assuming that both parties have enough storage space and computing capacity. Since we assume that TA is reliable in this paper, we can conclude that KGC and TRA are also reliable.

(1) Given the safety parameter ℓ, TAs use two large prime numbers p, q and an elliptic curve defined by $y^2 = x^3 + ax + b (\mathrm{mod}) p, a, b \in F_p$.

(2) The KGC chooses point P from E and generates group $\mathbb{G}$ through P. KGC selects the random number $\alpha \in Z_q^*$ and calculates

$$P_{pub} = \alpha P, \tag{1}$$

where α is the secret value stored in KGC and is the master key used to extract part of the key.

(3) The TRA picks point P from E and produces the group $\mathbb{G}$ through P. TRA selects the random number $\beta \in Z_q^*$ and calculates

$$T_{pub} = \beta P, \tag{2}$$

where β is the secret value stored in TRA and the master key for traceability.

(4) TAs choose four secure hash functions H_1: $\mathbb{G} \longrightarrow Z_q^*$, H_2: $\{0,1\}^* \longrightarrow \mathbb{G}$, H_3: $\{0,1\}^* \longrightarrow \mathbb{G}$, h_1: $\{0,1\}^* \longrightarrow \{0,1\}^n$, h_2: $\{0,1\}^* \longrightarrow Z_q^*$.

(5) They publish the system parameters *Para*:

$$Para = \left(P, p, q, E, \mathbb{G}, H_1, H_2, H_3, h_1, h_2, P_{pub}, T_{pub}\right). \tag{3}$$

When the system is initialized, these public system parameters, *Para*, are reloaded into the tamper-proof device in the vehicle node.

4.2. Node Registration. In this subsection, when each vehicle node V_i registers with the system (TA), it needs to rely on its unique real identity (RID_i. In addition, the public key can be authenticated using the aggregation list distributed by AP, thus achieving certificateless, trust-based authentication. Therefore, the proposed scheme does not need the public key certificate (Figure2).

Table 2: System notations.

Notations	Description
V_i	The ith vehicle
M	Messages from vehicles
t_iT_i	A timestamp
TV_i	The short-lived trust value of V_i
T_TV_i	The validity period of TV_i
AC_TV_i	The authentication code of TV_i
$Cert_i$	The certificate of ID_i issued by TA
P_{pub}, P_{pub}	The public key pair of KGC and TRA
RID_i	The real identity of V_i
(U_i, V_i)	The one-off public/private pair key of TV_i
sign_i	The signature from V_i
$\mathbb{G}$	A cycle additive group
P	A generator of the group $\mathbb{G}$
q	The order of $\mathbb{G}$
$H(\cdot)$	A MapToPoint hash function
$h(\cdot)$	The hash function
TA	Trusted authority
CRL	Certificate revocation list
OBU	On board unit
RSU	Road side unit

(1) The vehicle V_i selects a random number $k_i \in Z_q^*$ and calculates

$$ID_{i,1} = k_iP. \tag{4}$$

TRA receives $(RID_i, ID_{i,1})$ from the vehicle, and the communication channel between the two parties is safe, where the vehicle node V_i can be uniquely identified through RID_i.

(2) When TRA receives RID_i from vehicle V_i, where RID_i is the real identity of V_i, it first checks for RID_i and then calculates

$$ID_{i,2} = RID_i \oplus H_1\left(\beta \cdot ID_{i,1}, T_i, T_{pub}\right), \tag{5}$$

where T_i indicates the validity period of this pseudoidentity. The TAs choose random $u_i \in Z_q$, TAs also provide certificate $Cert_i = u_iP$. The node uses this certificate to request its trust value from TAs. Going down this, KGC can receive pseudoidentity ID_i and $Cert_i$ in a secure manner.

$$ID_i = \left(ID_{i,1}, ID_{i,2}, T_i\right). \tag{6}$$

(3) When KGC obtains the pseudoidentity ID_i, it calculates part of the private key psk_{ID_i} after selecting a random number $d_i \in Z_q^*$ and computing $Q_{ID_i} = d_iP$.

$$psk_{ID_i} = d_i + h_2\left(ID_i, Q_{ID_i}\right) \times \alpha \bmod q. \tag{7}$$

The vehicle receives $(ID_i, psk_{ID_i}, Cert_i, Q_{ID_i})$ from KGC in a secure manner, including the pseudoidentity, partial private key, and certificate.

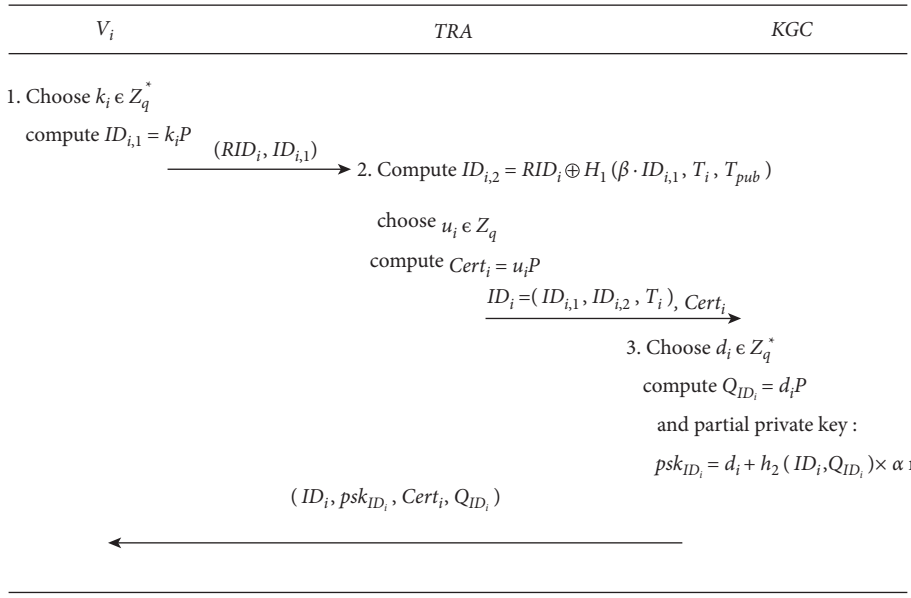


Figure 2: Registration of vehicle node.

4.3. Issue Trust Value and Aggregate List. First of all, each AP (executed by TA or IOV nodes) delivers an original trust value with a valid period and the aggregate list of INT hashes for node V_i in the system; the AP then notifies all vehicle nodes of the newly generated aggregation list. The AP first inspects the validity of the previous trust value before deciding whether to reissue the trust value. In this subsection, one of its essential components is nodes to verify the trust values of other nodes during communication. Nodes request and receive INT in a trustworthy way. In addition, AP will use its signature to distribute the latest INT summary list. Based on its current INT, the trusted processor can produce a one-time public and private key pair.

The trust value of the vehicle can be obtained by analyzing the message records issued by the vehicle collected by AP. At AP, the information collector saves the results in a database after collecting and processing message records from the vehicle nodes. The trust evaluator is used to evaluate the trust value of the vehicle node and detect the malicious vehicle node. The trusted publisher issues an aggregated list of INT hash values for all nodes on the IOV node on a regular or per request basis. When a vehicle node is registered, TAs issue an original trust value on the basis of the behavior of the vehicle node. The TAs collaborate with APs to determine the node's INT and track its true identity without revealing the node's true identity to any other IOV node. The TA database also holds the trust value of each node and its true identity. The AP can communicate with the TA more stably and reliably than a normal node.

After the AP reevaluates the trust, a new trust value is obtained, and it then stores the hash value of the new INT value to the appropriate location of Ha or Ha_AP_j. When the value of trust expires, the trust value is re-requested, and the AP deletes the old value. Its corresponding INT is saved to the appropriate location in the latest aggregation list. The AP then publishes the updated list to all vehicle nodes. All AP simultaneously broadcasts its latest INT summary list. The value of the node's trust can be verified through the presence and location $h_1(Q_i \| h_1(n_i))$ or $h_1(Q_i \| h_1(n_i \_ AP_j))$ of the aggregation list (Ha or Ha_AP_j). Because INT values are sorted in the list (for example, in ascending order), the node during the message authentication is easy to compare trust value. The following will be described separately in two cases, as described in detail below.

(1) AP is executed by TAs: in this phase, based on the true identity of the vehicle, the TAs construct an original or new INT value for the vehicle node. When the current period of trust value expires, a new trust value is requested, at which point TA reevaluates the trust value of V_i and publishes it to V_i using the authentication code AC_TV_i. The vehicle node transmits a random number $d1$ and its certificate $Cert_i$ to TAs to request a trust value. The shared session key between TA and V_i is established using the Diffie–Hellman key agreement protocol, and $d2$ is selected by TA. Afterwards TAs transmit parameters: $\{h_1(TV_i \| AC\_TV_i), T\_TV_i, s_i, Q_i = s_i \cdot P\}$, where TV_i is due at T_TV_i. The list Ha of INT hashes is produced periodically by TAs: $Ha = \{h_1(Q_1 \| h_1(n_1)), \cdot, h_1(Q_i \| h_1(n_i)), \cdot\}$, where $n_i = h_1(TV_i \| AC\_TV_i)$. And then all the nodes will receive $\{Ha \| sign_{TS}(Ha)\}$ from TAs.

(2) AP is carried out by the IOV node: AP (AP_j) can be played by node V_j to assess the others' trust value in IOV. In the same way, Diffie–Hellman key agreement protocol is adopted to establish the shared session key between AP and V_i. Afterwards AP_j transmits parameters: $\{h_1(TV_i \_ AP_j \| AC\_TV_i \_ AP_j), T\_TV_i \_ AP_j\}$ to V_i,

where $TV_i\_AP_j$ is due at $T\_TV_i\_AP_j$. In this case, V_j also can be authenticated with by node V_i. If there are multiple APs, Ha_AP_j is produced periodically by AP_j. And publish it to all nodes after signing: $\left\{Ha\_AP_j \| sign_{AP_j}(Ha\_AP_j)\right\}$ with his private key. Of which

$$Ha\_AP_j = \left\{ \begin{array}{l} h_1\left(Q_1 \| h_1\left(n_1\_AP_j\right)\right), \ldots \\ h_1\left(Q_i \| h_1\left(n_i\_AP_j\right)\right), \ldots \end{array} \right\}. \tag{8}$$

4.4. One-Off Key Pair Generation. In this subsection, vehicle nodes can construct its one-off key pair on INT to sign the messages it sends. Receivers can verify received messages individually or aggregately.

Be based on $n_i = h_1(TV_i \| AC\_TV_i)$, one-off anonymous public and private key pairs (Y_i and r_i) can be constructed by V_i. The production of one-off anonymous key pairs is depicted in Algorithm 1. By randomly changing the nonce a, V_i can produce a distinctive key pair for a new one-off public and private key pair. Therefore, if n_i is the same, different key pairs can be generated to achieve advanced privacy.

4.5. Signature Generation. In this subsection, the vehicle must sign the message with the one-off private key before sending the message, in order to authenticate and preserve the integrity of a message. Vehicle V_i first randomly selects pseudo ID_i from memory and selects the latest timestamp t_i. The updated timestamp t_i protects signature messages from replay attacks. Given the signature key (psk_{ID_i}, r_i) and message M_i, the following steps will be performed by vehicle V_i.

(1) The node sends the message M_i by calculating h_i and signing on M_i using the private key Y_i.

$$\begin{aligned} h_i &= H_3(M_i, ID_i, Y_i, t_i) \\ sign_i(M_i) &= h_i \cdot s_i + psk_{ID_i} \bmod q. \end{aligned} \tag{9}$$

(2) After that, V_i outputs the final message and uses the following format to send msg to other nodes

$$msg = (ID_i, Y_i, sign_i, M_i, t_i, Q_i). \tag{10}$$

4.6. Aggregate. If different nodes send many messages to the same node over a period of time, we can calculate multiple signature combinations as $S = \sum_{i=1}^{n} sign_i(M_i)$ for getting a collection of individual certificateless signatures at a receiver.

4.7. Verification. When adjacent vehicles communicate with each other and send messages, the receiving vehicle needs to check the signature of the message to ensure that the corresponding vehicle does not attempt to propagate a false message (Figure3).

(1) Individual verify: when the node receives the message, the receiver first extracted $h_1(n_i)$ from Y_i:

$$h_1(n_i) = Y2_i \oplus H_1(Y1_i), \tag{11}$$

and calculates $h_1(Q_i \| h_1(n_i))$ to verify the trust value of V_i according to the location in the list. Once the authenticity of the sender's trust value is verified, the recipient performs signature verification. The receiver uses system common parameters to validate the sender's signature by computing $h_{i,2} = h_2(ID_i, Q_{ID_i})$ and $h_i = H_3(M_i, ID_i, Y_i, t_i)$, then checks if the following equation is met, $sign_i(M_i) \cdot P = h_i \cdot Q_i + Q_{ID_i} + h_{i,2} \cdot P_{pub}$, and if satisfied, the recipient accepts this certificateless signature. Since $P_{pub} = \alpha P$, $psk_{ID_i} = d_i + h_2(ID_i, Q_{ID_i}) \times \alpha \bmod q$, $Q_{ID_i} = d_i P$, $Q_i = s_i \cdot P$, and $sign_i(M_i) = h_i \cdot s_i + psk_{ID_i} \bmod q$. We obtain

$$\begin{aligned} sign_i(M_i) \cdot P &= \left(h_i \cdot s_i + psk_{ID_i}\right) \cdot P \\ &= h_i \cdot s_i \cdot P + \left(d_i + h_{i,2} \times \alpha\right) \cdot P \\ &= h_i \cdot Q_i + Q_{ID_i} + h_{i,2} \cdot P_{pub}. \end{aligned} \tag{12}$$

(2) Aggregate verify: when the node receives the message, the receiver first calculates:

$$h_1(n_i) = Y2_i \oplus H_1(Y1_i). \tag{13}$$

Extract from Y_i and calculate $h_1(Q_i \| h_1(n_i))$ to verify the trust value of V_i according to the location in the list, in which $i =, 1, 2, \ldots, n$. Once the authenticity of the sender's trust value is verified, the recipient performs signature verification. The receiver uses system common parameters to validate the sender's signature by computing $h_{i,2} = h_2(ID_i, Q_{ID_i})$ and $h_i = H_3(M_i, ID_i, Y_i, t_i)$, which $i =, 1, 2, \ldots, n$, then check that the following equation is met, $sign_i(M_i) \cdot P = \sum_{i=1}^{n}(h_i \cdot Q_i) + \sum_{i=1}^{n}(Q_{ID_i}) + \sum_{i=1}^{n}(h_{i,2}) \cdot P_{pub}$, and if satisfied, the recipient accepts this certificateless signature. Since $P_{pub} = \alpha P$, $Q_{ID_i} = d_i P$, $psk_{ID_i} = d_i + h_2(ID_i, Q_{ID_i}) \times \alpha \bmod q$, $Q_i = s_i \cdot P$, and $sign_i(M_i) = h_i \cdot s_i + psk_{ID_i} \bmod q$. We can get

$$\begin{aligned} S \cdot P &= \sum_{i=1}^{n} sign_i(M_i) \cdot P \\ &= \sum_{i=1}^{n}\left(h_i \cdot s_i + psk_{ID_i}\right) \cdot P \\ &= \sum_{i=1}^{n} h_i \cdot s_i \cdot P + \sum_{i=1}^{n}\left(d_i + h_{i,2} \times \alpha\right) \cdot P \\ &= \sum_{i=1}^{n} h_i \cdot Q_i + \sum_{i=1}^{n} Q_{ID_i} + \sum_{i=1}^{n} h_{i,2} P_{pub}. \end{aligned} \tag{14}$$

4.8. Identity Tracking. Once a vehicle sends a malicious message, the TRA can track the identity of the vehicle. Through the pseudoidentity $ID_i = (ID_{i,1}, ID_{i,2}, T_i)$, TRA

Require:
$n_i = h_1(TV_i \| AC\_TV_i)$
Ensure:
(1) $Y_i(Y1_i$ and $Y2_i)$ and $r_i(r1_i$ and $r2_i)$
(2) (i) The one-off public key Y_i can be calculated as:
(3) $Y1_i = n_i \cdot a \cdot Q_i$; $Y2_i = h_1(n_i) \oplus H_1(n_i \cdot a \cdot Q_i)$
(4) Where α is a random number and $\oplus$ represent an XOR operation;
(5) (ii) The private key r_i can be calculated as
(6) $r1_i = Y1_i \cdot s_i$; $r2_i = H_2(Y1_i \| Y2_i) \cdot s_i$

Algorithm 1: Generation of one-off anonymous key pairs.

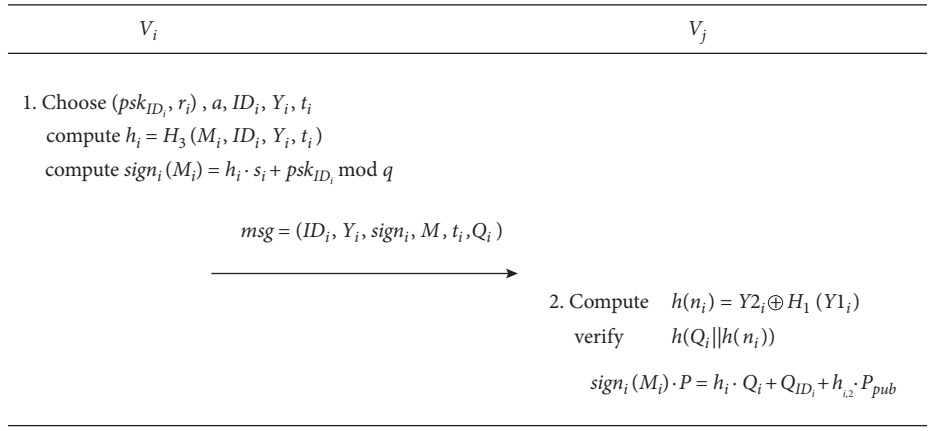


Figure 3: Anonymous authentication process based on trust.

calculates the equation $RID_i = ID_{i,2} \oplus H_1(\beta \cdot ID_{i,1}, T_i, T_{pub})$ to trace the vehicle's true identity. At the same time, the AP will reevaluate the trust of the vehicle and publish the updated list to all vehicle nodes. In addition, TA will update its database.

5. Security Proof and Analysis

Before we show that the proposed scheme has the security and privacy requirements, existential unforgeability of the signature, $sign_i(M_i)$, is proved in the random oracle model.

5.1. Security Model. The security model of the proposed scheme is to design a game between challenger $\mathscr{C}$ and adversary $\mathscr{A}$, that is, whether adversary $\mathscr{A}$ can win the challenge given by challenger $\mathscr{C}$ in polynomial time with a nonnegligible probability. Adversary $\mathscr{A}$ performs the query described below in the game.

(i) Setup: challenger $\mathscr{C}$ creates the public key and gives it to $\mathscr{A}$.

(ii) $h_2(\cdot)$ queries: in this query, challenger $\mathscr{C}$ chooses a random $v_i \in Z_q^*$ and then adds (ID_i, Q_{ID_i}, v_i) into the hash list h_2^{list}. Finally, $\mathscr{C}$ sends $v_i = h_2(ID_i, Q_{ID_i})$ to $\mathscr{A}$.

(iii) $Y(\cdot)$ queries: challenger $\mathscr{C}$ picks random $x_i \in Z_q^*$, inserts tuple $Y1_i$, $Y2_i$, x_i into Y^{list} and responds to $\mathscr{A}$ with $Y1_i$, $Y2_i$ in query i.

(iv) $H_2(\cdot)$ queries: in this query, challenger $\mathscr{C}$ picks random $y_i \in Z_q^*$, inserts the tuples $Y1_j$, $Y2_i$, y_i, H_{2i} into Y^{list} and responds to $\mathscr{A}$ with H_{2i} in query i.

(v) $H_3(\cdot)$ queries: in this query, challenger $\mathscr{C}$ picks random $u_i \in Z_q^*$, inserts the tuple u_i, m_i, H_{3i} to H_3^{list} and responds to $\mathscr{A}$ with $H_3(m_i, ID_i, Y_i, t_i) = H_{3i}$.

(vi) Partial private key queries: in this query, challenger $\mathscr{C}$ calculates psk_{ID_i} and then the value psk_{ID_i} is outputted to $\mathscr{A}$.

(vii) Sign queries: after receiving the message M_i, $\mathscr{C}$ generates the request message $(ID_i, Y_i, sign_i, M_i, t_i, Q_i)$ and sends it to $\mathscr{A}$.

The probability that $\mathscr{A}$ may violate the authentication of proposed scheme Γ is expressed as $A\,dv_\Gamma^{Auth}(\mathscr{A})$.

Definition 3. The proposed scheme Γ for IOV is secure if $A\,dv_\Gamma^{Auth}(\mathscr{A})$ is negligible for any polynomial adversary $\mathscr{A}$.

5.2. Security Proof. In this subsection, to prove unforgeability of the proposed scheme, we need to show that it is unforgeable against adversary $\mathscr{A}$. If and only if CDHP is difficult, our scheme is safe under adaptive selective message attack in the random prediction model.

Theorem 1. *Unforgeability: make the prime order group* $\mathbb{G}$ *into* (τ, t', ε')*– CDH group, which implies that no challenger*

$\mathscr{C}(t\prime, \varepsilon\prime)$ can destroy CDHP on it. Therefore, the proposal is that the existence of an attack on adaptive selection is $(t, \varepsilon, q_Y, q_{h_2}, q_{H_2}, q_{H_3}, q_{pk}, q_S)$-secure, and $\varepsilon = eq_S\varepsilon\prime$, and $c_{\mathscr{C}}$ and $t = t\prime - c_{\mathscr{C}}(q_Y + q_{h_2} + q_{H_2} + q_{H_3} + q_{pk} + q_S)$ is constant, where e is the basis of the natural logarithm.

Game: adversary $\mathscr{A}$ has the advantage of ε and time t. Suppose $\mathscr{A}$ queries q_Y times for Y queries, q_{h_2} times for h_2 queries, q_{H_2} times for H_2 queries, q_{H_3} times for H_3 queries, q_{pk} times for Partial private key queries, and q_S times for Sign queries. And then, a challenger $\mathscr{C}$ who has the advantage of at least ε/eq_S and runtime:

$$t + c_{\mathscr{C}}\left(q_Y + q_{h_2} + q_{H_2} + q_{H_3} + q_{pk} + q_S\right), \tag{15}$$

to solve CDHP.

Proof. Challenger $\mathscr{C}$ gives parameters q, $\mathbb{G}$, e and random instance of CDHP, which is P, aP, bP, whereas P is a random generator of $\mathbb{G}$ with order q, a and b are random in Z_q^*. Let $D = abP \in \mathbb{G}$ be the solution for CDHP. Challenger $\mathscr{C}$ interacts with $\mathscr{A}$ to find the solution through the following query.

Setup: challenger $\mathscr{C}$ creates $K_{pub} = q, \mathbb{G}, P, h_2, H_2, H_3$ and gives it to $\mathscr{A}$. This is h_2, H_2, and H_3, which is a random oracle controlled by $\mathscr{C}$, as follows:

$h_2(\cdot)$ queries: when $\mathscr{A}$ makes a h_2 query with parameter (ID_i, Q_{ID_i}), $\mathscr{C}$ checks whether tuples already exist in the hash list h_2^{list}. In that case, $\mathscr{C}$ transfers $v_i = h_2(ID_i, Q_{ID_i})$ to $\mathscr{A}$. If not, $\mathscr{C}$ selects a random $v_i \in Z_q^*$ and then adds (ID_i, Q_{ID_i}, v_i) into the hash list h_2^{list}. Finally, $\mathscr{C}$ sends $v_i = h_2(ID_i, Q_{ID_i})$ to $\mathscr{A}$.

$Y(\cdot)$ queries: challenger $\mathscr{C}$ can query the public key Y. In response to queries, challenger $\mathscr{C}$ keeps tuple list $Y1_j, Y2_j, x_j$ called Y^{list}. At first, it was empty. $\mathscr{C}$ selects random $x_i \in Z_q^*$, $Y2_i \in \{0,1\}^n$ and calculates $Y1_j = x_iP \cdot aP$. It then adds the tuples $Y1_i, Y2_i, x_i$ into Y^{list} and when querying i, it responds to $\mathscr{A}$ with $Y1_i$ and $Y2_i$.

$H_2(\cdot)$ queries: in response to queries, challenger $\mathscr{C}$ maintains list H_2^{list} in tuple $Y1_j$, $Y2_j$, y_j, H_{2j}. $\mathscr{C}$ picks random $y_i \in Z_q^*$ and sets $H_{2i} = H_2(Y1_i \| Y2_i) = y_iP$. Then, it adds the tuples $Y1_j, Y2_i, y_i, H_{2i}$ into Y^{list} and when querying i, it responds to $\mathscr{A}$ with H_{2i}.

$H_3(\cdot)$ queries: in response to queries, challenger $\mathscr{C}$ keeps tuple list u_j, m_j, H_{3j}, called H_3^{list}. At first, it was empty. To respond to the query m_i, challenger $\mathscr{C}$ will do the following:

(1) If it already exists in the tuple u_i, m_i, H_{3i} in H_3^{list} when m_i is queried, $\mathscr{C}$ responds to $H_3(m_i, ID_i, Y_i, t_i) = H_{3i}$

(2) Otherwise, $\mathscr{C}$ only produces random bit $i_b \in \{0,1\}$, which will be determined later for ξ in $P_r[b_i = 1] = \xi$

(3) $\mathscr{C}$ selects random number $u_i \in Z_q^*$. If $b_i = 0$, it then sets $H_3(m_i, ID_i, Y_i, t_i) = H_{3i} = u_iP$. If $b_i = 1$, it then sets $H_3(m_i, ID_i, Y_i, t_i) = H_{3i} = bP \cdot u_iP$. Afterwards, $\mathscr{C}$ adds the tuple u_i, m_i, H_{3i} to H_3^{list} and responds to $\mathscr{A}$ with $H_3(m_i, ID_i, Y_i, t_i) = H_{3i}$. Note that H_{3i} is homogeneous in $\mathbb{G}$ and independent of $\mathscr{A}$.

Partial private key queries: $\mathscr{A}$ queries partial private key for pseudoidentity ID_i, $\mathscr{C}$ calculates $Q_{ID_i} = d_iP$ and then examines if the tuple (ID_i, Q_{ID_i}, v_i) already exists in the hash list h_2^{list}, where d_i is a random number. When the corresponding tuple $(ID_i, Q_{ID_i}, v_i)\mathscr{C}$ is not found, $\mathscr{C}$ will output a failure and stop because the query cannot be answered coherently. Or else $\mathscr{C}$ evaluates $psk_{ID_i} = d_i + h_2(ID_i, Q_{ID_i}) \times \alpha \bmod q$ and outputs psk_{ID_i} to $\mathscr{A}$. It is worth noting that by calling this part of the partial private key query, $\mathscr{A}$ cannot obtain the psk_{ID_j} of the target user through ID_j.

Sign queries: the signature oracle is simulated by maintaining the list of tuples m_j, H_{3j}, σ_j in response to any message m_j signature query. We call this list S^{list}, which was initially empty. When $\mathscr{A}$ uses the message m_i to query oracle *Sign*, $\mathscr{C}$ responds to the query.

(1) If the query m_i already exists in the tuple m_i, H_{3i}, σ_i in S^{list}, challenger $\mathscr{C}$ responds with σ_i.

(2) Besides, $\mathscr{C}$ inspects whether (u_i, m_i, H_{3i}), $(Y1_i, Y2_i, x_i)$, (ID_i, Q_{ID_i}, v_i), and $(Y1_i, Y2_i, y_i, H_{2i})$ exist. Otherwise, $\mathscr{C}$ executes h_2-queries to obtain (ID_i, Q_{ID_i}, v_i), Y-queries to gain $(Y1_i, Y2_i, x_i)$, H_2-queries to obtain $(Y1_i, Y2_i, y_i, H_{2i})$, and H_3-queries to gain (u_i, m_i, H_{3i}). Next, $\mathscr{C}$ picks two random numbers r_i and h_i. If $b_i = 0$, $\sigma_i = h_i \cdot r_i + psk_{ID_i} \bmod q$. If $b_i = 1$, it sets $\sigma_i = *$, value of placeholder. Finally, it adds tuple m_i, H_{3i}, σ_i to list S^{list} and replies to σ_i.

Challenge: challenger $\mathscr{C}$ publishes the signature query m_i. Challenger $\mathscr{C}$ obtains $\sigma_i \in \mathbb{G}$ by running the above algorithm in response to Sign queries. Note that $\mathscr{C}$ can use the public key K_{pub} to run $\mathscr{A}$ to obtain P, aP, H_{3i}, σ_i, which can be converted into a valid Diffie–Hellman tuple.

Claim: $\mathscr{A}$ stops, admit defeat, or forged signature $m\prime$, $\sigma\prime$, where $m\prime = m_{i*}$, for some i^* where $\mathscr{A}$ does not query the signature. If $\mathscr{A}$ is successfully forged, it means that CDHP is solved. At this time, $\mathscr{C}$ outputs "success." Otherwise, the $\mathscr{C}$ output "fails." $\mathscr{A}$ performed exactly as expected in the game model. Thus,

$$\begin{aligned} A\ dv_{\mathscr{C}} &= P_r\left[\mathscr{C}^{\mathscr{A}}(P, aP, bP) = \text{success}: a, b \in Z_q^*\right] \\ &= \left[\text{Verify}(Y, m\prime, \sigma\prime) = \text{valid}: \begin{smallmatrix}(Y,r)\leftarrow \text{OneoffKeyGen} \\ (m', \dot{\sigma})\leftarrow \mathscr{A}(Y)\end{smallmatrix}\right] \\ &= \varepsilon. \end{aligned} \tag{16}$$

By modifying, if $\mathscr{A}$ cannot create forgery, $\mathscr{C}$ will also fail. But if $\mathscr{A}$ finds m_{i^*}'s forgery successfully, $\mathscr{C}$ claims success only at $b_{i^*} = 1$, and $\mathscr{A}$ use index $i_1, i_2, \ldots, i_{q_S}$ to q_S sign oracle query for messages with $b_i = 0$ (for $b_i = 1$, $\mathscr{A}$ will stop immediately after the failure is declared), then $A\ dv_{\mathscr{C}}' = A\ dv_{\mathscr{C}} \cdot \Pr[b_{i^*} = 1] \cdot \Pr[b_{i_j} = 0, j = 1, 2, \ldots, q_S] = \xi(1-\xi)^{q_S}\varepsilon$.

Therefore, challenger $\mathscr{C}$ uses signature forger $\mathscr{A}$ to solve CDHP, which has the advantage of $\varepsilon\prime$ and time $t\prime$. The maximization of function $\xi(1-\xi)^{q_S}\varepsilon$ is at $\xi = 1/(1+q_S)$, of which it has the following values:

$$\frac{1}{1+q_S}\left(1-\frac{1}{1+q_S}\right)^{q_S}\cdot\varepsilon=\frac{1}{q_S}\left(1-\frac{1}{1+q_S}\right)^{q_S+1}\cdot\varepsilon. \quad (17)$$

For large q_S, $(1-1/(1+q_S))^{q_S+1}\approx 1/e$.

Meanwhile, $\mathscr{C}$'s running time consists of $\mathscr{A}$'s running time and the additional overhead, in which the group multiplication to evaluate each signature and hash request from $\mathscr{C}$ is the main part. Any such multiplication can be done by using up to $c_{\mathscr{C}}$ time units on $\mathbb{G}$. $\mathscr{C}$ may have to answer a request like $q_Y+q_{h_2}+q_{H_2}+q_{H_3}+q_{pk}+q_S$. Therefore, its overall runtime is $t+c_{\mathscr{C}}(q_Y+q_{h_2}+q_{H_2}+q_{H_3}+q_{pk}+q_S)$.

If there is a forgery $\mathscr{C}$ that $(t,\varepsilon,q_Y,q_{h_2},q_{H_2},q_{H_3},q_{pk},q_S)$ breaks our proposed scheme on $\mathbb{G}$, then there is a challenger $\mathscr{C}(t\prime,\varepsilon\prime)$ that can destroy CDHP, where $\varepsilon\prime=\varepsilon/(eq_S)$ and $t\prime=t+c_{\mathscr{C}}(q_Y+q_{h_2}+q_{H_2}+q_{H_3}+q_{pk}+q_S)$. On the contrary, if the group $\mathbb{G}$ is a $(\tau,t\prime,\varepsilon\prime)$-CDH group, no challenger could break the proposed scheme, where $t=t\prime-c_{\mathscr{C}}(q_Y+q_{h_2}+q_{H_2}+q_{H_3}+q_{pk}+q_S)$ and $\varepsilon=eq_S\varepsilon\prime$. □

5.3. Security Analysis. We demonstrate that our proposal complies with all security and privacy requirements described in Section 3.3. As summarized by the comparison results in Table 3, we compared the proposal with other schemes for meeting security requirements, where SR1, SR2, SR3, SR4, and SR5, respectively, represent the authentication, anonymity, traceability, unlinkability, and replaying resistance. The comparison results show that the proposed scheme is superior. The security requirements of the proposed scheme are analyzed next.

5.3.1. Authentication. For the following reasons, the proposed scheme provides message integrity and validity of sender identity: signature $sign_i(M_i)\cdot P$ is used to verify the authenticity of the message sent from vehicle to verifier vehicle. And, as shown in Theorem 1, in the random oracle model, signature $sign_i(M_i)\cdot P$ is nonforgery for adaptive selection message and identity attack under the difficulty of CDHP.

5.3.2. Anonymity. The INT value n_i given by AP (node or TAs) produces the one-off public key Y_i used in message authentication, which cannot be linked to the real identity. Moreover, in order to distinguish a one-off key pair for each message, V_i changes the random number each time (Y_i,r_i) is produced. Therefore, for the reason that TA is completely trusted, node privacy can be securely protected. For trust-based anonymous authentication, TA periodically distributes Ha to IOV nodes and uses its private key to sign. Ha's internal position on behalf of the trust value of V_i, however, does not link to the real identity of V_i and the true value of trust. Therefore, the proposed scheme provides anonymous authentication of identity privacy-preserving based on trust.

5.3.3. Traceability. Through the equation $RID_i=ID_{i,2}\oplus H_1(\beta\cdot ID_{i,1},T_i,T_{pub})$, TRA can track the identity of a malicious vehicle. Accordingly, when a vehicle is marked as controversial, TRA can track malicious vehicles to meet traceability requirements. Hence, our proposed scheme provides conditional privacy-preserving authentication.

Table 3: Comparison of security between related schemes and ours.

Scheme	[15]	[44]	[45]	Proposed
SR1	√	√	√	√
SR2	√	√	√	√
SR3	√	√	√	√
SR4	√		√	√
SR5				√

5.3.4. Unlinkability. In our proposal, the INT values in aggregated lists (Ha or Ha_AP_j) are broken down into different levels. According to the INT value published by AP, each node generates n_i. We can set up an INT value range $n_i=h_1(TV_i\|AC\_TV_i)$ to represent a set of nodes that have the same trust level. Therefore, the trust value of many nodes may belong to the same level of trust. Even if the message receiver validates that the same n_i exists in Ha or Ha_AP_j, if during the period of authentication from the same node sent two or more messages, message receiver is indistinguishable. Specific vehicles cannot be linked to any two signatures, so the proposed scheme supports unlinkability.

5.3.5. Replaying Resistance. The time stamp t_i in the message $(ID_i,Y_i,sign_i,M_i,t_i,Q_i)$ is used to keep the message fresh. Vehicles will check the timestamp t_i freshness, so that they can detect the replay message. Therefore, our proposed scheme for IOV provides resistance against the replay attack.

6. Performance Evaluation

In this section, we will analyze the performance of the proposed scheme and compare it with the existing schemes proposed by Horng et al. [15], Bayat et al. [44], and Zhang et al. [45], respectively. The analysis of computation cost and communication overhead is highlighted below.

6.1. Computation Overhead and Comparison. The computational cost refers to the computational overhead of each entity in the authentication process. Table4provides the main operations of the four schemes in signing messages and authenticating a single signature, respectively.

The crypto-operations of Horng et al.'s scheme [15], Bayat et al.'s scheme [44] and Zhang et al.'s scheme [45] are established on bilinear pairings. Furthermore, the crypto-operations of the proposed scheme are established on ECC. In order to reach the 80-bit security level, we consider various parameters in pairing and ECC-based schemes, as given in Table 5.

Before the analysis of the computation cost, we define the time required for each cryptographic-related operation for signature and verification; a few notes to be used in comparison will be described below. In this paper, we use the experiment in Ref. [40] to learn the execution time of the

Table 4: Comparison of computation cost.

Scheme	Signing	Verification
Horng et al. [15]	$4T_{sm-bp} + 1T_{mtp}$	$2T_{bp} + 2T_{sm-bp} + 1T_{mtp}$
Bayat et al. [44]	$5T_{sm-bp} + 1T_{mtp}$	$3T_{bp} + 1T_{mtp}$
Zhang et al. [45]	$2T_{mtp}$	$2T_{bp} + 2nT_{sm-bp} + 2nT_{mtp}$
Proposed scheme	$1T_{mtp} + 1T_{sm-ecc}$	$3T_{sm-ecc} + 1T_{mtp}$

Table 5: Length of the group in bilinear pairing and ECC.

Type of the system	Type of curve	Cyclic group	$\lvert\bar{p}\rvert \lvert p\rvert$	$\lvert G\rvert$	Length of elements of the group
Bilinear pairing	E: $y^2 = x^3 + x \pmod{p}, a, b \in F_p$	$G_1(P)$	$\lvert\bar{p}\rvert$ = 512 bits (64 bytes)	q = 160 bits	$\lvert G_1\rvert$ = 1024 bits
ECC	E: $y^2 = x^3 + ax + b \pmod{p}, a, b \in F_p$	$G(P)$	$\lvert p\rvert$ = 160 bits (20 bytes)	q = 160 bits	$\lvert G\rvert$ = 320 bits

basic cryptographic operation by using the MIRACL library, running on the platform of 3.4 GHZ i7-4770. The following results are obtained from [40]: T_{sm-ecc} is 0.442 ms, $T_{sm-ecc-s}$ is 0.0276 ms, T_{sm-bp} is 1.709 ms, T_{mtp} is 4.406 ms, and T_{bp} is 4.211 ms. As a result of these, operating mainly determines the speed of signature verification, We're just going to talk about these five operations and ignore others, such as addition and one-way hash function.

(i) T_{sm-ecc}: the execution time of a scale multiplication operation $x \cdot P$ associated with ECC, where $x \in Z_q^*$ and $P \in \mathbb{G}$

(ii) $T_{sm-ecc-s}$: the execution time of a small scale multiplication operation $v_i \cdot p$ used in the small exponential test technique, where $P \in \mathbb{G}$, v_i is a small random integer in $[1, 2^t]$ and t is a small integer

(iii) T_{sm-bp}: the execution time of a scale multiplication operation $x \cdot P$ associated with the bilinear pairing, where $x \in Z_q^*$ and $P \in \mathbb{G}$

(iv) T_{mtp}: the execution time of a hash-to-point operation associated with the bilinear pairing, where the hash function maps a string to a point of $\mathbb{G}$

(v) T_{bp}: the execution time of a bilinear pairing operation $e(S, T)$, where $S, T \in \mathbb{G}$

First, we review the message signature time overhead. For Horng et al.'s b-SPECS + scheme [15], the vehicle needs to perform four scalar multiplication operations and one hash-to-point operation associated with the bilinear pairing. To sum up, the time overhead for this scheme is $4T_{sm-bp} + 1T_{mtp}$. For Bayat et al.'s scheme [44], the vehicle is required to perform five scalar multiplication operations and one hash-to-point operation associated with the bilinear pairing. To sum up, the time overhead for this scheme is $5T_{sm-bp} + 1T_{mtp}$. For Zhang et al.'s scheme [45], the vehicle needs to perform two hash-to-point operations related to the bilinear pairing. To sum up, the time overhead for this scheme is $2T_{mtp}$. For the proposed scheme, the vehicle needs to perform one scalar multiplication operation associated with the ECC and one hash-to-point operation associated with the bilinear pairing. To sum up, the time overhead for this scheme is $1T_{sm-ecc} + 1T_{mtp}$.

We observe the verification time of the signature through the verification equation. For Horng et al.'s scheme [15], the verifier is required to perform two bilinear pairing operations, two scalar multiplication operations, and one hash-to-point operation associated with the bilinear pairing. To sum up, the time overhead for this scheme is $2T_{bp} + 2T_{sm-bp} + 1T_{mtp}$. For Bayat et al.'s scheme [44], the verifier is required to perform three bilinear pairing operations, one scalar multiplication operation, and one hash-to-point operation associated with the bilinear pairing. To sum up, the time overhead for this scheme is $3T_{bp} + 1T_{mtp}$. For Zhang et al.'s scheme [45], the verifier is required to perform two bilinear pairing operations, two hash-to-point operations associated with the bilinear pairing, and two scalar multiplication operations. To sum up, the time overhead for this scheme is $2T_{bp} + 2T_{sm-bp} + 2T_{mtp}$. In our scheme, we evaluate the operation time of two parts of verification: trust authentication and signature verification. Thus, the verifier needs to perform three scalar multiplication operations associated with the ECC and one hash-to-point operation associated with the bilinear pairing. To sum up, the time overhead for this scheme is $3T_{sm-ecc} + 1T_{mtp}$.

The number of signatures during verification is then denoted by n. By batch verification of the equation, we can obtain that the verification time of n different signatures is $2T_{bp} + 2nT_{sm-bp} + nT_{mtp} = 7.824n + 8.422ms$ for Horng et al.'s scheme [15], $3T_{bp} + nT_{mtp} = 4.406n + 12.633ms$ for Bayat et al.'s scheme [44], and $2T_{bp} + 2nT_{sm-bp} + 2nT_{mtp} = 12.23n + 8.422ms$ for Zhang et al.'s scheme [45], respectively. For the authentication phase of n signatures of our proposed scheme, the execution time of the phase is $3nT_{sm-ecc} + nT_{mtp} = (5.732n)ms$.

Figure 4 shows the computational overhead of signing messages in each scheme. The linear relationship between the computation cost and the number of messages of four authentication schemes is given. Our proposed scheme has a slightly better performance time than Refs. [15, 44, 45]. The computational efficiency of our second scheme in this phase has been improved by 56.88% than Horng et al.'s scheme

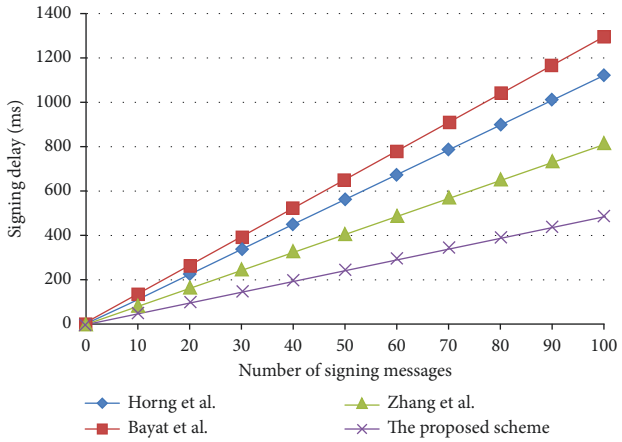


Figure 4: Delay in signing messages with respect to the number of messages.

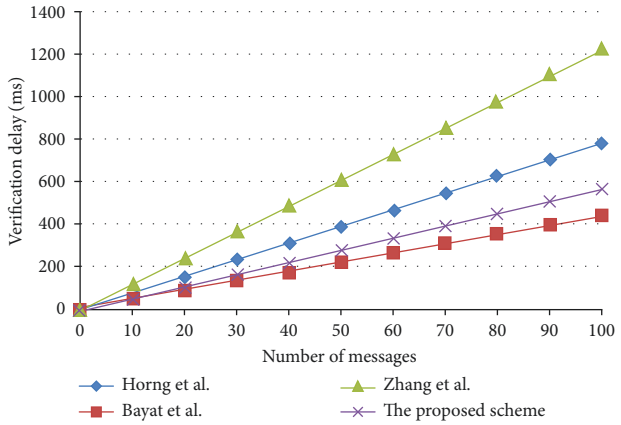


Figure 5: Delay in verifying messages with respect to the number of messages.

[15], by 62.57% than Bayat et al.'s scheme [44], and by 40.30% than Zhang et al.'s scheme [45].

Figure 5 shows the total execution time for verifying n messages, as the number of vehicles in each scenario is increasing. We can see from the figure that Bayat et al.'s scheme's [44] execution time is less than Horng et al.'s scheme [15], Zhang et al.'s scheme [45], and our scheme in the authentication phase.

6.2. Communication Overhead. In this subsection, we compare the communication overhead of the proposed scheme with other schemes, as given in Table 6.

Table 6: Comparison of communication cost.

Scheme	Message	Length (byte)
Horng et al. [15]	$(ID_i, M_i, \sigma i)$	63
Bayat et al. [44]	(AID_i, T_i, U_i)	388
Zhang et al. [45]	$(m_i, PPID_{i,t}, \sigma_{i,t})$	148
Proposed scheme	$(ID_i, Y_i, sign_i, M_i, t_i, Q_i)$	228

According to the analysis in Section 6.1, $|\bar{p}|$ and $|p|$ are 64 and 20 bytes, respectively. Consequently, bytes of elements in group $\mathbb{G}_1$ and group $\mathbb{G}$ are 128 bytes and 40 bytes, respectively. Assuming that the number of bytes of message time t_i is 4 bytes, the number of bytes of RID is 20 bytes, and the number of bytes of the general hash function's output is 20 bytes, the communication overhead of a complete verification in the authentication scheme of IOV usually consists of vehicle signatures, pseudoidentities, current time stamps, and public keys, while the message itself is not considered.

Because of identity-based encryption, Horng et al.'s scheme [15] does not require any signing certificate together with the message to send. Instead, send a 42 byte pseudoidentity, i.e., $|ID_i| = |ID_{i_1}| + |ID_{i_2}| = 42$ bytes, and the length of a signature is 21 bytes. Thus, the total transmission overhead is $42 + 21 = 63$ bytes. In Bayat et al.'s scheme [44], the verifier receives the broadcast anonymous identity and signature (AID_i, T_i, U_i) from the vehicle, where $AID_i = \{AID_i^1, AID_i^2\}, AID_i^1, AID_i^2, U_i \in G_1$ and T_i is the timestamp. To sum up, the communication cost is $128 \times 3 + 4 = 388$ bytes. In Zhang et al.'s scheme [45], the vehicle signs the message as $(m_i, PPID_{i,t}, \sigma_{i,t})$. The overhead of communication can also be calculated using the method shown above. For our proposed scheme, the vehicle signs the message as $ID_i, Y_i, sign_i, M, t_i, Q_i$ and broadcasts it to the verifier, where $ID_i = (ID_{i,1}, ID_{i,2}, T_i)$, Q_i, $sign_i$ both are elements in $\mathbb{G}$. $Y_i = (Y1_i, Y2_i)$ where $Y1_i$ is an element in $\mathbb{G}$, and $Y2_i$ is an array of 20 bytes. T_i and t_i are the timestamp. Thus, the proposed scheme has a communication overhead of $40 \times 5 + 20 + 4 \times 2 = 228$ bytes.

7. Conclusion and Future Work

In the proposal, we proposed a scheme to authenticate the trust of vehicle nodes in IOV. First, our scheme not only provided anonymous authentication of trust but also an effective conditional privacy tracking mechanism, which achieved identity authentication and conditional preserving of privacy, and improved the reliability of V2V communication messages. Next, our proposed scheme realized efficient certificateless authentication, which is based on ECC and replaced the trust on revocation list. Furthermore, we also proved that the proposed scheme is secure against existential forgery in the random oracle model under the CDHP. In future work, we will further consider the characteristics of IOV to design a more efficient scheme, such as high dynamics. In addition, testing the efficiency, adaptability, and robustness of the scheme in a real environment is also an issue to be addressed in the future.

Data Availability

The data used to support 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 study was funded by the Anhui first-class undergraduate talent demonstration and leading base (2019rcsfjd088).

References

[1] S. S. Manvi and S. Tangade, "A survey on authentication schemes in vanets for secured communication," *Vehicular Communications*, vol. 9, pp. 19–30, 2017.

[2] R. Lu, X. Lin, H. Zhu, P. H. Ho, and X. Shen, "Ecpp: efficient conditional privacy preservation protocol for secure vehicular communications," in *Proceedings of the INFOCOM 2008. The Conference on Computer Communications*, pp. 1229–1237, IEEE, Phoenix, AZ, U.S.A, April 2008.

[3] J. Cui, L. Wei, H. Zhong, J. Zhang, Y. Xu, and L. Liu, "Edge computing in vanets-an efficient and privacy-preserving cooperative downloading scheme," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 6, pp. 1191–1204, 2020.

[4] J. Zhang, J. Cui, H. Zhong, Z. Chen, and L. Liu, "Pa-crt: PA-CRT: Chinese remainder Theorem based conditional privacy-preserving authentication scheme in vehicular ad-hoc networks," *IEEE Transactions on Dependable and Secure Computing*, vol. 18, no. 2, pp. 722–735, 2021.

[5] L. Wei, J. Cui, Y. Xu, J. Cheng, and H. Zhong, "Secure and lightweight conditional privacy-preserving authentication for securing traffic emergency messages in vanets," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 1681–1695, 2021.

[6] S. J. Horng, S. F. Tzeng, T. Li, X. Wang, P. H. Huang, and M. K. Khan, "Enhancing security and privacy for identity-based batch verification scheme in vanet," *IEEE Transactions on Vehicular Technology*, vol. 66, p. 1, 2017.

[7] W. Xiong, R. Wang, Y. Wang, F. Zhou, and X. Luo, "Cppa-d: efficient conditional privacy-preserving authentication scheme with double-insurance in vanets," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 4, pp. 3456–3468, 2021.

[8] L. Zhang, Q. Wu, A. Solanas, and J. Domingo-Ferrer, "A scalable robust authentication protocol for secure vehicular communications," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 4, pp. 1606–1617, 2010.

[9] L. Wei, J. Cui, H. Zhong, I. Bolodurina, and L. Liu, "A lightweight and conditional privacy-preserving authenticated key agreement scheme with multi-ta model for fog-based vanets," *IEEE Transactions on Dependable and Secure Computing*, p. 1, 2021.

[10] S. Gyawali, Y. Qian, and R. Q. Hu, "A privacy-preserving misbehavior detection system in vehicular communication networks," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 6, pp. 6147–6158, 2021.

[11] M. Najafi, L. Khoukhi, and M. Lemercier, "A multidimensional trust model for vehicular ad-hoc networks," in *Proceedings of the 2021 IEEE 46th Conference on Local Computer Networks (LCN)*, pp. 419–422, IEEE, Canada, October 2021.

[12] K. Rostamzadeh, H. Nicanfar, N. Torabi, S. Gopalakrishnan, and V. C. M. Leung, "A context-aware trust-based information dissemination framework for vehicular networks," *IEEE Internet of Things Journal*, vol. 2, no. 2, pp. 121–132, 2015.

[13] M. Raya and J.-P. Hubaux, "Securing vehicular ad hoc networks," *Journal of Computer Security*, vol. 15, no. 1, pp. 39–68, 2007.

[14] K. A. Shim, "${\cal CPAS}$: an efficient conditional privacy-preserving authentication scheme for vehicular sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 4, pp. 1874–1883, 2012.

[15] S. J. Horng, S. F. Tzeng, Y. Pan et al., "b-specs+: b-SPECS+: batch verification for secure pseudonymous authentication in VANET," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 11, pp. 1860–1875, 2013.

[16] J. Shao, X. Lin, R. Lu, and C. Zuo, "A threshold anonymous authentication protocol for vanets," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 3, pp. 1711–1720, 2016.

[17] J. Li, K. K. R. Choo, W. Zhang et al., "Epa-cppa: an efficient, provably-secure and anonymous conditional privacy-preserving authentication scheme for vehicular ad hoc networks," *Vehicular Communications*, vol. 13, pp. 104–113, 2018.

[18] Z. Liu, Z. Liu, L. Zhang, and X. Lin, "Marp: a distributed mac layer attack resistant pseudonym scheme for vanet," *IEEE Transactions on Dependable and Secure Computing*, vol. 17, no. 4, pp. 869–882, 2020.

[19] X. Liu, Y. Yang, E. Xu, and Z. Jia, "An authentication scheme in vanets based on group signature," in *Proceedings of the International Conference on Intelligent Computing*, pp. 346–355, Springer, Berlin, Germany, July 2019.

[20] A. Boualouache, S. M. Senouci, and S. Moussaoui, "Privanet: an efficient pseudonym changing and management framework for vehicular ad-hoc networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 8, pp. 3209–3218, 2020.

[21] H. Du, Q. Wen, and S. Zhang, "An efficient certificateless aggregate signature scheme without pairings for healthcare wireless sensor network," *IEEE Access*, vol. 7, pp. 42683–42693, 2019.

[22] H. Zhong, S. Han, J. Cui, J. Zhang, and Y. Xu, "Privacy-preserving authentication scheme with full aggregation in vanet," *Information Sciences*, vol. 476, pp. 211–221, 2019.

[23] M. Bayat, M. Pournaghi, M. Rahimi, and M. Barmshoory, "Nera: a new and efficient rsu based authentication scheme for vanets," *Wireless Networks*, vol. 26, no. 5, pp. 3083–3098, 2020.

[24] G. K. Verma, B. B. Singh, N. Kumar, and V. Chamola, "Cb-cas: certificate-based efficient signature scheme with compact aggregation for industrial internet of things environment," *IEEE Internet of Things Journal*, vol. 7, no. 4, pp. 2563–2572, 2020.

[25] Z. Xu, D. He, N. Kumar, and K. K. R. Choo, "Efficient certificateless aggregate signature scheme for performing secure routing in vanets," *Security and Communication Networks*, vol. 2020, Article ID 5276813, 12 pages, 2020.

[26] Q. Mei, H. Xiong, J. Chen, M. Yang, S. Kumari, and M. K. Khan, "Efficient certificateless aggregate signature with conditional privacy preservation in iov," *IEEE Systems Journal*, vol. 15, no. 1, pp. 245–256, 2021.

[27] Y. Chen and J. Chen, "Cpp-clas: efficient and conditional privacy-preserving certificateless aggregate signature scheme for vanets," *IEEE Internet of Things Journal*, vol. 9, 2021.

[28] I. Ali, Y. Chen, N. Ullah, R. Kumar, and W. He, "An efficient and provably secure ecc-based conditional privacy-preserving authentication for vehicle-to-vehicle communication in vanets," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 2, pp. 1278–1291, 2021.

[29] B. Wiedersheim, Z. Ma, F. Kargl, and P. Papadimitratos, "Privacy in inter-vehicular networks: why simple pseudonym change is not enough," in *Proceedings of the International Conference on Wireless On-Demand*, pp. 176–183, Network Systems & Services, Slovenia, February 2010.

[30] M.-C. Chuang and J. F. Lee, "Team: Trust-extended authentication mechanism for vehicular ad hoc networks," *IEEE Systems Journal*, vol. 8, no. 3, pp. 749–758, 2014.

[31] R. Sugumar, A. Rengarajan, and C. Jayakumar, "Trust based authentication technique for cluster based vehicular ad hoc networks (vanet)," *Wireless Networks*, vol. 24, no. 2, pp. 373–382, 2016.

[32] Z. Yan, P. Wang, and W. Feng, "A novel scheme of anonymous authentication on trust in pervasive social networking," *Information Sciences*, vol. 445, pp. 79–96, 2018.

[33] J. Liang and M. Ma, "Ecf-mrs: an efficient and collaborative framework with markov-based reputation scheme for idss in vehicular networks," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 278–290, 2021.

[34] Y. Begriche, R. Khatoun, A. Rachini, and L. Khoukhi, "A reputation system using a bayesian statistical filter in vehicular networks," in *Proceedings of the 2020 Sixth International Conference on Mobile and Secure Services (MobiSecServ)*, pp. 1–7, IEEE, Miami Beach, FL, USA, February 2020.

[35] K. A. Awan, I. Ud Din, A. Almogren, M. Guizani, and S. Khan, "Stabtrust—a stable and centralized trust-based clustering mechanism for iot enabled vehicular ad-hoc networks," *IEEE Access*, vol. 8, pp. 21159–21177, 2020.

[36] A. Alnasser, H. Sun, and J. Jiang, "Recommendation-based trust model for vehicle-to-everything (v2x)," *IEEE Internet of Things Journal*, vol. 7, no. 1, pp. 440–450, 2020.

[37] X. Chen, J. Ding, and Z. Lu, "A decentralized trust management system for intelligent transportation environments," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 1, pp. 558–571, 2022.

[38] H. Gao, C. Liu, Y. Yin, Y. Xu, and Y. Li, "A hybrid approach to trust node assessment and management for vanets cooperative data communication: historical interaction perspective," *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[39] F. Ahmad, F. Kurugollu, C. A. Kerrache, S. Sezer, and L. Liu, "Notrino: a novel hybrid trust management scheme for internet-of-vehicles," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 9, pp. 9244–9257, 2021.

[40] J. Cui, L. Wei, J. Zhang, Y. Xu, and H. Zhong, "An efficient message-authentication scheme based on edge computing for vehicular ad hoc networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, 2018.

[41] N. Koblitz, "Elliptic curve cryptosystems," *Mathematics of Computation*, vol. 48, no. 177, pp. 203–209, 1987.

[42] V. S. Miller, *Use of elliptic curves in cryptography Advances in Cryptology-CRYPTO'85*, Springer, vol. 218pp. 173–1933, 1986.

[43] S. Biswas, J. Mišic, and V. Mišic, "Id-based safety message authentication for security and trust in vehicular networks," in *Proceedings of the International Conference on Distributed Computing Systems Workshops*, pp. 323–331, Minneapolis, MN, U.S.A, June 2011.

[44] M. Bayat, M. Barmshoory, M. Rahimi, and M. R. Aref, "A secure authentication scheme for vanets with batch verification," *Wireless Networks*, vol. 21, no. 5, pp. 1733–1743, 2015.

[45] L. Zhang, Q. Wu, J. Domingo-Ferrer, B. Qin, and C. Hu, "Distributed aggregate privacy-preserving authentication in vanets," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 3, pp. 516–526, 2017.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 4202084, 18 pages
https://doi.org/10.1155/2022/4202084

Research Article

Towards Fair and Decentralized Federated Learning System for Gradient Boosting Decision Trees

Shiqi Gao,[1] Xianxian Li,[1,2] Zhenkui Shi,[1,2] Peng Liu,[1,2] and Chunpei Li[1]

[1]*Guangxi Key Lab of Multi-source Information Mining and Security, Guangxi Normal University, Guilin, China*
[2]*College of Computer Science and Engineering, Guangxi Normal University, Guilin, China*

Correspondence should be addressed to Xianxian Li; lixx@gxnu.edu.cn and Zhenkui Shi; shizhenkui@gxnu.edu.cn

Received 9 April 2022; Accepted 21 June 2022; Published 2 August 2022

Academic Editor: Andrea Michienzi

At present, gradient boosting decision trees (GBDTs) has become a popular machine learning algorithm and has shined in many data mining competitions and real-world applications for its salient results on classification, ranking, prediction, etc. Federated learning which aims to mitigate privacy risks and costs, enables many entities to keep data locally and train a model collaboratively under an orchestration service. However, most of the existing systems often fail to make an excellent trade-off between accuracy and communication. In addition, they overlook an important aspect: fairness such as performance gains from different parties' datasets. In this paper, we propose a novel federated GBDT scheme based on the blockchain which can achieve constant communication overhead and good model performance and quantify the contribution of each party. Specifically, we replace the tree-based communication scheme with the pure gradient-based scheme and compress the intermediate gradient information to a limit to achieve good model performance and constant communication overhead in skewed datasets. On the other hand, we introduce a novel contribution allocation scheme named split Shapley value, which can quantify the contribution of each party with a limited gradient update and provide a basis for monetary reward. Finally, we combine the quantification mechanism with blockchain organically and implement a closed-loop federated GBDT system FGBDT-Chain in a permissioned blockchain environment and conduct a comprehensive experiment on public datasets. The experimental results show that FGBDT-Chain achieves a good trade-off between accuracy, communication overhead, fairness, and security under large-scale skewed datasets.

1. Introduction

Machine learning (ML) has achieved extensive success in many practical applications. However, a well-trained ML model heavily depends on massive data. In reality, there may be sensitive information in the data sets which may lead to growing concerns about personal privacy and even national security. And data is considered as a valuable asset and a critical strategic resource increasingly. All these constraints greatly motivate federated learning (FL) [1], which enables multiple entities to collaboratively train a model under an orchestration service for immediate aggregation and store data locally. The data in FL may be generated at different contexts. This may lead the data distribution to be unbalanced or Non-IID. The data sets' scale and quality may be different. These may lead to different intermediate computation and communication cost for different parties. And data is a significantly important asset to organizations, so a nice FL scheme could stimulate and incent the parties with high-quality datasets to join the training to form a better model and guarantee their rewards that match their contribution in addition to privacy preservation. In this context, it is necessary to consider factors such as privacy protection, unbalanced/skewed data distribution, fairness, to form a closed-loop federated learning system (FLS) [2]. On the other hand, gradient boosting decision trees (GBDTs) has become a popular machine learning algorithm and has shined in many machine learning and data mining competitions [3, 4] as well as real-world applications for its salient results on classification, ranking, prediction, etc., (especially for tabular data mining task) [5]. And several works have studied the horizontal federated GBDT system

[6, 7]. They focus on training and publishing a single decision tree among multiple federated parties to compose the global ensemble model. But in these systems, there are still some challenges as follows:

(i) Balance of efficiency, learning accuracy, and privacy-preserving. In most of the existing schemes, each party trains a single decision tree, and then shares the tree with the next participating party [6, 8]. And the global communication cost of building each tree is a multiple of the corresponding trainer's data. Other schemes may adopt cryptographic methods or differential privacy [7]. Cryptographic methods may bring prohibitive overhead. And the accuracy is relatively lower in the existing federated GBDT scheme with differential privacy in skewed data distribution.

(ii) Contribution quantification. Many data owners may not actively participate in federated learning, especially when the data owners are enterprises rather than devices [9]. As mentioned previously, a nice FL scheme could stimulate parties with high quality datasets to join the training to train a better model and guarantee their rewards that match their contribution. It is also essential to prevent participants from inflating their contributions. Most of the existing schemes overlooked this and failed to provide an outstanding quantifying mechanism.

(iii) Accuracy measurement and verification. In the FL setting, there is no guarantee that all parties are honest and trusted. To tackle these issues, [6] proposed to use MAE to measure the accuracy, and [8] adopted the blockchain for verification. However, it leads to additional communication overhead to achieve higher accuracy. It is necessary to consider two factors in accuracy measurement: (1) whether the feature with the most information gain is correctly selected; (2) whether the samples are in the correct sorting position [10]. To the best of our knowledge, there is no effective solution to measure and verify the accuracy contribution of each party.

In response to the above challenges, we propose a closed-loop federated GBDT system FGBDT-Chain which consists of two components: FV-tree and FQ-chain. More specifically, FV-tree is our federated GBDT framework. And we combine FV-tree with blockchain organically and design FQ-chain to quantify the contribution logic on the smart contract to attain a decentralized verification and auditability. Our scheme can achieve a relatively better balance of efficiency, learning accuracy, and privacy-preserving in skewed distribution of data. Particularly, it can also quantify parties' contribution for the global model, provide a value-driven incentive mechanism that encourages parties with different data sets to be honest, and suit to large-scale datasets.

Our contributions can be summarized as follows:

(1) We propose FV-tree, a federated GBDT framework that can achieve constant communication cost and less precision loss in skewed distributed data. FV-tree is based on the data-parallel algorithm of the decision tree to find the global top-2 candidate features and utilizes private spatial decomposition (PSD) to capture other parties' distribution and refits gradients to vote on the local most informative feature. We also design a scalable differential privacy mechanism in this process to enhance privacy-preserving.

(2) We design a contribution quantifying mechanism with a metric, namely, *split Shapley value* and a decentralized verification endorsement mechanism, namely, FQ-chain, which can reach a relatively fair and auditable federated GBDT. It can encourage and incent organizations with different datasets to train a better model.

(3) We implement the system FGBDT-Chain in a permissioned blockchain environment and conduct a comprehensive experiment on public datasets. The results show that FGBDT-Chain has high performance and can meet the practical application, especially for large-scale datasets.

The rest of the paper is organized as follows. Section 2 reviews the related work about federated GBDT systems. Section 4 introduces the design outline of our system. The technical details of FV-tree and FGBDT-Chain are introduced in Section 5. Section 6 presents the performance evaluation of our system in terms of accuracy and fairness. We give a brief discussion and analysis in Section 7. Section 8 summarizes the paper and puts forward the potential research directions in the future.

2. Related Work

In this section, we review the literature on the federated GBDT and fairness in federated learning.

2.1. Federated Gradient Boosting Decision Tree. Gradient boosting decision tree (GBDT) and its effective implementations such as XGBoost [3] and LightGBM [4] are widely used machine learning both in industry and academic applications [5, 11, 12]. In distributed GBDT, the training data is located in different machines and should be partitioned according to the sample level. Generally, the local histograms of features are broadcasted to all the parties to obtain the global distribution. Then each party chooses the most informative splitting points [13]. Among them, the parallel voting decision tree (PV-tree) [14] is a representative scheme. It performs full-granular histogram communication according to the features selected by each machine, then calculates the global split point. PV-tree can achieve a very low communication cost (independent of the total number of features/samples) in the context of uniform data distribution and has great scalability in the context of large datasets.

In recent years, with the growing concerns about data security and privacy, several horizontal federated GBDT systems have been developed. [6] designed a distributed GBDT scheme, in which each party trains a differential

privacy decision tree and uses Mean Absolute Error (MAE) to evaluate the accuracy of each decision tree. [8] took a similar approach and extended this learning process to the blockchain. However, in these tree-based sharing schemes, the quality of the shared tree is low. To solve this problem, [7] proposed Sim-FL, in which, each instance gathers similar instances' gradients of other parties through a local sensitive hash (LSH) to learn the distribution of other parties. This weighted gradient boosting strategy can significantly improve the accuracy of each decision tree, and achieve a primary level of privacy protection. Unfortunately, the communication overhead in each iteration is proportion to the number of local instances in the training party, which is not feasible in large-scale datasets learning. Intuitively, we summarize the existing federated GBDT system and compare them with our scheme in Table 1.

2.2. Fairness in Federated Learning. Many data owners may not actively participate in federated learning, especially when the data owners are enterprises rather than devices [9]. Therefore, the fairness of the federated learning system needs to be taken into account. In the existing federated learning research, fairness is mainly realized through an incentive mechanism. There are two main ideas: (i). All parties enjoy a global model; (ii). According to the contribution of parties, parties get different model rewards [15].

The goal of incentive mechanism is to make the party get a reward commensurate with its contribution. A number of literature focused on designing incentive mechanisms by clients' resources [16] and reputation [17]. Whereas, we concentrate on the incentive mechanism based on the contribution of data quality. Because data quality is a key factor that affects the model. In the scheme based on data quality contribution, Shapley value [18] has a wide range of applications, and [15, 19, 20] studied the Shapley Value of the data point contribution during ML training. In the training process of federated learning, [21] proposed to record the intermediate results (i.e. gradients and models), and then use them to reconstruct the model for approximate the contribution indexes. This approach is efficient and feasible in horizontal federated learning. Unfortunately, there is an essential difference between gradient-based distributed GBDT and Gradient Descent-based algorithms. Because reconstructed models are not always useful and internal nodes will not affect the prediction score. Therefore, we need a new contribution measurement mechanism for the scenario without an intermediate model.

In addition, some works use blockchain technology to record the training milestones of clients and ensure the security of the incentive mechanism [22–24]. These works do not promise a good balance of privacy-preserving, efficiency, and learning accuracy to form a practical federated GBDT.

3. Preliminaries

3.1. GBDT. GBDT is an ensemble model of sequential training for several decision trees. In each iteration, the following objective function is minimized to fit the residual of previous learners [25]:

$$\tilde{\mathscr{L}}^{(t)} = \sum_{i}^{n}\left[g_i f_t(\mathbf{x}_i) + \frac{1}{2} f_t^2(\mathbf{x}_i)\right] + \Omega(f_t), \tag{1}$$

where $g_i = \partial_{\hat{y}^{(t-1)}} l(y_i, \hat{y}^{(t-1)})$ is first-order gradient and $\Omega(f)$ is a regularization term. Let $I = I_L \cup I_R$, where I is the instance set of the father node, I_L and I_R are the instance sets of left and right nodes after a split. The gain of a split point is given by:

$$G(I_L, I_R) = \left(\frac{\left(\sum_{i\in I_L} g_i\right)^2}{|I_L| + \lambda}\right) + \left(\frac{\left(\sum_{i\in I_R} g_i\right)^2}{|I_R| + \lambda}\right). \tag{2}$$

To reduce the computational complexity of traversing all feature values, histogram-based algorithms like [4, 26] use discrete bins to find the approximate optimal split. The detail of the histogram-based algorithm as shown in Algorithm 1.

3.2. Private Spatial Decompositions (PSD). Generally, any dataset with ordered attributes or moderate to high cardinality (e.g. numerical features such as salary) can be considered as spatial data. In addition, if a dataset can be indexed through a tree structure (such as a B-tree, R-tree, kd-tree etc.), it can be implicitly treated as spatial [27]. Formally, a spatial decomposition is a hierarchical (tree) decomposition of a geometric space into smaller areas/hyperspaces, with data points partitioned among the leaves. Indexes are usually computed down to a level where the leaves either contain a small number of points, or have a small enough area, or a combination of the two. There have been many approaches to spatial decompositions. Some are data-independent, such as quadtrees which recursively divide the data space into equal quadrants. Other methods, such as the popular kd-trees, aim to better capture the data distribution, and they are data-dependent. [27] gives a full framework for privately representing spatial data. We use the PSD to share a coarse distribution summary with other data owners. And it is both used in collaborative learning and calculation verification under statistical heterogeneity scenarios.

3.3. Blockchain. Blockchain [28] is a kind of chained data structure that combines data blocks in order according to time sequence. The append-only data are ensured that they are tamperproof and unforgeable through cryptographic primitives. The main advantages of blockchain are decentralization, security, transparency, and traceability. Hyperledger Fabric [29] is a popular and efficient enterprise-level permissioned blockchain framework. And Fabric also realizes the modularization of consensus mechanism, authentication, and other components, which is more suitable for business cooperation between enterprise organizations. In summary, the fabric can provide a decentralized trust environment for a group of organizations to carry out complex business transactions for collaborative GBDT training tasks.

Table 1: Compare with existing federated GBDT systems.

	Accuracy[1]	DP[2]	Shared information[3]	Communication efficiency[4]	Blockchained[5]
[6]	×	✓	Model	✓	×
[8]	✓	×	Model + gradients	×	×
[7]	×	✓	Model	✓	✓
Our scheme	✓	✓	Gradients	✓	✓

[1] The accuracy of federated GBDT model performance well in skewed data distribution. Notice: "×" representative does not meet the requirement, "✓" meets the requirements. [2] The system has differential privacy extensibility. [3] The system's communication architecture, especially the shared training information in federated GBDT training. [4] The communication cost is independent of the number of samples in the local dataset. [5] In the absence of a third-party server (none of the above systems need it), the blockchain is used as an autonomous platform to coordinate the training process.

Input: I: instance set of the current node, F:feature set.
Output: bestSplit.
forall f in F **do**.
H ← new Histogram();
forall x in I **do**.
bin ← x[f].bin;
H[bin].g ← H[bin].$g + x$.gradient;
H[bin].n ← H[bin].$n + 1$;
forall bin in H **do**.
leftSum, rightSum = CalSumFromSplit(bin);
split.gain = SplitGain(leftSum, rightSum);//(2) ;
bestSplit = ChoiceBetterOne(split, bestSplit);
return bestSplit.

Algorithm 1: FindBestSplit.

4. The FGBDT-Chain Framework

This section describes the overall design of FGBDT-Chain, including the design objectives and system overview. We adopt the general assumption of federated learning, in which one model requester publishes a model request and multiple parties participant in the collaborative learning task. The problem description is included in Section 3-A. The system summary is shown in Section 3-B. The main symbols used in this paper are given in Table 2.

4.1. Design Objectives. We assume that there are M parties, and each party is denoted by P_m ($m \in [1, M]$). We use $I_m = \{(x_i^m, y_i^m)\}$ to denote the instance set of P_m, where $x_i^m \in \mathbb{R}^f$, $y_i^m \in \mathbb{R}$. We focus on the collaborative training of GBDT model, in which M parties (data owners) include one requestor cooperate to implement a federated GBDT training task. For example, as shown in Figure 1, due to the different distribution of patients, two private hospitals P_1, P_2 may prefer accurate test predictions for female and young patients, respectively [15]. Without relying on unrealistic public datasets and third-party central servers, they hope to achieve peer-to-peer collaborative learning and obtain high-quality models in a trusted environment. More importantly, they need to be guaranteed that they can get rewards corresponding to their own contributions. Out of this assumption, our federated GBDT system tries to meet the following three objectives:

(i) Model accuracy and efficiency. It is the basic requirement of all parties to build a high-quality global model in multiple skewed data sets. In addition, the geographical distance between parties may be far away, and the intermediate process can be stored in blockchain for the sake of fairness and security. The communication cost should be strictly reasonable. For this reason, we propose FV-tree, which can reduce the communication to a small range, and obtain good model performance in the case of skewed data distribution.

(ii) Fairness: As mentioned previously, data is considered a valuable asset and a critical strategic resource increasingly. In addition, participants need to invest tremendous of computation and storage in FL. Without any revenue, data owners may not voluntarily provide data and training resources. To encourage more parties to participate in a collaborative learning program, it is necessary to accurately calculate the cooperative contribution of each participant. We use the split gain generated by the party's updated gradients to calculate the split Shapley value of each party. In this way, we can fairly quantify the contribution of each party in the whole process, and provide the mechanism for the monetary reward of delayed payment.

(iii) **Security:** We assume that parties are curious, and they will not maliciously attack the federated model

Table 2: List of symbols.

Symbols	Meaning
P_m	m-th party in federated learning task;
M	Number of participating party;
I_m	Instance set of party P_m;
T	Number of decision trees in GBDT;
d	Maximum depth of decision tree;
Q	Total number of ensemble model split;
h_m	P_m's histogram of ordered gradients;
$\text{bin}_g, \text{bin}_n$	The sum of gradients and counts of each bin in one histogram;
gain^q	The split gain of q-th split in the GBDT model;
split^q	The split point of q-th split in the GBDT model, which includes the split feature and split threshold;
psd_m	Privacy spatial decomposition structure of P_m;
c_q^m, C_m	They represent the P_m's contribution index of the q-th split and the contribution index of total splits respectively;
ϕ_m^q	P_m's split indexes (split Shapley value) during the $q-th$ split;
κ_m^q	P_m's voting contribution indexes during the $q-th$ split;
$\mathbf{W^m}$	The P_m' distribution weight matrix;
$\mathbf{w^{m*}}$	The P_m' global distribution weight vector;
pk_m, sk_m	The P_m' key pair for signing and verification respectively;

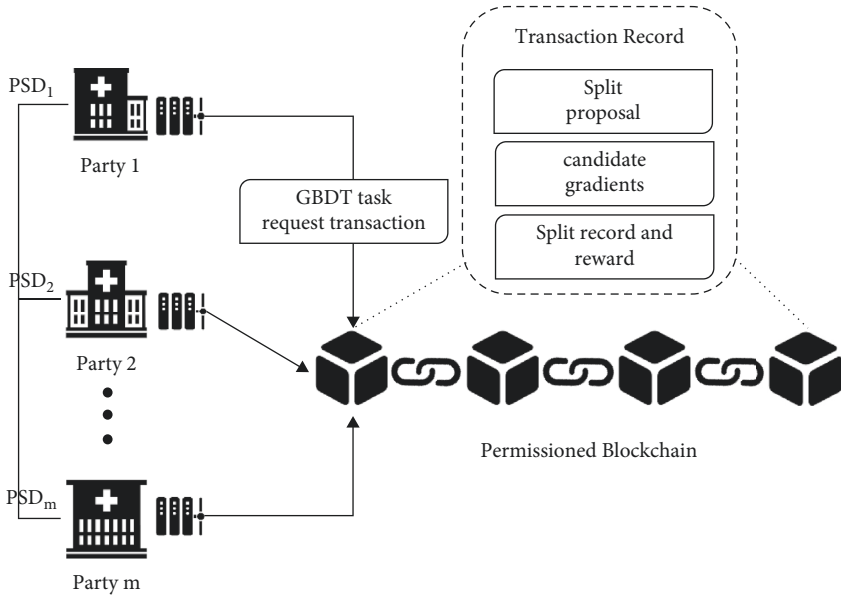


Figure 1: FGBDT-Chain system overview.

unless they can get higher income. This means that our system not only needs to avoid leaking the original data in the learning process but also needs to provide a necessary verification mechanism. We also have to eliminate the potential that greedy participants deliberately exaggerate contribution through updated information. Therefore, we propose FGBDT-Chain which can provide an extension of differential privacy, and a decentralized endorsement mechanism to filter distorted update information.

4.2. The Proposed Architecture. Our proposed system consists of two modules: permissioned blockchain module and federated GBDT module. The permissioned blockchain establishes secure connection channels among all nodes. FGBDT-Chain is based on the FV-tree training framework, which includes three stages: distribution preprocessing, features voting, and gradient histogram aggregation. Permissioned blockchain module includes four types of transactions: model request transaction, feature voting transaction, gradient histogram upload transaction, and contribution indexes allocation transaction. The contribution indexes assignment is implemented by smart contracts according to historical transactions. The stored information in the permissioned blockchain is shown in Figure 2.

Step 1. In the beginning, a model requester initializes the permissioned blockchain and specifies the requirements of the learning task, such as dataset requirements and model

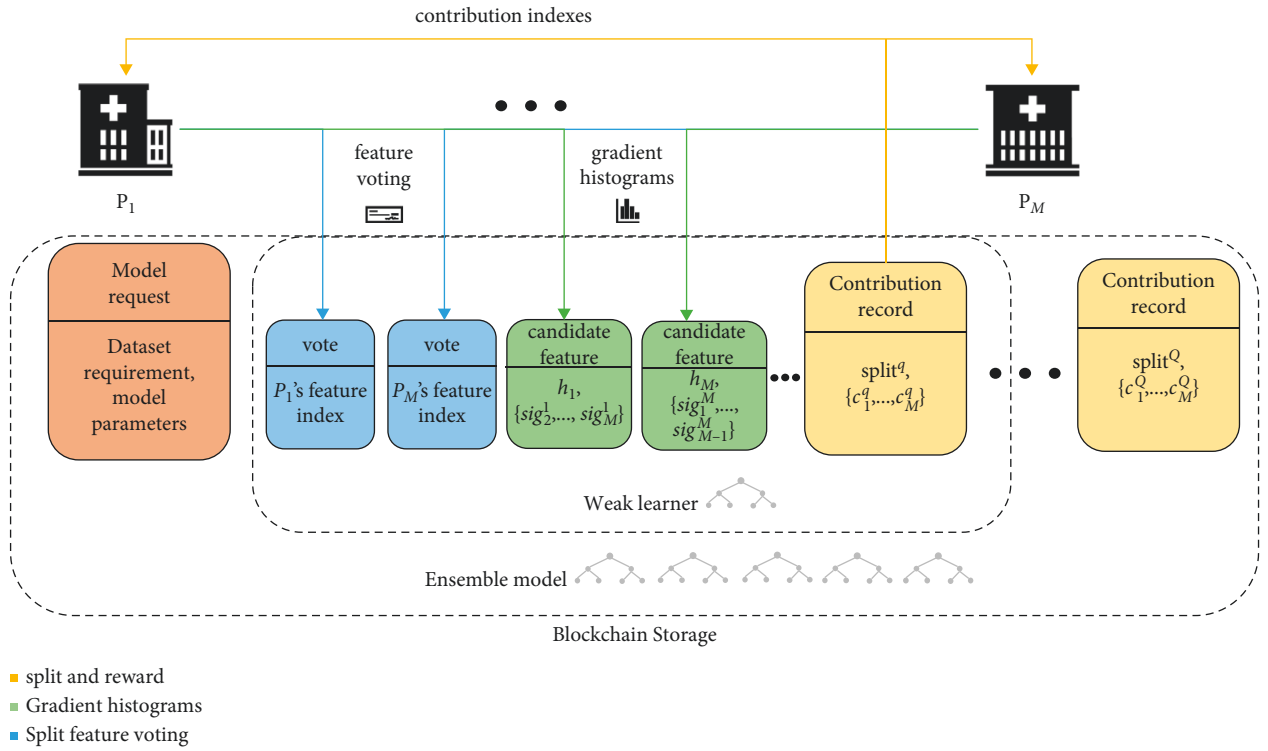


Figure 2: Blockchain-based ledger storage of FGBDT-Chain system.

parameters. Parties that wish to join the learning task or receive a request should be authenticated, then upload the rough distribution summary (i.e., PSD) of their datasets. The model requester has the right to refuse a party to become a federation member according to the observation of the distribution summary.

Step 2. After a specified number of organizations join the federated learning task, each party downloads all PSDs, and establishes the distribution matrix and global distribution vector. So far, the initialization work is completed.

Step 3. In the stage of collaborative training, each party uses the local dataset I_m and the global distribution vector to calculate the local most informative features and uploads the feature index through the voting transaction. At the same time, all parties can calculate the top-2 features with the highest number of votes as candidate features according to on-chained transactions.

Step 4. Parties broadcast the local original gradient histograms of candidate features. After one party receives most signatures corresponding to his histogram, the histograms and signature set are written into the transaction. With the help of the distribution matrix, the verification algorithm can detect malicious updates in skewed data distribution (Malicious update refers to the gradient histogram stretched by greedy participants to improve their contribution indicators).

Step 5. The smart contract will calculate the best split point and allocate contribution indexes according to the historical transactions. These two sub operations can be parallelized and the complexity is low. In addition, since the update records are stored in transactions, the contribution indexes can be calculated after the emergency task training process is completed.

The above 3–5 steps will form a loop that continues to execute until the stop training condition is met. When the learning task is finished, the federated GBDT model and parties' update/contribution records are stored in the blockchain's transactions. The whole learning process does not depend on any single party. In addition, because all the records created during the training of the decision tree are tamper-proof, the federated member can be audited at any time.

5. The Design Detail of FGBDT-Chain

FGBDT-Chain is a collaborative learning framework based on blockchain for GBDT. We will introduce the framework in two parts: FV-tree and FGBDT-Chain. Firstly, we will introduce the PSD-based preprocessing phase, which provides the basis for our framework (Section 5-A). Secondly, we will describe the GBDT training framework FV-tree in detail, which includes tree growth processes based on feature voting, gradient histograms publishing, and the expansion of differential privacy (Section 5-B). Finally, we introduce FGBDT-Chain's fairness assurance, including the fair guaranteed incentive mechanism based on a novel

contribution measurement algorithm, and the decentralized verification scheme on the blockchain (Section 5-C).

5.1. Preprocessing Stage. When a party receives the model request transaction, it first checks the dataset requirements and filters out the instances that meet the task description in the local instance, which is expressed as I_m. Then it starts the preprocessing operations. The main idea is to capture the data distribution of all other parties by generating a rough distribution matrix $W^m \in \mathbb{R}^{N_m \times M}$ and a global distribution vector $w^{m*} \in \mathbb{R}^{N_m}$. Where W_{ij}^m is the distribution weight of P_m's instance x_i^m in party P_j's instance set I_j, and w_i^{m*} is the distribution weight of the instance x_i^m in the global instance set I. In our scheme, w^{m*} is an optional term. When distributions are badly skewed, it will be used in the voting stage to select the most informative local feature (Section 5-B1), and W^m is used for verification subsequently (Section 5-C2).

More specifically, party P_m firstly calculates the psd_m by I_m, which has been well studied in previous research [27]. Let V_l^m be the value of l-th leaf in psd_m. Intuitively, the psd_m is a tree model represents the rough data distribution summary of P_m, where the value V_l^m is the number of instances corresponding to the hyper-space represented by the leave node l, and the count value V_l^m has been perturbed by differential privacy. Party P_m can upload psd_m with the blockchain's transaction, and download other parties' psd in the collaborative learning task. Then P_m maintains the distribution weight matrix W^m and the global distribution weight vector w^{m*}. The detail is shown in Algorithm 2. After party P_m downloads psd_j from P_j, it uses a local instance set I_m to query psd_j. Assuming that the query result of i-th instance (x_i^m, y_i^m) is l-th leaf in psd_j, then P_m pushes index i into the set S_l^j, where S_l^j is the set of P_m's instances falling in the hyperspace psd_l^j. After all instances have been queried, W^m can be assigned, where $W_{ij}^m = (|S_l^j|_{i \in S_l^j})/V_l^j$. Finally, after calculating the distribution vectors $W_1^m, \ldots, W_j^m$ of all other participants, P_m will further assign the global distribution vector w^{m*}, as follows:

$$w_i^{m*} = \delta \sum_{j=1}^{M} \left(W_{ij}^m \times \frac{N_j}{N} \right), = \delta \sum_{j=1}^{M} \left(W_{ij}^m \times \frac{\sum_{l=1}^{L_j} V_l^j}{\sum_{k=1}^{M} \sum_{l=1}^{L_k} V_l^k} \right), \tag{3}$$

where δ is a parameter of fitting distribution degree, N and N_j denote the number of instances of global and party P_j respectively, which is got from the accumulated leaves' value of different psd s. In addition, N_j/N represents a fitting budget of P_j. The more instances a party has, the larger fitting budget needs to be allocated. For Algorithm 2, we have the following observations. Firstly, the calculation of PSD only needs one time, and the distributed structure of tree model will greatly reduce the communication cost compared with the approach of sending each sample hash [7]. Secondly, the structure of psd s can be different, which means parties do not need to communicate in advance to use a unified structure of psd. In other words, parties can choose any tree model or inner nodes, whether it is a quad-tree or a kd-tree. It will not affect other parties to generate their weight matrix.

5.2. FV-Tree. When the local weight matrix W^m and global weight vector w^{m*} are established, parties can start to enter the training stage. In the training phase, each party does not train a complete tree, instead, it sends minimal update information. There are two types of update information: (i) parties' split feature voting and (ii) gradient histogram of candidate feature which is used to calculate global split points. In each node split, parties calculate the split feature with the most informative gain locally and vote on it. The top-2 features with majority votes in the global voting will become candidate features, and then parties send the gradient histograms of them. According to the above two kinds of update information, each party can update the global GBDT model synchronously.

However, this method may produce errors due to the split feature may be not globally optimal, especially in the context of decentralized data owners with different distributions/sizes. So, we consider *gradient refit* to alleviate this problem. The basic idea of gradient refit is to adjust gradients according to the global weights of the instances, then calculate the most informative feature according to the refitted gradients. When the global candidate features are selected, the two local original histograms are sent. The details of FV-tree are shown below.

At the beginning of an iteration, party P_m has a local instance set I_m, and the global distribution weight vector w^{m*}. First, P_m updates gradients and synchronizes the split information of each new node. Details are shown in the Algorithm 3 and Figure 3. For each new node generated in the decision tree, P_m calculates the local split gain of all the split points. The split gain is calculated as follows:

$$G_m(I_L, I_R) = \frac{\left(\sum_{i \in I_L} g_i^m w_i^{m*} \right)^2}{\sum_{i \in I_L} w_i^{m*} + \lambda} + \frac{\left(\sum_{i \in I_r} g_i w_i^{m*} \right)^2}{\sum_{i \in I_R} w_i^{m*} + \lambda}. \tag{4}$$

When the local split point with the highest split gain is selected, party P_m will publish the corresponding feature's index as a vote. And after receiving all the local votes, every party can sort features according to the number of votes. So far, each party can get the ranking of the same features, then select the top-2 features as candidate features, and upload the corresponding gradient histograms. It should be noted that the original uploaded gradients histogram is not the fitted one. After receiving the histograms from other parties, each party will traverse all the split points in the aggregated histograms to find the best split with the highest split gain. The gain of each split point is calculated as follows:

$$G_{\text{global}} = \frac{1}{2} \left[\frac{\left(\sum_{m=1}^{M} \sum_{i \in I_L^m} g_i^m \right)^2}{\sum_{m=1}^{M} \left| I_L^m \right| + \lambda} + \frac{\left(\sum_{m=1}^{M} \sum_{i \in I_R^m} g_i^m \right)^2}{\sum_{m=1}^{M} \left| I_R^m \right| + \lambda} \right], \tag{5}$$

where, $\sum_{i \in I_L^m} g_i^m$, $\sum_{i \in I_R^m} g_i^m$, $\sum_{m=1}^{M} |I_L^m|$, and $\sum_{m=1}^{M} |I_R^m|$ are calculated from the aggregated histograms. When the node

Input: PSD model set psd_1, psd_2..., psd_M, instance set I_m
Output: distribution weight matrix: W; global distribution vector: w^*
//establish distribution weight matrix
for $j \leftarrow 1$ to M **do**
for $i \leftarrow 1$ to N_m **do**
S $\leftarrow \text{psd}_j$.getLeafNode((x_i, y_i));
S.push(i);
//set hyperspace′s weight to matrix W
for $l \leftarrow 1$ to L_j **do**
S_j.weight $\leftarrow |S_l^j|_{i \in S_l^j} / V_l^j$;
forall i in S_l^j **do**
W [i][j] $\leftarrow S_l^j$.weight;
//establish global distribution vector
for $i \leftarrow 1$ to N_m **do**
for $j \leftarrow 1$ to M **do**
w^* [i] += W [i][j] $\times N_j / N$;
return W, w^*;

Algorithm 2: FVtree:DistributionMatrixEstablish.

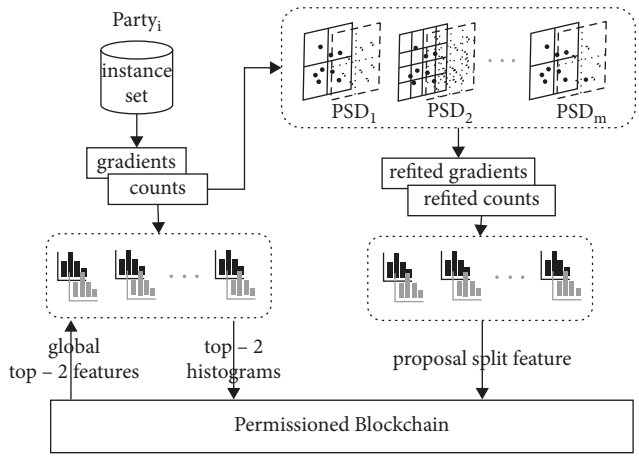


Figure 3: Training process of FV-tree.

Input: local gradients $g_1, \ldots, g_{N_m}$, global distribution weight vector w^{m*}
Output: bestSplit
localHistograms = ConstructHistograms($g_1, \ldots, g_{N_m}$);
localRefittedHistograms = ConstructHistograms($g_1, \ldots, g_{N_m}$, w^{m*});
//Local Voting
forall H in localRefittedHistograms **do**
splits.Push(H.FindBestSplit())//For details in Algorithm 1;
localVote = Max(splits).getFeatureID();
uploadVote(localVote);
//Global Voting
featureRanking ← gather other parties' localVote;
globalCandidate = featureRanking.Top2ByMajority();
uploadHistograms(globalCandidate, localHistograms);
//Merge global histograms
globalHistograms ← gather other parties' localHistograms;
bestSplit = globalHistograms.FindBestSplit();
return bestSplit;

Algorithm 3: FV-tree:FindBestSplit.

reaches the max depth, it becomes a leaf node and the value is calculated through the following equation:

$$\text{Value} = -\left(\frac{\sum_{m=1}^{M}\sum_{i\in I_m} g_i^m}{\sum_{m=1}^{M}|I_m| + \lambda}\right). \tag{6}$$

In the training process of FV-tree, a participant needs to update information from other parties to split none-leaf node, and the value of a leaf node is directly generated by the histograms of its parent node. So, we only need to allocate the privacy budget to the none-leaf nodes. In the communication process of FV-tree, local feature voting and histograms aggregation may lead to privacy leakage. For the local best split point selection, the information gain is used as the utility function, and the exponential mechanism is used to return the split point with the largest gain value. Let g^* be the gradient with the largest absolute value. By introducing the conclusion of previous work [13], the sensitivity is $\Delta G = ((3\lambda + 2)/((\lambda + 1)(\lambda + 2)))g^*$. Before updating histograms, the count of each bin is perturbed by Laplace noise [14]. The sensitivity of the gradient histogram is $2g^*$, and the sensitivity of the count histogram is 1. To maintain the effectiveness of boosting, we use the two-level boosting structure (EOE) to allocate the privacy budget for multiple decision trees [13], and our method satisfies the ϵ-differential privacy.

Proof. Assume that the privacy budget of a tree is ϵ_t, and the max depth of a decision tree is d. Since the nodes in one depth have disjoint inputs according to the parallel composition, each instance will go through at most $d - 1$ times node split. Further, each split will be regarded as five queries, namely, the best split feature voting and twice gradient histograms and count histograms updating respectively. The privacy budget for each split is $\epsilon_{\text{split}} = (\epsilon_t/5(d-1))$. Thence, the privacy budget of a single decision tree satisfies ϵ_t-differential privacy. In EOE, if there are a total of E ensembles, the privacy budget of each tree is $\epsilon_t = \epsilon/E$, and the whole FV-tree training process satisfies ϵ-differential privacy.

In summary, our scheme leverages voting split features and updating gradient histogram to make a tradeoff between accuracy, communication cost and security, and we give a brief discussion in section 7-A. □

5.3. FGBDT-Chain. To attract more institutions with high-quality data into the federal learning task, it is necessary to quantify the contribution of each party fairly and provide incentive mechanisms according to the contribution index. A widely used approach is to quantify the contribution of each participant's local model [9]. However, it is infeasible when the local model does not exist. For example, in our FV-tree scheme, there is no local model, and split points are decided by all parties. We should design a new approach and mechanism to quantify the contribution of federated parties. We first define the fairness of the federated GBDT task.

Definition 1. (Collaborative fairness in GBDT) In a collaborative GBDT learning task, multiple parties train a global model together. The party that provides more valuable information for the global model will get a higher contribution index. Specifically, fairness can be measured by the parties' split gain.

We define what is valuable information as follows.

Definition 2. (Valuable information in gradient-based collaborative GBDT): Suppose party P and P' participate in distributed GBDT learning. Once the global best split point is determined, we can informally say that party P provides more valuable information than P', if the gradients submitted by P bring more split gain than the gradients submitted by P' on the global split point.

The growing process of decision tree is to constantly find the split point which can bring the maximum split gain. The split gain provided by party's update information for the global model can reflect the corresponding contribution because split gain represents the reduced uncertainty in the selection process of the split point. Formally, let $C \triangleq \{P_1, \ldots, P_M\}$ denote a set of M parties. We call a subset B a coalition of parties if $B \subseteq C$. The histogram vector of $P_m \in B$ is represented by h_m, coalition B's histogram set is denoted by HB. And we denote the best splitting point as split^q, the global gain of split^q is G^q. Then, we define the utility function U_B:

$$U_B \triangleq G(\text{HB}; \text{split}_q) = \frac{\left(\sum_{m\in B}\sum_{bin\in h_L^m}\text{bin}_g\right)^2}{\sum_{m\in B}\sum_{\text{bin}\in h_L^m}\text{bin}_n + \lambda} + \frac{\left(\sum_{m\in B}\sum_{bin\in h_R^m}\text{bin}_g\right)^2}{\sum_{m\in B}\sum_{bin\in h_R^m}\text{bin}_n + \lambda}. \tag{7}$$

The above equation is the histogram form transformed from (5). Where h_L^m/ h_R^m denote the set of bins on the left/right parts segmented by split^q, bin_g and bin_n denote the sum of gradients and counts in the corresponding bin respectively. According to the observation of (7), two properties fulfill the standard assumptions of cooperative game theory:

Property 1. Histogram of the empty coalition has no utility: $U_\varnothing = 0$;

Property 2. Histogram of any coalition $B \subseteq C$ has nonnegative value: $\forall B \subseteq C, U_B \geq 0$;

Proof. The above two properties can be proved simply. For Property 1, when $B \subseteq \varnothing$, each bin in HB equals 0, so the $G(\text{HB}; \text{split}_q)$ equals 0. For Property 2, because $(\sum_{m\in B}\sum_{\text{bin}\in h_L^m}\text{bin}_g)^2 \geq 0$, and n_{bin} is a natural number, the minimum value of U_B is $(0/\lambda) + (0/\lambda) = 0$.

To guarantee that the histograms' contribution measurement is fair to all M parties, we use Shapley Value, which is the unique value division scheme that satisfies symmetry, null player, additivity, and efficiency properties. Next, we define the contribution of a federated party in a single split: □

Definition 3. (Split Shapley value) In the q-th node split split^q of federated GBDT model, given a utility function $U \triangleq G$ where G is the split gain function of GBDT algorithm,

and a histogram set $HC \triangleq \{h^m\}_{m\in\{1,M\}}$, the split Shapley value of a federated party $P_m \in C$ is defined as:

$$\phi(h_m; U, \mathrm{HC}) \triangleq \frac{1}{M}\sum_{j=1}^{M}\frac{1}{\binom{M-1}{j-1}}\sum_{\mathrm{HB}\subseteq \mathrm{HC}\setminus\{h_m\}:\ |\mathrm{HB}|=j-1}\left(U\left(\mathrm{HB}\cup\{h_m\};\mathrm{split}^q\right)-U\left(\mathrm{HB};\mathrm{split}^q\right)\right). \tag{8}$$

For simplicity, we use ϕ_m^q denotes the split Shapley value of P_m at the q-th splitting, it can be called as split contribution index.

In addition to the split contribution, the voting contributions are required to encourage parties to choose the most informative features. In the q-th split, the voting contribution κ of P_m is defined as:

$$\kappa_m^q = \begin{cases} 0, & \text{If } P_m\text{'s vote hits the split feature,} \\ G_q, & \text{If } P_m\text{'s vote does not hit the split feature,} \end{cases} \tag{9}$$

Finally, the party P_m's total contribution index of the q-th splitting of the federated GBDT model is defined as c_m^q:

$$c_m^q = \alpha\kappa_m^q + \phi_m^q, \tag{10}$$

where κ_m^q is the voting contribution, $\alpha \in (0, 1]$ is a variable parameter that controls the voting contribution, and ϕ_m^q is the split contribution comes from Equation[eq_split]. When the federated GBDT model training is complete, the contribution of party P_m is $C_m = \sum_{q=1}^{Q} c_m^q$, where Q is the total number of split (number of nonleaf nodes).

In the previous section, we described in detail how to quantify the contribution of a party. However, it is a challenge to calculate C when there is no trusted third party because C is directly related to the interests of each participant. To ensure the security of the logic of contribution measurement, we use a smart contract to retrieve historical transactions and record the contribution of each party.

Even smart contract can achieve the security of computing process, due to the sensitivity of split Shapley value, greedy parties can get a higher split contribution ϕ by tampering with the local histograms. As a concrete example, it is shown in Table 3. Suppose two parties P_1, and P_2 submitted their local histogram transactions h_1 and h_2 where $h_1 = \{\{1, 2\}, \{10, 10\}\}$, $h_2 = \{\{-1, 1\}, \{10, 10\}\}$. For simplicity, let $\lambda = 0$, we can get G is 0.45, and split contribution ϕ of P_1 and P_2 was 0.375 and 0.075, respectively. However, if P_2 tampers with its gradient histogram h_2 by doubling the magnification, the global G increases to 0.85. Accordingly, the split contribution ϕ_1, ϕ_2 is changed to 0.275 and 0.575. It can be seen that P_2 has increased his split contribution a lot.

Based on the above analysis, it is necessary to verify the updated information in our system to maintain fairness. In federated GBDT, the only existing verification scheme is to use local datasets to measure the performance of the updated model [6, 8]. Because it is difficult to generate public validation data sets, this scheme is considered as a minimized method in the federated scenario [30]. We inherit this idea of using a local dataset as the basis of verification. However, we cannot directly use the performance of the model, the reasons are as follows: First, updating information in FV-tree is gradients rather than models. Using gradients to reconstruct a model requires additional calculation; Secondly, the verification of model quality cannot fundamentally solve the above problem, because the contribution value of a histogram will be significantly higher after it is stretched proportionally. But the quality of the model using the stretched histogram may not be much different from the original one. In response to the above problems, we take the histogram overlap degree as the verification algorithm, in which the histogram used for verification is constructed by the distribution matrix W and the local histogram h. And we integrate this method into the endorsement mechanism of the permissioned blockchain to implement the FV-tree's decentralized verification scheme.

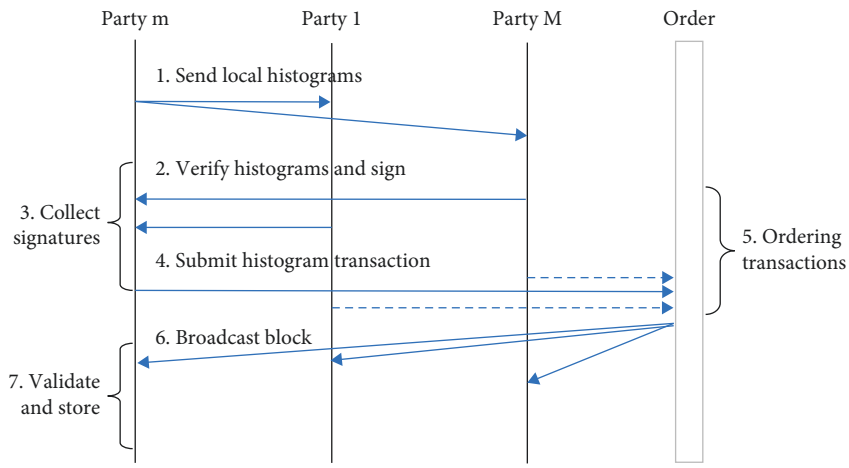
Specifically, as shown in Figure 4, before party P_m submits a histogram transaction, it first needs to broadcast the histogram h_m to other parties for signature. When $P_j \in C\setminus\{P_m\}$ received the signature request of h_m from P_m, the P_j' local gradients and the distribution vector W_m^j will be used to construct the refitted histogram h_m^{j*}, which denotes the histogram constructed by P_j to verify h_m. For P_j, there is only its histogram, which can simply denote as h_m^*. The details of this process are similar to Algorithm 1, except that $x \cdot \mathrm{gradient}, 1$ are replaced by $gW_{m,i}^j$ and $W_{m,i}^j$ in line 5 and line 6 respectively. Then h_m^* is used to calculate the overlapping degree with h_m:

$$\mathrm{Ver}\left(h_m, h_m^{j*}\right) = \sum_{bin\in h_m}\left(\frac{\left|\mathrm{bin}_g^m - \mathrm{bin}_g^{m*}\right|}{\max\left(\mathrm{bin}_g^m, \mathrm{bin}_g^{m*}\right)} + \sum_{bin\in h_m}\frac{\left|\mathrm{bin}_n^m - \mathrm{bin}_n^{m*}\right|}{\max\left(\mathrm{bin}_n^m, \mathrm{bin}_n^{m*}\right)}\right), \tag{11}$$

where bin_g^m, bin_n^m denote cumulative gradients and count respectively. The overlapping degree can verify the correlation of bin values and whether they are stretched. When the overlapping degree is less than the threshold, P_j will sign the histogram h_m, and send $\mathrm{sig}(h_m, sk_j)$ to P_m, where sk_j is a private key of P_j. When P_m obtains the signatures of most parties, it will write the histogram and signature set into the transaction and sends it to orderers, then the histogram transaction will be packaged into block.

Table 3: An example of the influence of local histogram h on split contribution ϕ.

	No tampering with histogram	P_2 tampered With his histogram
Local histogram h_1	$\{\{1, 2\}, \{10, 10\}\}$	$\{\{1, 2\}, \{10, 10\}\}$
Local histogram h_2	$\{\{-1, 1\}, \{10, 10\}\}$	$\{\{-2, 2\}, \{10, 10\}\}$
Gain of global split	0.45	0.85
P_1' split contribution ϕ_1	0.375	0.275
P_2' split contribution ϕ_2	0.075	0.575



3. Add the collected signatures to the histogram transaction, and commit the transaction when the number of signatures meets requirement.
7. When encountering a histogram transaction, verify the signature and check the number of signatures.

Figure 4: Blockchain-based histogram transactions working flow.

The above design is suitable for the overall architecture of our federated GBDT, which can detect the histogram with exaggerated contribution, and will not significantly affect the efficiency of the system. Firstly, to consider the data distribution of parties, we can avoid misjudging the correctly calculated update information as malicious by using the refitted histogram to a certain extent, and the stretched histogram can be easily discovered. For the efficiency of the verification scheme, the whole decentralized verification process is very similar to Fabric's high-level transaction flow [29]. The only difference is that the party uses the local data set under blockchain instead of simulating the execution of the smart contract. In addition, this process is also different from the processing method of Proof of Quality (PoQ) [8], where they suggest checking the quality of all models after block generation. If there is a malicious transaction, the block needs to be repackaged, which means retraining the whole GBDT model. In our scheme, orderers can filter out the transactions that are not recognized by the majority of participants when ordering transactions.

6. Implementation and Evaluation

6.1. Experiment Setup. We implement FV-tree based on LightGBM (https://github.com/microsoft/LightGBM). For PSD, we use a data-independent tree model. Each time of the PSD's node splitting, we randomly select a feature in the unused feature set and divide it according to the average of the global maximum and minimum values (the maximum and minimum values are specified in the task initialization transaction), we also treat the label as a feature. The maximum depth of PSD is 8, the maximum value of each leaf node is 500. Laplace noises are injected into the leaf nodes, where the privacy budget $\epsilon = 1$. For the GBDT model, the maximum depth of each tree is 8, the number of iterations is 500, the regulation parameter λ is set to 0.1, and the maximum number of bin in the feature histogram is 16 (more bin will bring higher accuracy, but this small accuracy difference is not significant for the federated GBDT framework).

We used three public datasets to evaluate our scheme (https://www.csie.ntu.edu.tw/cjlin/libsvmtools/datasets/), as shown in Table 4 And 75% of these datasets are used for training, the rest are used for testing. To allocate skewed local datasets, as the realistic scenario requires, we used the partition method of previous work [31], which allocates the datasets for each party according to the unbalanced ratio $\theta \in \{0, 1\}$. After allocation, half of parties got $(\theta * N_{\text{class0}})/M$ instances of class 0, and $((1 - \theta) * N_{\text{class0}})/M$ of instances of class 1, the other parties are just the opposite. This partition method well represents the data distribution in the federation scene. Specifically, in addition to label skewed, there is also feature skewed between local datasets [32]. As shown in Figure 5, we use kernel density estimation (KDE) to

Table 4: Dataset description.

Dataset	Cardinality	Dimension
a9a	32615	123
SUSY	1000000	18
HIGGS	1000000	28

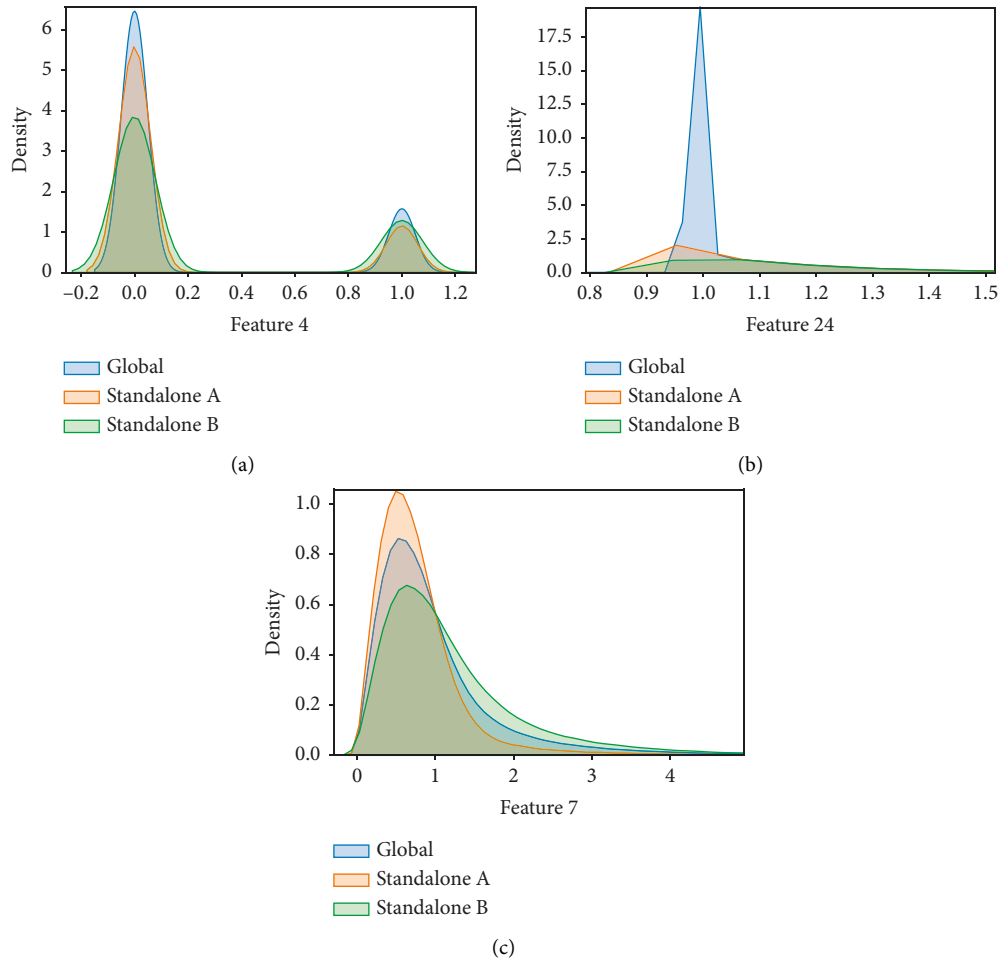


Figure 5: Compare feature distributions between local and global datasets by using Kernel density estimation (KDE). (a) a9a feature 4 (b) HIGGS feature 24 (c) SUSY feature 7.

intuitively show the skew degree of feature distribution between local and global datasets.

We compare our federated GBDT system with the other two frameworks: Standalone framework. This framework assumes that the parties training integration model only use their local dataset. The standalone setting shows the performance of the local training model of the party. In addition, there are two types of local dataset distributions in the unbalanced partition. We represent one part of the parties with more positive samples as Standalone A, and the other part as Standalone B. Centralized framework: This framework assumes that there is a trusted server accessing all parties' data, and uses global data to train the ensemble model without any privacy concerns. The centralized framework is high-precision, but it is hindered to implement in practice due to various restrictions. In addition, we also compare our scheme with other advanced federated GBDT frameworks in several same settings, such as TFL based on tree model communication and SimFL based on both tree model and gradients communication.

6.2. Experimental Results

6.2.1. Voting by Refitted Gradients. We first show the accuracy of FV-tree without considering differential privacy. To evaluate the effect of gradient refit, we compare FV-tree and PV-tree by convergence speed. Without losing generality, the number of parties is set to 4, and the ratio θ is set to 80%. The default parameters are used in all frameworks. The experimental results are shown in Figure 6. We can observe the following points. First, FV-tree performs better than PV-tree and Standalone models in all datasets. And because of

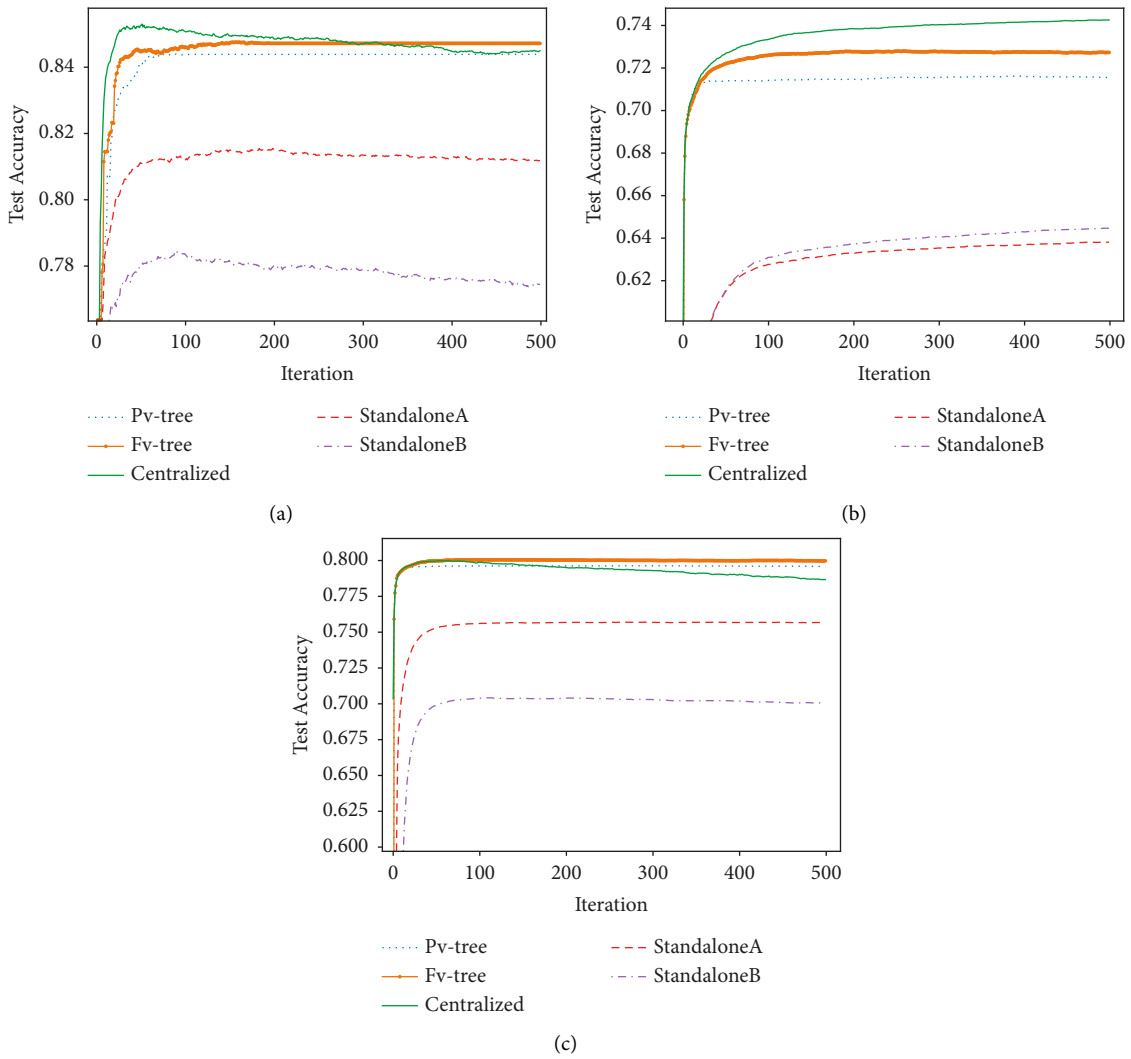


Figure 6: Comparison of the convergence speed, where the number of parties is set to 4, and the ratio θ is set to 80%. (a) a9a (b) HIGGS (c) SUSY.

the data skew, the accuracy of standalone mode is greatly reduced. This is because each party is affected by the data distribution bias in the learning process. And FV-tree uses a gradient to refit through PSDs, so it has a greater probability to select the most informative feature. Second, in the datasets a9a and SUSY, the centralized framework may lead to overfitting, while there is no such problem in the schemes based on FV-tree and PV-tree. Finally, the accuracy of PV-tree is significantly higher than the Standalone mode. This means that when considering differential privacy, we can get a tighter sensitivity without using the gradient refit.

6.2.2. The Impact of Unbalanced Ratio θ. To show the influence of different skew degrees on the FV-tree, we simply set the number of parties to 2. The experimental results are compared with SimFL, an advanced work without differential privacy. We observe the influence of different unbalanced distribution degrees on the prediction accuracy, as shown in Figure 7. We can observe that the accuracy of the standalone model decreases greatly with the skew of distribution. Secondly, although the accuracy of our framework and SimFL can be higher than local training when the unbalanced ratio is greater than 70%, FV-tree is much less affected than SimFL. This may be because the model accuracy is only affected by the feature selection in the FV-tree framework. While SimFL is affected by the feature selection and calculation of leaf weight. This means FV-tree is more suitable for skewed data distribution.

6.2.3. The Impact of the Number of Parties **M**. The number of different parties will also affect the accuracy of the model. We set a different number of parties when the unbalanced ratio θ is set to 80%. The experimental results are shown in Figure 8. Firstly, we can observe that FV-tree outperforms Standalone and SimFL in different number of parties settings, even the test error on dataset SUSY is less than that of over fitted centralized model. Secondly, with increasing number of parties, it does not have too much impact on FV-tree. This advantage may also come from the fact that FV-tree is not affected by the calculation of leaf weight.

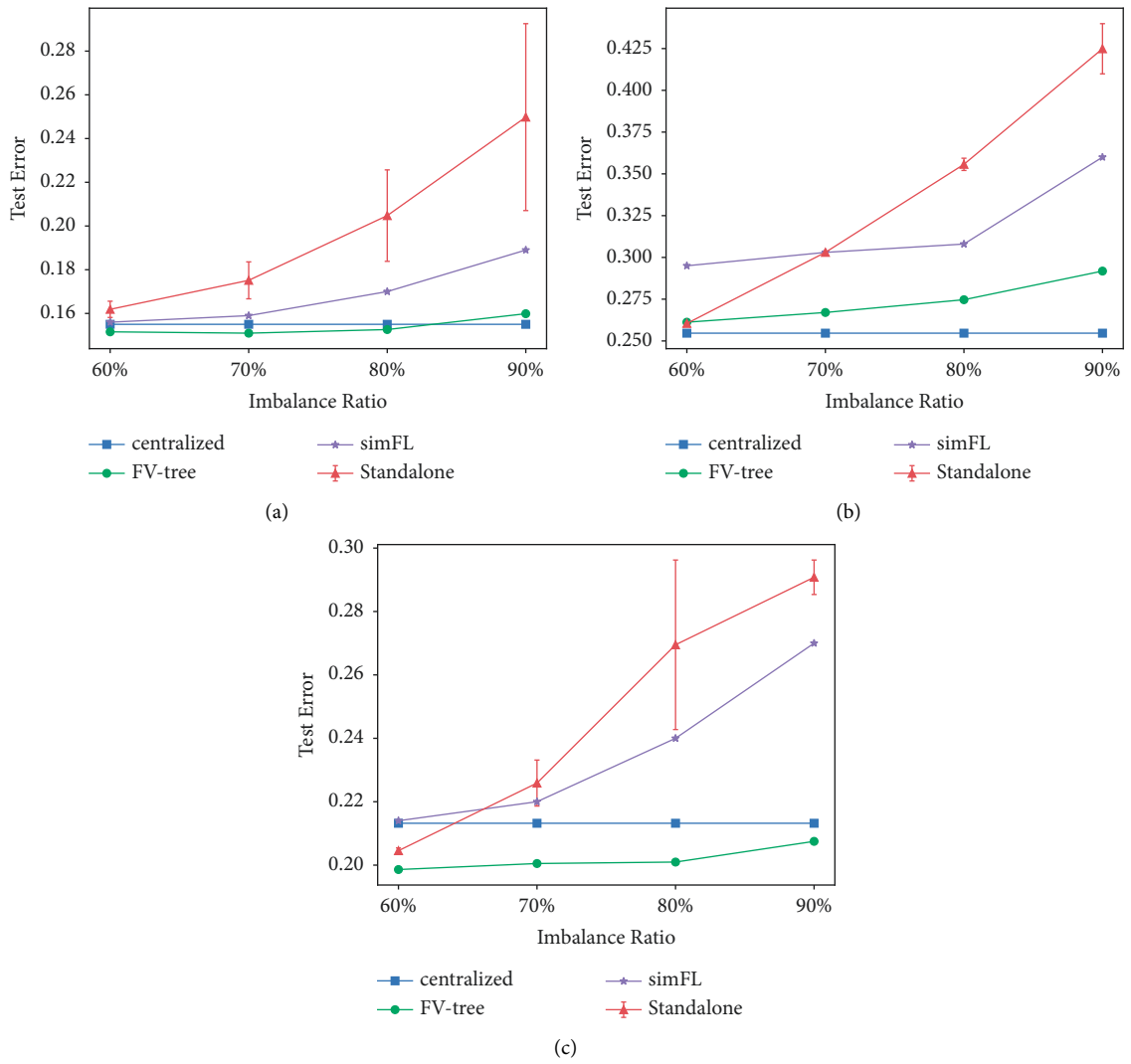


FIGURE 7: Comparison of the test errors given different unbalanced ratio θ, where the number of parties is set to 2. (a) a9a (b) HIGGS (c) SUSY.

6.2.4. The Impact of Differential Privacy. Based on the above experimental evaluation, FV-tree can achieve almost the same accuracy in distributed settings as centralized settings. Then, we test the FV-tree with differential privacy. Generally, we set the number of parties M to 4, and the unbalanced ratio θ is still set to 80%. To control the consumption of privacy budget, we set the maximum depth d of a single decision tree to 3. For dataset a9a, which has a small number of instances, is set as two ensembles, and each ensemble contains 20 trees. Dataset SUSY and HIGGS, which have a large number of instances, are set as one ensemble. To ensure a strict total privacy budget, PSD is not used. We evaluated the test error for different privacy budgets ϵ, as shown in Figure 9. Due to the randomness of differential privacy, we conducted 10 experiments and showed the maximum, minimum and average values (To be fair, the default parameter settings are still used in centralized and standalone models. Because there is no need to consider the consumption of the privacy budget, the iterations T and depth d can be increased to achieve higher accuracy).

We can observe that the accuracy of the FV-tree can still be higher than that of local training after using differential privacy on large-scale HIGGS and SUSY datasets. However, in the a9a dataset, due to the small amount of data, too much noise is added to the histogram, which reduces the accuracy of the model, but it is still comparable to the best training effect of local training. This means that our scheme has a good performance in large-scale datasets, and can meet the needs of practical applications.

7. Discussion

7.1. Accuracy Loss and Communication Overhead

7.1.1. Accuracy Loss. The accuracy loss of the FV-tree comes from the selection of the best split features. In the balanced data partition, we assume that the feature values of each dimension are i.i.d. uniform random variables, and assign the same number of instances to each party. Then, the possibility of selecting the best feature is as same as PV-tree

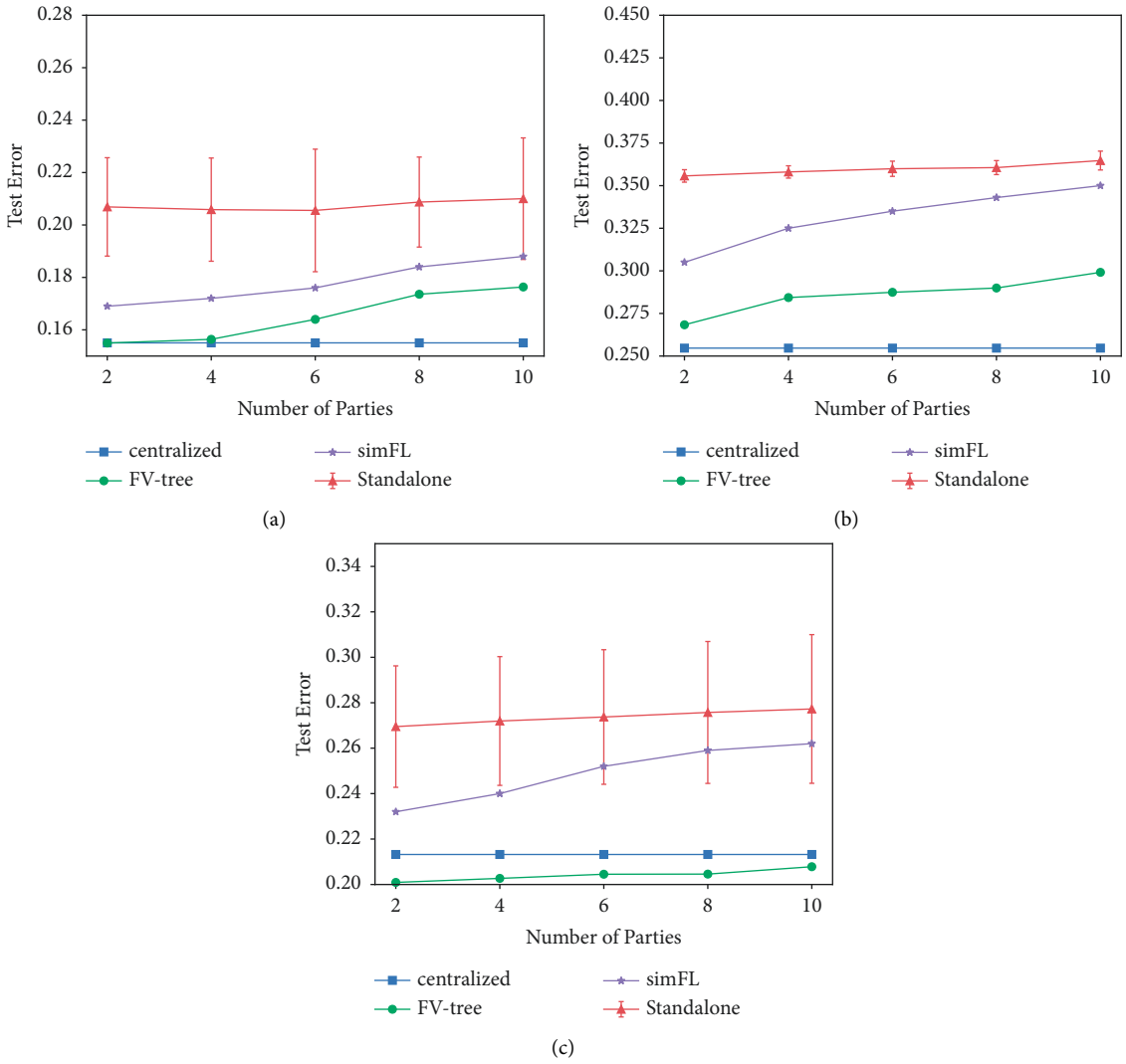


FIGURE 8: Comparison of the test errors given a different number of parties M, where the unbalanced ratio θ is set to 80%. (a) a9a (b) HIGGS (c) SUSY.

[33]. In the scenario of the skewed data partition, the experiment shows that FV-tree still has high accuracy. Moreover, in the case of significantly skewed data distribution, we can use the weight distribution calculated by PSDs to refit feature distribution, which can improve the possibility of selecting the best feature. However, the global distribution weight vector is used may cause high gradient values, which will make the privacy boundary loose. Under these circumstances, gradient cutting may be a feasible choice [34]. In addition, our scheme is not effective for small and continuous feature data sets. This obstacle is mainly due to adding a lot of noise to histograms, which reduces the effectiveness of the gradient histogram. Therefore, in small-scale dataset scenarios, we still need to use other federated GBDT frameworks.

7.1.2. Communication Overhead. The communication cost of our federated GBDT system is constant. First, in the pretraining phase, assuming that the depth of a PSD is d_{psd}, each party has to send one PSD model and receive $M-1$ PSD models, so the cost is $M(2^{d_{\text{psd}}}-1)$. In the training phase, assuming that there are T trees, and the depth of each tree is d, $2^{d-1}-1$ times node splitting is needed. Because each inner node needs to communicate three times, including one voting and two histograms uploading, where the voting communication is a real number. And the cost of a party sending $M-1$ times histogram to communicate histogram is $2(M-1)n_{\text{bin}}$. When two 2/3 of the signatures are received, the transaction can be sent. Let L_{sign} be the length of signature, then the cost of receiving the signatures is $2/3ML_{\text{sign}}$. In addition, they need to receive other parties' histograms and sign them, where the cost is $2(M-1)n_{\text{bin}}+(M-1)L_{\text{Sign}}$. Therefore, the communication overhead of a histogram aggregation is $(4M-3)N_{\text{bin}}+((7/3)M-1)L_{\text{sign}}$. Because there are T trees, the total communication overhead is $(2^{d-1}-1)[(4M-3)N_{\text{bin}}+((7/3)M-1)L_{\text{sign}}]T$, where d, M, L_{Sign}, T, N_{bin} are constants. So total communication cost of FV-tree is $\#O(1)$, which is less than other $\#O(|I_m|)$

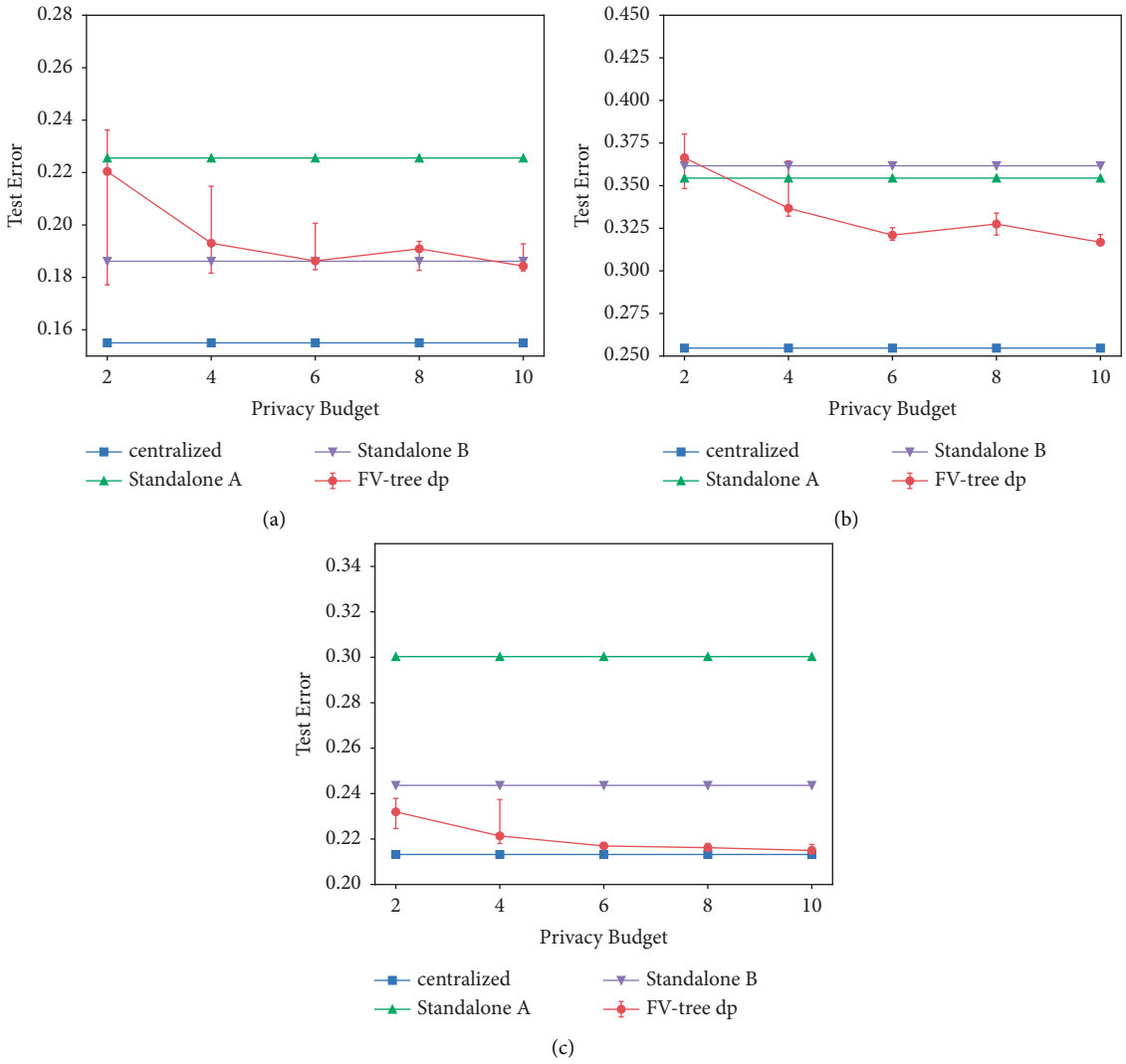


FIGURE 9: Comparison of the test errors given different total privacy budgets ϵ, the unbalanced ratio θ is set to 80%, where the maximum depth d of a single decision tree to 3. Dataset a9a is set as two ensembles, and each ensemble contains 20 trees. Dataset SUSY and HIGGS, are set as one ensemble with 50 trees. (a) a9a (b) HIGGS (c) SUSY.

federated GBDT framework [7]. In addition, the storage cost in the permissioned blockchain can reach an acceptable level to ensure fairness and tamper-proof.

7.2. Fairness and Efficiency. We regard the growth process of the decision tree as multiple cooperative games. Shapley value is used to measure the individual contribution in cooperation, the fairness of Shapley value is widely recognized. In our design, every node segmentation is fair, and the details can be obtained from Section 5-C. In addition, because the benefits obtained by the participants each time directly come from the gain value, it is also fair for the whole training process. For example, in the early stage of training, each split will produce a great gain, and each party will get more contribution value from it. On the other hand, the computational complexity of split Shapley value is acceptable. We can see only M is variable through (8), and in organization-cross federated scenes, M is usually a relatively small value. Besides, we do not need to traverse all the split points in histograms to calculate of U, because the global best split has been determined in split_q.

7.3. Security. It is assumed that all parties will aim at maximizing revenue and act honestly in the stage of voting characteristics because in the absence of any data of other parties, they can only choose the feature with the highest gain value to vote according to their real data to obtain voting awards. Similarly, in the phase of communicating gradient histogram, if the modified gradient histogram is detected, the histogram transaction cannot be published because of the need for a similarity test. Hence, a party can only get the histogram contribution reward if it publishes the real histograms.

Further, if there are malicious participants in the alliance, our system is still robust. Firstly, suppose that in the voting feature stage, if multiple malicious participants

conspire to select a feature f' with less gain to enter the global candidate features. At the same time, as long as one honest party selects another feature f, f' is still likely not to be the split point, because the gain value of f may be greater than it. On the contrary, if the gain value of f is less than f', it means that, f' is a good segmentation feature, and dividing nodes according to f', f' will not cause great harm to the model. Secondly, in the histogram aggregation stage, because the gradient histogram of the malicious party needs to be verified by two-thirds of the parties, it is necessary for the malicious parties involved in the conspiracy to reach two-thirds of the total number to make the histogram of the damage model accepted by the federation.

8. Conclusion

In this paper, we aim to present a closed-loop federated GBDT system. In our scheme, each party can get a good performance model and be allocated to a fair contribution index. At the same time, with the help of blockchain and decentralized verification mechanism, the calculation of the contribution index will remain secure, the results cannot be tampered with, and provide additional functions such as delayed payment or audit for any need. Besides, the communication overhead is constant which enables our method to fit federated GBDT tasks with large-scale datasets very well. Due to privacy constraints, this scheme may not be suitable for small-scale data sets, which is the direction we plan to study in our future work. [35].

Data Availability

The experiment source data used to support the findings of this study have been deposited in the https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/. And the experimental results 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 work was supported by the National Natural Science Foundation of China (no. U21A20474), the Guangxi "Bagui Scholar" Teams for Innovation and Research Project, the Guangxi Science and Technology Plan Projects (no.AD20159039), the Guangxi Young and Middle-aged Ability Improvement Project (no. 2020KY02032), and the Innovation Project of Guangxi Graduate Education (no. YCBZ2021038).

References

[1] H. B. McMahan, E. Moore, D. Ramage, and S. Hampson, B. A. Y. Arcas, Communication-efficient learning of deep networks from decentralized data," in *Artificial Intelligence and Statistics*, pp. 1273–1282, PMLR, 2017.

[2] Q. Li, Z. Wen, Z. Wu et al., "A Survey on Federated Learning Systems: Vision, Hype and Reality for Data Privacy and protection," 2019, https://arxiv.org/abs/1907.09693.

[3] T. Chen and C. Guestrin, "Xgboost: a scalable tree boosting system," in *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, San Francisco California USA, March 2016.

[4] G. Ke, Q. Meng, T. Finley et al., "Lightgbm: a highly efficient gradient boosting decision tree," *Advances in Neural Information Processing Systems*, vol. 30, pp. 3146–3154, 2017.

[5] A. Callens, D. Morichon, S. Abadie, M. Delpey, and B. Liquet, "Using random forest and gradient boosting trees to improve wave forecast at a specific location," *Applied Ocean Research*, vol. 104, Article ID 102339, 2020.

[6] L. Zhao, L. Ni, S. Hu et al., "Inprivate digging: enabling tree-based distributed data mining with differential privacy," in *Proceedings of the IEEE INFOCOM 2018-IEEE Conference on Computer Communications*, pp. 2087–2095, IEEE, Honolulu, HI, USA, April 2018.

[7] Q. Li, Z. Wen, and B. He, "Practical federated gradient boosting decision trees," in *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 4642–4649, NY, USA, February 2020.

[8] Y. Lu, X. Huang, Y. Dai, S. Maharjan, and Y. Zhang, "Blockchain and federated learning for privacy-preserved data sharing in industrial iot," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 6, pp. 4177–4186, 2020.

[9] Y. Zhan, J. Zhang, Z. Hong, L. Wu, P. Li, and S. Guo, "A survey of incentive mechanism design for federated learning," *IEEE Transactions on Emerging Topics in Computing*, vol. 10, no. 2, pp. 1035–1044, 2022.

[10] L. Zhao, Q. Wang, C. Wang, Q. Li, C. Shen, and B. Feng, "Veriml: enabling integrity assurances and fair payments for machine learning as a service," *IEEE Transactions on Parallel and Distributed Systems*, vol. 32, no. 10, pp. 2524–2540, 2021.

[11] J. Mateo, J. Rius-Peris, A. Maraña-Pérez, A. Valiente-Armero, and A. Torres, "Extreme gradient boosting machine learning method for predicting medical treatment in patients with acute bronchiolitis," *Biocybernetics and Biomedical Engineering*, vol. 41, no. 2, pp. 792–801, 2021.

[12] S. Lee, T. P. Vo, H.-T. Thai, J. Lee, and V. Patel, "Strength prediction of concrete-filled steel tubular columns using categorical gradient boosting algorithm," *Engineering Structures*, vol. 238, Article ID 112109, 2021.

[13] R. Kufrin, "Decision Trees on Parallel Processors," *Machine Intelligence and Pattern Recognition*, vol. 20, pp. 279–306, 1997.

[14] Q. Meng, G. Ke, T. Wang et al., "A Communication-Efficient Parallel Algorithm for Decision Tree," in *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS, Barcelona, Spain, December 2016.

[15] R. H. L. Sim, Y. Zhang, M. C. Chan, and B. K. H. Low, "Collaborative machine learning with incentive-aware model rewards," in *Proceedings of the 37th International Conference on Machine Learning*, pp. 8927–8936, PMLR, New York City, NY, USA, July 2020.

[16] Y. Zhao, J. Zhao, L. Jiang, R. Tan, and D. Niyato, "Mobile Edge Computing, Blockchain and Reputation-Based Crowdsourcing Iot Federated Learning: A Secure, Decentralized and Privacy-Preserving System," pp. 2327–4662, 2019, https://arxiv.org/abs/1906.10893%20.

[17] L. U. Khan, S. R. Pandey, N. H. Tran et al., "Federated learning for edge networks: resource optimization and incentive

mechanism," *IEEE Communications Magazine*, vol. 58, no. 10, pp. 88–93, 2020.

[18] H. W. Kuhn and A. W. Tucker, *Contributions to the Theory of Games*, Princeton University Press, Princeton, New Jersey, 1953.

[19] R. Jia, D. Dao, B. Wang et al., "Towards efficient data valuation based on the shapley value," in *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics*, pp. 1167–1176, PMLR, New York City, NY, USA, February 2019.

[20] A. Ghorbani, M. Kim, and J. Zou, "A distributional framework for data valuation," in *Proceedings of the International Conference on Machine Learning*, pp. 3535–3544, PMLR, New York City, NY, USA, February 2020.

[21] T. Song, Y. Tong, and S. Wei, "Profit allocation for federated learning," in *Proceedings of the 2019 IEEE International Conference on Big Data (Big Data)*, pp. 2577–2586, IEEE, Los Angeles, CA, USA, December 2019.

[22] H. Kim, J. Park, M. Bennis, and S.-L. Kim, "Blockchained on-device federated learning," *IEEE Communications Letters*, vol. 24, no. 6, pp. 1279–1283, 2020.

[23] L. Lyu, J. Yu, K. Nandakumar et al., "Towards fair and privacy-preserving federated deep models," *IEEE Transactions on Parallel and Distributed Systems*, vol. 31, no. 11, pp. 2524–2541, 2020.

[24] J. Weng, J. Weng, J. Zhang, M. Li, Y. Zhang, and W. Luo, "Deepchain: Auditable and Privacy-Preserving Deep Learning with Blockchain-Based Incentive," *IEEE Transactions on Dependable and Secure Computing*, vol. 18, no. 5, 2019.

[25] S. Si, H. Zhang, S. S. Keerthi, D. Mahajan, I. S. Dhillon, and C.-J. Hsieh, "Gradient boosted decision trees for high dimensional sparse output," in *Proceedings of the International Conference on Machine Learning*, pp. 3182–3190, PMLR, New York City, NY, USA, August 2017.

[26] K. Alsabti, S. Ranka, and V. Singh, "Clouds: a decision tree classifier for large datasets," in *Proceedings of the 4th knowledge discovery and data mining conference*, AAAI Press, New York, NY, August 1998.

[27] G. Cormode, C. Procopiuc, D. Srivastava, E. Shen, and T. Yu, "Differentially private spatial decompositions," in *Proceedings of the 2012 IEEE 28th International Conference on Data Engineering*, pp. 20–31, IEEE Computer Society, Los Alamitos, CA, USA, April 2012.

[28] S. Nakamoto, "Bitcoin: a peer-to-peer electronic cash system," *Decentralized Business Review*, Article ID 21260, 2008.

[29] E. Androulaki, A. Barger, V. Bortnikov et al., "Hyperledger fabric: a distributed operating system for permissioned blockchains," in *Proceedings of the Thirteenth EuroSys Conference*, pp. 1–15, Porto, Portugal, April 2018.

[30] Y. Li, C. Chen, N. Liu, H. Huang, Z. Zheng, and Q. Yan, "A blockchain-based decentralized federated learning framework with committee consensus," *IEEE Network*, vol. 35, no. 1, pp. 234–241, 2021.

[31] M. Yurochkin, M. Agarwal, S. Ghosh, K. H. Greenewald, T. N. Hoang, and Y. Khazaeni, "Bayesian nonparametric federated learning of neural networks,"vol. 97, pp. 7252–7261, in *Proceedings of the 36th International Conference on Machine Learning*, vol. 97, pp. 7252–7261, ICML, Long Beach, California, USA, June 2019.

[32] K. Hsieh, A. Phanishayee, O. Mutlu, and P. Gibbons, "The non-iid data quagmire of decentralized machine learning," in *Proceedings of the International Conference on Machine Learning*, pp. 4387–4398, PMLR, Shanghai, China, November 2020.

[33] Q. Meng, G. Ke, T. Wang et al., "A Communication-Efficient Parallel Algorithm for Decision Tree," 2016, https://arxiv.org/abs/1611.01276.

[34] Q. Li, Z. Wu, Z. Wen, and B. He, "Privacy-preserving gradient boosting decision trees," in *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 784–791, NY, USA, February 2020.

[35] C. Dwork, "Differential privacy: A Survey of Results," *Theory and Applications of Models of Computation*, vol. 4978, pp. 1–19, 2008.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 8998339, 18 pages
https://doi.org/10.1155/2022/8998339

Research Article

An Enhanced RFID-Based Authentication Protocol using PUF for Vehicular Cloud Computing

Vikas Kumar,[1] **Rahul Kumar**,[1] **Srinivas Jangirala**,[2] **Saru Kumari**,[3] **Sachin Kumar**,[4] **and Chien-Ming Chen**[5]

[1]*Department of Mathematics, SSV College Hapur, Hapur, Uttar Pradesh, India*
[2]*Jindal Global Business School, O. P. Jindal Global University, Sonipat, Haryana 131001, India*
[3]*Department of Mathematics, Ch. Charan Singh University, Meerut, U P, India*
[4]*Department of Computer Engineering, Ajay Kumar Garg Engineering College, Ghaziabad 201009, India*
[5]*College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao, Shandong, China*

Correspondence should be addressed to Chien-Ming Chen; chienmingchen@ieee.org

Received 8 March 2022; Revised 2 May 2022; Accepted 19 May 2022; Published 30 July 2022

Academic Editor: Jie Cui

RFID (radio frequency identification) is an Internet of Things (IoT) enabling technology. All physical devices can be connected to the Internet of Things thanks to RFID. When RFID is extensively utilized and fast increasing, security and privacy concerns are unavoidable. Interception, manipulation, and replay of the wireless broadcast channel between the tag and the reader are all possible security threats. Unverified tags or readers provide untrustworthy messages. IoT requires a safe and consistent RFID authentication system. PUFs are also physical one-way functions made up of the unique nanoscopic structure of physical things and their reactivity to random occurrences. PUF includes an unclonable feature that takes advantage of physical characteristics to boost security and resistance to physical attacks. We analyze the security of the RSEAP2 authentication protocol that has been recently proposed by Safkhani et al., a hash-based protocol, and elliptic curve cryptosystem-based protocol. Our security analysis clearly shows important security pitfalls in RSEAP2 such as mutual authentication, session key agreement, and denial-of-service attack. In our proposed work, we improved their scheme and enhanced their version using physically unclonable function (PUF), which are used by the proposed protocol in tags. This research proposes a cloud-based RFID authentication technique that is both efficient and trustworthy. To decrease the RFID tag's overhead, the suggested authentication approach not only resists the aforementioned typical assaults and preserves the tag's privacy, but also incorporates the cloud server into the RFID system. According to simulation results, our approach is efficient. Moreover, according to our security study, our protocol can withstand a variety of attacks, including tracking, replay, and desynchronization assaults. Our scheme withstands all the 18 security features and further consumes the computation cost as 14.7088 ms which is comparable with the other schemes. Similarly, our scheme consumes the communication cost as 672 bits during the sending mode and 512 bits during the receiving mode. Overall, the performance of our proposed method is equivalent to that of related schemes and provides additional security features than existing protocols. Mutual authentication, session key generation, and ephemeral session security are all achieved. Using the real-or-random concept, we formalize the security of the proposed protocol.

1. Introduction

Recognition technologies are deserving of our attention as they are both essential parts of the Internet of Things. Recognition of barcodes, optical characters, biometric identity, and magnetic card identification and contact IC card identification are all examples of traditional automated identification technologies. However, when employed in the IoT, they have a number of drawbacks. Bar codes, for example, can only hold a limited amount of data; optical character recognition is too expensive; biological recognition is flawed; and magnetic card and contact IC card

identification need intimate touch, which is inflexible. Currently, some of these identification methods are unable to protect personal information [1]. In contrast, RFID is a noncontact automatic identification technology that does not need mechanical or visual contact between the system and the target, and security protections can help keep user information private. Because of these advantages, RFID has emerged as one of the most promising IoT technologies [2].

An RFID system consists of RFID tags, RFID readers, and a database server. Tag-affixed objects are uniquely identifiable, and their identifying information is saved. They communicate with the reader using radio waves. In a typical RFID system, the database server is a local back-end server.

When RFID devices generate a large number of data, back-end servers' performance is limited. Cloud computing overcomes this problem in the IoT context. As a result, the integration of the cloud platform with the RFID system is required [2, 3]. RFID systems' reliability and data processing capabilities have dramatically enhanced since the introduction of cloud computing. Almost all of the data acquired by RFID sensors are processed on the cloud, which can aid in the resolution of issues such as data loss and latency [4]. In the IoT, the most commonly used public cloud servers are only semi-trustworthy. Because of the properties described above, the RFID system is vulnerable to attack. As a result, IoT necessitates the use of a secure and reliable RFID authentication system.

Similarly, a number of protocols based on physically unclonable functions (PUFs) have been proposed [12–14]. PUFs are, in reality, physical one-way functions derived from the unique nanoscopic structure of physical things (e.g., integrated circuits, crystals, magnets, lenses, solar cells, or papers) and their reactivity to random occurrences. The quirks in the manufacturing process of the items are responsible for the innate uniqueness of the structure and reactivity. It enables for both the unique identification and authentication of an object. Furthermore, it is considered that copying an object's PUF (and hence the object itself) is impossible, which might be seen as a security-by-design feature that prevents impersonation and cloning attacks. As a result, PUFs are regarded as a trustworthy and well-known physical security method for developing IoT authentication protocols. Physical devices are protected by PUF-based protocols, which are resistant to physical attacks and provide multilayer protection. Furthermore, even if the device is stolen, the attacker will not be able to use the PUF. However, the majority of proposed VANET solutions are still subject to different security concerns such as replay attacks, impersonation attacks, forgery attacks, and non-repudiation attacks. As a result, it is critical to build a viable VANET solution to address the existing issues.

2. Literature and Related Works

Several RFID authentication schemes have used elliptic curve cryptography in recent years (ECC). Due to the difficulties of resolving the discrete logarithm problem (DLP), ECCs have demonstrated their efficiency in assuring security and privacy. The state-of-the-art of ECC-based RFID, mobile computing, and VCC authentication protocols are reviewed in this section and are shown in Table 1. Also, the details of PUF-based recent works are given in Table 2.

2.1. Problem Definition. Security protocols, such as authentication methods, are supposed to ensure the confidentiality, integrity, and availability (CIA triangle) of security. The parties to the protocols must be able to authenticate and synchronize with one another at any moment. Desynchronization attacks can break this condition by blocking protocol messages or forcing protocol parties to modify their shared secret values to different values, preventing the parties from authenticating each other and destroying service availability. Many protocols have been developed in the literature to satisfy CIA security standards; however, multiple instances of attacks [2, 10–14] against them show that they have failed to achieve the needed security. As a result, attempts to build a secure protocol are still continuing, and new attacks are emerging that provide designers fresh insight into how to (not) design a protocol. As a result of these assaults and security evaluations, the protocols have progressed.

2.2. Motivation and Contributions. In recent years, a number of key agreement and authentication techniques have been created. Most of these protocols have a greater calculation cost, making them unsuitable with devices with limited resources. We also noticed that the literature reviewed above did not take into account the physical factors of security for vehicle RFID communication systems in VCC situations. However, in the automotive RFID communication environment, the necessity of PUF receives a lot of attention in the literature.

A PUF-based protocol is capable of dealing with physical security risks. Even stealing the PUF from the on-board memory will not allow an attacker to obtain it. As a result, for VCC, we developed a PUF-enabled RFID-based authentication protocol. The following are some of the many contributions made by this research:

(1) To build an authentication protocol for VCC communication, the system and threat models are defined first.

(2) We created a PUF-enabled RFID-based authentication mechanism using the hypothesized attack model.

(3) To keep the proposed protocol's cost minimal, only fundamental cryptographic operations such as ECC, XOR, concatenation, and hash function are used. PUF is also used to protect against recognized physical security risks.

(4) Our approach ensures that possible security threats are avoided, based on formal and informal security assessments.

(5) The results of the performance study show that our protocol is superior to other similar protocols.

Table 1: Summary of cryptographic techniques applied and limitations of previous existing user authentication mechanisms.

Scheme	Year	Cryptographic techniques	Advantages	Drawbacks/limitations
Jiang et al. [3]	2018	(i) Uses "one-way cryptographic hash function"	(i) Fits for vehicular cloud networking environment	(i) Fails to preserve "revocability" (ii) Prone to "replay attack"
Alamr et al. [5]	2018	(i) Based on "RFID" (ii) Applies "ECC cryptographic technique" (iii) Uses "to support IoT"	(i) Applicable in IoT environment	(i) Does not support "revocability and password/biometric update" (ii) Vulnerable to "data integrity and key compromise"
Dinarvand and Barati [6]	2019	(i) Based on "RFID technology" uses "one-way cryptographic hash function" (ii) Based on "ECC cryptographic technique"	(i) Does not fit for generic IoT networking environment	(i) Fails to preserve "impersonation and key compromise" (ii) No "formal security" analysis
Bagga et al. [1]	2018	(i) Based on "three factors (user mobile device, user password, and personal biometrics" (ii) Applies "ECC cryptographic technique" (iii) Uses "fuzzy extractor for biometric verification"	(i) Applicable in industrial IoT environment	(i) Does not support "revocability and password/biometric update" (ii) Vulnerable to "known session key attack"
Kumar et al. [7]	2020	(i) Based on "three factors (smart card, user password, and biometrics)" uses "one-way cryptographic hash function" (ii) Based on "fuzzy extractor for biometric verification"	(i) Fits for generic IoT networking environment	(i) Fails to preserve "revocability" (ii) No "formal security" analysis
Jiang et al. [4]	2018	(i) Based on "three factors (user mobile device, user password, and personal biometrics" (ii) Applies "ECC cryptographic technique" (iii) Uses "fuzzy extractor for biometric verification"	(i) Applicable in cloud environment	(i) Does not support "revocability and password/biometric update" (ii) Vulnerable to "known session key attack"
Hosseinzadeh et al. [8]	2020	(i) Based on "RFID systems" uses "one-way cryptographic hash function"	(i) Fits for IoT networking environment	(i) Fails to preserve "revocability" (ii) No "session key agreement"
Zhu [9]	2020	(i) Based on "RFID systems and quadratic residue" uses "Gong-Needham-Yahalom (GNY) logic" (ii) Confined to "healthcare system"	(i) Fits for healthcare environment	(i) Fails to preserve "revocability" (ii) Desynchronization issues
Gabsi et al. [10]	2021	(i) Based on "RFID systems" uses "arithmetic calculation of ECC" (ii) Based on "ECC cryptographic system"	(i) Fits for communicating reader to reader environment	(i) Does not have to freedom to connect with the cloud server (ii) Not suitable for cloud environment
Mishra et al. [11]	2018	(i) Based on "three factors (user mobile device, user password, and personal biometrics" (ii) Applies "ECC cryptographic technique" (iii) Uses "fuzzy extractor for biometric verification"	(i) Applicable in industrial IoT environment	(i) Does not support "revocability and password/biometric update" (ii) Vulnerable to "known session key attack"
Safkhani et al. [2]	2021	(i) Based on "RFID and ECC cryptosystem" Uses "one-way cryptographic hash function"	(i) Fits for IoT networking environment	(i) Fails to establish "mutual authentication" (i) No proper "session key agreement" (ii) Could not resist "denial-of-service"

2.3. Roadmap of Article. The rest of the article is structured as follows: The preliminaries are presented in Section 3. The RSEAP2 system is described in detail in Section 4. We give a security study of the RSEAP2 protocol as well as various efficient and strong attacks against it in Section 5. The improved protocol is presented in Section 6. In Section 7, we provide a verifiable security analysis of our approach. The performance analysis is presented in Section 8. Section 9 concludes the article.

3. Definitions and Mathematical Preliminaries

The key size comparison between the public-key cryptosystems like ECC and RSA shows that the communication

Table 2: Summary of cryptographic techniques applied using PUF authentication mechanisms.

Scheme	Year	Cryptographic techniques and environment	Advantages	Drawbacks/limitations
Xu et al. [15]	2021	Based on PUF is applicable for RFID healthcare systems	Fits for healthcare systems	Does not support "revocability." Vulnerable to "know session key attack"
Gope and Sikdar [16]	2019	Based on smart grid communication systems. The lightweight cryptographic primitives such as physically unclonable functions and one-way hash function is utilized	A novel privacy-aware authenticated key agreement scheme which can not only ensure secure communication between smart meters and the service providers, but also the physical security of smart meters	(i) Does not support "revocability and password/biometric update" (ii) Vulnerable to "known session key attack"
Cao et al. [17]	2021	(i) Based on "three factors (user mobile device, user password, and personal biometrics" (ii) Applies "ECC cryptographic technique" (iii) Uses "fuzzy extractor for biometric verification"	(i) Applicable in smart grid environment and data collection scheme	(i) Does not support "revocability and password/biometric update" (ii) Vulnerable to "mutual authentication attack" and "known session key attack"
Zhang et al. [18]	2020	Key distribution in wireless sensor networks	It did not only save the storage overhead, but also provided perfect resilience against sensor capture attacks	This cannot resist anonymity, traceability, and forward secrecy attacks
Mall et al. [19]	2022	This approach is a survey on PUF-based authentication and key agreement protocols for IoT, WSN, and smart grids	(i) This survey paper can be utilized to understand the technologies such as IoT, WSN, and smart grids and the way to address the AKA in these technologies (ii) Systematically and taxonomically examine and discuss with pros and cons of AKA applications to the fast-growing areas of IoT, WSNs, and smart grids based on a meticulous survey of existing literature	This study fails to address the security pitfalls which can integrate all these technologies
Liu et al. [20]	2021	Key distribution for dynamic sensor networks	Compared with traditional key predistribution schemes, the proposal reduces the storage overhead and the key exposure risks and thereby improves the resilience against node capture attacks	This study cannot be applied to the current technologies such as IoT and cloud computing
Mukhopadhyay [21]	2016	PUFs as promising tools for security in Internet of Things. This article discusses about security violation in the authentication of a commercial IoT	(i) Studied the lightweight construction of PUFs (ii) Proof context test-bed simulations were presented for commercially available tools to show how PUFs can interact with other IoT nodes to provide overall security	This study fails to address the security features and how they can be applied for the AKA protocols
Wang et al. [22]	2021	Blockchain and lightweight authentication protocol for wireless medical sensor networks. Applies "fuzzy extractor for biometric verification"	Incorporated for blockchain and wireless medical sensor networks	(i) Desynchronization attacks (ii) Excess communication cost

Table 2: Continued.

Scheme	Year	Cryptographic techniques and environment	Advantages	Drawbacks/limitations
Lee and Chen [23]	2021	Lightweight fog computing-based authentication protocols using physically unclonable functions for Internet of medical Things	(i) The proposed protocols use lightweight cryptographic operations, including a one-way cryptographic hash function, the barrel shifter physically unclonable function (BS-PUF) (ii) This study ensures the security of the sensors and fog nodes and to avoid a computational burden on devices	This study is restricted to fog environment
Hassija et al. [24]	2021	A survey on supply chain security: application areas, security threats, and solution architectures	(i) This article discusses the supply chain's security critical application areas and presents a detailed survey of the security issues in the existing supply chain architecture (ii) Various emerging technologies, such as blockchain, machine learning (ML), and physically unclonable functions (PUFs) as solutions to the vulnerabilities in the existing infrastructure of the supply chain	This study is a survey work and fails to address the security features and how they can be applied for the AKA protocols

messages can utilize the elliptic curve cryptosystem to reduce the communication bandwidth. The key size comparison between ECC and RSA is given in Table 3.

3.1. Background of ECC. "Let $\mathscr{E}$ denotes an elliptic curve over the prime finite field F_q, where q be the large prime number. An equation of elliptic curve over F_q is given by $v^2 = u^3 + \alpha u + \beta \bmod q$, where $\alpha, \beta \in F_q$. The elliptic curve is said to be nonsingular if $4\alpha^3 + 27\beta^2 \bmod q \neq 0$. The additive elliptic curve group G is defined as $G = \{(u, v)\colon u, v \in F_q (u, v) \in \mathscr{E}\} \bigcup \{\Phi\}$, where the point Φ is known as asymptotic point which work as the identity element or zero element in G."

Some operations on the group G are as follows [2, 7]:

(1) Let $\vee = (u, v) \in G$, then define $-\vee = (u, -v)$ and $\vee + (-\vee) = \Phi$.

(2) If $\underset{1}{\vee} = (u_1, v_1)$ and $\underset{2}{\vee} = (u_2, v_2) \in G$, then $\underset{1}{\vee} + \underset{2}{\vee} = (u_3, v_3)$, where $u_3 = \rho^2 - u_1 - u_2 \bmod q v_3 = \rho(u_1 - u_2) - v_1 \bmod q$ and

$$\rho = \begin{cases} (v_2 - v_1)/(u_2 - u_1) \bmod q, & \text{if } \underset{1}{\vee} \neq \underset{2}{\vee}, \\ (3u_1^2 + \alpha)/2v_1 \bmod q, & \text{if } \underset{1}{\vee} = \underset{2}{\vee} \end{cases}$$

(3) Let $\vee = (u, v) \in G$, then scalar multiplication in G is defined as: $\eta.\vee = \vee + \vee + \vee + \cdots + \vee (\eta - \text{times})$.

(4) If g is the generator of G with order η, then $\eta.g = \Phi$.

(a) "Elliptic curve discrete logarithm problem (ECDLP)": Finding $\mu \in Z_q^*$ such that $\underset{2}{\vee} = \mu.\underset{1}{\vee}$, for a given $\underset{1}{\vee}, \underset{2}{\vee} \in G$ is difficult.

(b) "Elliptic curve computational Diffie–Hellman problem (ECCDHP)": If g is the generator of G and $\alpha.g$, $\beta.g$ are supplied $(g, \alpha.g, \beta.g)$, then computing $\alpha.\beta.g$ in G is difficult.

3.2. Physically Unclonable Function. The PUF hardware primitive accepts a challenge $\mathscr{C}$ and generates the matching response R from the physical properties of its integrated chip (IC) and C. A PUF may easily be thought of as a one-way function $R = \mathrm{PUF}(C)$ since both the accepted challenge C and the produced answer R are bit strings [14].

In essence, PUF security is based on the fact that, even if various ICs use the same production processes, each IC will be somewhat different owing to manufacturing variances. The following are the characteristics of PUF [15]:

(i) Uniqueness: A PUF cannot be duplicated;

(ii) Unidirectionality: In the real manufacturing circuit, the variances between input and output function mapping are both fixed and unpredictable. It is the hardware counterpart of the one-way function in this regard;

(iii) Invulnerability: Any effort to tamper with the device containing the PUF will cause the PUF to modify its behaviour and, as a result, it will be destroyed [14];

3.3. Network Model. Figure 1 represents the architecture which we applied for the design of communication among the participants. The RFID tag communicates with the

Table 3: Key size comparison between ECC and RSA.

S. No	ECC key size (bits)	RSA key size (bits)	Key size ratio
1	163	1024	1 : 6
2	256	3072	1 : 12
3	384	7680	1 : 20
4	512	15360	1 : 30

roadside RFID reader and thereby the communication passes through the vehicular cloud server. In order to communicate efficiently, the communication parties have to undergo the authentication and key agreement phase to establish a session key. More details regarding how the participants actually take part in the authentication and key agreement and communication process is discussed in the next section.

3.4. Threat Model. The "CK-adversary model" is widely regarded as the "current de facto standard model in modeling key-exchange protocols." Using the "CK-adversary model," the adversary A can "deliver messages (as in the DY model)," and in addition, A can also "compromise other information, such as session state, private keys, and session keys." "Since the sessions as procedures run inside a party, the internal state of a session is well-defined. An important point here is that what information is included in the local state of a session. For instance, the information revealed in this way may be the exponent used by a party. Typically, the revealed information will include all the local state of the session and its subroutines, except for the local state of the subroutines that directly access the long-term secret information." Therefore, it is important that "the leakage of some forms of secret information, such as session ephemeral (short-term) secrets or session key, should have the least possible effect on the security of other secret credentials of the communicating entities in an authenticated key-exchange protocol." We demonstrate that the proposed technique is secure against well-known attacks and offers session key security and strong credentials' privacy under the CK-adversary model through a comprehensive formal security analysis.

3.5. Security Requirements for an IoT-Based RFID Communication System. To the best of our knowledge and based on the available literature, many authentication algorithms for RFID communication systems have been proposed in recent years. The best ways for making RFID systems appropriate for a wide variety of applications are authentication and key agreement. Several forms of security threats might arise during the transfer of messages between RFID tags and readers.

Any authentication mechanism attempting to secure a viable RFID-based system should meet the following security requirements: Impersonation attack: By repeating a message recorded from the channels, an attacker might try to imitate genuine protocol participants (such as the cloud

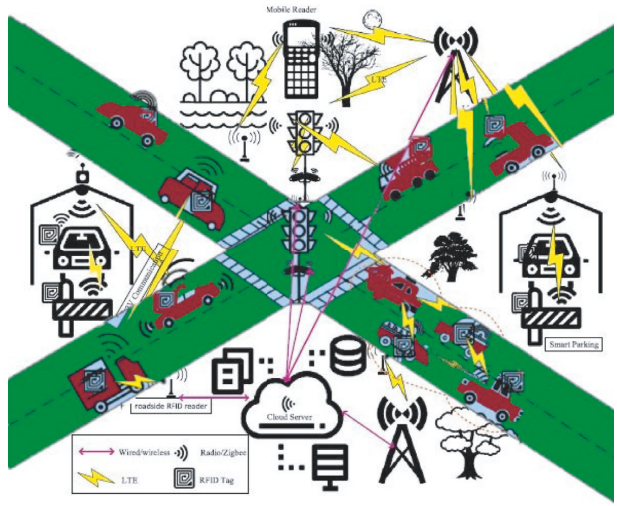


Figure 1: Communication architecture (source: this architecture was adopted from [2, 7]).

database server, RFID reader, or RFID tag). At all costs, any impersonation should be avoided.

Replay attack: In this attack, an outsider tries to deceive other certified participants by restating intercepted data. This attack is aimed at a user whose data have been intercepted by an untrustworthy third party. Mutual authentication: The authentication procedure takes place between the RFID tag and the back-end database server. Messages are exchanged across an unprotected communication route between the tag, reader, and server. This is the most crucial feature of any authentication system. Mutual authentication must also be accomplished with all three RFID system players present.

Tag anonymity: This is the most critical and required security criterion to reduce forgeries and assure security. Furthermore, the RFID authentication method retains its anonymity if an opponent is unable to trace an RFID tag during message transmission over a public channel. There are two types of anonymity, namely strong anonymity and weak anonymity. Furthermore, in order to protect their security and privacy, participants in IoT communication do not reveal their true identities.

Man-in-the-middle attack: In this attack, an adversary listens to the transmitted data before attempting to remove or change the data supplied to recipients.

Insider attack: Any insider can play the role of adversary in the RFID communication system.

Desynchronization attack: If a protocol's authentication is reliant on shared values, an adversary may cause desynchronization difficulties. If the shared data are updated by the server but the tag is not, the server might be unable to validate the tag in the future. Attempts to desynchronize should be avoided at all costs.

Untraceability: Untraceability in the RFID communication system means that no one can track the participants' activity patterns or their relayed messages.

Session key agreement: A session key agreement will be made between users and their mobile devices, as well as the network control centre, following the successful deployment of the proposed protocol.

Confidentiality: The security of RFID communications between the tag and the reader is ensured by encrypting shared secrets on the public channel.

Perfect forward secrecy: This is utilized in the authentication protocol architecture to keep previously transmitted messages private, so that an adversary who obtains the entities private and public keys will be unable to deduce a past session key.

Availability: The authentication and key agreement mechanism between the RFID tag and the RFID back-end database server operates continuously in an RFID system. To accomplish the characteristic of accessibility, the shared secret information between the RFID tag and the RFID back-end database server must be updated in most authentication procedures. However, security issues such as denial-of-service (DoS) or desynchronization attacks may cause this process to be disrupted. As a result of these problems, the RFID system's efficiency may be jeopardized. Hence, this issue should be considered while creating an authentication mechanism.

4. RSEAP2 Protocol

We offer a brief explanation of RSEAP2 [2] in this section. The tag T_i and the cloud database server S interact through the reader R_j to establish a session key SK_{ST} in this protocol. It is divided into two parts. The tag enrollment or startup phase is the first step, in which the tag talks with S via a secure connection to provide the needed data. The login and authentication phase is the second phase of the protocol, and it is used to perform mutual authentication and share the session key $SK_{ST} = SK_{TS}$. This part of the communication takes place via a public network. We have made use of the notations as shown in Table 4.

In the initialization phase of RSEAP2, the server S chooses an elliptic curve $E(F_q)$ over F_q and a generator g over G. It also selects $x_s F_q^*$ as its secret key and its public key will be $x_s.g$. Any tag T_i which aims to register with S inputs its ID_{Ti} and pw_{Ti}, generates a random value $R_{Ti}F_q$, computes $\mathrm{PWT} = h(ID_{Ti} \| (pw_{Ti} \oplus R_{Ti}))$, and sends the tuple $M_{R1} = \{\mathrm{PWT}, ID_{Ti}, TS_{R1}\}$ to S. Once S received M_1, verifies the timestamp, that is $|TS_{R2} - TS_{R1}|^? \leq t_{TS} x \Delta T$ at the first. Next, it generates $sn_i F_q$ and sets it as the T_i's serial number, computes $X_{Ti} = h(sn_i \| ID_{Ti} \| x_S)$, $A_{Ti} = h(\mathrm{PWT} \| (X_{Ti} \oplus ID_{Ti}))$, $B_{Ti} = X_{Ti} \oplus \mathrm{PWT}$, and stores sn_i corresponding to ID_{Ti}. It then sends tuple $M_{R2} = \{A_{Ti}, B_{Ti}, sn_i, g, x_s.g, G, h(.)\}$ to T_i. The tag T_i stores $\{A_{Ti}, B_{Ti}, sn_i, g, x_s.g, G, h(.)\}$.

The description of the protocol is as follows:

L1. T_i uses it credentials $(ID_{Ti}, pw_{Ti}, R_{Ti})$, computes $\mathrm{PWT} = h(ID_{Ti} \| (pw_{Ti} \oplus R_{Ti}))$, $X_{Ti}^* = B_{Ti} \oplus \mathrm{PWT}$, and verifies $A_{Ti}^* \stackrel{?}{=} h(\mathrm{PWT} \| (X_{Ti}^* \oplus ID_{Ti}))$. If verification is successful, T_i generates $\alpha \in F_q^*$, calculates $\alpha.g$ and $W_1 = h((\alpha.g) \oplus (X_{Ti}^* \| ID_{Ti}) \| TS_{LA1})$, and sends $M_1 = \{(ID_{Ti} \| W_1) \oplus \alpha.x_s.g, \alpha.g, TS_{LA1}\}$ to the reader R_j.

L2. The reader checks the timestamp, that is, $|TS_{LA2} - TS_{LA1}|^? \leq t_{\mathrm{TR}} \Delta T$, generates $\beta \in F_q^*$, computes $\beta.g$, and then sends $M_2 = \{M_1, TS_{LA3}, \beta.g, (h((TS_{LA1} \| TS_{LA3}) \oplus (x_{Ri}.g)) \| ID_{Rj}) \oplus \beta.x_s.g\}$ to S.

L3. Once S received M_2, it verifies the timestamps, that is, $|TS_{LA4} - TS_{LA1}|^? \leq t_{TS} x \Delta T$ and $|TS_{LA4} - TS_{LA3}|^? \leq t_{RS} \times \Delta T$. Next S extracts $h^*(TS_{LA1} \| TS_{LA3} \oplus (x_{Ri}.g)) \| RID_i^* = h((TS_{LA1} \| TS_{LA3} \oplus (x_{Ri}.g)) \| RID_i$, retrieves $x_{Ri}.g$ from the database, and evaluates $h((TS_{LA1} \| TS_{LA3} \oplus (x_{Ri}.g))$ to authenticate R_j. After the successful authentication on R_j parameters, S extracts $ID_T^* \| W_1^*$, verifies $W_1^* \stackrel{?}{=} h((\alpha.g) \oplus (X_T^* \| ID_T^*) \| TS_{LA1})$, retrieves the related sn_i using ID_T^*, computes $X_T^* = h(sn_i \| ID_T^* \| x_s)$, and verifies $W_1^* \stackrel{?}{=} h((((\alpha.g) \oplus (X_T^* \| ID_T^*)) \| TS_{LA1})$. Further generating $\gamma \in \mathcal{F}_q^*$, computing the session key $SK_{ST} = h((ID_T^* \oplus X_T) \| (\alpha.\beta.\gamma.g \oplus x_s.\alpha.g) \| (sn_i \oplus (TS_{LA1} \| TS_{LA5})))$, $W_2 = h(ID_T^* \| SK_{ST})$, and sending $M_3 = \{W_2 \oplus h(RID_i \| \beta.\gamma.g), \gamma.g, TS_{LA5}, h(RID_i \| TS_{LA5} \| \beta.\gamma.g)\}$ to R_j.

L4. R_j verifies the timestamp, that is, $|TS_{LA5} - TS_{LA3}|^? \leq t_{SR} x \Delta T$ and $h(RID_i \| TS_{LA5} \| \beta.\gamma.g)$ to authenticate S. Subsequently, it extracts W_2 and then sends $M_4 = \{W_2, \beta.\gamma.g, TS_{LA5}\}$ to T_i.

L5. Similarly, T_i verifies the timestamp, that is, $|TS_{LA7} - TS_{LA1}|^? \leq t_{ST} x \Delta T$ and $|TS_{LA5} - TS_{LA1}|^? \leq t_{RT} x \Delta T$, and then computes $SK_{TS} = h((ID_T \oplus X_T) \| (\alpha.\beta.\gamma.g \oplus x_s.\alpha.g) \| (sn_i \oplus (TS_{LA1} \| TS_{LA5})))$ and checks $W_2^* \stackrel{?}{=} h(ID_T \| SK_{TS})$. If so, it sets SK_{TS} as the session key.

5. Security Analysis of RSEAP2

5.1. Inefficient Mutual Authentication Attack. On receiving the message M_2 from the reader R_j, the cloud database server S extracts and computes to validate the user and reader. The details are as follows:

(1) The cloud server performs the computations and validates the timestamps such as

Table 4: Notations along with their descriptions.

Symbol	Description
T_i, R_j, S	i^{th} tag, j^{th} reader, and cloud server
ID_{Ti}, ID_{Rj}	Unique identities of T_i and R_j
$pwTi$	Password of T_i
x_s	Long-term private secret key of the S
$\|\oplus$	Operations of bitwise concatenation and bitwise XOR
SK_{TS}/SK_{ST}	Session key established between T_i and S
PUF	Physically unclonable function
$h(\cdot)$	Cryptographic collision-resistant one-way hash function
α_i, β_j	Random numbers
TS_{TAi}	Timestamps used at i^{th} transmission
ΔT	Maximum threshold transmission delay allowed
$i \stackrel{?}{=} j$	Validation check, if expression i matches j or not
A	An adversary

$|TS_{LA4} - TS_{LA1}|^? \leq t_{TS} x \Delta T$ and $|TS_{LA4} - TS_{LA3}|^? \leq t_{RS} \times \Delta T$.

(2) Next S extracts $h^*((TS_{LA1}\|TS_{LA3} \oplus (x_{Ri}.g))\|RID_i^* = h((TS_{LA1}\|TS_{LA3} \oplus (x_{Ri}.g))\|RID_i$, retrieves $x_{Ri}.g$ from the database, and evaluates $h((TS_{LA1}\|TS_{LA3} \oplus (x_{Ri}.g))$ to authenticate R_j.

(3) After the successful authentication on R_j parameters, S extracts $ID_T^*\|W_1^*$, verifies $W_1^* \stackrel{?}{=} h((\alpha.g) \oplus (X_T^*\|ID_T^*)\|TS_{LA1})$, retrieves the related sn_i using ID_T^*, computes $X_T^* = h(sn_i\|ID_T^*\|x_s)$, and verifies $W_1^* \stackrel{?}{=} h(((\alpha.g) \oplus (X_T^*\|ID_T^*))\|TS_{LA1})$ the authenticity of the user.

(4) It further generates $\gamma \in \mathcal{F}_q^*$ and computes the session key $SK_{ST} = h((ID_T^* \oplus X_T)\|(\alpha.\beta.\gamma.g \oplus x_s.\alpha.g)\|(sn_i \oplus (TS_{LA1}\|TS_{LA5})))$.

But the conflict here is that the cloud server fails to compute the proper session key to pass it on to the tag for the validation. The reason is that the cloud server could not retrieve the random values generated by the tag and reader such as $\alpha, \beta \in \mathcal{F}_q^*$, and in the session key the cloud server uses $(\alpha.\beta.\gamma.g \oplus x_s.\alpha.g)$ value without the knowledge of the random numbers. Though the cloud server performs this computation, it would be certainly a garbage value which the tag cannot validate at any given point of time. Thus, this scheme holds the inefficiency to perform mutual authentication.

5.2. Inefficient Session Key Establishment Attack. On receiving the message m_3 from the cloud server, the tag performs the mutual authentication verification. But, the verification gets fails. The details are as follows:

(1) As discussed in the above Section 5.1, we understood that the cloud server fails to compute the authentic session key. However, on receiving the message from the cloud server, T_i verifies the timestamp, that is, $|TS_{LA7} - TS_{LA1}|^? \leq t_{ST} x \Delta T$ and $|TS_{LA5} - TS_{LA1}|^? \leq t_{RT} x \Delta T$, and then computes $SK_{TS} = h((ID_T \oplus X_T)\|(\alpha.\beta.\gamma.g \oplus x_s.\alpha.g)\|(sn_i \oplus (TS_{LA1}\|TS_{LA5})))$ and checks $W_2^* \stackrel{?}{=} h(ID_T\|SK_{TS})$. If so, it sets SK_{TS} as the session key.

(2) Now you can see that the tag T_i did not retrieve or has the potential to draw out the value $(\alpha.\beta.\gamma.g \oplus x_s.\alpha.g)$ but still perform the computation to validate the session key.

This validation never gets successful as it is a known fact that without the proper parameters and values the verification fails and the tag and the cloud server cannot establish the session key for the future communications. Thus, this scheme holds the inefficiency to perform session key establishment.

5.3. Denial-of-Service Attack. According to RSEAP2's scheme, the legitimate participants tries to communicate to each other and get the services as and when required, but from the security flaw as shown above in Sections 5.1 and 5.2, we understood that the scheme fails to establish the session key and mutual authentication. This shows the enough conclusive evidence that the scheme fails to provide services to the participants thought the tag and readers are the legitimate participants in the system. Hence, this scheme is prone to the denial-of-service attack.

6. Our Proposed Scheme

This section presents the proposed secure authentication protocol and the program architecture which is divided into a tag, a reader, and a cloud server for parallel processing, with each component working independently. In this architecture as shown in Figure 2, the tag initiates the communication by computing the validating message and transmits the validating message with a virtual ID to the reader. Upon receiving the message, it challenges the reader to validate the message. Thus, the reader computes the validating message and transmits the validating message with the virtual ID to the cloud server for further process. Once the message is received by the cloud server, it validates the reader message thereby the cloud server authenticates the reader and tag. After the successful authentication, it computes the session key to establish the key. Further, at the next stage, the reader receives the Ack1 and Ack2 from the cloud server as an acknowledgment. Then the check happens in the next stage, where the tag receives Ack1 from the reader

and simultaneously the reader checks the received Ack2. Finally, once the check is successful, the tag establish the session key and end the process (see process flow diagram in Figure 2).

In this section, we present our proposed scheme. In the initialization phase, the server S chooses an elliptic curve $E(\mathcal{F}_q)$ over $\mathcal{F}_q$ and a generator g over G. It also selects $x_s \in \mathcal{F}_q^*$ as its secret key and its public key will be $x_s \cdot g$. Any tag T_i which aims to register with S, inputs its ID_{Ti}, pw_{Ti}, generates challenge C_{Ti}, computes $\alpha_{Ti} = \mathrm{PUF}(C_{Ti})$, $pa_{Ti} = \alpha_{Ti} \oplus h(ID_{Ti} \| pw_{Ti})$, and sends the tuple $M_{R1} = \{ID_{Ti}, \alpha_{Ti}, C_{Ti}\}$ to S. Once S received M_{R1}, verifies in the records whether ID_{Ti} exists or not. If the ID_{Ti} is new, it generates $sn_i \in \mathcal{F}_q$, computes $\mathrm{pid}_{Ti} = sn_i \cdot x_s \cdot g$, $A_{Ti} = h((\alpha_{Ti} \oplus ID_{Ti}) \| \mathrm{pid}_{Ti})$, and stores $\{\mathrm{pid}_{Ti}, C_{Ti}, sn_i, \alpha_{Ti}\}$ corresponding to ID_{Ti}. It then sends tuple $M_{R2} = \{\mathrm{pid}_{Ti}, A_{Ti}\}$ to T_i. The tag T_i computes $\mathrm{PWT} = h(\mathrm{ID}_{Ti} \| (pw_{Ti} \oplus \alpha_{Ti}) \| pa_{Ti})$ and stores $\{\mathrm{PWT}, \mathrm{pid}_{Ti}, pa_{Ti}, A_{Ti}\}$. Similarly, reader R_j aims to register with S, generates challenge C_{Rj}, computes $\alpha_{Rj} = \mathrm{PUF}(C_{Rj})$, $pa_{Rj} = C_{Rj} \oplus ID_{Rj}$), and sends $M_{R3} = \{ID_{Rj}, \alpha_{Rj}, pa_{Rj}\}$ to the cloud database server S. S computes pid_{Rj} by its private key and $C_{Rj} = pa_{Rj} \oplus ID_{Rj}$, $\alpha_{Rj}^* = \mathrm{PUF}(C_{Rj})$, $A_{Rj} = h((\alpha_{Rj}^* \oplus ID_{Rj}) \| \mathrm{pid}_{Rj})$; sends $M_{R4} = \{\mathrm{pid}_{Rj}, A_{Rj}\}$; and stores $\{\mathrm{pid}_{Rj}, ID_{Rj}, C_{Rj}, \alpha_{Rj}\}$ in its database. Further R_j also stores $\{\mathrm{pid}_{Rj}, pa_{Rj}, A_{Rj}\}$. The illustration of the tag registration and reader registration is shown in Table 5 and Table 6, respectively.

6.1. Login and Authentication Phase. To access the services from S, T_i needs to establish a session key with S. The following steps are followed by T_i, R_j, and S during this phase. The illustration is shown in Table 7.

LA1: The tag logs on by (ID_{Ti}, pw_{Ti}), computes $\alpha_{Ti}^* = pa_{Ti} \oplus h(ID_{Ti} \| pw_{Ti})$, verifies $\mathrm{PWT} \stackrel{?}{=} h(ID_{Ti} \| (pw_{Ti} \oplus \alpha_{Ti}^*) \| pa_{Ti})$, generates $\alpha \in \mathcal{F}_q^*$ to compute $W_1 = h((\alpha.g.\alpha_{Ti}) \oplus (A_{Ti}^* \| ID_{Ti}) \| TS_{LA1})$, and sends $M_1 = \{(pid_{Ti} \| A_{Ti} \| W) \oplus \alpha.x_s.g, \alpha.g, TS_{LA1}\}$ to R_j.

LA2: On receiving the request, R_j verifies TS_{LA1}; computes $C_{Rj} = pa_{Rj} \oplus ID_{Rj}$, $\alpha_{Rj} = \mathrm{PUF}(C_{Rj})$, and $W_2 = h(\alpha_{Rj} \| A_{Rj} \| TS_{LA3})$; and sends $M_2 = \{M_1, TS_{LA3}, (W_2 \| pid_{Rj})\}$ to S.

LA3: On receiving the request, S verifies TS_{LA1} and TS_{LA3}; extracts $M_1, TS_{LA3}, (W_2 \| pid_{Rj})$; validates $W_2 \stackrel{?}{=} h(\alpha_{Rj} \| h((\alpha_{Rj} \oplus ID_{Rj}) \| pid_{Rj}) \| TS_{LA3})$; and extracts $(pid_{Ti}^* \| A_{Ti}^* \| W_1^*)$ to verify $W_1^* \stackrel{?}{=} h((\alpha.g.\alpha_{Ti}) \oplus (A_{Ti}^* \| ID_{Ti}) \| TS_{LA1})$ and on success, generates $\beta \in \mathcal{F}_q^*$, computes $SK_{ST} = h((ID_{Ti}^* \oplus A_{Ti}) \| (\alpha.\beta.g \| x_s.\alpha.g) \| (sn_i \oplus (TS_{LA1} \| TS_{LA5})))$, $W_3 = h(\alpha.\beta.g \| SK_{ST} \| A_{Ti} \| pid_{Ti})$, $W_4 = h(ID_{Rj} \| A_{Rj} \| TS_{LA5} \| pid_{Rj})$, and sends $M_3 = \{W_3, W_4, TS_{LA5}, \beta.g\}$ as a response to R_j.

LA4: After receiving the response from S, R_j checks TS_{LA5}, verifies $W_4 \stackrel{?}{=} h(ID_{Rj} \| A_{Rj} \| TS_{LA5} \| pid_{Rj})$, and sends $M_4 = \{W_3, \beta.g, TS_{LA5}\}$ to T_i.

LA5: On receiving the response from R_j, T_i verifies TS_{LA5}, computes $SK_{TS} = h((ID_{Ti} \oplus A_{Ti}) \| (\alpha.\beta.g \| \alpha.x_s.g) \| t(sn_i \oplus (TS_{LA1} \| TS_{LA5})))$, and checks $W_3^* \stackrel{?}{=} h(\alpha.\beta.g \| SK_{TS} \| A_{Ti} \| pid_{Ti})$. On successful verification, T_i sets SK_{TS} as the session key.

6.2. Revocation and Reissue Phase. To revoke the access of T_i, S checks for the availability of ID_{Ti} during the subsequent login attempts. The tag will be given or refused access on the basis of the check. Since all dynamic identities have a finite lifetime, it is also impossible to continuously use the same dynamic identity.

In addition, the next steps to get new credentials are crucial when a tag T_i from an approved registered user is stolen/lost.

RR1: The tag keeps the same ID_{Ti}, but chooses a password pw_{Ti}^R and generates challenge C_{Ti}^R to compute $\alpha_{Ti}^R = \mathrm{PUF}(C_{Ti}^R)$, $pa_{Ti}^R = \alpha_{Ti}^R \oplus h(ID_{Ti} \| pw_{Ti}^R)$. Further submitting the revocation request $M_{RR1} = \{ID_{Ti}, \alpha_{Ti}^R, C_{Ti}^R\}$ to the cloud database server S through secure channel.

RR2: On receiving the request, S checks the database for the availability of $A_{Ti}^R = h((\alpha_{Ti}^R \oplus ID_{Ti}) \| pid_{Ti}^R)$ where pid_{Ti}^R is computed by private key of S. If A_{Ti}^R is not available, the cloud database server computes and sends $M_{RR2} = \{pid_{Ti}^R, A_{Ti}^R\}$ to T_i over the secure channel.

RR3: Finally, for each tag, the cloud server S issues the new credentials.

RR4: After receiving the new credentials, T_i completes the registration process as processed in the registration phase.

6.3. Tag's Password/Update Phase. A registered tag T_i can update his/her current password and follow the steps without contacting S:

PU1: The tag logs on by (ID_{Ti}, pw_{Ti}), computes $\alpha_{Ti}^* = pa_{Ti} \oplus h(ID_{Ti} \| pw_{Ti})$, and verifies $\mathrm{PWT} \stackrel{?}{=} h(ID_{Ti} \| (pw_{Ti} \oplus \alpha_{Ti}^*) \| pa_{Ti})$. Upon unsuccessful verification, this process gets terminated by T_i. Otherwise, T_i uses new password.

PU2: T_i picks pw_{Ti}^{new}; computes $\alpha_{Ti} = \mathrm{PUF}(C_{Ti})$, $pa_{Ti}^{\mathrm{new}} = \alpha_{Ti} \oplus h(ID_{Ti} \| pw_{Ti}^{\mathrm{new}})$, and $\mathrm{PWT}^{\mathrm{new}} = h(ID_{Ti} \| (pw_{Ti}^{\mathrm{new}} \oplus \alpha_{Ti}) \| pa_{Ti}^{\mathrm{new}})$; and stores $\{\mathrm{PWT}^{\mathrm{new}}, pid_{Ti}, pa_{Ti}^{\mathrm{new}}, A_{Ti}\}$ to complete the process.

7. Formal Security Analysis

Formal security examination strategies are usually used to inspect and evaluate diverse check plans. According to literature [25], various security assessment systems can be used

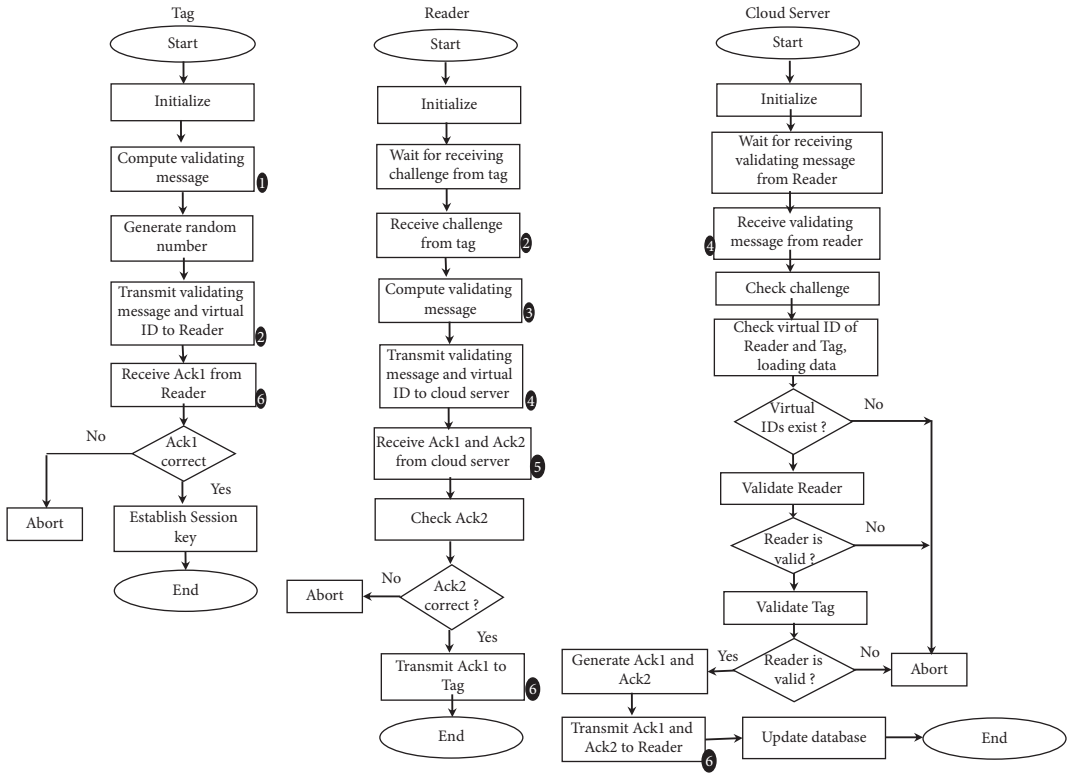


Figure 2: Communication flowchart.

Table 5: Tag registration phases of our scheme.

TagT_i	Cloud database server S
Inputs (ID_{Ti}, pw_{Ti}) Generates challenge C_{Ti} Computes $\alpha_{Ti} = \mathrm{PUF}(C_{Ti})$	Verifies ID_{Ti}
$pa_{Ti} = \alpha_{Ti} \oplus h(ID_{Ti} \| pw_{Ti})$	Computes pid_{Ti} by private key of S
$\Rightarrow_{\text{SecureChannel}}^{M_{R1}=\{ID_{Ti}, \alpha_{Ti}, C_{Ti}\}}$	$A_{Ti} = h((\alpha_{Ti} \oplus ID_{Ti}) \| \mathrm{pid}_{Ti})$ $\Leftarrow_{\text{SecureChannel}}^{M_{R2}=\{\mathrm{pid}_{Ti}, A_{Ti}\}}$
$\mathrm{PWT} = h(ID_{Ti} \| (pw_{Ti} \oplus \alpha_{Ti}) \| pa_{Ti})$, deletes $(ID_{Ti}, C_{Ti}, \alpha_{Ti})$ and stores $\{\mathrm{PWT}, \mathrm{pid}_{Ti}, pa_{Ti}, A_{Ti}\}$	Stores $\{\mathrm{pid}_{Ti}, ID_{Ti}, C_{Ti}, sn_i, \alpha_{Ti}\}$

Table 6: Reader registration phases of our scheme.

Reader R_j	Cloud database server S
Generates challenge C_{Rj} Computes $\alpha_{Rj} = \mathrm{PUF}(C_{Rj})$ $pa_{Rj} = C_{Rj} \oplus ID_{Rj}$	Computes pid_{Rj} by private key of S
$\Rightarrow_{\text{SecureChannel}}^{M_{R3}=\{ID_{Rj}, \alpha_{Rj}, pa_{Rj}\}}$	$C_{Rj} = pa_{Rj} \oplus ID_{Rj}$ $\alpha^*_{Rj} = \mathrm{PUF}(C_{Rj})$ $A_{Rj} = h((\alpha^*_{Rj} \oplus ID_{Rj}) \| \mathrm{pid}_{Rj})$ $\Leftarrow_{\text{SecureChannel}}^{M_{R4}=\{\mathrm{pid}_{Rj}, A_{Rj}\}}$
Deletes $(ID_{Rj}, C_{Rj}, \alpha_{Rj})$ Stores $\{\mathrm{pid}_{Rj}, pa_{Rj}, A_{Rj}\}$	Stores $\{\mathrm{pid}_{Rj}, ID_{Rj}, C_{Rj}, \alpha_{Rj}\}$

Table 7: Login and authentication phases of our scheme.

TagT_i	Reader R_j	Cloud database server S
Logs on by (ID_{Ti}, pw_{Ti}) Computes $\alpha^*_{Ti} = pa_{Ti} \oplus h(ID_{Ti} \| pw_{Ti})$ Verifies PWT$\stackrel{?}{=} h(ID_{Ti} \| (pw_{Ti} \oplus \alpha^*_{Ti}) \| pa_{Ti})$ Generates $\alpha \in \mathcal{F}^*_q$ Computes $W_1 = h((\alpha.g.\alpha_{Ti}) \oplus (A^*_{Ti} \| ID_{Ti}) \| TS_{LA1})$ $\longrightarrow^{M_1} = (pid_{Ti} A_{Ti}) W_1 \oplus \alpha.x_s.g, \alpha.g, TS_{LA1}$	Verifies TS_{LA1}, compute $C_{Rj} = pa_{Rj} \oplus ID_{Rj}$ Computes $\alpha_{Rj} = \mathrm{PUF}(C_{Rj})$ $W_2 = h(\alpha_{Rj} \| A_{Rj} \| TS_{LA3})$ $\longrightarrow^{M_2 = \{M_1, TS_{LA3}, (W_2 \| \mathrm{pid}_{Rj})\}}$	Verifies TS_{LA1} and TS_{LA3} Extracts M_1, TS_{LA3}, $(W_2 \| \mathrm{pid}_{Rj})$ Verifies $W_2 \stackrel{?}{=}$ $h(\alpha_{Rj} \| h((\alpha_{Rj} \oplus ID_{Rj}) \| pid_{Rj}) \| TS_{LA3})$, Extracts $(\mathrm{pid}^*_{Ti} \| A^*_{Ti} \| W^*_1)$ Verifies $W^*_1 \stackrel{?}{=}$ $h((\alpha.g.\alpha_{Ti}) \oplus (A^*_{Ti} \| ID_{Ti}) \| TS_{LA1})$ Generates $\beta \in \mathcal{F}^*_q$, computes $SK_{ST} = h((ID^*_{Ti} \oplus A_{Ti}) \| (\alpha.\beta.g \| x_s.\alpha.g) \|$ $(sn_i \oplus (TS_{LA1} \| TS_{LA5}))$ $W_3 = h(\alpha.\beta.g \| SK_{ST} \| A_{Ti} \| \mathrm{pid}_{Ti})$ $W_4 = h(ID_{Rj} \| A_{Rj} \| TS_{LA5} \| \mathrm{pid}_{Rj})$ $\stackrel{M_3 = (W_3, W_4, TS_{LA5}, \beta.g)}{\longleftarrow}$
Verifies TS_{LA5} and computes $SK_{TS} = h(ID_{Ti} \oplus A_{Ti}) \| (\alpha.\beta.g \| \alpha.x_s.g) \|$ $(sn_i \oplus (TS_{LA1} \| TS_{LA5}))$ Checks $W^*_3 \stackrel{?}{=} h(\alpha.\beta.g \| SK_{TS} \| A_{Ti} \| \mathrm{pid}_{Ti})$ to set SK_{TS} as the session key	Checks TS_{LA5} and Verifies $W_4 \stackrel{?}{=} h(ID_{Rj} \| A_{Rj} \| TS_{LA5} \| \mathrm{pid}_{Rj})$ $\stackrel{M_4 = (W_3, \beta.g, TS_{LA5})}{\longleftarrow}$	

to evaluate authentication methods. In this article, we used ROR security theories.

7.1. ROR Model-Based Proof. Under this model, adversaries say that $\mathcal{A}$ has access to a set of executing entity queries including CorruptTi (T_i), Test (P^t), , Execute (T_i, S_j), and Reveal (P^t), which perform simulation to check the real attack. The query descriptions of such queries are given in Table 8. The ROR model components are as follows:

(i) Participants: The associated participants with the proposed scheme are the tag T_i, reader R_j, or a cloud server S_j. The instances t_1 and t_s of T_i and S_j are marked as $\mathrm{P}^{t_1}_{T_i}$ and $\mathrm{P}^{t_2}_{S_j}$ which are known as oracles.

(ii) Accepted state: If the peer points achieve an accepted status when the final communication has been authenticated, the instance "P^t" comes under "accepted state." For the ongoing session, sid is a P^t session ID created in a sequence by P*Pt* after the sent and received messages were rearranged.

(iii) Partnering: The following things must be accomplished to be partnered between P^{t_1} and P^{t_2}:

(1) They are in "accepted states."
(2) They possess the same sid. Further also "authenticate mutually with each other."
(3) They are also "mutual partners of each other."

(iv) Freshness: $\mathrm{P}^{t_1}_{T_i}$ or $\mathrm{P}^{t_2}_{S_j}$ is fresh when the constructed session key between T_i and S_j is not leaked to $\mathcal{A}$ using the Reveal (P^t) query listed in Table 8.

The proposed scheme undergoes "semantic security" as defined in Definition 1.

Definition 1. If $\mathrm{Adv}^{Rfid-\mathrm{PUF}}_{\mathcal{A}}(t_p)$ is the "advantage of an adversary $\mathcal{A}$ running in polynomial time t_p in breaching the semantic security of $Rfid-\mathrm{PUF}$ to extract the session key (SK_{TS}) among a tag T_i and a cloud server S_j," $\mathrm{Adv}^{Rfid-\mathrm{PUF}}_{\mathcal{A}}(t_p) = |2\Pr[c\prime = c] - 1|$, where c are the correct bits and c' indicate the guessed bits.

Furthermore, Definition 2 is about "collision-resistant one-way hash function" and Definition 3 is about "elliptic curve decisional Diffie–Hellman problem (ECDDHP)," for briefing $Rfid-\mathrm{PUF}$.

Definition 2. A "deterministic function," say h: $\{0,1\}^* \longrightarrow \{0,1\}^{l_b}$, is a "one-way collision-resistant hash function" if it produces fixed length of l_b bits output string $h(m) \in \{0,1\}^{l_b}$ as "hash value or message digest" upon an arbitrary length input string $m \in \{0,1\}^*$. Let an adversary $\mathcal{A}$ want to find a hash collision. Then, the "advantage" of $\mathcal{A}$ in attacking "hash collision" is provided by $\mathrm{Adv}^{\mathrm{Hash}}_{\mathcal{A}}(t_h) = \Pr[(m_1, m_2) \leftarrow_r \mathcal{A}: m_1 \neq m_2, h(m_1) = h(m_2)]$. $\Pr(X)$ here shows the chance that the pair will be randomly picked by $\mathcal{A}$ in the case of "random event X" and

Table 8: Various queries with their descriptions.

Query	Significance
CorruptTi (T_i)	$\mathcal{A}$ can extract the stored credentials by compromised tag T_i's memory
Execute(T_i , S_j)	This supports $\mathcal{A}$ in intercepting communications between T_i and S_j
Reveal (P^t)	This allows $\mathcal{A}$ to obtain the SK_{ST} (= SK_{TS}) session key from P^t and its partner
Test (P^t)	It allows $\mathcal{A}$ to request P^t for the session key SK_{TS} (= SK_{ST}) and is probably a consequence of a flickered "unbiased coin c" P^t output

$(m1, m2) \leftarrow_r \mathcal{A}$. The attack of (η, t)-adversary of $\mathcal{A}$ to the resistance of collision of $h(\cdot)$ indicates that the maximum runtime of t_h to the $\mathrm{Adv}_{\mathcal{A}}^{\mathrm{Hash}}(t_h) \le \eta$.

Definition 3. Consider an elliptic curve $E_q(u, v)$ and a point P, the ECDDHP is "for a quadruple $\langle P, uv_1.P, uv_2.P, uv_3.P\rangle$, decide whether $uv_3 = uv_1 \cdot uv_2$ or it is a uniform value," where $uv_1, uv_2, uv_3 \in Z_q^*$ $(=\backslash\{1, 2, \ldots, q-1\backslash\})$.

To make ECDDHP intractable, the chosen prime q needs to be at least 160-bit number.

Theorem 1. *Suppose our scheme* $(Rfid - PUF)$ *runs in "polynomial time* t_p*" and the adversary* $\mathcal{A}$ *is working to gain advantage on* $Rfid - PUF$*. If* $query_h$*,* $|Hash|$*, and* $Adv_{\mathcal{A}}^{ECDDHP}(t_p)$ *indicate the "cardinality of hash queries," "size of one-way hash function* $h(\cdot)$*," and "*$\mathcal{A}$*'s advantage in breaching ECDDHP in t ime* t_p *(see Definition III-A)," respectively, and chosen passwords follow the Zipf's law [26], then the bit-lengths of the PUF key* $PUF(C*)$ *where* $*$ *refers to* T_i/R_j *and the tag identity* ID_{T_i} *are* l_1 *and* l_2*, respectively,* γ' *and* $s\gamma'$ *are the Zipf's parameters [26] respectively,* $\mathcal{A}$*'s advantage in compromising the semantic security of the proposed scheme* $Rfid - PUF$ *is* $Adv_{\mathcal{A}}^{Rfid-PUF}(t_p) \le 2Adv_{\mathcal{A}}^{ECDDHP}(t_p) + (query_h^2/|Hash|) + 2\max\{(query_s/2^{l_1}), (query_s/2^{l_2}), \gamma' \cdot query_s^{s\gamma}\}$.

Proof. This proof is presented in the similar way as presented by authentication protocols. Here four games are played, such as G_k, $(k = 0, 1, 2, 3)$ related to the evidence where G_0 is the starting and G_3 is the finishing game. We define $\mathrm{Succ}_{\mathcal{A}}^{G_k}$ as "an event wherein $\mathcal{A}$ can guess the random bit c in the game G_k correctly" and also the "advantage of $\mathcal{A}$ in winning the game G_k as $\mathrm{Adv}_{\mathcal{A},G_k}^{Rfid-\mathrm{PUF}} = \Pr[\mathrm{Succ}_{\mathcal{A}}^{G_k}]$." The detailed study of these games is as follows:

G_0: $G0$ is the same as the real ROR model protocol. Therefore, the semantic security of $Rfid - \mathrm{PUF}$ is defined in Definition 1.

$$\mathrm{Adv}_{\mathcal{A}}^{Rfid-\mathrm{PUF}}(t_p) = \left|2 \cdot \mathrm{Adv}_{\mathcal{A},G_0}^{Rfid-\mathrm{PUF}} - 1\right|, \quad (1)$$

G_1: In this game, we model for the "eavesdropping attack" in which $\mathcal{A}$ can intercept all the communicated messages $M_1 = \{(pid_{Ti} \| A_{Ti} \| W_1) \oplus \alpha.x_s.g, \alpha.g, TS_{LA1}\}$, $M_2 = \{M_1, TS_{LA3}, (W_2 \| pid_{Rj})\}$, $M_3 = \{W_3, W_4, TS_{LA5}, \beta.g\}$, and $M_4 = \{W_3, \beta.g, TS_{LA5}\}$ while executing "authentication and key agreement phase" in Section A using *Execute* query as discussed in Table 8. To confirm whether the "calculated session key SK_{TS} between T_i and S is real or a random number," $\mathcal{A}$ can execute both *Reveal* and *Test* queries. The established session key is $SK_{ST} = h((ID_{Ti}^* \oplus A_{Ti}) \| (\alpha.\beta.g \| x_s.\alpha.g) \| t(sn_i \oplus (TS_{LA1} \| TS_{LA5}))) = SK_{TS}$. It is worth noting that the key to session security is dependent on both α and β "temporary secrets" and T_i' and S' for long-term secretions that cannot be disregarded by eavesdrops of the messages M_1, M_2, M_3, and M_4. Therefore, this "eavesdropping attack" does not give any advantage/increase of winning probability of $\mathcal{A}$ in G_1. This shows G_0 and G_1 games become "indistinguishable," and thus obtains the following result:

$$\mathrm{Adv}_{\mathcal{A},G_1}^{Rfid-\mathrm{PUF}} = \mathrm{Adv}_{\mathcal{A},G_0}^{Rfid-\mathrm{PUF}}. \quad (2)$$

G_2: In this game, the hash searches are simulated. Both A_{Ti} and TS_{LA1} are altered in the M_1 message. Similarly, M_2, M_3, and M_4 are also equally unexpected, as they include random timestamps and random numbers, such as A_{Rj}, α_{Rj}, TS_{LA3}, pid_{Ti}^*, sn_i, and TS_{LA5} are equally unforeseeable. So, no collision occurs when $\mathcal{A}$ does hash queries. Since both G_1 and G_2 are "indistinguishable" except for the inclusion of the G_2 simulations, we obtain birthday paradox outcomes as

$$\left|\mathrm{Adv}_{\mathcal{A},G_2}^{Rfid-\mathrm{PUF}} - \mathrm{Adv}_{\mathcal{A},G_1}^{Rfid-\mathrm{PUF}}\right| \le \frac{\mathrm{query}_h^2}{2|\mathrm{Hash}|}. \quad (3)$$

G_3: The CorruptTi (T_i) query was implemented in this final game. Therefore, the opponent $\mathcal{A}$ is extracted depending on the performance of the query for the credentials A_{Ti}, pid_{Ti}, α_{Ti}, α_{Rj}, A_{Rj}, pid_{Rj} from a compromised tag T_i. The $\mathcal{A}$ probability to properly guess the $\mathrm{PUF}(C*)$ physically unclonable function secret key of l_1 bit-length and ID_{Ti} user identity of l_2 bit-length are $\mathrm{query}_s/2^{l_1}$ and $\mathrm{query}_s/2^{l_2}$, respectively. The advantage of $\mathcal{A}$ is more than 0.5, if $\mathrm{query}_s = 10^7$ or 10^8, since the passwords of the users selected tend to obey the law of Zipf's, by using assaults via trawling. If $\mathcal{A}$ can exploit user's personal data for a targeted assault, then $\mathrm{query}_s \le 10^6$ gives him an edge over 0.5.

Furthermore, $\mathcal{A}$ will have all the intercepted messages M_1, M_2, M_3, and M_4. To derive the session key $SK_{ST} = h((ID_{Ti}^* \oplus A_{Ti}) \| (\alpha.\beta.g \| x_s.\alpha.g) \| t(sn_i \oplus (TS_{LA1} \| TS_{LA5}))) = SK_{TS}$ shared between T_i and S, $\mathcal{A}$ needs to calculate $h(\alpha_{Rj} \oplus ID_{Rj})$, $(A_{Ti}^* \| ID_{Ti})$ which in a polynomially restricted time t_p is computationally costly owing to the intractability of ECDDHP. Since G_2 and G_3 games are "indistinguishable," the following is excepted to include the question and ECDDHP of CorruptTi (T_i)

$$\left|\mathrm{Adv}_{\mathscr{A},G_3}^{Rfid-\mathrm{PUF}}-\mathrm{Adv}_{\mathscr{A},G_2}^{Rfid-\mathrm{PUF}}\right|\le \mathrm{Adv}_{\mathscr{A}}^{\mathrm{ECDDHP}}\left(t_p\right)+\max\left\{\frac{\mathrm{query}_s}{2^{l_1}},\frac{\mathrm{query}_s}{2^{l_2}},\beta'\cdot\mathrm{query}_s^{s\beta}\right\}. \tag{4}$$

Now, all the relevant queries related to the above games are executed, and then the Reveal query is executed along with Test query to guess the random bit c. Thus, we get

$$\mathrm{Adv}_{\mathscr{A},G_3}^{Rfid-\mathrm{PUF}}=\frac{1}{2}. \tag{5}$$

Combining equations (1), (2), and (5) derives:

$$\begin{aligned}\frac{1}{2}\cdot\mathrm{Adv}_{\mathscr{A}}^{Rfid-\mathrm{PUF}}\left(t_p\right)&=\left|\mathrm{Adv}_{\mathscr{A},G_0}^{Rfid-\mathrm{PUF}}-\frac{1}{2}\right|\\&=\left|\mathrm{Adv}_{\mathscr{A},G_1}^{Rfid-\mathrm{PUF}}-\mathrm{Adv}_{\mathscr{A},G_3}^{Rfid-\mathrm{PUF}}\right|\\&\le\left|\mathrm{Adv}_{\mathscr{A},G_1}^{Rfid-\mathrm{PUF}}-\mathrm{Adv}_{\mathscr{A},G_2}^{Rfid-\mathrm{PUF}}\right|\\&\quad+\left|\mathrm{Adv}_{\mathscr{A},G_2}^{Rfid-\mathrm{PUF}}-\mathrm{Adv}_{\mathscr{A},G_3}^{Rfid-\mathrm{PUF}}\right|.\end{aligned} \tag{6}$$

Next, combining equations (3), (4), and (6) provide the following result:

$$\frac{1}{2}\cdot\mathrm{Adv}_{\mathscr{A}}^{Rfid-\mathrm{PUF}}\left(t_p\right)\le\frac{\mathrm{query}_h^2}{2|\mathrm{Hash}|}+\mathrm{Adv}_{\mathscr{A}}^{\mathrm{ECDDHP}}\left(t_p\right)+\max\left\{\frac{\mathrm{query}_s}{2^{l_1}},\frac{\mathrm{query}_s}{2^{l_2}},\beta'\cdot\mathrm{query}_s^{s\beta}\right\}. \tag{7}$$

Finally, the equation (7) is multiplied by 2 on both sides to get

$$\mathrm{Adv}_{\mathscr{A}}^{Rfid-\mathrm{PUF}}\left(t_p\right)\le 2\mathrm{Adv}_{\mathscr{A}}^{\mathrm{ECDDHP}}\left(t_p\right)+\frac{\mathrm{query}_h^2}{|\mathrm{Hash}|}+2\max\left\{\frac{\mathrm{query}_s}{2^{l_1}},\frac{\mathrm{query}_s}{2^{l_2}},\beta'\cdot\mathrm{query}_s^{s\beta}\right\}. \tag{8}$$

□

7.2. Informal Security Analysis

Proposition 1. *Location privacy (non-traceability)*

Proof. The tag Ti simply transmits the message $M_1=\{(pid_{Ti}\|A_{Ti}\|W_1)\oplus\alpha.x_s.g,\alpha.g,TS_{LA1}\}$, with $W_1=h((\alpha.g)\oplus(A_{Ti}^*\|ID_{Ti})\|TS_{LA1})$. Only $(pid_{Ti}\|A_{Ti}\|W_1)$ of this message can be utilized to identify the tag. On each session, the variables alpha and TS_{LA1} masked and randomized the token described above. The attacker has no control over any of these values. If a collision happens on the specified value by T_i in the worst-case scenario, the adversary could detect it by monitoring the alpha.g fraction of M_1, and then T_i could be monitored. However, the adversary's advantage in finding a collision after N protocol sessions is $O(N^2/|F_q^*|)$, which is modest enough in practice. Furthermore, M_1 makes no mention of R_j or S.

The reader R_j delivers $M_2=\{M_1,TS_{LA3},(W_2\|pid_{Rj})\}$ to S, where $(W_2\|pid_{Rj})$ may be used to monitor the reader and determine whether the W_2 fraction has a collision. Similarly, after N protocol executions, the adversary has an advantage of $O(N^2/|F_q^*|)$ in detecting a collision. As a result, the opponent's chances of success are slim.

$M_3=\{W_3,W_4,TS_{LA5},\beta.g\}$ is sent by the server S, where $W_3=h(\alpha.\beta.g\|\|SK_{ST}\|A_{Ti}\|pid_{Ti})$, $W_4=h(ID_{Rj}\|A_{Rj}\|TS_{LA5}\|pid_{Rj})$. The reader R_j and the server S, on the other hand, in each of the W_3 and W_4 tokens are randomized in each session. As a result, an adversary is unable to retrieve data that could aid in the breach of the protocol's location privacy.

Finally, R_j sends $M_4=\{W_3,\beta.g,TS_{LA5}\}$ to T_i. The adversary's only target in this communication could be W_3. This token is a function of $SK_{TS}=h((ID_{Ti}\oplus A_{Ti})\|(\alpha.\beta.g\|x_s.\alpha.g)\|t(sn_i\oplus(TS_{LA1}\|TS_{LA5})))$, which is randomized by T_i,R_j, and S on each session.

Overall, the location privacy of all of our entities (i.e., T_i,R_j, and S) is guaranteed by our protocol. □

Proposition 2. *Mutual authentication and session key agreement*

Proof. It is obvious that the pairs (S,T_i) and (S,R_j) are mutually authenticated if a legitimate tag T_i connects with an honest server S through a valid reader R_j and within acceptable time thresholds. However, we do not require mutual authentication between the reader R_j and the tag T_i in this protocol. In more detail, S is the source of trust for T_i, while R_j is only a gateway to S. The following is a list of the session key's correctness and mutual agreement:

Correction Proof:

$$\begin{aligned}W_3&=h(\beta.(\alpha.g)\|SK_{ST}\|A_{Ti}\|pid_{Ti})\\&=h(\beta.(\alpha.g)\|(h(ID_{Ti}^*\oplus A_{Ti})\|(\beta.(\alpha.g)\|x_s.(\alpha.g))\|(sn_i\oplus(TS_{LA1}\|TS_{LA5})))\|A_{Ti}\|pid_{Ti})\\&=h\big(\alpha.\beta.g\|(h((ID_{Ti}^*\oplus A_{Ti})\|(\alpha.\beta.g\|x_s.\alpha.g)\|(sn_i\oplus(TS_{LA1}\|TS_{LA5}))))\big\|A_{Ti}\big\|pid_{Ti}\big)\\&=h\big(\alpha.\beta.g\|(h((ID_{Ti}\oplus A_{Ti})\|(\alpha.\beta.g\|\alpha.x_s.g)\|(sn_i\oplus(TS_{LA1}\|TS_{LA5}))))\big\|A_{Ti}\big\|pid_{Ti}\big)\\&=h(\alpha.\beta.g\|SK_{TS}\|A_{Ti}\|pid_{Ti})=W_3^*.\end{aligned} \tag{9}$$

Because the tag and the server have mutual authentication, S has already authenticated R_j, and T_i may trust the reader R_j. As a result, our technique ensures mutual authentication and establishes suitable session key agreement. □

Proposition 3. *Physical security*

Proof. . Any alteration or damage to the device with built-in PUF will cause PUF to respond differently or the device to become unavailable, according to PUF's characteristics. It is impossible to collect any relevant information in an accessible environment since car sensors do not preserve any information. Physical attacks, aside from rendering the hardware components in the proposed protocol ineffective, are unable to extract any relevant information. As a result, the suggested protocol can ensure the system's physical security. □

Proposition 4. *Achieving forward secrecy*

Proof. In our proposed scheme, the session key is computed as $SK_{TS} = h((ID_{Ti} \oplus A_{Ti}) \| (\alpha.\beta.g \| x_s.\alpha.g) \| t(sn_i \oplus (TS_{LA1} \| TS_{LA5})))$. This session key is established between the tag T_i and the server S. If $\mathscr{A}$ wishes to compromise the session key, $\mathscr{A}$ requires the knowledge of the session-specific random values $\{\alpha, \beta\}$, fixed value α_{Ti}, and the identities of the participants involved in the session key establishment. Now, even if pw_{Ti}, pa_{Ti} are compromised by $\mathscr{A}$, due to the lack of knowledge of C_{Ti} or random values $\{\alpha, \beta\}$ and fixed value α_{Ti}, attacker fails to compute W_1. Thus, $\mathscr{A}$ does not gain any advantage even if he compromises pw_{Ti}, pa_{Ti}. Therefore, $\mathscr{A}$ cannot compute the previous/current/future session keys. □

Proposition 5. *Message authentication*

Proof. In this protocol, the server authenticates M_1 and M_2. The reader R_j authenticates S, M_3 partially and the tag T_i totally. The use of random integers and the one-way hash function ensure the integrity of all messages. Any alteration to the conveyed message causes the receiver to reject the message.

For instance, consider $M_1 = \{(pid_{Ti} \| A_{Ti} \| W_1) \oplus \alpha.x_s.g, \alpha.g, TS_{LA1}\}$ message, where $W_1 = h((\alpha.g) \oplus (A^*_{Ti} \| ID_{Ti}) \| TS_{LA1})$, which should be authenticated by S. $TS_{LA4} - TS_{LA1}{}^{?} \leq \Delta T$ is checked by the server S first. As a result, if the adversary replicates the message, S will reject it. Then, S extracts $(pid^*_{Ti} \| A^*_{Ti} \| W^*_1)$, retrieves the related sn_i value using ID_{Ti} and α_{Ti}, and computes and verifies $W^*_1 \stackrel{?}{=} h((\alpha.g.\alpha_{Ti}) \oplus (A^*_{Ti} \| ID_{Ti}) \| TS_{LA1})$ to accept the message. It is clear that any modification in TS_{LA}, $\alpha.g$, or $(pid^*_{Ti} \| A^*_{Ti} \| W^*_1)$ renders the probability of $W^{*}_{1}{}^{?} = W_1$ to 2^{-n}, where n is the hash length, for example 256 – bit for SHA – 256. The other messages in the protocol can be reasoned about in the same way. As a result, our protocol ensures message authentication between the parties involved. □

Proposition 6. *Replay attack*

Proof. In a replay attack, the adversary attempts to use a previously traded message at a later time t'. Any message received outside of the threshold time (a preset factor of ΔT) is likely to be rejected in our protocol. Aside from that, the one-way hash function ensures the integrity of timestamps. As a result, replay attacks against our protocol are impossible. Finally, the adversary may break the tag's anonymity if he extracted $xs.g$ from the $\alpha.xs.g$ and $\alpha.g$ pair. It is most likely the same as solving ECCDHP, which is known to be a difficult task (see Section 3.1). □

Proposition 7. *Impersonation attack*

Tag:

Proof. Due to the integrity of TS_{LA1}, the only way to spoof the tag is to construct a valid M_1. It is not possible, however, without guessing or computing a valid $W_1 = h((\alpha.g) \oplus (A^*_{Ti} \| ID_{Ti}) \| TS_{LA1})$, where TS_{LA1} is the attack time's timestamp. The enemy also lacks A_{Ti} and ID_{Ti}. As a result, the adversary's chance of successfully impersonating the tag is 2^{-n}, where n is the hash function's bit-length. To put it another way, the repeat attack is a waste of time.

Reader: □

Proof. Because the integrity of TS_{LA3} is guaranteed in our protocol, the adversary cannot replay messages to impersonate a reader. As a result, generating a legitimate $W_2 \| pid_{Rj}$ is the only way to impersonate the R_j in front of S. The opponent, on the other hand, lacks $W_2 = h(\alpha_{Rj} A_{Rj} TS_{LA3})$, W_2, and A_{Rj}. Even if she/he obtains the values W_2, pid_{Rj}, and A_{Rj} in some other way, she/he must extract α_{Rj} from M_2 in order to determine ID_{Rj}. It necessitates reverse engineering of the one-way hash function, which is a difficult challenge that makes the assault impracticable. As a result, impersonating R_j to S is not feasible under this protocol.

Server: □

Proof. To impersonate the server S in front of R_j, the adversary would have to compute $W_3, W_4, TS_{LA5}, \beta.g$, where $W_3 = h(\alpha.\beta.g \| SK_{ST} \| A_{Ti} \| pid_{Ti})$ and $W_4 = h(ID_{Rj} \| A_{Rj} \| TS_{LA5} \| pid_{Rj})$. $pid_{Rj}, ID_{Rj}, x_s.\alpha.g$ would be required. Aside from $x_s.\alpha.g, \beta.\alpha.g$, which is contributed by T_i through sending $\alpha.g$, this token is randomized by $x_s.\alpha.g, \beta.\alpha.g$. Solving a ECCDHP problem, which is a difficult problem, would be required for the disclosure of α and x_s. Even if the adversary reveals the band and adapts it appropriately, the adversary still needs to know ID_{Ti} due to TS_{LA5} in $h(\alpha.\beta.g \| SK_{ST} \| A_{Ti} \| pid_{Ti})$, which is not the case. As a result, cheating R_j and successfully mimicking S gives the opponent a 2^{-n} advantage. Furthermore, impersonating S in front of R_j is a prerequisite for impersonating S in front of T_i. As a result, the attacker cannot effectively impersonate the server S in front of T_i using R_j. Only $M_4 = \{W_3, \beta.g, TS_{LA5}\}$, where $W_3 = h(\alpha.\beta.g \| SK_{ST} \| A_{Ti} \| pid_{Ti})$. Unlikely as it may seem, the attacker lacks ID_{Ti}. As a result, the adversary's

advantage in committing this impersonation attack is negligible (i.e., 2^{-n}). □

Proposition 8. *Offline password guessing attack*

Proof. The rationale for security against this attack is nearly comparable to that of RSEAP2. In a nutshell, PWT = $h(ID_{Ti} \| (pw_{Ti} \oplus \alpha_{Ti}) \| pa_{Ti})$ calculates the tag's temporary password. Even if the adversary could estimate PWT, the value α_{Ti}, which is a random integer created by the tag T_i, is still required. As a result, the opponent who could not foresee α_{Ti} will be defeated by this assault. □

Proposition 9. *Desynchronization attack*

Proof. Because there is no updating phase of shared parameters after the protocol execution concludes, our proposed technique is immune to desynchronization assaults. The attacker may only block the M_4 message if the tag T_i is used to set the session key SK_{TS}/SK_{ST}. Because T_i has not received M_4 in a timely manner, this entity may need to restart the login and authentication step in order to reestablish the session key. We wish to underline that the aforementioned situation is distinct from an impersonation assault—as previously stated, an adversary cannot impersonate a valid tag. In addition, the tag T_i must start the protocol; otherwise, the server S would reject the request. □

Proposition 10. *Insider attack*

Proof. In the initialization phase of our scheme, T_i sends $M_{R1} = \{ID_{Ti}, \alpha_{Ti}, C_{Ti}\}$ to S and receives $M_{R2} = \{pid_{Ti}, A_{Ti}\}$ in return. Further computes, where PWT = $h(ID_{Ti} \| (pw_{Ti} \oplus \alpha_{Ti}) \| pa_{Ti})$. Likely, the chances for an insider attacker to disclose pw_{Ti} are almost null (i.e., 2^{-n}). □

Proposition 11. *Man-in-the-middle attack*

Proof. To carry out a successful man-in-the-middle attack, an adversary must be able to impersonate a protocol entity or modify a message without being discovered. Nonetheless, the aforementioned attack will fail in our suggested protocol for the following reasons. For starters, as we explained in Section 7, the adversary's advantage in impersonating the tag, the reader, or the server is insignificant. Second, we have shown (5) that any change to the transmitted message causes the receiver to reject the received message. Finally, we demonstrated how an opponent cannot properly relay a message to deceive about his distance or replay an earlier message in Sections 6. As a result, the suggested protocol is impenetrable to a man-in-the-middle assault. □

Proposition 12. *Ephemeral secret leakage (ESL) attack:*

Proof. As described in the Proposition 2, both T_i and S establish a common session key during the execution of the proposed scheme. The session key is computed as $SK_{TS} = h((ID_{Ti} \oplus A_{Ti}) \| (\alpha.\beta.g \| x_s.\alpha.g) \| t(sn_i \oplus (TS_{LA1} \| TS_{LA5})))$. The SK-security of the proposed scheme relies on the secret credentials as discussed in the following two cases:

Case 1. Let us consider $\mathscr{A}$ knows the ephemeral (short-term) secret credentials α and β. It is computationally infeasible for $\mathscr{A}$ to create the valid session key SK_{TS} without the knowledge of the long-term secrets ARj, A_{Ti}, α_{Rj}, and x_s.

Case 2. We assume that the long-term secrets ARj, A_{Ti}, α_{Rj}, and x_s some or all of them are revealed to $\mathscr{A}$, and the attacker $\mathscr{A}$'s task to generate SK_{TS} without the ephemeral secret credentials α and β this again turns out to be computationally infeasible task.

This shows that $\mathscr{A}$ can generate a valid session key SK_{TS} only if both the ephemeral and long-term secret credentials are revealed. Furthermore, if a particular session is compromised, the session key established in previous/future sessions are completely different to the compromised session key due to the application of both long-term secrets and newly generated random nonces, which are secret and not revealed to $\mathscr{A}$. Therefore, both forward as well as backward secrecy along with the SK-security are preserved in the proposed scheme. Moreover, in the proposed scheme, with the help of the session hijacking attack, a session key is leaked in a particular session; it has no affect to compromise the security of other previous as well as future sessions. By summing up all these cases, the proposed scheme is secure against the ESL attack. □

8. Observations and Performance Analysis

We use the implementation results in [2] "(CPU: Intel(R) Core(TM)2T6570 2.1 GHz, Memory: 4G, OS: Win7 32-bit, Software: Visual C++ 2008, MIRACL C/C++ Library)" to estimate the computation time. Because SHA-2 occupies 15.8 cycles per bytes [27], it takes $T_h^{\text{fun}} = 0.0004 * (15.8/11.4) = 0.0005$ milliseconds to compute. To be clear, the number T_h^{fun} corresponds to a single call to the SHA-2 compression function (fun). The SHA-2 compression function has a message-block length of 512 bits. We built the new protocol in detail to reduce the amount of calls to this compression function, particularly on the tag side, which is the most limited device. Finally, the time required to calculate scalar multiplication on ECC-160, represented by T^{EMP}, is 7.3529 milliseconds, whereas the time required to calculate a chaotic map is $T^{CH} = T^{\text{EMP}}$[28]. The needed time for encryption/decryption of a symmetric scheme T^{Sym} varies depending on the employed symmetric encryption method; however, the stated time for AES is $T^{\text{Sym}} = 0.1303$ milliseconds. The details are shown in Table 9.

The hash function output, nonces, timestamps, tag/reader identities, a symmetric encryption output block, and elliptic curve points all have bit widths of 160, 160, 32, 160, 128, and 320 bits, respectively, for the performance analysis. We compare the computational and communication expenses of RSEAP2 with our method in Table 10. Because tags are the most limited

Table 9: Approximate time required for various operations.

Notation	Description (Time to compute)	Approximate computation Time (in milliseconds)
T_h^{fun}	Hash function	0.0005
T^{EMP}	ECC point multiplication	7.3529
$T^{\text{Sym}_{\text{Enc/Dec}}}$	Symmetric encryption/decryption	0.1303
$T_{QR} \approx T^{\text{EMP}}$	QR code	7.3529
$T^{CH} \approx T^{\text{EMP}}$	Chaotic map	7.3529
$T_{\text{MAC}} \approx T_h^{\text{fun}}$	Message authentication code	0.0005

Table 10: Comparison of communication costs.

Scheme	Communication cost (sending mode)	Communication cost (receiving mode)	Computation (in milliseconds)	Time (ms)
Jiang et al. [3]	768	768	$9T_h^{\text{fun}} + 5T^{\text{Sys}} + 3T^{\text{EMP}}$	37.4205
Kumar et al. [7]	832	544	$9T_h^{\text{fun}} + 3T^{\text{EMP}}$	22.0632
Mishra et al. [11]	672	224	$4T_h^{\text{fun}} + 2T^{CH}$	14.7078
Jiang et al. [4]	1280	800	$9T_h^{\text{fun}} + T^{\text{Sys}} + 5T^{\text{EMP}}$	22.1935
Safkhani et al. [2]	672	512	$6T_h^{\text{fun}} + 3T^{\text{EMP}}$	22.0617
Our scheme	672	512	$6T_h^{\text{fun}} + 2T^{\text{EMP}}$	14.7088

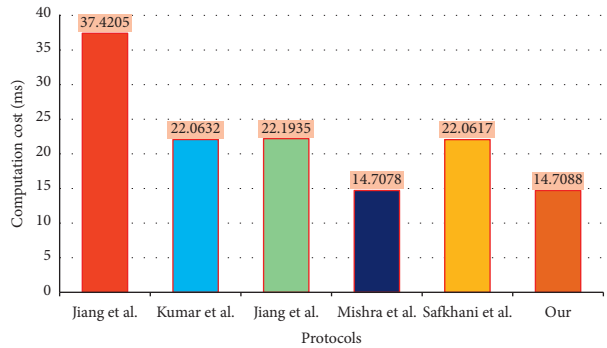


Figure 3: Computation cost comparison.

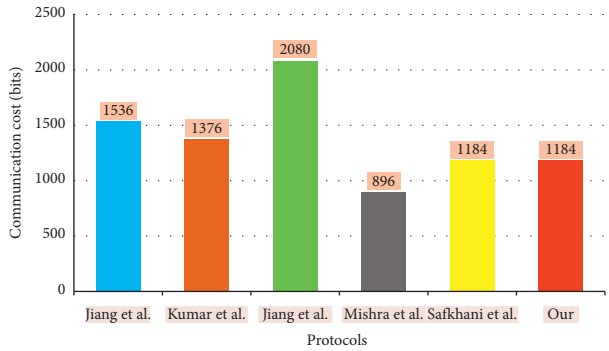


Figure 4: Communication cost comparison.

devices in the system, we focus our investigation on them. There are no major changes in consuming time when compared to RSEAP2, as shown in Figure 3, simply a minor improvement in our approach. Our scheme is much more efficient than RSEAP2 in terms of bits sent (and received), as shown in Figure 4. It entails a significant reduction in power consumption, which is a critical metric in such devices. Finally, in Table 11, we compare and contrast the security qualities afforded by comparable systems with our scheme (see Figure 5 for an instance). To summarize, the new protocol is more efficient and secure than the old one.

Table 11: Comparison of security features.

Security attributes	[7]	[3]	[4]	[2]	[11]	Our
Traceability preservation	×	√	√	√	√	√
Suitable for cloud environments	√	√	√	√	√	√
Password guessing attack	√	√	√	√	√	√
Privileged-insider attack	√	√	√	√	√	√
User anonymity preservation	×	√	√	√	√	√
Relay attack	×	×	×	√	×	√
Replay attack	√	√	√	√	√	√
Impersonation attacks	×	√	√	×	×	√
Denial-of-service attack	×	√	√	×	√	√
Message authentication	×	√	√	×	×	√
Mutual authentication	√	√	√	×	√	√
Man-in-the-middle attack	×	√	√	√	×	√
ESL attack	√	√	√	√	√	√
Session key agreement	√	√	√	×	√	√
Desynchronization attack	√	√	√	√	√	√
Revocability	×	×	×	×	×	√
Free password/biometric change	√	√	√	×	√	√
Security attributes achieved	9	15	15	10	12	18

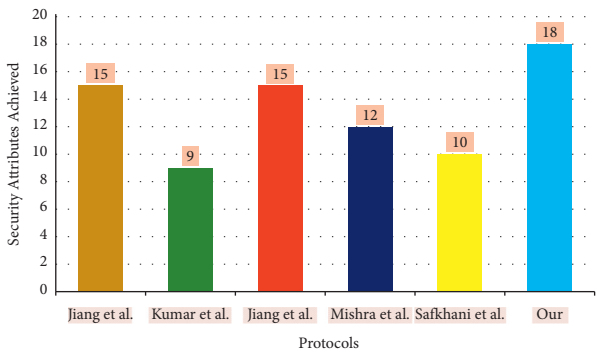


Figure 5: Security attribute comparison.

9. Concluding Remarks

In this article, we designed a PUF and RFID-based authentication protocol for vehicular cloud computing environment which ensure the secure communication among the participating entities such as tag, reader, and the cloud server. The uniqueness property of PUF and ECC allows significant functional advantages in ensuring and designing the secure key establishment and communication. Our proposed protocol efficiently supports for the revocation and reissue features and tag's friendly password update/change mechanism. Using the provable random oracle model, we presented the advantages of an adversary in violating the security features. Moreover, through the informal security analysis, we have shown that the proposed scheme successfully prevents all the well-known security attacks for authentication protocols. Our scheme withstands all the 18 security features and further consumes the computation cost of $6T_h^{\text{fun}} + 2T^{\text{EMP}} = 14.7088$ ms which is comparable with the other schemes. Similarly, our scheme consumes the communication cost as 672 bits during the sending mode and 512 bits during the receiving mode. Overall, the performance of our proposed scheme is comparable with the related schemes and provides more security features compared to the other related existing protocols.

Data Availability

No data collection method is applied.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study did not receive any funding in any form.

References

[1] P. Bagga, A. K. Das, M. Wazid, J. J. P. C. Rodrigues, K.-K. R. Choo, and Y. Park, "On the design of mutual authentication and key agreement protocol in internet of vehicles-enabled intelligent transportation system," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 2, pp. 1736–1751, 2021.

[2] M. Safkhani, C. Camara, P. Peris-Lopez, and N. Bagheri, "Rseap2: an enhanced version of RSEAP, an RFID based authentication protocol for vehicular cloud computing," *Vehicular Communications*, vol. 28, Article ID 100311, 2021.

[3] Q. Jiang, J. Ni, J. Ma, L. Yang, and X. Shen, "Integrated authentication and key agreement framework for vehicular cloud computing," *IEEE Network*, vol. 32, no. 3, pp. 28–35, 2018.

[4] Q. Jiang, N. Zhang, J. Ni, J. Ma, X. Ma, and K.-K. R. Choo, "Unified biometric privacy preserving three-factor authentication and key agreement for cloud-assisted autonomous vehicles," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 9, pp. 9390–9401, 2020.

[5] A. A. Alamr, F. Kausar, J. Kim, and C. Seo, "A secure ECC-based RFID mutual authentication protocol for internet of things," *The Journal of Supercomputing*, vol. 74, no. 9, pp. 4281–4294, 2018.

[6] N. Dinarvand and H. Barati, "An efficient and secure RFID authentication protocol using elliptic curve cryptography," *Wireless Networks*, vol. 25, no. 1, pp. 415–428, 2019.

[7] V. Kumar, M. Ahmad, D. Mishra, S. Kumari, and M. K. Khan, "RSEAP: RFID based secure and efficient authentication protocol for vehicular cloud computing," *Vehicular Communications*, vol. 22, Article ID 100213, 2020.

[8] M. Hosseinzadeh, O. H. Ahmed, S. H. Ahmed et al., "An enhanced authentication protocol for RFID systems," *IEEE Access*, vol. 8, 2020.

[9] F. Zhu, "Secmap: a secure RFID mutual authentication protocol for healthcare systems," *IEEE Access*, vol. 8, p. 192, 2020.

[10] S. Gabsi, Y. Kortli, V. Beroulle, Y. Kieffer, A. Alasiry, and B. Hamdi, "Novel ECC-based RFID mutual authentication protocol for emerging IoT applications," *IEEE Access*, vol. 9, 2021.

[11] D. Mishra, V. Kumar, D. Dharminder, and S. Rana, "SFVCC: chaotic map-based security framework for vehicular cloud computing," *IET Intelligent Transport Systems*, vol. 14, no. 4, pp. 241–249, 2020.

[12] U. Chatterjee, R. S. Chakraborty, and D. Mukhopadhyay, "A PUF-based secure communication protocol for IoT," *ACM Transactions on Embedded Computing Systems*, vol. 16, no. 3, pp. 1–25, 2017.

[13] M. N. Aman, K. C. Chua, and B. Sikdar, "Mutual authentication in IoT systems using physical unclonable functions," *IEEE Internet of Things Journal*, vol. 4, no. 5, pp. 1327–1340, 2017.

[14] Q. Jiang, X. Zhang, N. Zhang, Y. Tian, X. Ma, and J. Ma, "Three-factor authentication protocol using physical unclonable function for IoV," *Computer Communications*, vol. 173, pp. 45–55, 2021.

[15] H. Xu, X. Chen, F. Zhu, and P. Li, "A novel security authentication protocol based on physical unclonable function for RFID healthcare systems," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 8844178, 14 pages, 2021.

[16] P. Gope and B. Sikdar, "Privacy-aware authenticated key agreement scheme for secure smart grid communication," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 3953–3962, 2018.

[17] Y.-N. Cao, Y. Wang, Y. Ding, H. Zheng, Z. Guan, and H. Wang, "A PUF-based lightweight authenticated metering data collection scheme with privacy protection in smart grid," in *Proceedings of the 2021 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom)*, pp. 876–883, IEEE, New York City, NY, USA, October 2021.

[18] Z. Zhang, Y. Liu, Q. Zuo, L. Harn, S. Qiu, and Y. Cheng, "PUF-based key distribution in wireless sensor networks," *Computers, Materials & Continua*, vol. 64, no. 2, pp. 1261–1280, 2020.

[19] P. Mall, R. Amin, A. K. Das, M. T. Leung, and K.-K. R. Choo, "PUF-based authentication and key agreement protocols for IoT, WSNS and smart grids: a comprehensive survey," *IEEE Internet of Things Journal*, vol. 9, 2022.

[20] Y. Liu, Y. Cui, L. Harn et al., "PUF-based mutual-authenticated key distribution for dynamic sensor networks," *Security and Communication Networks*, vol. 2021, Article ID 5532683, 13 pages, 2021.

[21] D. Mukhopadhyay, "PUFs as promising tools for security in internet of things," *IEEE Design & Test*, vol. 33, no. 3, pp. 103–115, 2016.

[22] W. Wang, C. Qiu, Z. Yin et al., "Blockchain and PUF-based lightweight authentication protocol for wireless medical sensor networks," *IEEE Internet of Things Journal*, vol. 9, 2021.

[23] T.-F. Lee and W.-Y. Chen, "Lightweight fog computing-based authentication protocols using physically unclonable functions for internet of medical things," *Journal of Information Security and Applications*, vol. 59, Article ID 102817, 2021.

[24] V. Hassija, V. Chamola, V. Gupta, S. Jain, and N. Guizani, "A survey on supply chain security: application areas, security threats, and solution architectures," *IEEE Internet of Things Journal*, vol. 8, no. 8, pp. 6222–6246, 2020.

[25] X. Li, J. Niu, S. Kumari, F. Wu, A. K. Sangaiah, and K.-K. R. Choo, "A three-factor anonymous authentication scheme for wireless sensor networks in internet of things environments," *Journal of Network and Computer Applications*, vol. 103, pp. 194–204, 2018.

[26] D. Wang, H. Cheng, P. Wang, X. Huang, and G. Jian, "Zipf's law in passwords," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 11, pp. 2776–2791, 2017.

[27] W. Dai, "Crypto++ 5.6. 0 benchmarks," 2009, https://www.cryptopp.com/benchmarks.html.

[28] J. Srinivas, A. K. Das, N. Kumar, and J. J. P. C. Rodrigues, "Tcalas: temporal credential-based anonymous lightweight authentication scheme for internet of drones environment," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 7, pp. 6903–6916, 2019.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 1336094, 15 pages
https://doi.org/10.1155/2022/1336094

Research Article

Privacy-Aware Task Assignment for IoT Audit Applications on Collaborative Edge Devices

Linyuan Liu,[1] **Haibin Zhu**,[2] **Shenglei Chen**,[1] **and Zhiqiu Huang**[3]

[1]*Department of E-Commerce, Nanjing Audit University, Nanjing 211815, China*
[2]*Collaborative Systems Laboratory, Nipissing University, North Bay ON P1B8L7, Canada*
[3]*College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China*

Correspondence should be addressed to Linyuan Liu; liulinyuang@nau.edu.cn

Received 7 January 2022; Revised 18 April 2022; Accepted 1 June 2022; Published 21 June 2022

Academic Editor: Ke Gu

To meet the rapidly increasing demand for Internet of Things (IoT) applications, edge computing, as a novel computing paradigm, can combine devices at the edge of the network to collaboratively provide computing resources for IoT applications. However, the dynamic, heterogeneous, distributed, and resource-constrained nature of the edge computing paradigm also brings some problems, such as more serious privacy leakages and performance bottlenecks. Therefore, how to ensure that the resource requirements of the application are satisfied, while enhancing the protection of user privacy as much as possible, is a challenge for the task assignment of IoT applications. Aiming to address this challenge, we propose a privacy-aware IoT task assignment approach at the edge of the network. Firstly, we model the resource and privacy requirements for IoT applications and evaluate the resource satisfaction and privacy compatibility between edge devices and tasks. Secondly, we formulate the problem of privacy-aware IoT task assignment on edge devices (PITAE) and develop two solutions to the PITAE problem based on the greedy search algorithm and the Kuhn–Munkres (KM) algorithm. Finally, we conduct a series of simulation experiments to evaluate the proposed approach. The experimental results show that the PITAE problem can be solved effectively and efficiently.

1. Introduction

With the development of the Internet of Things (IoT), various IoT applications emerge as the times require, such as disaster relief, public safety, and face recognition [1–3]. According to Garner [4], the global IoT-enabled applications and infrastructure market will represent a 33 billion US dollar opportunity in 2025. IoT applications usually have a large amount of data that need to be processed in time. Hence, they have strict requirements on computing resources, response time, and privacy [5–7]. The traditional cloud-centric task processing model fails to meet these requirements, because it often needs to transmit a large amount of data to the cloud, which increases network transmission delay and network traffic [8].

To address the shortcomings of the cloud model, researchers have proposed edge computing [9, 10]. As a novel computing paradigm, edge computing can combine the resources of multiple devices at the edge of the network to provide task processing for IoT applications [11, 12]. With the wide adoption of wireless sensing and communication technology, a large number of IoT devices are emerging at the network edge, such as closed-circuit television (CCTV) cameras, smartphones, tablets, smart watches, smart home devices, and smart vehicles. Due to limited resources, these devices are generally only responsible for data collection and preprocessing, while complex data analysis work is offloaded to edge servers or cloud servers.

Supported by the advances in hardware and networking technologies, IoT devices are constantly increasing in resources and processing capabilities. They communicate with each other to collect and share data, and immediately process tasks near the data source [13, 14]. Edge computing has recently moved beyond the initial principle of utilizing IoT devices to collect and preprocess sensory data and is now able to combine and coordinate multiple IoT devices to

provide processing for IoT applications [1, 14, 15]. Therefore, the advantages of edge computing such as low latency and local data processing are further highlighted [2, 16]. In this paper, we refer to these resource-constrained IoT devices with data collection and task processing capabilities as edge devices.

Due to the dynamic, heterogeneous, and distributed nature of the edge computing paradigm, edge devices are generally owned by individuals with different interests and affiliations [17]. As a result, the owner of edge devices may illegally use and disclose the user privacy information hidden in IoT data, e.g., faces, motions, locations, etc., resulting in serious privacy leakages [18, 19].

Consider a data-intensive IoT application consisting of multiple interrelated tasks, where each task has different resource requirements, e.g., CPU, memory, storage, bandwidth, etc. To protect user privacy, each private data in the task specify a set of privacy requirements. Correspondingly, each edge device has a set of available resources and provides a set of privacy policies. An important prerequisite for an edge device to be qualified to execute an IoT task is that it must satisfy the resource and privacy requirements of the task. Moreover, a single edge device is difficult to process relatively complex computations due to limited resources. Consequently, multiple tasks of an IoT application need to be assigned to multiple edge devices for execution. In summary, how to assign tasks to multiple edge devices that satisfy resource and privacy requirements is an important challenge in task assignment for IoT applications.

In the research of task assignment for IoT applications, some useful approaches were proposed to offload tasks to cloud, fog, and edge [1, 20–22]. However, most of them regard the task assignment from the perspective of resources and quality of service (QoS), while ignoring the privacy requirements of the users. Moreover, some researches focus on the privacy-aware IoT task assignment. They mainly adopt various privacy technologies like differential privacy, data generalization, task fragmentation, and privacy conflict avoidance to control data access [23–26], but they are inadequate to address the issue of how private data will be used after being accessed, such as the purpose of using the data, the retention time of the data, and the operations executed on the data.

Inspired by these works, in this paper, we propose a privacy-aware IoT task assignment approach at the edge of the network, which assigns IoT tasks to multiple edge devices close to the data source. These devices do not rely on a central coordinator and collaborate to process IoT tasks in a distributed manner. Specifically, we first model the resource and privacy requirements of the IoT tasks and evaluate whether the edge devices can satisfy these requirements. Then, we formulate the problem of privacy-aware IoT task assignment on edge devices (PITAE) as an optimization problem to maximize the privacy compatibility degree between IoT tasks and edge devices. Furthermore, we develop two solutions based on the greedy search algorithm and the KM algorithm [27, 28] to solve the problem. The main contributions of this paper are as follows:

(1) An integer programming optimization model is used to formulate the PITAE problem considering both the resource and privacy constraints.

(2) A privacy model is presented to specify the privacy requirements and privacy policies, and the weighted Euclidean distance is employed to measure the privacy compatibility degree between edge devices and tasks.

(3) Two solutions based on greedy search and KM algorithm are developed to solve the PITAE problem. The experimental results demonstrate that the proposed approaches can significantly improve the privacy compatibility degree of the solution compared with the benchmark approach.

The rest of this paper is structured as follows. Section 2 describes the motivation and framework of the PITAE problem. Section 3 formally specifies the PITAE problem. Section 4 presents two solutions to solve the PITAE problem. The experiments and results are illustrated in Section 5. The related work is reviewed in Section 6. Finally, the conclusion and further works are given in Section 7.

2. Motivation and Framework

In this section, we show an audit example of emergency supply distribution in a disaster relief scenario. In such a scenario, emergency supplies are usually ample in quantity and variety, and the distribution time is urgent. Therefore, it is a very complicated task for traditional manual audit methods to handle. An IoT-based audit application can quickly and automatically execute this process. Such a process captures emergency supply distribution videos stored in CCTV cameras and uses nearby edge devices to analyze the videos to automatically identify some violations, e.g., fake or erroneous emergency supply distribution.

As shown in Figure 1, the workflow of the IoT audit application includes six tasks (t_0-t_5): data collection, object detection, face recognition, supply recognition, violation analysis, and alarm and report. Firstly, task t_0 collects data required for subsequent tasks, e.g., supply distribution video, supply distribution location, and supply application form. Secondly, t_1 uses video data as input to execute object detection and sends the detected face and supply images to t_2 and t_3, respectively. Thirdly, t_2 recognizes the face image to obtain personal identity information (PII), t_3 recognizes the type and quantity of supplies, and t_4 conducts violation analysis based on the recognition results, location, and supply application form. Finally, t_5 issues an alert based on the violation result and generates an audit report. In Figure 1, the rectangular boxes represent tasks, the arrows represent the invocation of the tasks within the application workflow, the vertical solid lines mean that all the previous tasks should be accomplished before the next task is initiated, and the workflow starts from the left and ends at the right.

This example is a typical data-intensive IoT application. The input data of each task may involve the user's private

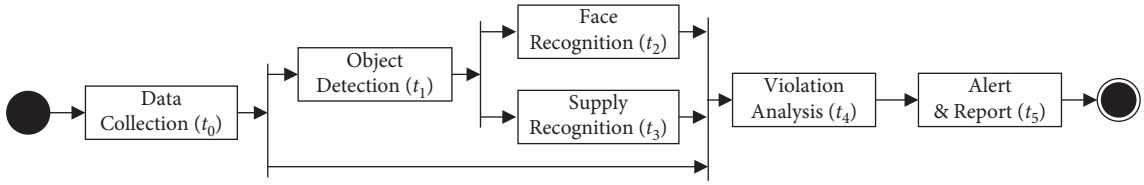


Figure 1: An example IoT audit application.

data, e.g., face, location, application form, etc. In addition, each task needs to be assigned to edge devices with different available resources, e.g., CPU, memory, storage, bandwidth, etc. The resources and private data required for each task are shown in Table 1. There are 10 available edge devices (d_0-d_9) in the demonstration scenario, and the available resources of each device are shown in Table 2.

Before assigning IoT tasks to edge devices, it is necessary to evaluate whether these devices can satisfy the resource requirements of the tasks [29, 30]. As shown in Tables 1 and 2, d_0 only satisfies the resource requirements of t_5, while d_9 can satisfy the resource requirements of all the tasks.

According to the General Data Protection Regulation (GDPR) [31], data consumers can only collect private data for legal purposes. At the same time, the GDPR also requires data consumers not to use the collected data for other purposes, and the retention time of the data and the operations executed on the data must be consistent with those necessary for the stated purpose. To comply with GDPR, private data in Table 1 have a set of privacy requirements, e.g., the sensitivity of the data, the purpose of using the data, the retention time of the data, and the operations executed on the data. Correspondingly, each edge device also provides a set of privacy policies. Therefore, another prerequisite for assigning tasks to edge devices is that the privacy policies of the edge devices should be compatible with the privacy requirements of the tasks. The higher the privacy compatibility degree between the edge device and the task, the more suitable the edge device is to undertake the task.

For example, a privacy requirement of the video data in task t_0 is <video, 0.8, data collection, {read, transfer}, 1>. It means that the sensitivity degree of video is 0.8, and an edge device can only execute read and transfer operations on the video for the purpose of data collection. At the same time, it also requires that the trust degree of the device must be greater than or equal to 0.8 (sensitivity degree), and video cannot be retained more than 1 month. Correspondingly, a privacy policy of edge device d_2 for the video data is <video, 0.6, data collection, {read, write, transfer, profiling}, 12>. It indicates that the trust degree of d_2 is 0.6, d_2 will execute read, write, transfer, and profiling operations on the video for the purpose of data collection, and d_2 will retain the video for at least 12 months. As can be seen from this example, the privacy policy of d_2 is incompatible with the privacy requirement of t_0 in terms of sensitivity degree, operations executed, and retention time.

In summary, the task assignment problem of IoT audit applications is to assign multiple tasks to suitable edge devices, so as to satisfy the resource requirements of the tasks while maximizing the overall privacy compatibility of the assigned edge devices.

Based on the above example, the privacy-aware IoT task assignment framework at the edge of the network is shown in Figure 2. In Figure 2, the developer designs an IoT application based on resource requirements, privacy requirements, tasks, and their dependencies. The tasks of the IoT application need to be deployed to qualified edge devices for execution. Each edge device contains an available resource description file, a privacy policy description file, and is equipped with a task assignment manager responsible for device discovery, qualification evaluation, task assignment, and coordination.

The framework in Figure 2 does not depend on a central coordinator and supports distributed task assignment. Therefore, each participating edge device of IoT applications can generally play the role of coordinator or collaborator. To protect privacy and reduce network transmission, all edge devices participating in the application should be as close as possible to the data source. The IoT application shown in Figure 1 is a typical stream data processing application, and the data collection device (e.g., CCTV camera) is the data production source of the application. Therefore, the application developer selects it as the coordinator of the application, which delivers offloading requests to nearby edge devices. If there are multiple data collection devices (i.e., multiple data sources) in an application, the application developer will select an edge device with large data volume and high privacy protection requirements from these devices as the coordinator.

The coordinator is responsible for discovering a set of qualified edge devices from nearby and forming a collaborative group with these devices as its collaborators, and offloading tasks to these collaborators at the same time. Specifically, once the coordinator receives the deployment request of an IoT application, it will advertise the task processing request to nearby edge devices. The edge devices that are willing to participate in the collaboration accept the request and reply their available resources status and privacy policies to the coordinator. Then, the coordinator evaluates the resource satisfaction and privacy compatibility of each collaborative device. More specifically, the application developer sets a privacy compatibility threshold for tasks. During the privacy evaluation process, if the privacy compatibility between the task and all its candidate devices fails to satisfy the threshold constraints, the coordinator will request the application developer to relax the privacy threshold to ensure that the task has enough devices to perform its function. Finally, the coordinator assigns tasks to the most suitable set of devices according to the evaluation

Table 1: Resource requirements and private data request for tasks.

Tasks	Resource requirements				Private data
	CPU (GHz)	Memory (GB)	Storage (TB)	Bandwidth (Mbps)	
t_0	1.4	4	0.6	18	Video, location, application form
t_1	1.8	6	0.5	18	Video
t_2	1.8	8	0.4	15	Face image
t_3	1.8	8	0.4	15	
t_4	1.6	6	0.6	12	PII, location, application form
t_5	1.2	2	0.2	10	Violation result

Table 2: Available resources provided by edge devices.

Edge devices	Available resources			
	CPU (GHz)	Memory (GB)	Storage (TB)	Bandwidth (Mbps)
d_0	1.2	2	0.2	12
d_1	1.4	4	0.4	18
d_2	1.6	6	0.6	18
d_3	1.8	8	0.8	20
d_4	2.0	10	1.0	22
d_5	1.6	8	0.6	18
d_6	1.8	10	0.8	20
d_7	2.0	12	1.0	22
d_8	2.2	14	1.2	25
d_9	2.5	16	1.5	28

results to maximize the overall privacy compatibility degree of the collaboration group.

After the collaboration group is established, each device starts to execute the assigned tasks. The coordinator is responsible for managing and coordinating the execution of all tasks, and periodically scanning the network to discover new edge devices. Once an edge device leaves the collaboration group, the coordinator will invite a new device to join the collaboration group and assign a task to it. Considering that a task may have multiple candidate new devices, the coordinator first evaluates the resource satisfaction and privacy compatibility between these devices and the task, and then selects the one with the highest privacy compatibility degree for the task from the qualified devices.

3. Problem Description

3.1. Application Model. A typical IoT application is defined by the developer at design time. It specifies the functional and nonfunctional requirements. Formally, it is described by a directed acyclic graph $G = (T, E)$, nodes $T = \{t_0, t_1, .., t_{n-1}\}$ represent a set of tasks where t_j $(0 \le j < n)$ is the jth task, and edges $E = \{(t_g, t_h) | t_g, t_h \in T\}$ are a set of links between tasks, which represent data and task dependencies. Each task t_j is characterized by a set of inputs $\mathrm{IN}_j = \{\mathrm{in}_j^0, \mathrm{in}_j^1, \ldots\}$, a set of outputs $\mathrm{OUT}_j = \{\mathrm{out}_j^0, \mathrm{out}_j^1, \ldots\}$, and a set of resource and privacy requirements.

(1) Resource requirements RR_j: RR_j represents a set of resources required to execute task t_j, such as CPU, memory, storage, and bandwidth. $RR_j = \{rr_j^0, rr_j^1, \ldots, rr_j^o\}$, where rr_j^c $(0 \le j < n, 0 \le c < o)$ is the requirement of task t_j for the cth resource.

(2) Privacy requirements PR_j: Let PD be a set of private data of the user in an IoT application. $PR_j = \{pr_j^0, pr_j^1, \ldots, pr_j^p\}$ specifies a set of privacy requirements for task t_j, where pr_j^k $(0 \le j < n, 0 \le k < p)$ is the kth privacy requirement of t_j, it is defined as a tuple $< pd_j^k, sd_j^k, pu_j^k, OP_j^k re_j^k >$, where $pd_j^k \in PD$ is a private data item of the user, $sd_j^k \in [0, 1]$ is the sensitivity degree of pd_j^k, it specifies the trust degree that an edge device must have when it uses pd_j^k, $sd_j^k = 0$ indicates the lowest sensitivity and 1 the highest, pu_j^k specifies the purpose for which the pd_j^k can be used, OP_j^k specifies a set of operations that can be executed on the pd_j^k, and re_j^k specifies the longest time that the edge device can retain pd_j^k.

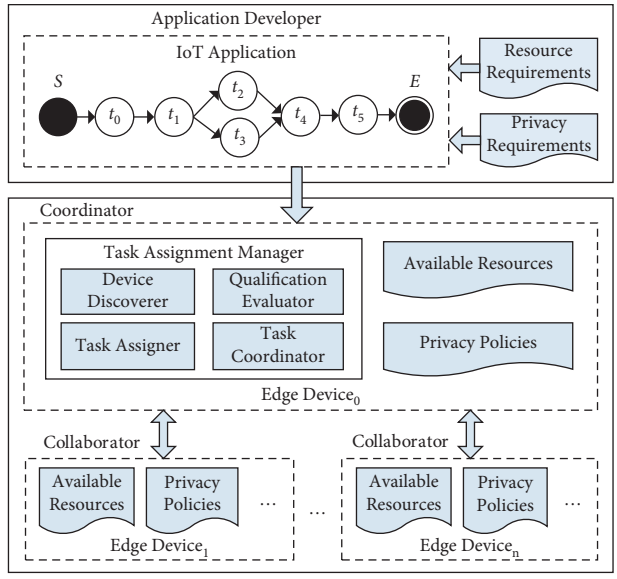


Figure 2: Privacy-aware IoT task assignment framework.

3.2. System Model. A PITAE scenario usually consists of multiple heterogeneous edge devices that communicate with each other and collaborate to execute multiple tasks of an IoT application. Let $D = \{d_0, d_1, .., d_{m-1}\}$represent a set of edge

devices, where d_i $(0 \le i < m)$ is the ith edge device. Each edge device d_i is characterized by a set of available resources and a set of privacy policies.

(1) Available resources AR_i: $AR_i = \{ar_i^0, ar_i^1, \ldots ar_i^{o-1}\}$ represents a set of available resources of d_i, where ar_i^c $(0 \le i < m,\ 0 \le c < o)$ is the cth resource of d_i.

(2) Privacy policies PP_i: $PP_i = \{pp_i^0, pp_i^1, \ldots, pp_i^{q-1}\}$ represents a set of privacy policies of d_i, where pp_i^l $(0 \le i < m,\ 0 \le l < q)$ is the lth privacy policy of d_i. Each privacy policy pp_i^l is defined as a tuple $< pd_i^l, td_i^l, pu_i^l, OP_i^l, re_i^l >$, where $pd_i^l \in PD$ is a private data item for which the policy is defined, $td_i^l \in [0, 1]$ is the trust degree of d_i, where 0 indicates complete no-trust and 1 complete trust, the larger the value of td_i^l, the stronger is the privacy protection provided by the d_i, pu_i^l is the purpose for d_i using pd_i^l, OP_i^l is a set of operations executed by d_i on the pd_i^l, and re_i^l is the time for d_i to retain pd_i^l.

Example 1. Figure 3 demonstrates a privacy-aware IoT task assignment model including 3 tasks and 6 edge devices. That is, $T = \{t_0, t_1, t_2\}$ and $D = \{d_0, d_1, d_2, d_3, d_4, d_5\}$. In Figure 3, circles represent IoT tasks, rectangles represent edge devices, and dashed lines represent potential assignments between tasks and edge devices. The dashed rectangles show the resources requirements and privacy requirements of each task, and the available resources and privacy policies of each device. The prerequisite for whether a task can be assigned to an edge device is that the device can satisfy the resource and privacy requirements of the task.

3.3. Qualification Evaluation Model. To determine whether the edge device d_i is qualified to execute the task t_j, it is necessary to evaluate the resource satisfaction and privacy compatibility between d_i and t_j. The specific evaluation process is as follows:

(1) Resource satisfaction evaluation. Considering that RR_j is a set of minimum resources required to fulfill task t_j, if the edge device d_i is a qualified edge device for t_j, then the available resources AR_i of d_i must satisfy the requirements RR_j. The resource satisfaction evaluation $f_{i,j}^R$ is obtained by

$$f_{i,j}^R = \begin{cases} 1, & \text{if } \forall c \in o, ar_i^c \ge rr_j^c \\ 0, & \text{otherwise,} \end{cases} \quad \text{where } ar_i^c \in AR_i, rr_j^c \in RR_j \tag{1}$$

(2) Privacy compatibility evaluation. The privacy compatibility degree between the edge device d_i and the task t_j is measured by the average compatibility degree of the privacy requirements of t_j with the corresponding privacy policies in d_i, and it is evaluated by

$$f_{i,j}^P = \frac{\sum_{k=0}^{p-1} f_{i,j}^k}{p}, \tag{2}$$

where $f_{i,j}^k \in [0, 1]$; it represents the privacy compatibility degree between the kth privacy requirement pr_j^k of t_j and the corresponding privacy policy pp_i^l in d_i, and p expresses the number of privacy requirements of t_j, which is an integer greater than or equal to 0.

To evaluate the compatibility degree between pr_j^k and pp_i^l, firstly, it is necessary to ensure that the private data and its usage purpose are consistent, e.g., $pd_j^k = pd_i^l$, $pu_j^k = pu_i^l$; secondly, it is necessary to measure the compatibility degree between pr_j^k and pp_i^l in terms of the sensitivity attribute, operation attribute, and retention time attribute. Accordingly, we express pr_j^k's privacy attributes sd_j^k, OP_j^k, and re_j^k and pp_i^l's privacy attributes td_i^l, OP_i^l, and re_i^l as two three-dimensional vectors. The work in [32] adopts Euclidean distance to evaluate the Security Service-Level Agreement (Security-SLA) between cloud users and cloud service providers. Inspired by this work, we employ the Euclidean distance to measure the compatibility degree between the two privacy attribute vectors. More specifically, considering that the different privacy attributes play different roles in the measurement process, we use the weighted Euclidean distance to reflect the difference in the importance of different attributes. Based on the above analysis, the privacy compatibility degree $f_{i,j}^k$ is calculated by

$$f_{i,j}^k = \begin{cases} \sqrt{w_1 \times \left(f_{i,j}^{k,sd}\right)^2 + w_2 \times \left(f_{i,j}^{k,OP}\right)^2 + w_3 \times \left(f_{i,j}^{k,re}\right)^2}, & \text{if } pd_j^k = pd_i^l \wedge pu_j^k = pu_i^l, \\ 0, & \text{otherwise,} \end{cases} \tag{3}$$

where w_1, w_2, and w_3 are three weight parameters, $w_1 + w_2 + w_3 = 1$. $f_{i,j}^{k,sd}$, $f_{i,j}^{k,OP}$, and $f_{i,j}^{k,re}$ are the compatibility degrees of the sensitivity attribute, operation attribute, and retention time attribute, respectively. The compatibility degree of sensitivity attribute is obtained by

$$f_{i,j}^{k,sd} = \begin{cases} td_i^l - sd_j^k, & \text{if } td_i^l \ge sd_j^k, \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

The compatibility degree of operation attribute is obtained by

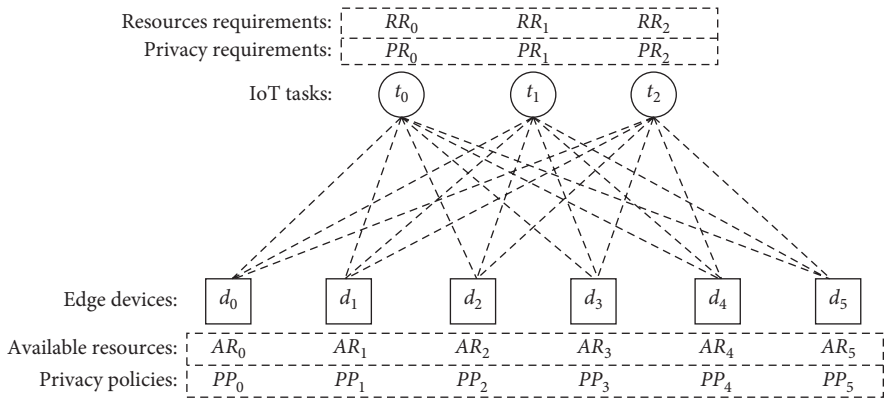


Figure 3: An illustration of the privacy-aware IoT task assignment model.

$$f_{i,j}^{k,OP} = \begin{cases} \dfrac{\left(\left|OP_j^k\right| - \left|OP_i^l\right|\right)}{\left|OP_j^k\right|}, & \text{if } OP_i^l \subseteq OP_j^k, \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

The compatibility degree of retention time attribute is obtained by

$$f_{i,j}^{k,re} = \begin{cases} \dfrac{\left(re_j^k - re_i^l\right)}{re_j^k}, & \text{if } re_i^l \le re_j^k, \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

Example 2. Assume that the evaluation results of resource satisfaction and privacy compatibility between tasks and edge devices in Figure 3 are shown in Figures 4(a) and 4(b), respectively. Figure 4(c) shows the potential task assignments that satisfy qualification requirements, where the values on the dotted line represent the degree of privacy compatibility.

3.4. Problem Definition. Despite the ever-increasing resources of edge devices, they are still considered resource-constrained and often unable to execute complex data processing workflow [1]. Hence, the tasks of an IoT application need to be assigned to multiple edge devices for execution. During the task assignment process, if multiple tasks are assigned to an edge device, the available resources of the device may not be able to meet the resource requirements of these tasks. Furthermore, when the device undertakes multiple tasks at the same time, it will collect multiple pieces of private data from different tasks and may infer more privacy information through data mining and machine learning techniques [33]. To meet resource constraints and protect user privacy, in this paper, we assign only one task to each edge device.

Due to the dynamic and distributed nature of edge environments, unpredictable link/device failures and churn of mobile and portable devices often result in IoT applications that are not able to run stably and reliably [34, 35]. To enhance the reliability of the IoT applications, we consider assigning each task to multiple edge devices. that is, the task is backed up to multiple edge devices. When a device that undertakes the task cannot work, the backup device can also ensure the task is executed properly.

In summary, whether a task can be assigned to an edge device is a big issue. If and only if the device satisfies the task's resource requirements and privacy compatibility degree constraint, then the task can be assigned to this device. Given n tasks and m edge devices, the PITAE problem aims to find a solution with maximum privacy compatibility degree by assigning IoT tasks to qualified edge devices. To illustrate the PITAE problem, specific data structures can be formalized as follows:

(1) Lower bound vector of tasks B: It is an n-dimensional vector, where $B[j]$ $(0 \le j < n)$ expresses how many edge devices must be assigned to task t_j. $B[j] > 1$ means that t_j requires multiple edge devices for execution.

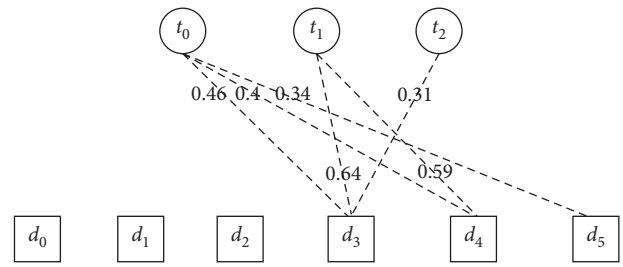
It is worth noting that the application developer does not know the failure and churn rates of edge devices when designing applications. Hence, how to properly set $B[j]$ is nontrivial, which is out of the scope of this paper. We may need to conduct a thorough investigation of this topic in the future. Here, we point out a few initial considerations that require attentions. To enhance the reliability of the IoT applications, each task generally needs to create 2-3 instances: a main task and 1-2 task replicas, and the main task and task replicas are assigned to different edge devices, i.e., $B[j] \le 3$. We present a $B[j]$ setting scheme as follows: firstly, the application developer preliminarily estimates the average failure and churn rates of edge devices based on experience. Secondly, the application developer determines $B[j]$ by comprehensively considering the average failure and churn rates of the devices, and the criticality of the task t_j. Thirdly, during the task assignment process, if a feasible task assignment solution cannot be found due to some tasks being restricted by B, the application developer will adjust B for these tasks and start a new round of task assignment.

	t_0	t_1	t_2
d_0	0	0	0
d_1	0	0	0
d_2	1	0	0
d_3	1	1	1
d_4	1	1	1
d_5	1	0	0

(a)

	t_0	t_1	t_2
d_0	0.18	0.82	0.26
d_1	0.16	0.34	0.23
d_2	0.22	0.34	0.35
d_3	0.46	0.64	0.31
d_4	0.40	0.59	0.08
d_5	0.34	0.32	0.75

(b)



(c)

Figure 4: (a) Resource evaluation results. (b) Privacy evaluation results. (c) Potential task assignments that satisfy qualification requirements.

(2) Privacy compatibility matrix C: It is an $m \times n$ matrix, where $C[i, j] = f_{i,j}^{P}$ $(0 \le i < m,\ 0 \le j < n)$ denotes the privacy compatibility degree between the edge devices d_i and the task t_j.

(3) Evaluation matrix E: It is an $m \times n$ matrix, where $E[i, j](0 \le i < m,\ 0 \le j < n)$ expresses whether the edge device d_i satisfies the resource and privacy compatibility threshold constraints of task t_j, and $E[i, j] = 1$ means yes and 0 no. $E[i, j]$ is obtained by

$$E[i, j] = \begin{cases} 1, & \text{if } f_{i,j}^{P} \ge th \wedge f_{i,j}^{R} = 1, \\ 0, & \text{otherwise}, \end{cases} \tag{7}$$

where $th \in [0, 1]$ is the privacy compatibility threshold, which specifies the minimum privacy compatibility degree that the edge devices must have when executing tasks.

(4) Assignment matrix A: It is an $m \times n$ matrix, where $A[i, j](0 \le i < m, 0 \le j < n) \in \{0, 1\}$ expresses whether t_j is assigned to the edge device d_i ($A[i, j] = 1$) or not ($A[i, j] = 0$).

Given B, C, and E, the PITAE problem is to find a matrix A to Max:

$$\sum_{i=0}^{m-1} \sum_{j=0}^{n-1} C[i, j] \times A[i, j]. \tag{8}$$

subject to

$$A[i, j] \in \{0, 1\}\,(0 \le i < m, 0 \le j < \mathrm{n}), \tag{9}$$

$$\sum_{i=0}^{m-1} A[i, j] = B[j]\,(0 \le j < n), \tag{10}$$

$$\sum_{j=0}^{n-1} A[i, j] \le 1\,(0 \le i < m), \tag{11}$$

$$E[i, j] \times A[i, j] > 0\,(0 \le i < m, 0 \le j < \mathrm{n}), \tag{12}$$

where Constraint (9) specifies that the decision variables are binary; Constraint (10) guarantees that each task is assigned $B[j]$ edge devices; Constraint (11)ensures that each edge device can only be assigned to one task; and Constraint (12) ensures that each assigned edge device satisfies the resource and privacy compatibility threshold constraints.

Example 3. In an IoT audit application, the resource requirements of tasks and the available resources provided by edge devices are shown in Tables 1–2. Assume that the privacy compatibility threshold th is specified as 0.3, the lower bound vector of tasks $B = [1, 1, 2, 2, 1, 1]$, and the privacy compatibility matrix is shown in Figure 5(a). The evaluation matrix is obtained by Equation (7), as shown in Figure 5(b). Based on B and Figure 5(a) and 5(b), the assignment solution with the maximal privacy compatibility degree (5.03) should be $\{d_6, d_3, \{d_7, d_9\}, \{d_4, d_8\}, d_2, d_1\}$, and the assignment matrix is demonstrated in Figure 5(c).

4. Solutions to the PITAE Problem

The PITAE problem is a typical one-to-many task assignment problem. If the exhaustive search method is used to solve this problem, the solution space can be up to $O(m^n)$ [36]. Therefore, we first develop a task assignment solution based on the greedy search to solve this problem. Then, to improve the effectiveness of task assignment, we propose a task assignment solution based on the KM algorithm to find the optimal solution to the PITAE problem.

4.1. Greedy Search-Based Task Assignment (GSTA) Solution. The GSTA solution selects $B[j]$ the most qualified edge devices for each task in the task set T according to the privacy compatibility matrix C and the evaluation matrix E. Specifically, for each t_j belonging to T and d_i belonging to D, it first evaluates whether d_i satisfies the resource requirements and privacy compatibility threshold constraints of task t_j, e.g., $E[i, j] = 1$. Then, it determines whether d_i has been assigned a task, e.g., $S[i] = 1$. If yes, it skips d_i and examines the next edge device; otherwise, it adds the privacy compatible degree $C[i, j]$ to the candidate edge device vector V of t_j. Subsequently, it reversely sorts all candidate edge devices in V according to their privacy compatibility degrees and selects top $B[j]$ candidate edge devices for t_j from sorted candidate edge device vector SV. Finally, it sets the assignment $A[i, j]$ corresponding to the edge device d_i and task

$$\begin{bmatrix} 0.18 & 0.82 & 0.26 & 1.00 & 0.45 & 0.05 \\ 0.16 & 0.34 & 0.23 & 1.00 & 0.55 & 0.58 \\ 0.22 & 0.34 & 0.35 & 1.00 & 0.47 & 0.16 \\ 0.46 & 0.64 & 0.31 & 1.00 & 0.37 & 0.58 \\ 0.40 & 0.59 & 0.08 & 1.00 & 0.48 & 0.73 \\ 0.34 & 0.32 & 0.75 & 1.00 & 0.20 & 0.39 \\ 0.62 & 0.35 & 0.44 & 1.00 & 0.39 & 0.62 \\ 0.40 & 0.52 & 0.39 & 1.00 & 0.60 & 0.24 \\ 0.43 & 0.40 & 0.27 & 1.00 & 0.50 & 0.45 \\ 0.39 & 0.41 & 0.33 & 1.00 & 0.57 & 0.38 \end{bmatrix} \quad \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \quad \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

(a) (b) (c)

Figure 5: Matrixes. (a) The privacy compatibility matrix. (b) The evaluation matrix. (c) The assignment matrix.

t_j to 1, and updates the edge device selection vector S. The details of GSTA are described in Algorithm 1.

The time complexity of Algorithm 1 is $O(n \times m + n \times m \times \log_2 m)$, where $O(m \times \log_2 m)$ is the time complexity of a sorting operation.

4.2. KM Algorithm-Based Task Assignment (KMTA) Solution. In a data-intensive IoT application, its workflow is usually composed of tens of tasks, rarely hundreds or thousands [1]. The IoT audit application is a typical data-intensive IoT application, and we estimate its number of tasks to be on the order of tens of magnitudes. Moreover, to enhance the reliability of the application, each task generally needs to be assigned to 2-3 edge devices, e.g., $B[j] \leq 3$. Therefore, the total number of edge devices required for an IoT audit application should be around tens to two hundred. On the other hand, with the widespread application of the IoT technology, there are often hundreds of IoT devices connected to the edge network near the data source. Based on the above considerations, in the PITAE scenario, we believe that the number of edge devices can meet the needs of IoT tasks, e.g., $m > n$, and each task requires $B[j]$ edge devices to execute it, but each edge device can only be assigned to one task.

The well-known KM algorithm can quickly solve standard task assignment problems, i.e., one-to-one task assignment problems, and the time complexity is $O(m^3)$ [27, 28]. In addition, the KM algorithm always finds the solution with the smallest sum [29]. However, the PITAE problem needs to find a solution with the maximum privacy compatibility degree. Furthermore, the KM algorithm can always find a result for the PITAE problem, but the result may not be a feasible solution. For example, when the edge device d_i cannot satisfy the resource requirements or the privacy compatibility threshold constraints of t_j, i.e., $E[i, j] = 0$, the KM algorithm may produce incorrect task assignments, leading to an infeasible solution.

To deal with the limitations of the KM algorithm, the KMTA solution improves the KM algorithm to solve the PITAE problem by adding virtual tasks and adjusting the privacy compatibility degrees between tasks and edge devices. Concretely, first of all, for each d_i belonging to D and t_j belonging to T, it evaluates whether d_i satisfies the resource requirements and privacy compatibility threshold constraints of t_j, and adjusts the privacy compatibility value $C[i,j]$ according to the evaluation result. More specially, if d_i passes the evaluation, e.g., $E[i, j] = 1$, it adjusts $C[i, j]$ to mpc-$C[i, j]$; otherwise, it adjusts $C[i, j]$ to $\sum_{j=0}^{n-1} B[j]$. The adjustment operation ensures that KMTA can find the solution with the maximum privacy compatibility degree, because mpc is the maximum privacy compatibility value in C, $C[i, j] \in [0, 1]$, and the privacy compatibility degree of a solution never exceeds $\sum_{j=0}^{n-1} B[j]$. Secondly, it extends matrix C into an m rows and $\sum_{j=0}^{n-1} B[j]$ columns matrix C^*, where for each column j in C, there are $B[j]$ corresponding copy columns in C^*. If the number of rows of C^* is greater than the number of columns, i.e., $m > \sum_{j=0}^{n-1} B[j]$, it adds $m - \sum_{j=0}^{n-1} B[j]$ virtual columns to C^* and sets their privacy compatibility value to 0. Thirdly, it calls the KM algorithm to obtain a temporary matrix H and forms the assignment matrix A according to H. Finally, it checks whether A is a feasible assignment solution. If each assignment in A is correct and each task is assigned $B[j]$ edge devices, it returns success; otherwise, it returns failure. The details of KMTA are shown in Algorithm 2.

The time complexity of Algorithm 2 is determined by the following: (1) the time complexity of adjusting the C matrix is $O(m \times n)$; (2) the time complexity of extending the C matrix is $O(m \times n \times B[j]) + O(m \times (m - \sum_{j=0}^{n-1} B[j]))$; (3) the time complexity of calling the KM algorithm and forming the assignment solution is $O(m^3) + O(m \times n)$; and (4) the time complexity of judging the feasibility of the solution is $O(m \times n) + O(n)$. Thus, the overall complexity of Algorithm 2 is $O(m^3) + O(m \times n \times B[j]) + O(m^2) + O(m \times n) + O(m - \sum_{j=0}^{n-1} B[j]) + O(n)$. In the presented scenarios, $B[j]$ is a constant (typically less than 10), and $m > n$. Consequently, the time complexity of Algorithm 2 can be simplified as $O(m^3)$.

Input:
T: the tasks set; D: the edge devices set; B: the lower bound vector;
C: the compatibility matrix; E: the evaluation matrix; S: the edge device selection vector.
Output:
A: the task assignment matrix.

```
for each task t_j in T do
    for each edge device d_i in D do
        if E[i, j] = 1 then
            if S[i] = 1 then;
                skip it and examine the next edge device;
            else
                V ← C[i, j];
            end if
        end if
    end for
    SV ← sorting V based on privacy compatibility degree;
    Select Top-B[j] edge devices from SV;
    Update A[i, j] and S[i];
end for
return A;
```

Algorithm 1: Greedy search-based task assignment.

Input:
T: the tasks set; D: the edge devices set; B: the lower bound vector;
C: the privacy compatibility matrix; E: the evaluation matrix.
Output:
Success: A; failure: no feasible A is obtained.

```
for each edge device d_i in D do
    for each task t_j in T do
        if E[i, j] = 1 then
            C[i, j] ← mpc − C[i, j];
        else
            C[i, j] ← ∑_{j=0}^{n−1} B[j];
        end if
    end for
end for
for each edge device d_i in D do
    cindex ← 0;
    for each task t_j in T do
        while B[j] > 0 do
            C*[i, cindex++] ← C[i, j];
            B[j] ← B[j] − 1;
        end while
    end for
end for
if m > ∑_{j=0}^{n−1} B[j] then
    Add m − ∑_{j=0}^{n−1} B[j] virtual columns to C*, and set their corresponding element values to 0;
end if
H ← KM(C*);
Form the assignment matrix A based on H;
if there is any incorrect assignment in A then
    return Failure
end if
if for all columns of matrix A satisfy ∑_{i=0}^{m} A[i, j] = B[j] then
    return Success
else
    return Failure
end if
```

Algorithm 2: KM algorithm-based task assignment.

5. Experiments

In this section, we conducted four sets of simulation experiments to evaluate the effectiveness and efficiency of KMTA and GSTA. As far as we know, there is no other research directly related to our study. Hence, we implement a "Random (RNDM)" approach as a benchmark to compare with KMTA and GSTA. Given a set of tasks and a set of edge devices, RNDM randomly assigns each task to $B[j]$ edge devices that satisfy the resource requirements and privacy compatibility threshold constraints. All the experiments are performed on a Windows platform equipped with Intel Core i7-4790 @ 3.60 GHz and 8 GB RAM.

5.1. Experimental Setting. To comprehensively evaluate GSTA and KMTA, we have simulated various PITAE scenarios by changing the following parameters: (1) the number of edge devices (m); (2) the number of tasks (n); and (3) the privacy compatibility threshold (th). Specifically, in set #1, m changes from 30 to 300 with a step of 30, $n = m/3$, and th is set to 0.1. In set #2, m changes from 50 to 500 with a step of 50, $n = m/5$, and th is set to 0.1. In set #3, m and n are fixed at 150 and 50, respectively, and th changes from 0.1 to 0.5 with a step of 0.1. In set # 1.4, m is fixed at 250, and the other parameters are set as in set # 1.3. Each experiment is repeated 100 times, and the results are averaged. The detailed experimental settings are shown in Table 3.

In sets #1–4, $B[j]$ is randomly assigned from 1 to 3, and the resource requirements of each task and the available resources provided by each edge device are randomly generated following the uniform distribution. The details are shown in Table 4.

In sets #1–4, each task is randomly assigned 0-10 pieces of private data, and the privacy requirements and the privacy policies are randomly generated for private data. Specifically, for a privacy requirement $pr_j^k = < pd_j^k, sd_j^k, pu_j^k, OP_j^k, re_j^k >$, sd_j^k is assigned randomly with a value in [0.00, 1.00], pu_j^k is assigned randomly from 10 different purposes, OP_j^l is randomly generated from an operation set containing 5 different operations, and re_j^l is assigned randomly from 1 to 12 months. For a privacy policy $pp_i^l = < pd_i^l, td_i^l, pu_i^l, OP_i^l, re_i^l >$, the td_i^l, pu_i^l, OP_i^l, and re_i^l are the same as the setting of corresponding privacy attributes in pr_j^k.

5.2. Effectiveness Evaluation. Through comparison with RNDM, Figures 6 and 7 show the effectiveness of KMTA and GSTA in experiment sets #1-4 and the influence of three parameters, i.e., n, m, and th. On the whole, KMTA can find the optimal solution for the PITAE problem, and with the changes of n, m, and th, KMTA is significantly better than GSTA and RNDM in terms of privacy compatibility degree. Compared to KMTA, GSTA's privacy compatibility degree is lower than that of KMTA, especially in the case of stricter th constraints, but it is still significantly higher than RNDM in all cases.

Table 3: Experimental setting.

	m	n	th
Set #1	30, 60, . . ., 300	10, 20, . . ., 100	0.1
Set #2	50, 100, . . ., 500		
Set #3	150	50	0.1, 0.2, 0.3, 0.4, 0.5
Set #4	250		

Figure 6 illustrates the effect of increasing m on privacy compatibility degree. As shown in Figure 6(a), as m increases, the privacy compatibility degrees of all the approaches increase rapidly. In all cases, KMTA shows the highest privacy compatibility degree, RNDM shows the lowest privacy compatibility degree, and GSTA's privacy compatibility degree is slightly lower than that of KMTA. The reason is that KMTA always assigns $B[j]$ qualified edge devices to each task globally to obtain the highest privacy compatible solution. Hence, it can find the optimal solution to the PITAE problem. GSTA always assigns $B[j]$ qualified edge devices with the highest privacy compatibility for each task locally, resulting in the privacy compatibility degree of the solution it finds slightly lower than that of KMTA. However, RNDM always randomly assigns each task to $B[j]$ qualified edge devices. Consequently, the solution it finds has the lowest privacy compatibility degree. For example, in Figure 6(a), the average privacy compatibility degrees of KMTA, GSTA, and RNDM are 75.59, 74.33, and 42.28, respectively.

In Figure 6(b), as m/n increases from 3 to 5, the average range of candidate edge devices for each task also enlarges. As a result, the privacy compatibility degrees of all the approaches have improved to varying degrees, and KMTA is still higher than GSTA and RNDM. For example, comparing Figure 6(b) with Figure 6(a), the average privacy compatibility degrees of KMTA, GSTA, and RNDM increase by 4.11%, 3.87%, and 1.31%, respectively.

Figure 7 demonstrates the effect of th on the privacy compatibility degree after fixing m and n. It can be seen from Figure 7(a) that when th increases from 0.1 to 0.5, the privacy compatibility degrees of KMTA and RNDM remain basically unchanged, but the privacy compatibility degree of GSTA shows a clear downward trend. It is because as th increases, the number of qualified edge devices for each task decreases. Due to that GSTA always selects edge devices locally for each task, it is most affected by th. For example, in Figure 7(a), the privacy compatibility degrees of KMTA and RNDM are kept at about 45 and 18, respectively, in all cases. However, GSTA's privacy compatibility degree is reduced from 44.89 to 17.56. When m/n increases from 3 to 5, and we compare Figure 7(b) with Figure 7(a), the privacy compatibility degrees of all the approaches show different degrees of improvement, but the privacy compatibility degree of GSTA still decreases with the increases of th. For example, in Figure 7(b), the privacy compatibility degrees of KMTA and RNDM maintains at about 47 and 19, respectively, in all cases. However, GSTA's privacy compatibility degree is reduced from 46.91 to 23.43.

Table 4: Resource requirements and available resources settings.

	CPU (GHz)	Memory (GB)	Storage (TB)	Bandwidth (Mbps)
Resources requirements	[1, 2]	[2, 8]	[0.2, 1]	[10, 20]
Available resources	[1, 3]	[2, 16]	[0.5, 2]	[10, 30]

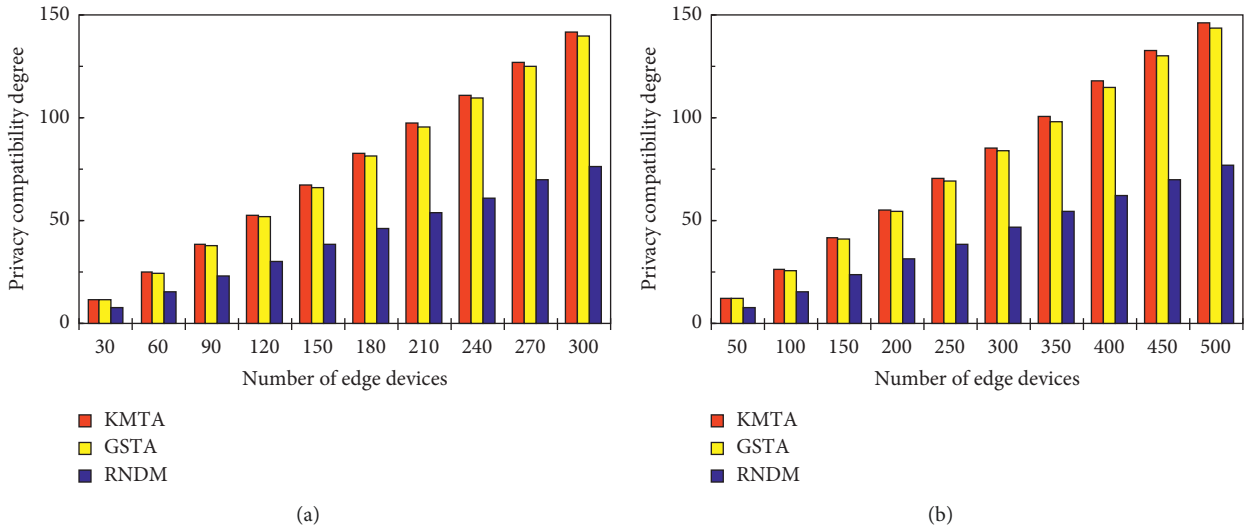


Figure 6: Effectiveness vs. number of edge devices. (a) Set #1. (b) Set #2.

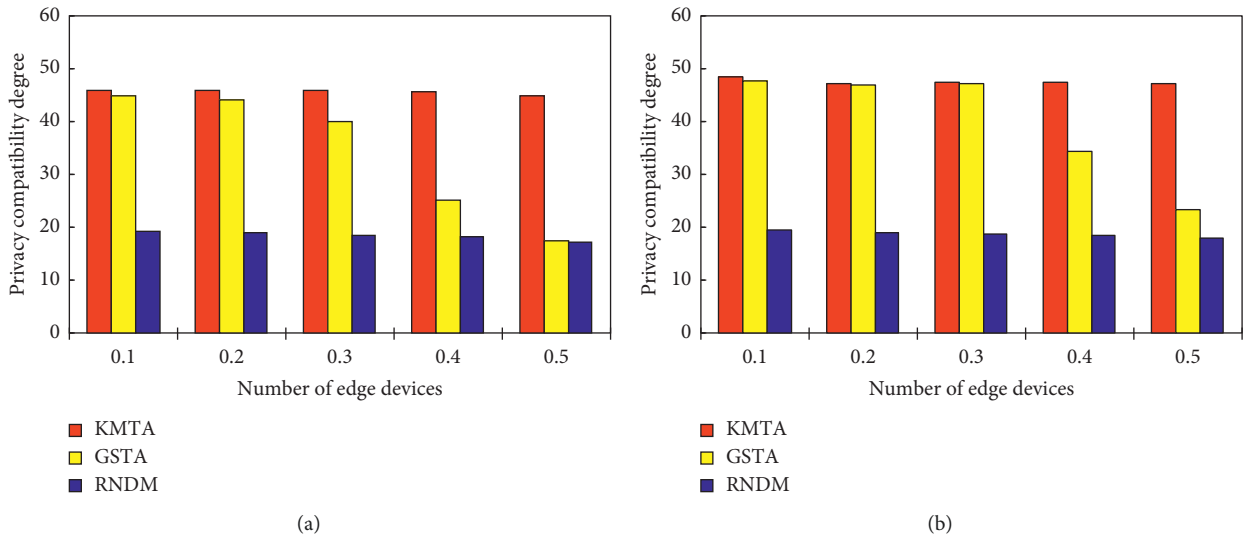


Figure 7: Effectiveness vs. number of edge devices. (a) Set #3. (b) Set #4.

5.3. Efficiency Evaluation. Figure 8 shows the times taken by KMTA, GSTA, and RNDM to find a solution. Since the solving time of the PITAE problem is mainly affected by n and m, we only compare the average execution time of all the approaches in experiment sets #1-2. In general, because KMTA is an optimal approach to solve the PITAE problem, it takes more execution time than GSTA and RNDM. Especially, when m and n are relatively large, this trend becomes more obvious.

As shown in Figure 8(a), when m is relatively small, e.g., $m < 120$, all the approaches consume basically the same time and increase slowly. However, when $m \geq 120$, KMTA consumes more time than GSTA and RNDM, and the consumed time by KMTA increases rapidly. For example, when m rises from 120 to 300, the execution time of KMTA increases from 10.91 ms to 320.77 ms, while the execution time of GSTA and RNDM is less than KMTA and remains below 15 ms. The results observed from Figure 8(b) show the influence of increasing m/n on time consumption. If we compare Figure 8(b) with Figure 8(a), we notice that the consumed time of all the approaches increases to different degrees. In addition, similar to Figure 8(a), in Figure 8(b), when m is

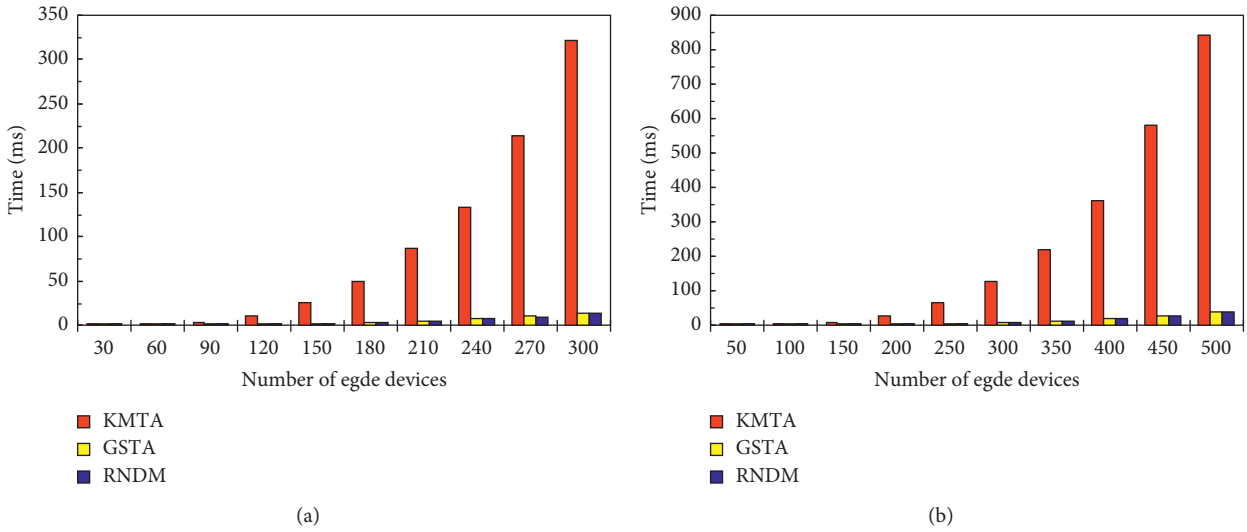


Figure 8: Average time consumption vs. number of edge devices. (a) Set #1. (b) Set #2.

relatively small, all the approaches consume basically the same time and increase slowly, while when $m \geq 200$, KMTA consumes more time than GSTA and RNDM, and the consumed time by KMTA increases rapidly. For example, in Figure 8(b), when m rises from 200 to 500, the time taken by KMTA increases from 29.29 ms to 840.79 ms, while GSTA and RNDM take less time than KMTA and keep the consumed time below 40 ms.

5.4. Discussion. From the above experimental results, we can make the following conclusions.

(1) In terms of effectiveness, KMTA and GSTA have significant advantages over RNDM. In addition, in all cases, KMTA can find a solution with a higher privacy compatibility degree than GSTA, especially in cases with stricter privacy constraints; e.g., th is relatively large, and the advantages of KMTA are more obvious.

(2) In terms of performance, the execution time of GSTA and RNDM is basically the same in all cases. In the case where m and n are relatively small, the execution time of KMTA is basically the same as that of GSTA and RNDM. However, in the case where m and n are relatively large, the execution time of KMTA is much longer than that of GSTA and RNDM.

(3) Although expanding m/n can improve the privacy compatibility degrees of all the approaches, it also brings more time consumption.

(4) In cases where m and n are relatively small or th is relatively large, KMTA outperforms GSTA and RNDM significantly. However, when m and n are relatively large, the overall performance of GSTA is better than that of KMTA and RNDM. In short, KMTA and GSTA can beat RNDM in different cases. Therefore, we can choose KMTA or GSTA to assign tasks according to different m, n, and th scenarios.

6. Related Work

With the emergence of a large number of edge devices with sensing, actuation, and computing capabilities in the urban environment, it has become more complicated to assign IoT tasks to edge devices for execution [8, 12]. Many research efforts have been focusing on task assignment based on vertical offloading technology and horizontal offloading technology. The former relies on a centralized coordinator to place simple task processing on local edge devices, while offloading complex data analysis tasks to fog/cloud nodes. The latter offloads tasks to multiple edge devices that are as close as possible to the data source, and these devices execute tasks in a distributed manner.

To serve IoT applications at the edge, Farhadi et al. [22] proposed a joint optimization method for service placement and request scheduling, and developed polynomial time algorithms to solve the placement and scheduling problems. Aiming at the task allocation problem in collaborative edge and cloud environment, Long et al. [21] proposed a non-cooperative game model between multiple agents and solved the task allocation problem with QoS constraints through a series of algorithms. Considering the latency and bandwidth requirements of IoT applications, Antonio et al. [20] proposed a QoS-aware application deployment method in fog computing. The proposed method models the deployment requirements of IoT applications, describes the available resources and quality of fog nodes, and develops optimization algorithms for the application deployment problem. Cheng et al. [37] proposed a task assignment method in a data sharing mobile edge computing system and designed three algorithms to deal with the holistic and divisible task assignment problem.

The above work uses vertical offloading technology to assign tasks for IoT applications. Recently, some new work has also emerged in the aspect of horizontal task offloading. The work in [1] clusters heterogeneous edge devices to

process data-intensive IoT applications. The proposed method first decomposes an IoT application into a set of simple tasks, then automatically discovers qualified edge devices, and finally assigns tasks to appropriate edge devices. Similarly, Avasalcai et al. [2] proposed a decentralized resource management framework for deploying delay-sensitive IoT applications at the edge of the network and found deployment solutions that meet the requirements through satisfiability modulo theory (SMT) technology.

The above work mainly focuses on the task allocation problem of resource and QoS constraints, and rarely considers user privacy requirements. With the widespread adoption of IoT applications, users are increasingly concerned about the privacy of their personal data. Some research contributions focus on the privacy-aware task assignment for IoT applications.

Aiming at the privacy protection problem in socially aware edge computing, Zhang et al. [23] proposed a privacy-aware task allocation method. The proposed method uses generalization techniques to reduce the accuracy of private data and develops a game theory model to optimize the QoS of the application while ensuring that the user's privacy requirements are satisfied. To protect user privacy in IoT data, Mian et al. [24] proposed a privacy-aware task offloading method in fog computing. The method first divides the IoT tasks into different small fragments according to the security requirements of the data, then these task fragments are offloaded to multiple fog nodes that meet security requirements, and finally a dynamic programming algorithm is used to obtain the task offloading solution that meets the security and delay requirements. Considering the privacy leakage of sensing data in mobile crowd sensing systems, Dai et al. [25] proposed a privacy preservation task assignment scheme and designed a user location privacy protection algorithm based on the differential privacy method. To avoid the privacy disclosure of the datasets due to data acquisition by different operators, Xu et al. [26] took the privacy conflict of different datasets as the optimization goal, formulated the application deployment problem in cyber-physical cloud systems as a multi-objective optimization problem, and used an improved differential evolution technology to solve it.

Although the above work has advantages, the privacy-aware task assignment for IoT applications is still an open issue. The above work employs various privacy technologies to control access to private data, but does not consider how the data will be used after being accessed, such as the purpose of data use, the retention time of the data, and the operations executed on the data. Our approach can fully support these requirements and can also measure the compatibility degree between privacy requirements and privacy policies.

Group Role Assignment (GRA) [29, 30, 36, 38, 39] has been proposed for modeling general assignment problems by solving different engineering problems. The solution to the GRA provides inspiration to this research. The creation of a qualification matrix of GRA is a prerequisite way to model various assignment problems in edge computing.

7. Conclusion

The edge computing paradigm has a great potential to support a wide variety of IoT applications. In this paper, we propose a privacy-aware task assignment approach for IoT applications, which assigns tasks to edge devices close to the data source in a distributed manner, thereby reducing latency and effectively protecting user privacy. Firstly, we model the resource and privacy requirements of the tasks and assess whether the edge devices satisfy the resource and privacy constraints. Secondly, we formalize the PITAE problem as an integer programming optimization problem and propose two task assignment solutions to solve the PITAE problem. Finally, we compare the proposed approaches with the baseline approach. Experimental results show that (1) when m and n are relatively small or th is relatively large, KMTA outperforms GSTA and RNDM significantly; and (2) when m and n are relatively large, the overall performance of GSTA is better than that of KMTA and RNDM. In short, KMTA and GSTA can beat RNDM in different cases.

For future work, we intend to extend our work with QoS constraints, such as response time (communication latency between edge devices and processing latency on edge devices) and energy consumption (transmission energy between edge devices and processing energy on edge devices), in order to provide a more effective task assignment solution that can meet diverse requirements. In addition, considering the privacy protection requirements of edge devices for various resource information and willingness to undertake tasks, we also plan to integrate these requirements into our current privacy model, so as to achieve privacy protection for users and edge devices at the same time.

Another direction is to specify and solve problems related to privacy protection in edge computing along with the development of GRA with constraint (GRA+) model [36, 38, 39], which provides different ways in modeling various constraints, such as time, space, and coupling between agents (resources) and roles (tasks).

Data Availability

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

Conflicts of Interest

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

Acknowledgments

This research was funded by the Natural Sciences and Engineering Research Council of Canada (Grant no. RGPIN-2018-04818), National Natural Science Foundation of China (Grant no. 61772270), and Jiangsu Province Planning Subject for the 13th Five-Year Plan of Education Sciences (Grant no. 2016-GH0303-00022).

References

[1] R. Dautov and S. Distefano, "Automating IoT data-intensive application allocation in clustered edge computing," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 1, pp. 55–69, 2021.

[2] C. Avasalcai, C. Tsigkanos, and S. Dustdar, "Resource management for latency-sensitive IoT applications with satisfiability," *IEEE Transactions on Services Computing*, p. 1, 2021.

[3] D. Liu, Q. Jiang, H. Zhu, and B. Huang, "Distributing UAVs as wireless repeaters in disaster relief via group role assignment," *International Journal of Cooperative Information Systems*, vol. 29, no. 01n02, Article ID 2040002, 2020.

[4] Gartner, "Gartner Forecasts Worldwide IoT-Enabled Software in 2019-2025," 2021, https://www.gartner.com/en/documents/4009207-forecast-iot-enabled-software-worldwide-2019-2025.

[5] E. Ahmed, I. Yaqoob, A. Gani, M. Imran, and M. Guizani, "Internet-of-Things-Based smart environments: state of the art, taxonomy, and open research challenges," *IEEE Wireless Communications*, vol. 23, no. 5, pp. 10–16, 2016.

[6] S. Eugene, T. Thanassis, and H. Wendy, "Analytics for the Internet of Things: a survey," *ACM Computing Surveys*, vol. 51, no. 4, pp. 74:1–74:36, 2018.

[7] R. Casadei, G. Fortino, D. Pianini, W. Russo, C. Savaglio, and M. Viroli, "Modelling and simulation of opportunistic IoT services with aggregate computing," *Future Generation Computer Systems*, vol. 91, pp. 252–262, 2019.

[8] G. Fortino, C. Savaglio, G. Spezzano, and M. Zhou, "Internet of Things as system of systems: a review of methodologies, frameworks, platforms, and tools," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 1, pp. 223–236, 2021.

[9] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: vision and challenges," *IEEE Internet of Things Journal*, vol. 3, no. 5, pp. 637–646, 2016.

[10] M. Satyanarayanan, "The emergence of edge computing," *Computer*, vol. 50, no. 1, pp. 30–39, 2017.

[11] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: the communication perspective," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 2322–2358, 2017.

[12] M. Dias de Assunção, A. da Silva Veith, and R. Buyya, "Distributed data stream processing and edge computing: a survey on resource elasticity and future directions," *Journal of Network and Computer Applications*, vol. 103, pp. 1–17, 2018.

[13] R. Dautov and S. Distefano, "Stream processing on clustered edge devices," *IEEE Transactions on Cloud Computing*, vol. 10, no. 2, pp. 885–898, 2020.

[14] M. Sun, Z. Zhou, X. Xue, W. Zhang, and W. Gaaloul, "Adaptive configuration of service-based smart sensors in edge networks," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 4, pp. 2674–2683, 2022.

[15] N. Fernando, S. W. Loke, and W. Rahayu, "Computing with nearby mobile devices: a work sharing algorithm for mobile edge-clouds," *IEEE Transactions on Cloud Computing*, vol. 7, no. 2, pp. 329–343, 2019.

[16] E. Ahmed, A. Ahmed, I. Yaqoob, and J. A. M. M. Shuja, "Bringing computation closer toward the user network: is edge computing the solution?" *IEEE Communications Magazine*, vol. 55, no. 11, pp. 138–144, 2017.

[17] C. Tsigkanos, C. Avasalcai, and S. Dustdar, "Architectural considerations for privacy on the edge," *IEEE Internet Computing*, vol. 23, no. 4, pp. 76–83, 2019.

[18] A. Alwarafy, K. A. Al-Thelaya, M. Abdallah, J. Schneider, and M. Hamdi, "A survey on security and privacy issues in edge-computing-assisted Internet of Things," *IEEE Internet of Things Journal*, vol. 8, no. 6, pp. 4004–4022, 2021.

[19] A. Atheer, B. Masoud, R. Omer, and P. Charith, "Privacy laws and privacy by design schemes for the Internet of Things: a developer's perspective," *ACM Computing Surveys*, vol. 54, no. 5, pp. 102:1–102:38, 2021.

[20] A. Brogi and S. Forti, "QoS-aware deployment of IoT applications through the fog," *IEEE Internet of Things Journal*, vol. 4, no. 5, pp. 1185–1192, 2017.

[21] S. Long, W. Long, Z. Li, K. Li, Y. Xia, and Z. Tang, "A game-based approach for cost-aware task assignment with QoS constraint in collaborative edge and cloud environments," *IEEE Transactions on Parallel and Distributed Systems*, vol. 32, no. 7, pp. 1629–1640, 2021.

[22] V. Farhadi, F. Mehmeti, T. He, and T. F. L. H. S. K. S. K. Porta, "Service placement and request scheduling for data-intensive applications in edge clouds," *IEEE/ACM Transactions on Networking*, vol. 29, no. 2, pp. 779–792, 2021.

[23] D. Zhang, Y. Ma, X. Sharon Hu, and D. Wang, "Toward privacy-aware task allocation in social sensing-based edge computing systems," *IEEE Internet of Things Journal*, vol. 7, no. 12, Article ID 11384, 2020.

[24] M. R. Mian, T. Byungchul, L. Peng, and G. Mohsen, "Privacy-aware collaborative task offloading in fog computing," *IEEE Trans. Comput. Soc. Syst.*vol. 9, 2022.

[25] M. Dai, J. Li, Z. Su, W. Chen, Q. Xu, and S. Fu, "A privacy preservation based scheme for task assignment in Internet of Things," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 4, pp. 2323–2335, 2020.

[26] X. Xu, R. Mo, X. Yin et al., "PDM: privacy-aware deployment of machine-learning applications for industrial cyber-physical cloud systems," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 8, pp. 5819–5828, 2021.

[27] H. W. Kuhn, "The Hungarian method for the assignment problem," *Naval Research Logistics Quarterly*, vol. 2, no. 1-2, pp. 83–97, 1955.

[28] J. Munkres, "Algorithms for the assignment and transportation problems," *Journal of the Society for Industrial and Applied Mathematics*, vol. 5, no. 1, pp. 32–38, 1957.

[29] H. Zhu, M. Zhou, and R. Alkins, "Group role assignment via a kuhn-munkres algorithm-based solution," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 42, no. 3, pp. 739–750, 2012.

[30] H. Zhu, *E-CARGO and Role-Based Collaboration: Modeling and Solving Problems in the Complex World*, Wiley-IEEE Press, Hoboken, NJ, USA, 2021.

[31] E. Union, *GeneralData Protection Regulation*, European Union, Maastricht, Netherlands, 2018.

[32] T. Halabi and M. Bellaiche, "A broker-based framework for standardization and management of cloud security-SLAs," *Computers & Security*, vol. 75, pp. 59–71, 2018.

[33] L. Liu, H. Zhu, S. Chen, and Z. Huang, "Privacy regulation aware service selection for multi-provision cloud service composition," *Future Generation Computer Systems*, vol. 126, pp. 263–278, 2022.

[34] D. T. Nguyen, H. T. Nguyen, N. Trieu, and V. K. Bhargava, "Two-stage robust edge service placement and sizing under demand uncertainty," *IEEE Internet of Things Journal*, vol. 9, no. 2, pp. 1560–1574, 2022.

[35] J. Xu, B. Palanisamy, and Q. Wang, "Resilient stream processing in edge computing," in *Proceedings of the 21st IEEE/*

ACM International Symposium on Cluster, Cloud and Internet Computing, (CCGrid), Melbourne, Australia, May 2021.

[36] H. Haibin Zhu and M. MengChu Zhou, "Role transfer problems and algorithms," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 38, no. 6, pp. 1442–1450, 2008.

[37] S. Cheng, Z. Chen, J. Li, and H. Gao, "Task assignment algorithms in data shared mobile edge computing systems," in *Proceedings of the IEEE Int. Conf. Distributed Comput. Syst., (ICDCS)*, Dallas, TX, USA, July 2019.

[38] H. Zhu, "Avoiding conflicts by group role assignment," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 46, no. 4, pp. 535–547, 2016.

[39] H. Zhu, D. Liu, S. Zhang, S. Teng, and Y. Zhu, "Solving the group multirole assignment problem by improving the ILOG approach," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 12, pp. 3418–3424, 2017.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 6911463, 12 pages
https://doi.org/10.1155/2022/6911463

Research Article

Toward Privacy-Preserving Blockchain-Based Electricity Auction for V2G Networks in the Smart Grid

Weijian Zhang,[1] **Wen Yang**,[2] **Cen Chen**,[2] **Nuannuan Li**,[2] **Zijian Bao**,[3] **and Min Luo**[3]

[1]*State Grid Henan Electric Power Company, Zhengzhou, China*
[2]*State Grid Henan Electric Power Research Institute, Zhengzhou, China*
[3]*Key Laboratory of Aerospace Information Security and Trusted Computing Ministry of Education, School of Cyber Science and Engineering, Wuhan University, Wuhan, China*

Correspondence should be addressed to Min Luo; mluo@whu.edu.cn

Received 26 January 2022; Accepted 19 May 2022; Published 16 June 2022

Academic Editor: Jie Cui

With the development of electric vehicle (EV) technology, EV has become a key component in the future smart grid. Due to the sheer large number of EVs on the road, the emerging vehicle-to-grid (V2G) technology, which allows for two-way electrical flows between EVs and the power grid, is gaining traction. However, establishing a fair and private electricity exchange scheme has gradually become a critical challenge. The emergence of blockchain technology offers a novel approach for resolving this issue. In this study, we overview the opportunities and challenges of blockchain in the smart grid. Then, we provide a privacy-preserving blockchain-based electricity auction scheme for V2G networks in smart grid. In particular, we exploit PS group signatures to keep the privacy of EVs or charging stations and leverage blockchain to provide automated auction execution. With our mechanism, the identity of EV/charging station is conditionally protected. In case of an emergency, the trusted authority (i.e., the group manager) can open the identity. Meanwhile, we present the security analysis to prove our scheme's security. Finally, we implement the experiment to evaluate efficiency. The experimental results show that our proposal is efficient and suitable for V2G networks.

1. Introduction

The smart grid [1,2], also known as "power grid 2.0," is the intellectualization of the electrical grid. It uses sophisticated sensing and measurement technology, advanced equipment technology, control methods, and decision support system technology to create an integrated, high-speed two-way communication network. It has the tremendous potential of making the electricity system more secure, dependable, cost-effective, and efficient. It can dispatch and control all components of the network according to its own needs and realize the intelligence, transparency, automation, and controllability of the grid.

Predictably, the smart grid will be widely used in all aspects of people's life. Among them, vehicle-to-grid (V2G) system is envisaged as a key component of the smart grid [3]. The research on EVs is in full bloom, reconstructing the ecological chain of the automobile industry. For example, Tesla, the largest electric vehicle company in the United States, has a market value of trillion [1].

In particular, our paper adopts the V2G network model as shown in Figure 1. The EVs can get electricity from the charging stations or other EVs. They can also sell their surplus power resources to get paid. The aggregators are responsible for the interaction and power arrangement of EVs and charging stations in the smart grid. Electricity resources can be flexibly scheduled between *vehicle-to-vehicle* and *vehicle-to-charging* stations to maximize energy use. Both approaches are helpful for providing available and cheap renewable energy sources.

However, electricity distribution is a difficult problem for smart grid applications. An auction scheme is expected to alleviate this problem [4,5]. Using an auction is a straightforward idea and seems to be easy. However, we have to face new challenge #1: in a normal auction system, there is always a centralized manager to conduct an auction scheme.

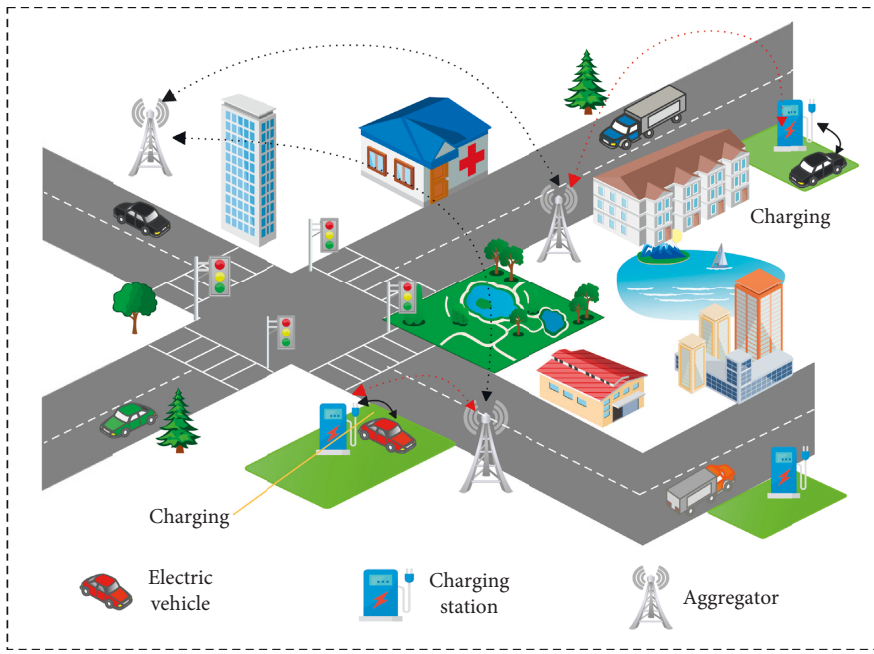


Figure 1: V2G architecture.

Although this way is efficient and convenient, there exist two problems: (1) there may be a single point of failure and 2) the auctioneer may be malicious. The auctioneer can unite users and infringe on the interests of the compliant user, thereby destroying the trust between the sellers, buyers, and auctioneers. We ask the question that "is there exist a decentralized auction scheme without a single point of failure?"

Fortunately, the answer is yes. With the advent of blockchain technology, a decentralized auction is possible. The blockchain technology [6,7], which has been rapidly rising in recent years, has been used in a variety of applications, including medical, Internet of things, and digital finance. It can bring new opportunities and challenges to the auction scheme and the smart grid. Blockchain is a decentralized shared ledger and database, which stores and verifies data using a chain-based data structure. Each block is made up of a set of transactions that are committed by network peers using a predetermined consensus procedure. All participants work together to maintain and supervise the data storage. It has the characteristics of decentralization, transparency, immutability, and security properties. The smart contract [8] supports a program to execute securely in a decentralized environment, making blockchain a powerful tool for building a self-organized system. In particular, Hyperledger Fabric [9], as a widely known project based on blockchain technology, allows users to complete data calculations securely and reliably on the chain through the deployment of smart contracts.

However, things are not as simple as they seem. Challenge #2 emerges: the openness and transparency of blockchain data often conflict with privacy protection. For example, in the auction process, we do not want the identity of users, the dealing price of the auction, and other information to be disclosed to others. For instance, the Australian Information Commissioner released the survey results, confirming that Uber violated the privacy of more than one million Australians [10]. A series of privacy-preserving measures should be implemented. However, existing solutions, such as Zerocash [11] and Monero [12], cannot be directly applied. Thanks to the emergence of anonymous authentication [13], for example, *CL signature* [14], *BBS + signature* [15], and *PS signature* [16], they can greatly alleviate the privacy contradiction between users and servers. In particular, the user is authorized by a trusted third party, and then, the user can request services from the server. The server does not need to know a user's identity but only needs to know that the user has a legal identity.

1.1. Our Contributions. Our contributions are listed as follows:

(1) Although there have been several discussions about the use of blockchain in the smart grid, few literature studies summarize the topic comprehensively. One principal goal of this study was to investigate the role of blockchain in the smart grid and summarize its usage in the smart grid.

(2) By leveraging the PS group signature, we first propose a privacy-preserving blockchain-based electricity auction for V2G networks. In particular, the bidder can send the auction request to the manager anonymously. Then, the manager invokes the smart contracts on the blockchain to automate the auction protocol. Finally, the bidder and auctioneer complete the electricity transaction.

(3) We analyze the scheme's security properties and provide experiments to show our system's computation costs of the offchain and onchain parts.

1.2. Organization. The structure of this study is as follows: in Section 2, we review the related work. In Section 3, we provide the related building blocks. In Section 4, we summarize the opportunities and challenges of the combination of blockchain and the smart grid. In Section 5, we give the system model, threat model, and design goals. In Section 6, we present the detailed construction. In Section 7, we analyze the security of our scheme. In Section 8, we point out some points that are not considered in this study and give the possible work direction in the future. In Section 9, we present the experimental results, including computation costs of the offchain and onchain parts. In the end, we give the conclusion in Section 10.

2. Related Work

At present, there have been several research works on auction schemes in the smart grid. Hahn et al. [17] provided a smart contract-based decentralized transactive energy auctions. The auction method uses a second-price auction, ensuring that bidders make honest bids. They implemented the contracts on the Ethereum blockchain. Wang et al. [18] presented a decentralized electricity transaction mode of microgrid based on blockchain and the double auction mechanism. They designed an adaptive aggressiveness strategy to allow traders to alter their bids in real time in response to the market changes. Ramachandran et al. [19] introduced a hybrid optimization method for decentralized energy resource management. Based on risk and competitive equilibrium price prediction, they implemented a profit-maximizing adaptive bidding technique. To cut the cost, a hybrid immune system-based particle swarm optimization is utilized to generate the model, which assumes actual power market pricing. Stubs et al. [20] proposed multitier double auctions for smart energy distribution grids using blockchain technology. The scheme can reduce blockchain workload by aggregating energy usage and generation. Edge computing is also used to improve reliability and response time. Ma et al. [21] proposed an efficient pricing method that can prevent users' cheating. They provided an enhanced Arrow-d'Aspremont-Gerard-Varet (AGV), a complex auction mechanism, to ensure truthfulness. Using an incentive mechanism, the user's payment is linked to their credit for consumption. Wen et al. [22] provided an effective search encryption auction system for marketing in smart grid. The scheme employs public key encryption with keyword search technologies to allow energy sellers to query relevant offers while preserving the anonymity of energy purchasers. For convenience, their system also supports conjunctive keyword search. Li et al. [23] concentrated on the problem of creating a secure online power market. They suggested an online double auction technique with differential privacy based on two building blocks: a Laplace-based winner selection mechanism and an exponential-based allocation algorithm. Zhou et al. [24] presented an auction mechanism for the geo-distributed cloud. By combining principles from the Gibbs sampling method and the alternating direction approach of the multiplier, they proposed a decentralized social welfare maximization algorithm.

3. Preliminary

In this section, we review the building blocks of our scheme, such as *bilinear pairing*, *PS signature*, *commitment*, *hashed ElGamal encryption*, *blockchain*, *smart contract*, *Hyperledger Fabric*, and *auction scheme*.

3.1. Notions

Definition 1 (Bilinear Pairing). Let $(\mathbb{G}, \mathbb{G}_T)$ be a bilinear map such that $\tilde{e}$: $\mathbb{G}_1 \times \mathbb{G}_2 \longrightarrow \mathbb{G}_T$, where p is the order for both $(\mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T)$. Note that we use type 3 bilinear pairing, where $\mathbb{G}_1 \neq \mathbb{G}_2$, and there is no efficient computable homomorphism between them. The bilinear map should satisfy the following properties:

(1) Bilinear: given any two elements $a, b \in Z_q^*$ and $\forall x \in \mathbb{G}_1, y \in \mathbb{G}_2$, $e(x^a, y^b) = e(x, y)^{ab}$.
(2) Nondegenerate: for $\forall x \in \mathbb{G}_1, y \in \mathbb{G}_2$, $e(x, y) \neq 1_{\mathbb{G}_T}$, where $1_{\mathbb{G}_T}$ represents the identity element in $\mathbb{G}_T$.
(3) Efficient Computability: for $\forall x \in \mathbb{G}_1, y \in \mathbb{G}_2$, $e(x, y)$ is efficiently computable.

3.2. PS Signature. PS signature was proposed by Pointcheval et al. in [16]. It utilizes type 3 bilinear pairing to construct a randomized signature. In particular, this original signature σ can be randomized to a new randomized signature $\sigma\prime$, which can be applied to many privacy-preserving application scenarios, and achieve well performance simultaneously. The detailed algorithms are as follows.

Definition 2 (PS Signature). *It consists of 4 probabilistic polynomial time ($\mathcal{PPT}$) algorithms.*

(1) Setup(1^n): given a system security parameter 1^n, a set of public parameters $\text{pp} = \{\mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T, e, p\}$ are outputs. We denote $\mathbb{G}_1^* = \mathbb{G}_1 / \{1_{\mathbb{G}_1}\}$, and p is the order of $\mathbb{G}_1, \mathbb{G}_2$, and $\mathbb{G}_T$.
(2) Keygen(pp): this algorithm randomly chooses $\tilde{g} \in \mathbb{G}_2$ and $(x, y) \in \mathbb{Z}_p^2$, then computes $(\tilde{X}, \tilde{Y}) \leftarrow (\tilde{g}^x, \tilde{g}^y)$, sets sk as (x, y), and sets pk as $(\tilde{g}, \tilde{X}, \tilde{Y})$.
(3) Sign(m, sk): given a message $m \in \{0, 1\}^*$, it randomly chooses $h \in \mathbb{G}_1^*$ and then computes $\sigma = (h, h^{(x+y \cdot m)})$.
(4) Verify(m, σ, pk): given a message $m \in \{0, 1\}^*$, a signature σ, and a public key pk , it parses σ as (σ_1, σ_2) and checks whether $\sigma_1 \neq 1_{\mathbb{G}_1}$ and $e(\sigma_1, \tilde{X} \cdot \tilde{Y}^m) = e(\sigma_2, \tilde{g})$. If true, it outputs 1; otherwise, 0.

PS signature is EUF-CMA under the LRSW assumption [25]. Meanwhile, a group signature can be easily obtained from the PS signature [16].

Definition 3 (Group Signature Based on PS Signature). It consists of 6 $\mathcal{PPT}$ algorithms.

(1) GSetip(1^n): the group manager runs Setup and Keygen to obtain (sk, pk), where $\text{sk} = (x, y)$,

$\mathrm{pk} = (\widetilde{g}, \widetilde{X}, \widetilde{Y})$, and then, it sets $\mathrm{gsk} := \mathrm{sk}$ and $\mathrm{gpk} := (\mathrm{pk}, g)$.

(2) KPI join $(i, 1^n)$: the user i generates its private/public key pair $(\mathrm{sk}_i, \mathrm{pk}_i)$ and then sends pk_i to the certificate authority.

(3) G Join: the user randomly chooses s_i, generates $(\delta, \widetilde{\delta}) \leftarrow (g^{s_i}, \widetilde{Y}^{s_i})$ and a signature $\theta \leftarrow \mathrm{Sign}(\mathrm{sk}_i, \delta)$, and then sends them to the group manager. The group manager checks whether θ is valid and $e(\delta, \widetilde{Y}) = e(g, \widetilde{\delta})$. Then, the user gives the zero-knowledge proof that he owns the s_i. After that, the group manager generates a random number r and computes $\sigma \leftarrow (\sigma_1, \sigma_2) \leftarrow (g^r, (g^x \cdot \delta^y)^r)$, which is a valid signature on s_i. In the end, the group manager stores $(i, \delta, \theta, \widetilde{\delta})$ in a secret register and sends $\widetilde{\sigma}$ and $e(\sigma_1, \widetilde{Y})$ to the user, where $\mathrm{gsk}_i = (s_i, \sigma, e(\sigma_1, \widetilde{Y}))$.

(4) G Sign (gsk_i, m): the user needs to randomize σ using a random number t and computes $(\sigma_1', \sigma_2') \leftarrow (\sigma_t^1, \sigma_2^t)$ along with a signature of knowledge of s_i. The detailed steps are as follows: the user randomly chooses $k \in \mathbb{Z}_p$ and computes $c \leftarrow \mathscr{H}(\sigma_1', \sigma_2', e(\sigma_1, \widetilde{Y})^{k \cdot t}, m)$, where $\mathscr{H}$ is a secure hash function. Finally, the user computes $s \leftarrow k + c \cdot s_i$ and outputs $(\sigma_1', \sigma_2', c, s)$ as the group signature μ on the message m.

(5) G Verify (gpk_i, m, μ): to verify whether the signature $(\sigma_1', \sigma_2', c, s)$ is valid, the verifier computes $T \leftarrow e(\sigma_{1\prime}, \widetilde{X})^c \cdot e(\sigma_{2\prime}, \widetilde{g})^{-c} \cdot e(\sigma_{1\prime}, Y)^s$ and $c = \mathscr{H}(\sigma_1', \sigma_2', T, m)$. If it is valid, it outputs 1; otherwise, 0.

(6) G Open (gmsk, m, μ): when we need to open one user's identity, the group manager searches in the list $(i, \delta_i, \theta_i, \widetilde{\delta}_i)$ and checks whether $e(\sigma_{2\prime}, \widetilde{g}) \cdot e(\sigma_{1\prime}, \widetilde{X})^{-1} = e(\sigma_{1\prime}, \widetilde{\delta}_i)$ until he gets a match. He then outputs a corresponding (i, δ_i, θ_i) with a proof of knowledge $\widetilde{\delta}_i$.

3.3. Cryptographic Commitment. A commitment scheme enables a user to commit to a specific statement, which is hidden from others during the *commit* phase, but visible to the users during the ***open*** phase. The following two properties belong to a commitment scheme:

(1) *Binding*: after committing to a statement, the committer is unable to alter it.

(2) *Hiding*: before the committer opens the commitment, the receiver knows nothing about the committed statement.

We give the Pedersen commitment [26] as follows:

(1) Setup (1^n): given a system security parameter 1^n, a set of public parameters $\mathrm{pp} = \mathscr{G}, p, g, h$ is outputs, where p is the order of $\mathscr{G}$, and g, h are the generators of $\mathscr{G}$.

(2) Commit $(m; r)$: on inputs a message $m \in \mathbb{Z}_p$, this algorithm randomly chooses $r \in \mathbb{Z}_p$ and outputs $c \leftarrow g^m h^r$.

(3) Open (c, m, r): if $c = g^m h^r$, this algorithm outputs 1; otherwise, 0.

3.4. Hashed ElGamal Encryption. The ElGamal encryption system is a asymmetric key encryption system based on the Diffie–Hellman key exchange [27]. Here, we use a variant of ElGamal encryption, called hashed ElGamal encryption [28]. It includes the 4 $\mathscr{PPT}$ algorithms listed as follows:

(1) *Setup* $((1^n))$: given a security parameter 1^n, it outputs $\mathrm{pp} = (\mathbb{G}, g, h, p, \mathscr{H})$, where g, h are generators of the cyclic group $\mathbb{G}$ of prime order p, and $\mathscr{H}$ is a hash function $\{0,1\}^* \longrightarrow (0,1)^n$.

(2) *Keygen* (pp): it outputs $(\mathrm{sk}, \mathrm{pk})$, where $\mathrm{sk} \leftarrow \mathbb{Z}_p$ and $\mathrm{pk} = g^{\mathrm{sk}}$.

(3) *Encrypt (pk,m)*: it chooses $r \leftarrow \mathbb{Z}_p$, computes $c_1 = g^r$, $c_2 = \mathscr{H}(\mathrm{pk}^r) \oplus m$, and outputs $c = (c_1, c_2)$.

(4) *Decrypt (sk,c)*: it computes $m = c_2 \oplus \mathscr{H}(c_1^{sk})$.

3.5. Blockchain and Smart Contract. Blockchain is a peer-to-peer decentralized ledger that is based on the Bitcoin concept [29]. It is a multi-technology application paradigm that includes encryption, game theory, decentralized systems, and other technologies. With a certain consensus algorithm, all nodes in the blockchain retain a consistent record. The ledger, as illustrated in Figure 2, is a series of data blocks that include various transactions sent by users in a peer-to-peer network, with the last block always including the hash of the preceding block. The following are the key characteristics of blockchain.

(1) *Decentralization*: as blockchain technology adopts a decentralized structure, there is no centralized management organization. Every node in it has the same rights and obligations. They jointly maintain the data ledger stored in the system.

(2) *Immutability*: once the information is verified and added to the blockchain, it will be stored permanently. It is impossible to update the data in a single block without affecting all following blocks. For example, in the Bitcoin system, unless the attacker has more than 50% of the whole network computing power, it is impossible to regenerate blocks to tamper with the data. Generally, we assume that the data in the blockchain cannot be tampered with.

(3) *Openness and Transparency*: once the transaction is packaged into a block, the block will be broadcast to all nodes, achieving data synchronization. Each node can trace back all the transaction information of any parties in the past.

(4) *Security*: the security of all entries in the blockchain is guaranteed using cryptographic algorithms, such as digital signatures and encryption algorithms.

Furthermore, thanks to Ethereum [8], a well-known blockchain project, the notion of smart contracts has been revitalized with the advent of the *Blockchain 2.0* era. Smart contracts are computer programs that are recorded on a blockchain and run automatically when certain criteria are met. For example, when a stock price is less than a certain value, a predefined smart contract can automatically execute

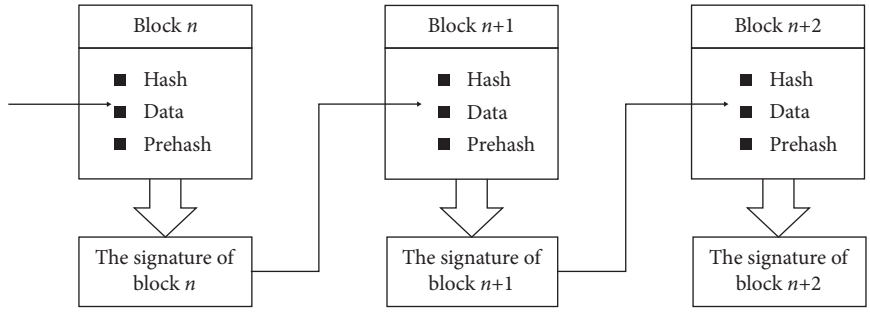


Figure 2: Blockchain structure.

the operation of buying the stock. The smart contract has been extensively used in the Internet of things, product traceability, supply chain finance, insurance, and so on. Based on the access mechanism, blockchain can be generally categorized into public, private, and consortium blockchains. In a public blockchain, each node is free to enter or leave at will. However, nodes without authorization cannot access the data in either a private or consortium blockchain. Consortium blockchain offers a stronger decentralized nature, since there are multiple institutions involved, rather than one in a private blockchain.

3.6. Hyperledger Fabric. Hyperledger Fabric is an open-source, enterprise-level, permission-based consortium blockchain platform [9]. It is underpinned by the modular architecture and offers excellent confidentiality, scalability, flexibility, and extensibility. There are three kinds of nodes in Hyperledger Fabric: the endorsement nodes, the order nodes, and the normal nodes. The endorsement nodes are responsible for endorsing and executing transactions. The order nodes are in charge of packaging transactions into blocks, and the normal nodes always publish the transactions to endorse nodes and receive new blocks from order nodes. This architecture avoids a bottleneck, thanks to the decoupling of node functions, which makes Fabric more efficient. Meanwhile, to protect privacy, private "subnets" are used to communicate among multiple specific network members, which are defined as a channel. Channel contains one or more organizations, which are the interest entities with collaborative relationships. Moreover, in Hyperledger Fabric, the consensus algorithm is designed to be pluggable. Fabric provides some alternative algorithms, such as Solo, Kafka, and Raft. Because there is just one order in which to sort messages and construct blocks in Solo mode, it is most commonly utilized in a testing environment. Raft is a sorting service that supports crash fault tolerance (CFT), which means it can only tolerate half of the fault nodes. Kafka is similar to Raft; however, it has a higher computational cost.

3.7. Auction Scheme. The auction can achieve an effective allocation of electricity resources and ensure the transparency and fair of the process [30]. There are some mainstream auction mechanisms.

(1) *The First-Price Sealed Auctions (FSAs)*: the bidder delivers the bid to the auctioneer in a sealed envelope. After that, the auctioneer opens the envelope and identifies the highest bidder.

(2) *The Second-Price Sealed Auctions (SSAs)*: the process is similar to FSA. The winner only needs to pay the second highest bid, which eliminates the bidder's concerns about the difference between the first and second prices.

(3) *The Open Ascending Bid Auctions (English Auctions)*: the bidders increase the bids gradually until no one wants to pay more than the current highest bid. The highest bidder gets the auction item at his price.

(4) *The Open Descending Bid Auctions (Dutch Auctions)*: the auctioneer gradually reduces the price from a preset high price until there is a bidder willing to pay the current price.

4. Opportunities and Challenges of Blockchain in the Smart Grid

Blockchain technology has been widely concerned in industry and academia and has become one of the new infrastructures in the digital age. It is expected to resolve some issues in the smart grid, promoting some research for the combination of blockchain and the smart grid.

We summarize the possible usage of blockchain as follows:

A Decentralized Database with Immutability: compared with the traditional database, blockchain can be regarded as a special kind of database, which only supports the adding operation. We can use blockchain to store critical data in the smart grid.

Automated Smart Contracts in Decentralized Environment: a smart contract is a piece of code that can be executed automatically by multiple consensus nodes in a decentralized environment, opening up new possibilities for electricity management in the smart grid. For example, a self-organized electricity auction system can be built by smart contracts, the electricity sellers and buyers only need to submit the demand information to the chain, and the smart contracts can automatically match the demand of both sides according to a predetermined algorithm.

Incentive Mechanisms: traditional electricity management requires a centralized organization, which may

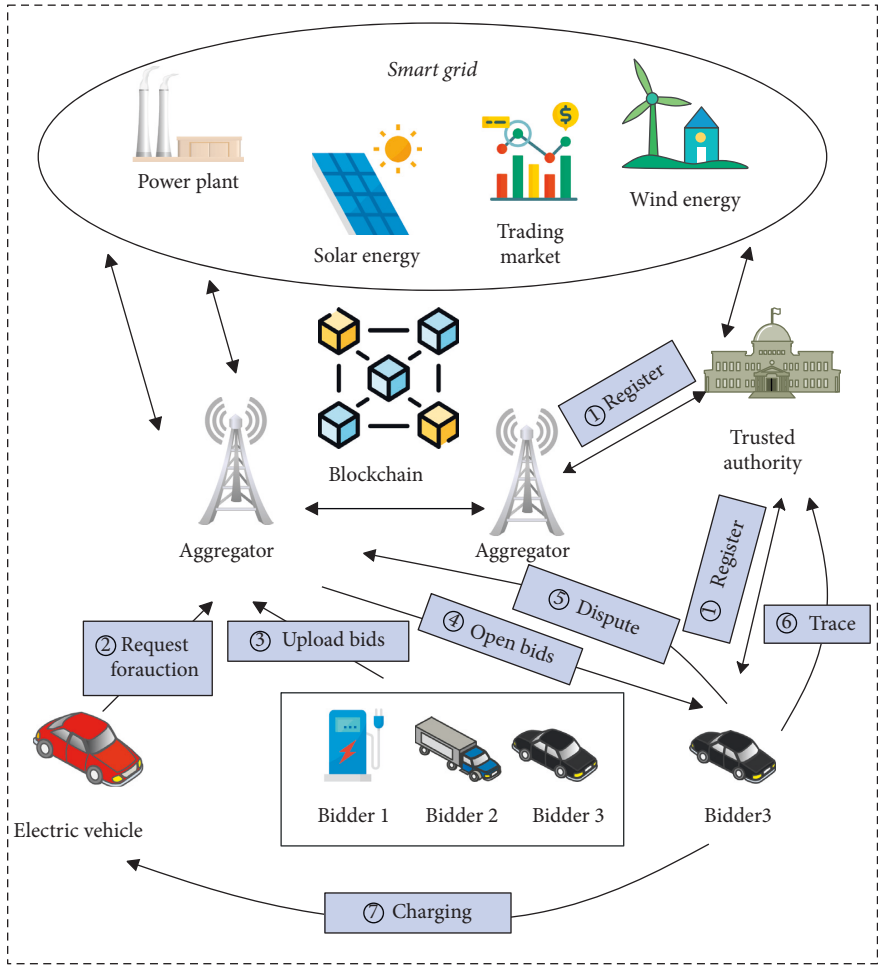


Figure 3: System model (*Note*. Steps ∗∗∗∗5 and ∗∗∗∗6 are only performed when there is a dispute or the identity needs to be opened).

cause a single point of failure. The management may corrupt for the sake of interests, which leads to unfair distribution of electricity. Thus, the current electricity management lacks a well-designed incentive mechanism to promote the benign behavior of electricity management organizations. Therefore, creating a blockchain-based electricity management system with economic incentives could be a promising direction.

Then, we list the points that need to be considered when using blockchain in the smart grid.

Efficiency: compared with the centralized structure, the decentralized structure and consensus process of blockchain will dramatically reduce the efficiency. The most extreme example is Bitcoin's processing speed of *7 transactions/per second (TPS)*. Generally speaking, we have certain requirements for transaction processing speed. As a result, how to choose a suitable form of blockchain based on the actual needs of real contexts and how to configure the blockchain's underlying data structure are challenges that need our attention.

Supervision: the allocation, sharing, and use of electricity in the smart grid should be recorded in the blockchain. To supervise the rational use of electricity and prevent some malicious acts, we must design appropriate mechanisms.

Privacy Protection: we need to protect the privacy of users, such as hiding the identity of the bidders, the auctioneers, and the auction price.

Selection of Blockchain Types: we have mentioned that blockchains can be divided into three types: public, private, and consortium. The public chain data are completely open, and any node can access the blockchain at any time, while the consortium chain is only for members of a specific group, which is more suitable for the scenario with permission control. Hence, it is more suitable for the scenarios in the smart grid.

5. Problem Overview

In this section, we present the specific system architecture in our paper. Then, we give the threat models and design goals of our system. Last, we show the detailed construction.

5.1. System Model. Figure 3 illustrates the system model in our paper. There are 5 entities in our model: *EV*, *charging station*, *trusted authority*, *aggregator*, and *blockchain*.

EV. EVs, owned by users, are mobile and geographically separated. Users want to charge their EVs or sell surplus electricity to other EVs in the smart grid. That is, an EV can be either a seller or a buyer of electricity. They communicate with the aggregator through the privacy-preserving method, publish their own needs on the blockchain, match the supplier through the auction mechanism, and complete the electricity transaction.

Charging Station. The charging station can charge EVs in their region. They are scattered all over the city.

Trusted Authority (TA). The TA is responsible for initializing the whole system and provides charging services for EVs and charging stations. TA will be offline, except in case of an emergency when we need to trace the identity of an entity (*i.e.*, EVs or charging stations).

Aggregator. The aggregators are responsible for coordinating EVs and charging stations. They act as the decision center to dispatch energy for the V2G network. They have sufficient computing power and storage capacity and jointly maintain a blockchain. In particular, they play the role of auctioneer. They can assist EVs in releasing requirements on the blockchain and matching transactions through auction protocols.

Blockchain. Blockchain is regarded as a tamper-resistant ledger in which smart contracts can provide decentralized program execution. We deploy functions for auctions on smart contracts, thereby ensuring the security of the auction protocol.

5.2. Threat Model. In this study, we assume that (1) the trusted authority is fully trusted and (2) the aggregators are honest but curious. That is, they can execute the protocol honestly but may infer the EV's private information. (3) We also consider that there exists an external adversary (abbreviated as $\mathcal{A}$) that can eavesdrop on the communication channels between the parties. They can access and record transactions on the blockchain. Moreover, (4) an $\mathcal{A}$ may impersonate the EV or the charging station to trick other parties in the electricity auction phase.

5.3. Design Goals. The design goals in our study are listed as follows:

(1) *EV/Charging Station Authentication*. The EV/charging station should be authenticated when they participate in the auction scheme. There is no $\mathcal{PPT}$ $\mathcal{A}$ who can forge their identities.

(2) *EV/Charging Station Privacy*. The EV/charging station should be protected. Even a malicious aggregator or external $\mathcal{A}$ cannot know their true identity, except the statement that they have legal identities.

(3) *Auction Privacy*. The bidding information is hidden, and only the bidding information of the final winner is published on the blockchain.

(4) *Traceability*. The TA can trace the identity of a malicious EV/charging station when needed.

(5) *Accountability*. The scheme needs to ensure the accountability of the auction agreement. That is, the bidder with the cheapest bid must become the winner, and the auctioneer cannot maliciously modify the results. Anyone who doubts the auction results can question the results and draw conclusions.

6. Our Construction

6.1. High-Level Description. Let us briefly give a high-level overview of the scheme. We adopt the *first-price sealed auction*, since its satisfactory performance in real life, and it can be easily combined with privacy-preserving technologies.

(1) *System Initialization*. The TA initializes the whole system and generates public parameters. All aggregators jointly generate the initialization parameters of blockchain and then maintain such a blockchain.

(2) *Register*. EVs, charging stations, and aggregators need to register with TA. Each entity needs to have an account and corresponding amount on the blockchain, namely account:=$\{\mathrm{pk}_{\mathrm{address}}, \mathrm{value}\}$.

(3) *Request for Auctions*. The EV sends its request (*i.e.*, buying the electricity) to the aggregator through the anonymous authentication method. After receiving the message, the aggregator sends the request information to the smart contracts on the blockchain.

(4) *Upload Bids*. When charging stations find the request on the blockchain, they commit their supply requests, including the commitment value of electricity bid, to the aggregator. Note that we use the charging station as the representative for the convenience of expression. Other EVs with surplus power can also participate in the bidding.

(5) *Open Bids*. After a specific time node, the aggregator (also as the auctioneer) needs to open all bids, select the bid which is the cheapest, and send it to the blockchain. Then, the auctioneer helps the EV and the winner to establish a secure channel for electricity exchange.

(6) *Dispute*. To ensure accountability, anyone who needs to spend a small part of the token can question the result of an auction. That is, the winner's bid is not the cheapest. Accordingly, the auctioneer needs to give the corresponding zero-knowledge proof that the winner's bid is cheaper than the challenger's bid and send it to the blockchain. If the proof is not given, the challenger can obtain a certain token as a reward from the auctioneer's deposit.

(7) *Trace*. The TA can trace the identity of a malicious EV/charging station when needed.

6.2. Detailed Process

6.2.1. System Initialization. The TA chooses a system security parameter 1^n and outputs a set of public parameters $\mathrm{pp} = \{\mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T, e, g, h, \tilde{g}, p, \mathscr{H}\}$, where g and h are generators of cyclic group $\mathbb{G}_1$, $\tilde{g}$ is the generator of $\mathbb{G}_2$, and p is

the order of $\mathbb{G}_1, \mathbb{G}_2$, and $\mathbb{G}_T$. $\mathscr{H}$ is the hash function. TA randomly chooses $\tilde{g} \in \mathbb{G}_2$ and $(x, y) \in \mathbb{Z}_p{}^2$ and then computes $(\tilde{X}, \tilde{Y}) \leftarrow (\tilde{g}^x, \tilde{g}^y)$, and sk is (x, y), and pk is $(\tilde{g}, \tilde{X}, \tilde{Y})$. Aggregators jointly determine blockchain parameters, such as *blockchain type selection*, *block generation speed,* and *block size*. Aggregators can jointly maintain a blockchain due to their considerable computing and storage capacity.

6.2.2. Register. The EV $\mathrm{id}_{\mathrm{ev}}$ generates its private/public key pair $(\mathrm{sk}_{\mathrm{ev}}, \mathrm{pk}_{\mathrm{ev}})$ and then sends $\mathrm{pk}_{\mathrm{ev}}$ to the TA. The EV randomly chooses s_{ev}, generates $(\delta, \tilde{\delta}) \leftarrow (g^{s_{\mathrm{ev}}}, \tilde{Y}^{s_{\mathrm{ev}}})$ and a signature $\theta \leftarrow \mathrm{Sign}(\mathrm{sk}_{\mathrm{ev}}, \delta)$, and then sends them along with $\mathrm{id}_{\mathrm{ev}}$ to the TA. The TA checks whether θ is valid and then whether $e(\delta, \tilde{Y}) = e(g, \tilde{\delta})$. Then, the EV gives the zero-knowledge proof that he owns the s_{ev}. After that, the TA generates a random number r and computes $\sigma \leftarrow (\sigma_1, \sigma_2) \leftarrow (g^r, (g^x \cdot \delta^y)^r)$. It is a valid signature on s_{ev}. In the end, the TA stores $(\mathrm{id}_{\mathrm{ev}}, \delta, \theta, \tilde{\delta})$ in a secret register and sends σ and $e(\sigma_1, \tilde{Y})$ to the EV. Finally, the EV's group public key is δ, and group private key is $\mathrm{gsk}_{\mathrm{ev}} = (s_{\mathrm{ev}}, \sigma, e(\sigma_1, \tilde{Y}))$.

Similarly, the charging station $\mathrm{id}_{\mathrm{cs}}$ generates its private/public key pair $(\mathrm{sk}_{\mathrm{cs}}, \mathrm{pk}_{\mathrm{cs}})$, randomly chooses s_{cs}, and generates $(\delta_{\mathrm{cs}}, \tilde{\delta}_{\mathrm{cs}}) \leftarrow (g^{s_{\mathrm{cs}}}, \tilde{Y}^{s_{\mathrm{cs}}})$. Through the similar interaction with TA, the charging station obtains its own group public key is $\delta\imath$, and group private key is $\mathrm{gsk}_{\mathrm{cs}} = (s_{\mathrm{cs}}, \sigma_{\mathrm{cs}}, e(\sigma_{1\,\mathrm{cs}}, \tilde{Y}))$. The TA stores $(\mathrm{id}_{\mathrm{cs}}, \delta_{\mathrm{cs}}, \theta_{\mathrm{cs}}, \tilde{\delta}_{\mathrm{cs}})$ in his secret register. The aggregator generates its private/public key pair $(\mathrm{sk}_{\mathrm{ag}}, \mathrm{pk}_{\mathrm{ag}})$ and then sends $\mathrm{pk}_{\mathrm{ag}}$ to the TA. The TA stores it in the secret register.

Meanwhile, any entity that wants to participate in electricity trading (*i.e.*, auction) needs to have an account on the blockchain, and there are a certain number of tokens in the account for purchasing or paying electricity. We use account $:= \{\mathrm{pk}_{\mathrm{address}}, \mathrm{value}\}$ to abstract the account. The aggregator, as an auctioneer, also needs to deposit enough tokens in the smart contract of the blockchain, which is in charge of the TA. When the illegal behavior of the auctioneer is detected, it can be punished.

6.2.3. Request for Auctions. When an EV needs to be charged, it sends a charging request CR to the aggregator, using the anonymous authentication based on PS group signatures. The specific operations are as follows:

The EV sets the required parameters $\mathrm{PA} := \{\mathrm{id}, \mathrm{eq}, \mathrm{model}, t_1, t_2, \mathrm{pr}\}$: id ID represents the unique serial number of the auction, eq is the required electric quantity, model is the charging model, t_1 is the deadline for accepting bids, t_2 is the deadline for the whole auction, and pr is the maximum price of a kilowatt-hour accepted by himself.

The EV first randomly chooses a k_{ev} and the public key $K = \tilde{g}^{k_{ev}}$, where $(k_{\mathrm{ev}}, K_{\mathrm{ev}})$ is the temporary public-private key pair for the future usage.

Then, the EV uses hashed ElGamal encryption to encrypt PA, and the ciphertext is (c_1, c_2), where $c_1 = \tilde{g}^{r\prime}$ and $c_2 = \mathscr{H}(\tilde{X}^{r\prime}) \oplus (\mathrm{PA} \| K_{\mathrm{ev}})$. Then, the EV uses the $GSign((\mathrm{gsk}_{\mathrm{ev}}))$ to sign the message $\mathrm{PA} \| K_{\mathrm{ev}}$. In particular, the EV needs to randomize σ using a random number t and computes $(\sigma_1', \sigma_2') \leftarrow (\sigma_1^t, \sigma_2^t)$.

Then, he randomly chooses $k \in \mathbb{Z}_p$ and computes $c \leftarrow \mathscr{H}(\sigma_1', \sigma_2', e(\sigma_1, \tilde{Y})^{k \cdot t}, \mathrm{PA} \| K_{\mathrm{ev}})$, where $\mathscr{H}$ is a secure hash function. Finally, the EV computes $s \leftarrow k + c \cdot s_{\mathrm{ev}}$ and outputs $(\sigma_1{}', \sigma_2{}', c, s)$ as the group signature μ on the message $\mathrm{PA} \| K_{\mathrm{ev}}$.

Then, the EV keeps the $(k_{\mathrm{ev}}, K_{\mathrm{ev}})$ in his secret register and sends to the aggregator the request tuple for auctions $\mathrm{CR} := \{\sigma_1', \sigma_2', c_1, c_2, c, s, \mathrm{eq}, \mathrm{model}, t_1, \mathrm{pr}\}$.

When the aggregator receives the request CR, he decrypts the ciphertext (c_1, c_2) and checks whether the PS group signature is valid. If it is valid, he releases the auction information to the blockchain. The specific operations are as follows:

The aggregator decrypts the ciphertext (c_1, c_2), by computing $\mathrm{PA} = c_2 \oplus \mathscr{H}(c_1^x)$.

To verify whether the signature $(\sigma_1', \sigma_2', c, s)$ is valid, the aggregator computes $T \leftarrow e(\sigma_{1\prime}, \tilde{X})^c \cdot e(\sigma_{2\prime}, \tilde{g})^{-c} \cdot e(\sigma_{1\prime}, \tilde{Y})^s$ and $c = \mathscr{H}(\sigma_1', \sigma_2', T, \mathrm{PA})$. If it is valid, the aggregator sends the auction information $\{\mathrm{eq}, \mathrm{model}, t_1, \mathrm{pr}\}$ to the smart contracts deployed on the blockchain; otherwise, the aggregator rejects the request.

6.2.4. Upload Bids. The charging stations find the request on the blockchain, and they commit their supply requests, including the commitment value of electricity bid, to the aggregator before t_1. The specific operations are as follows:

The charging station chooses the price v of a kilowatt-hour, randomly chooses r, computes the commitment $\mathrm{cm}: = g^r h^v$, and sets the bid as $\mathrm{bid}: = \{\mathrm{id}, \mathrm{pk}_{\mathrm{address}}, \mathrm{cm}\}$, where $i\ d$ represents the auction number participating in the bidding and $\mathrm{pk}_{\mathrm{address}}$ indicates the address of the charging station on the blockchain. At the same time, some funds need to be sent to the aggregator as deposits.

6.2.5. Open Bids. All bidders open the committed value to the auctioneer, and the auctioneer gets the highest bid and uploads the winner's identity (*i.e.*, *address*) to the blockchain before t_2. The winner can sell his electricity with the auction requester. Suppose the requester has α tokens, the smart contract will transfer β tokens $((\beta = v * \mathrm{eq}))$ in the requester's deposit to the winner's account, and the rest $((\alpha - \beta))$ will be returned to the requester's account. Then, the winner needs to prove that he has a legal identity; *i.e.*, he needs to sign the auction results.

The charging station first randomly chooses a k_{cs} and the public key $K_{\mathrm{cs}} = \tilde{g}^{k_{\mathrm{cs}}}$, where $(k_{\mathrm{cs}}, K_{\mathrm{cs}})$ is the temporary public-private key pair for the future usage.

Then, the charging station uses hashed ElGamal encryption to encrypt $\mathrm{bid} \| K_{\mathrm{cs}}$, and the ciphertext is $(c_{1\prime}, c_{2\prime})$, where $c_{1\prime} = \tilde{g}^{r\prime\prime}$ and $c_{2\prime} = \mathscr{H}(\tilde{X}^{r\prime\prime}) \oplus (\mathrm{bid} \| K_{\mathrm{cs}})$.

Then, the charging station gives a zero-knowledge proof that $\mathcal{SOK} = \{(\mathrm{sk}_{\mathrm{address}}): \mathrm{pk}_{\mathrm{address}} = g^{\mathrm{sk}_{\mathrm{address}}}\}(\mathrm{bid}\|K_{\mathrm{cs}})$.

Then, the charging station uses the *GSign* $((\mathrm{gsk}_{\mathrm{cs}}))$ to sign the message K_{cs}. In particular, the charging station needs to randomize σ_{cs} using a random number $t\prime$ and computes $(\sigma_{1\,\mathrm{cs}}{}', \sigma_{2\,\mathrm{cs}}{}') \leftarrow (\sigma_{1\,\mathrm{cs}}^{t\prime}, \sigma_{2\,\mathrm{cs}}^{t\prime})$. Then, he randomly chooses $k\prime \in \mathbb{Z}_p$ and computes $c\prime \leftarrow \mathcal{H}(\sigma_{1\,\mathrm{cs}}{}', \sigma_{2\,\mathrm{cs}}{}', e(\sigma_{1\,\mathrm{cs}}, \widetilde{Y})^{k\prime \cdot t\prime}, \mathrm{bid}\|K_{\mathrm{cs}})$, where $\mathcal{H}$ is a secure hash function. Finally, the EV computes $s\prime \leftarrow k\prime + c\prime \cdot s_{\mathrm{cs}}$ and outputs $(\sigma_{1\,\mathrm{cs}}{}', \sigma_{2\,\mathrm{cs}}{}', c\prime, s\prime)$ as the group signature $\mu\prime$ on the message $\mathrm{bid}\|K_{\mathrm{cs}}$.

Then, the charging station keeps the $(k_{\mathrm{cs}}, K_{\mathrm{cs}})$ in his secret register and sends to the aggregator the tuple $\{\sigma_{1\,\mathrm{cs}}{}', \sigma_{2\,\mathrm{cs}}{}', c_{1\prime}, c_{2\prime}, c\prime, s\prime, \mathcal{SOK}\}$.

When the aggregator receives the tuple, he decrypts the ciphertext $(c_{1\prime}, c_{2\prime})$ and checks whether the PS group signature is valid. If it is valid, he sends K_{cs} to the EV and sends K_{ev} to the charging station. Then, the charging station and the EV can establish trusted communication for power supply operation. The specific operations are as follows:

The aggregator decrypts the ciphertext $(c_{1\prime}, c_{2\prime})$, by computing $\mathrm{PA} = c_{2'} \oplus \mathcal{H}(c_1^x)$.

To verify whether the signature $(\sigma_{1\,\mathrm{cs}}', \sigma_{2\,\mathrm{cs}}', c\prime, s\prime)$ is valid, the aggregator computes $T\prime \leftarrow e(\sigma_{1\prime\,\mathrm{cs}}, \widetilde{X})^{c\prime} \cdot e(\sigma_{2\prime\,\mathrm{cs}}, \widetilde{g})^{-c\prime} \cdot e(\sigma_{1\prime\,\mathrm{cs}}, \widetilde{Y})^{s}$ and $c = \mathcal{H}(\sigma_{1\,\mathrm{cs}}', \sigma_{2\,\mathrm{cs}}', T\prime, \mathrm{bid}\|K_{\mathrm{cs}})$.

If it is valid, the aggregator sends K_{cs} to the EV and sends K_{ev} to the charging station. The EV computes $k = K_{\mathrm{cs}}^{k_{\mathrm{ev}}}$, and the aggregator computes $k = K_{\mathrm{ev}}^{k_{\mathrm{cs}}}$ for secure communications. They can conduct offline power supply operation after negotiation.

Handling Malicious Winners. There is a situation that when a malicious charging station wins the auction, it does not carry out subsequent operations. To prevent this situation, the aggregator can call the smart contract to deduct the deposit of the charging station.

6.2.6. Dispute. To ensure the accountability of the auction scheme, we allow anyone to question the auction results. That is, the price of the winner's bid is not the lowest price.

One can choose any bid participating in the auction to compare with the winner's price and upload the request to the smart contract by consuming a small amount of token.

The auctioneer generates a zero-knowledge proof to prove that the value in the designated bid commitment is higher than the winner's price. If the auctioneer is unable to make the proof within the specified time, the smart contract will deduct a certain proportion of the auctioneer's deposit as a punishment.

6.2.7. Trace. When we need to open someone, the TA searches in the list $(\mathrm{id}_i, \delta_i, \theta_i, \widetilde{\delta}_i)$ and checks whether $e(\sigma_2, \widetilde{g}) \cdot e(\sigma_1, \widetilde{X})^{-1} = e(\sigma_1, \widetilde{\delta}_i)$ until he gets a match. He then outputs a corresponding (i, δ_i, θ_i) with a proof of knowledge $\widetilde{\delta}_i$.

7. Security Analysis

In this section, we briefly analyze the properties of the scheme.

EV/Charging Station Authentication. As the PS group signature is EUF-CMA under LRSW assumption, all anonymous authentications of EVs and charging stations use PS group signatures, namely $(\sigma_1', \sigma_2', c, s)$ and $(\sigma_{1\,\mathrm{cs}}{}', \sigma_{2\,\mathrm{cs}}{}', c\prime, s\prime)$. So, we can reduce the authentication security to the signature's security.

EV/Charging Station Privacy. The EVs and charging stations use the anonymous method to send auction requests and supply requests to the aggregator. Each signature will be randomized with random numbers, so their identity information will not be disclosed.

Auction Privacy. We use the Pedersen commitment $\mathrm{cm} := g^r h^v$ to hide the information of the biddings. As the hiding properties of a cryptography commitment scheme, there is no $\mathcal{PPT}$ $\mathcal{A}$ who can know the hidden value v of a commitment. Only the bidding information of the final winner is published on the blockchain. Therefore, thanks to the FSA model, our scheme can protect the privacy of bidders as much as possible.

Traceability. The TA can trace the identity of a malicious EV/charging station when needed. The regulatory authority can send the signature that needs to be traced to the TA, and then, we use the *GOpen* algorithm to trace it. That is, the TA searches in the list $(i, \delta_i, \theta_i, \widetilde{\delta}_i)$ and checks whether $e(\sigma_{2\prime}, \widetilde{g}) \cdot e(\sigma_{1\prime}, \widetilde{X})^{-1} = e(\sigma_{1\prime}, \widetilde{\delta}_i)$ until he gets a match. Finally, he finds the corresponding (i, δ_i, θ_i).

Accountability. The accountability of our scheme is based on the premise of rational participants. One can choose any bid participating in the auction to compare with the winner's price and upload the request to the smart contract by consuming a small amount of token. The auctioneer needs to prove that the value in the designated bid commitment is higher than the winner's price. If he cannot, he will be deducted from a certain deposit. We assume that all auctioneers are rational. They do not want to be deducted because of cheating, so they execute the agreement honestly.

8. Future Work

In this section, we discuss some shortcomings of this study and possible future work directions.

8.1. Privacy-Preserving Payment on Blockchain. Although the scheme in this study takes into account the privacy of EV/charging station's identity, the payment process on the blockchain is not private, which means that it is possible to

Table 1: Experimental results on Miracle Library.

Notions	Description	Values (ms)
T_{bp}	Bilinear pairing	8.34
$T_{add\,1}$	Point addition in $\mathbb{G}_1$	0.01
$T_{mul\,1}$	Point multiplication in $\mathbb{G}_1$	2.82
$T_{add\,2}$	Point addition in $\mathbb{G}_2$	0.02
$T_{mul\,2}$	Point multiplication in $\mathbb{G}_2$	2.31
T_{mul}	Multiplication operation in $\mathbb{G}_T$	0.01
T_{exp}	Exponentiation operation in $\mathbb{G}_T$	0.58
$T_{\mathscr{H}}$	Hash function	0.03

Table 2: Computation costs of offchain part.

Stages		Computation costs
System initialization		$2\ T_{mul\,2}$
Register	EV	$1\ T_{add\,1} + 5\ T_{mul\,1}\ 1\ T_{mul\,2} + 3\ T_{bp}$
	Charging station	$1\ T_{add\,1} + 5\ T_{mul\,1}\ 1\ T_{mul\,2} + 3\ T_{bp}$
	Aggregator	$1\ T_{mul\,1}$
Request for auctions		$2\ T_{mul\,1} + 4\ T_{mul\,2} + 4\ T_{exp} + 2\ T_{mul} + 4\ T_{bp} + 4\ T_{\mathscr{H}}$
Upload bids		$t \cdot (1\ T_{add\,1} + 2\ T_{mul\,1})$, where t is number of bidders
Open bids		$1\ T_{add\,1} + 5\ T_{mul\,1} + 6\ T_{mul\,2} + 4\ T_{exp} + 2\ T_{mul} + 4\ T_{bp} + 3\ T_{\mathscr{H}}$
Trace		$d \cdot (1\ T_{exp} + 1\ T_{mul} + 3\ T_{bp})$, where d is the average number of matching queries

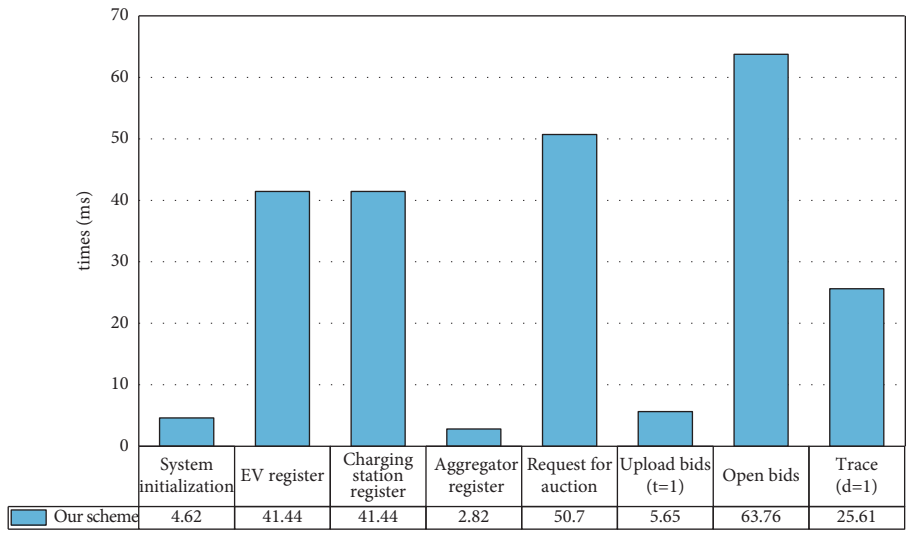


Figure 4: Computation costs of offchain part.

obtain EV or aggregator identity through the correlation of payment. There is some work in this area [31], and our scheme can adopt these strategies. In the future, we can also study the privacy protection payment system suitable for V2G and smart grid.

8.2. Light Client. In this study, the blockchain is jointly maintained by the aggregator. EVs and charging stations can view the blockchain information without writing information to the blockchain. Considering the storage and computing power of EV and charging station, as well as the development of vehicle networking, this assumption is feasible. However, the light client mode may be more suitable for the existing scheme. EVs and charging stations, as light clients, only need to maintain a small amount of data (*i.e.*, block header information) to verify the correctness of auction information on the blockchain. A series of studies on light clients [32,33] can be transplanted into our scheme.

8.3. More Efficient Auction Protocol. Although the FSA model is used in this scheme, there are actually more efficient auction protocols [30]. Combined with the scenarios of smart grid and V2G, considering the dynamic mobility of EVs, how to design a more efficient and secure auction scheme is also a potential future research direction.

9. Implementation

In this section, we evaluate the scheme by testing the onchain and offchain parts, respectively.

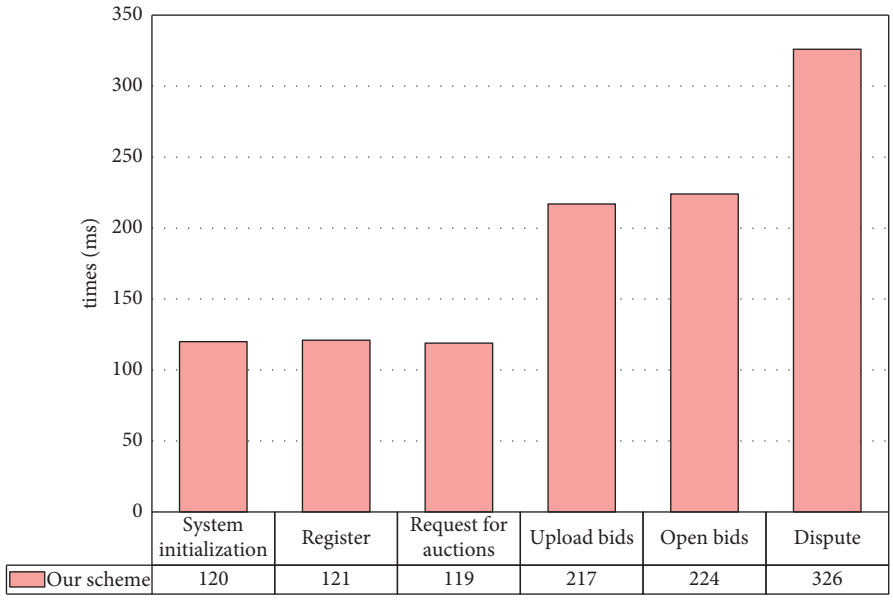


Figure 5: Experimental results of onchain part.

9.1. Experimental Environments. The experiment is carried out on a laptop with an i5-10400 Processor, 8G bytes RAM, and Windows 10 operating system. We utilize the Miracle Library [34] to implement cryptographic primitives. Hyperledger Fabric v1.4 is used in our experiment, and it is a stable version with long-term support. The smart contracts are developed in GoLang (v1.15.2) and tested using Fabric's dev-mode.

9.2. Evaluation Findings of Offchain Part. For time costs of anonymous authentication, we test the time overhead of the primary operations in Table 1 using the Miracle Library. We use a supersingular elliptic curve, with order divisible by $q = 2^{159} + 2^{17} + 1$ and security multiplier $k = 2$. The prime p is 512 bits. We count the number of main operations of our scheme in Table 2 and give the computation costs of the offchain part in our scheme in Figure 4.

9.3. Evaluation Findings of Onchain Part. For time costs on the blockchain, we evaluate the time costs of the steps in a local network with Raft modes. We set the same local private network configuration for each mode, including one channel and two organizations, and each organization has two peers.

We implement the steps on the blockchain, which correspond to our previous definitions. The results are shown in Figure 5. To ensure the accuracy of the experimental data, all the time consumption is obtained by executing the code 100 times to get the average value. In *system initialization* phase, we need to deploy smart contracts on the blockchain. In *register* phase, we need to initialize the accounts of EV, charging station, and aggregator and store certain tokens in the account. In *request for auction* phase, the aggregator uploads the information for bidding, *i.e.*, the auction request $\mathrm{PA} := \{\mathrm{id}, \mathrm{eq}, \mathrm{model}, t_1, t_2, \mathrm{pr}\}$. In *upload bid* phase, the charging station uploads the commitment for bidding. In *open bid* phase, the aggregator uploads the winner's address to the smart contract. In *dispute* phase, the challenger needs to send a challenge request to the blockchain, while the auctioneer needs to generate a zero-knowledge proof and then send it to the smart contract for verification.

In summary, we perform the experimental results of the offchain and onchain parts in our scheme. The results show that our proposal is efficient and suitable for V2G networks in the smart grid.

10. Conclusion

In this study, we systematically summarized the opportunities and challenges of blockchain in the smart grid. Then, we proposed a privacy-preserving blockchain-based electricity auction scheme for V2G networks in the smart grid under the FSA model, which inspires the applications of blockchain in the smart grid. Our scheme can ensure the privacy of EVs and charging stations, while using smart contracts to provide reliability. The experimental results showed the feasibility and practicality of our scheme.

Data Availability

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

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

The work was supported by the State Grid Henan Electric Power Company Science and Technology Project (No. 52170220009S).

References

[1] M. Z. Gunduz and R. Das, "Cyber-security on smart grid: threats and potential solutions," *Computer Networks*, vol. 169, p. 107094, 2020.

[2] Y. Guo, Z. Wan, and X. Cheng, "When Blockchain Meets Smart Grids: A Comprehensive Survey," *High-Confidence Computing*, vol. 2, no. 2, Article ID 100059, 2022.

[3] W. Han and Y. Xiao, "Privacy preservation for v2g networks in smart grid: a survey," *Computer Communications*, vol. 91-92, pp. 17–28, 2016.

[4] N. Fabra, N.-H. Fehr, and D. Harbord, "Designing electricity auctions," *The RAND Journal of Economics*, vol. 37, no. 1, pp. 23–46, 2006.

[5] J. Nicolaisen, V. Petrov, and L. Tesfatsion, "Market power and efficiency in a computational electricity market with discriminatory double-auction pricing," *IEEE Transactions on Evolutionary Computation*, vol. 5, no. 5, pp. 504–523, 2001.

[6] X. Li, P. Jiang, T. Chen, X. Luo, and Q. Wen, "A survey on the security of blockchain systems," *Future Generation Computer Systems*, vol. 107, pp. 841–853, 2020.

[7] Q. Feng, D. He, S. Zeadally, M. K. Khan, and N. Kumar, "A survey on privacy protection in blockchain system," *Journal of Network and Computer Applications*, vol. 126, pp. 45–58, 2019.

[8] G. Wood, "Ethereum: A secure decentralised generalised transaction ledger," *Ethereum Project Yellow Paper*, vol. 151, pp. 1–32, 2014.

[9] E. Androulaki, A. Barger, V. Bortnikov et al., "Hyperledger Fabric: a distributed operating system for permissioned blockchains," in *Proceedings of the European Conference on Computer Systems (EUROSYS)*, pp. 1–15, Crete, Greece, 17 April 2018.

[10] A. I. Commissioner, "Australia: Oaic Finds Uber Interfered with Privacy Rights," 2021, https://www.dataguidance.com/news/australia-oaic-finds-uber-interfered-privacy-rights.

[11] E. B. Sasson, A. Chiesa, C. Garman et al., "Zerocash: decentralized anonymous payments from bitcoin," in *Proceedings of the IEEE Symposium on Security and Privacy (SP)*, pp. 459–474, IEEE, Berkeley, CA, USA, 18 May 2014.

[12] N. Van Saberhagen, "CryptoNote v2," 2013, https://en.bitcoinwiki.org.

[13] L. Nguyen and R. Safavi-Naini, "Dynamic k-times anonymous authentication," in *Applied Cryptography and Network Security,* pp. 318–333, Springer, Berlin, Heidelberg, 2005.

[14] J. Camenisch and A. Lysyanskaya, "Signature schemes and anonymous credentials from bilinear maps," in *Advances in Cryptology - CRYPTO 2004,* pp. 56–72, Springer, Berlin, Heidelberg, 2004.

[15] M. H. Au, W. Susilo, and Y. Mu, "Constant-size dynamic k-taa," in *Proceedings of the International Conference on Security and Cryptography for Networks (SCN)*, pp. 111–125, Springer, Maiori, Italy, 6 September 2006.

[16] D. Pointcheval and O. Sanders, "Short randomizable signatures," in *Proceedings of the Cryptographers' Track at the RSA Conference (CTRSA)*, pp. 111–126, Springer, San Francisco, CA, USA, 29 February 2016.

[17] A. Hahn, R. Singh, C.-C. Liu, and S. Chen, "Smart contract-based campus demonstration of decentralized transactive energy auctions," in *Proceedings of the IEEE Power & Energy Society Innovative Smart Grid Technologies Conference*, pp. 1–5, IEEE, Washington, DC, USA, 23 April 2017.

[18] J. Wang, Q. Wang, and N. Zhou, "A Decentralized Electricity Transaction Mode of Microgrid Based on Blockchain and Continuous Double Auction," in *Proceedings of the IEEE Power & Energy Society General Meeting*, pp. 1–5, IEEE, Portland, OR, USA, Agust 2018.

[19] B. Ramachandran, S. K. Srivastava, C. S. Edrington, and D. A. Cartes, "An intelligent auction scheme for smart grid market using a hybrid immune algorithm," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 10, pp. 4603–4612, 2010.

[20] M. Stübs, W. Posdorfer, and S. Momeni, "Blockchain-based multi tier double auctions for smart energy distribution grids," in *Proceedings of the IEEE International Conference on Communications Workshops*, pp. 1–6, IEEE, Dublin, Ireland, 7 June 2020.

[21] J. Ma, J. Deng, L. Song, and Z. Han, "Incentive mechanism for demand side management in smart grid using auction," *IEEE Transactions on Smart Grid*, vol. 5, no. 3, pp. 1379–1388, 2014.

[22] M. Wen, R. Lu, J. Lei, H. Li, X. Liang, and X. S. Shen, "SESA: an efficient searchable encryption scheme for auction in emerging smart grid marketing," *Security and Communication Networks*, vol. 7, no. 1, pp. 234–244, 2014.

[23] D. Li, Q. Yang, W. Yu, D. An, Y. Zhang, and W. Zhao, "Towards differential privacy-based online double auctioryjrtn for smart grid," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 971–986, 2019.

[24] Z. Zhou, F. Liu, Z. Li, and H. Jin, "When smart grid meets geo-distributed cloud: an auction approach to datacenter demand response," in *Proceedings of the IEEE Conference on Computer Communications (INFOCOM)*, pp. 2650–2658, IEEE, Hong Kong, China, 26 April 2015.

[25] A. Lysyanskaya, R. L. Rivest, A. Sahai, and S. Wolf, "Pseudonym systems," in *Proceedings of the International Workshop on Selected Areas in Cryptography*, pp. 184–199, Springer, Kingston, ON, Canada, 9 August 1999.

[26] Q. Gang, W. Hong, W. Shimin, and X. Guozhen, "Information-theoretic secure verifiable secret sharing over rsa modulus," *Wuhan University Journal of Natural Sciences*, vol. 11, no. 6, pp. 1849–1852, 2006.

[27] T. ElGamal, "A public key cryptosystem and a signature scheme based on discrete logarithms," *IEEE Transactions on Information Theory*, vol. 31, no. 4, pp. 469–472, 1985.

[28] M. Abdalla, M. Bellare, and P. Rogaway, "The oracle diffie-hellman assumptions and an analysis of dhies," in *Proceedings of the Cryptographers' Track at the RSA Conference (CTRSA)*, pp. 143–158, Springer, San Francisco, CA, USA, 8 April 2001.

[29] S. Nakamoto, "Bitcoin: a peer-to-peer electronic cash system," *Decentralized Business Review*, p. 21260, 2008.

[30] V. Krishna, *Auction Theory*, Academic Press, Cambridge, CB, USA, 2009.

[31] S. Jain, N. J. Ahuja, P. Srikanth et al., "Blockchain and Autonomous Vehicles: Recent Advances and Future Directions," *IEEE Access*, vol. 9, 2021.

[32] B. Bünz, L. Kiffer, L. Luu, and M. Zamani, "Flyclient: super-light clients for cryptocurrencies," in *Proceedings of the IEEE Symposium on Security and Privacy (SP)*, pp. 928–946, IEEE, San Francisco, CA, USA, 18 May 2020.

[33] P. Chatzigiannis, F. Baldimtsi, and K. Chalkias, *SoK: Blockchain Light Clients*, Cryptology ePrint Archive, 2021.

[34] S. software ltd, "Miracle," 2021, https://github.com/miracl/MIRACL.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 2151645, 10 pages
https://doi.org/10.1155/2022/2151645

Research Article

Dynamic Detection and Placement for VSFs over Edge Computing Scenarios: An ACO-Based Approach

Chao Bu,[1] Xinyang Zhang,[1] Jianhui Lv,[2] and Jinsong Wang[1]

[1]*School of Computer Science and Engineering, Tianjin University of Technology, Tianjin 300384, China*
[2]*Pengcheng Laboratory, Shenzhen 518000, China*

Correspondence should be addressed to Jianhui Lv; lvjianhui2012@163.com

Received 21 January 2022; Accepted 22 March 2022; Published 11 April 2022

Academic Editor: Ke Gu

As an extension of cloud computing, the edge computing has become an important pattern to deal with novel service scenarios of the Internet of Everything (IoE), especially for the rapidly increasing different kinds of service requests from edge equipment. It is still a great challenge to satisfy the demands of delay-sensitive applications, so as to optimize the service provision delay for edge equipment under 5G. In this paper, by introducing virtualized service functions (VSFs) into edge computing pattern, the mechanism of service function detection and placement among multiple Edge Computing Servers (ECSs) is proposed. We firstly improve the Ant colony optimization (ACO) method to adapt to the situation that the service requests may frequently change from different edge network domains. Based on the improved ACO, a scheme of searching for the locations (i.e., ECSs) of the requested service functions is devised, so as to optimize the service searching delay. Then, a service function placement scheme is presented, and it deploys most of appropriate service functions in each ECS by predicting the future requested frequencies of functions, which further reduces the overall service provision delay. In addition, it also improves the ECS computing capacity utilization. The simulation results show that the proposed mechanism is feasible and effective.

1. Introduction

The 5th-generation mobile system (5G) is gradually integrating into people's daily life. The requirements of the novel service scenarios brought by the 5G have changed significantly [1]. For example, the delay sensitivity has become one of the most important service demands of edge equipment. Many types of research have been done on providing services mainly by the cloud computing center due to its powerful computing capacity, which enables almost all kinds of virtualized service functions (VSFs) [2, 3] to be placed and performed there. However, the cloud computing servers are usually located far away from most of network edges. With the rapid increasing service requests from mobile equipment that mainly locates at the network edges, the service provision based on the conventional cloud computing may lead to problems such as much higher transmission delay and serious congestion [4]. As an extension of cloud computing, the edge computing deploys the service provision capability near to network edges where most of service requests generate [5, 6]. Its distributed edge computing servers (ECSs) locate much closer to edge equipment so as to adapt to the service provision delay demand.

The 5G is able to support high-speed data transmission, which significantly decreases the service delivery delay of ECSs [7]. However, comparing to the service cluster (SC) in the cloud computing center, each ECS's limited computing capacity means that it can only provide some kinds of services due to the limited number of VSFs placed in it [8]. It is hard for an ECS to satisfy all kinds of service requests for edge equipment, because the corresponding VSFs may have not yet been placed in this ECS when such requests arrive. And the method, by which all such service requests are dealt with by the remote SC or the corresponding VSFs are instantly migrated from the remote SC to this ECS, is obviously unsustainable, especially when the network load is heavy [9]. Considering the fact that other ECSs locating much closer than the SC may have already been placed such

VSFs, the corresponding services can be provided by one of these ECSs. How to effectively find an ECS that currently contains a certain VSF becomes a key issue. In practice, a tremendous amount of different kinds of service requests are constantly generated and delivered, which leave the changing traces of recent service provision information in corresponding equipment [10]. In this paper, the VSF concentration on links is presented by leveraging the above changing traces, based on which an improved ACO-inspired service function detection scheme is devised further to effectively search for the placed locations (i.e., nearby ECSs) of the requested VSFs. It optimizes the concentration updating efficiency to overcome the high delay problem caused by several rounds of iteration of the conventional ACO method.

In addition, an ECS should adapt to the frequent changes in service requests by deploying appropriate VSFs in it in time. It is impossible for an ECS to contain all VSFs due to its limited computing capacity [9]. In fact, a lot of service requests are regional and periodic for an ECS [11]. That is, an ECS may often provide some certain kinds of services due to the working features of its nearby equipment [12]. However, the ECS also may be requested the VSFs that have not yet been placed in it because new applications become popular recently. We take the recent frequencies of the VSFs being performed and requested into account, and the VSF placement scheme is devised in this paper. It enables an ECS to retain the already deployed VSFs with higher being performed frequencies in it and migrate the not yet deployed VSFs with higher being requested frequencies to it, respectively. In this approach, the VSF detection delay and the service delivery delay can be further optimized.

Some types of research have been done on service placement and migration over edge computing scenarios (e.g., [13–16]). In [13], it introduces a framework for optimal placement of service replicas proactively in the 5G edge network. It deploys the multimedia service instances on the trajectory edge nodes by integrating the user's path prediction model, so as to provide an optimal deployment technique that traded off between maximizing the QoE and minimizing the deployment cost. In [14], it studies the container-based service migration problem in edge computing and proposes a service migration mechanism based on mobility awareness. Its service migration mechanism firstly triggers the service migration according to the service density of the current node and then selects the service and the corresponding destination node for the optimization object to minimize the service delay. In [15], it proposes a novel service migration policy method based on deep reinforcement learning and dynamic adaptation in multiaccess edge computing. It innovatively analyzes the different states of edge network quantitatively and applies deep Q-learning to migration methods, which can adjust the learning rate adaptively to implement rapid convergence in the learning process. In [16], it combines prediction mechanism and migration research together to optimize the migration of VNF. It built a system cost evaluation model integrating bandwidth utilization and migration time, and devised a heuristic algorithm to obtain the near-optimal solution. However, they mainly focus on instantly deploying functions on real-time service demands or migrating functions based on long-term iterative learning, which cannot well adapt to the service requests with frequent changes and delay sensitivity.

In this paper, the mechanism of ACO-based VSF detection and placement (AVDP) is proposed, so as to minimize the service provision delay to the edge equipment and optimize the ECS computing capacity utilization. The major contribution and innovations can be summarized as follows. The VSF concentration on links is presented to reflect the frequencies of VSFs being detected recently, and the improved ACO-inspired VSF detection scheme based on the VSF concentration is devised to efficiently search for the placed locations of the requested VSFs. It adapts to the novel service scenarios that the service requests for edge equipment and the deployed locations of VSFs may frequently change under 5G. Furthermore, the scheme of dynamically deploying appropriate VSFs in each ECS is devised, and it places VSFs by predicting the future requested frequencies of VSFs according to the variations of VSFs concentration on links and VSFs requested number in ECSs. Thus, the overall service provision delay is significantly optimized and the ECS computing capacity utilization is efficiently improved.

The remainder of this paper is structured as follows. In Section 2, we present the system framework and workflow of the proposed AVDP. In Section 3, we define the VSF detection ant and the VSF concentration, and devise the scheme of searching for VSFs among multiple ECSs based on them. In Section 4, we devise the scheme of placing appropriate VSFs in suitable ECSs with the ECS computing capacity utilization considered. In Section 5, we present simulation experiments and results. Finally, Section 6 concludes the paper.

2. System Framework

The system framework is shown in Figure 1. The conventional cloud computing center is usually located far away from network edges. It can provide almost all kinds of services due to the powerful computing capacity of the service cluster (SC), and VSFs placed in it can be dispatched to ECSs. ECSs are distributed in end network domains (ENDs) where most of service requests generate to efficiently meet the delay sensitivity demand of edge equipment. The service function pool (SFP) in each ECS only contains some VSFs because of the limited computing capacity. The VSFs already deployed in an ECS can be replaced by other VSFs due to the changing service scenarios in practice, and VSFs can be migrated from the SC or other ECSs. In addition, each switching equipment (SE) updates a table named VSFs detection record (VDR). The VDR is to reflect the current probabilities of successfully finding the locations of the requested VSFs by different next hops, so as to efficiently determine the suitable ECSs that can provide corresponding services.

The overall workflow of the system is shown in Figure 2; here, the number is the action order.

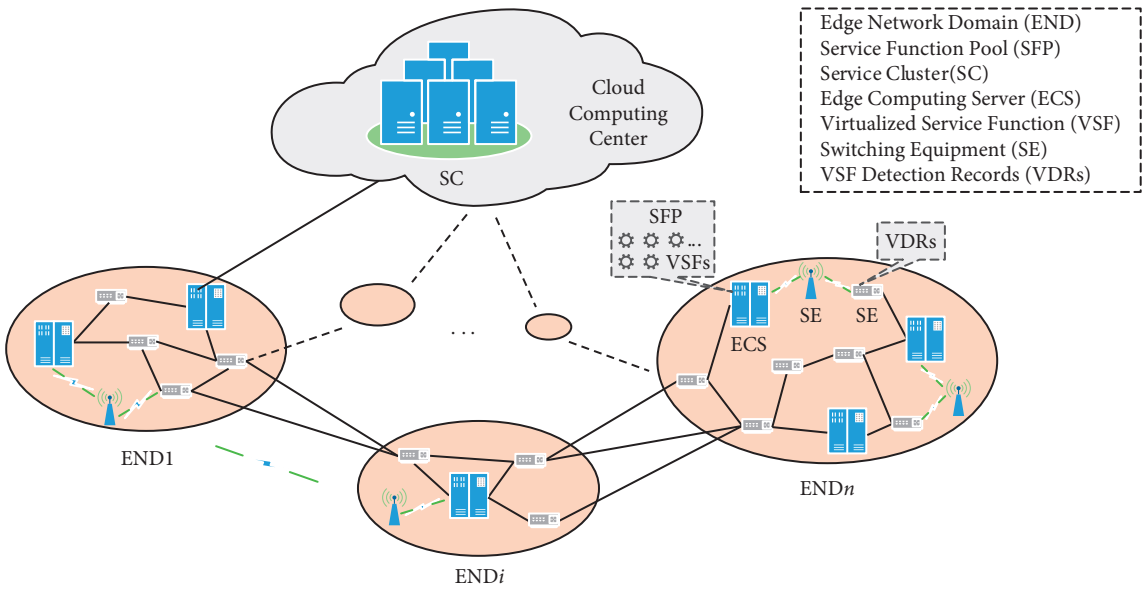


Figure 1: The system framework.

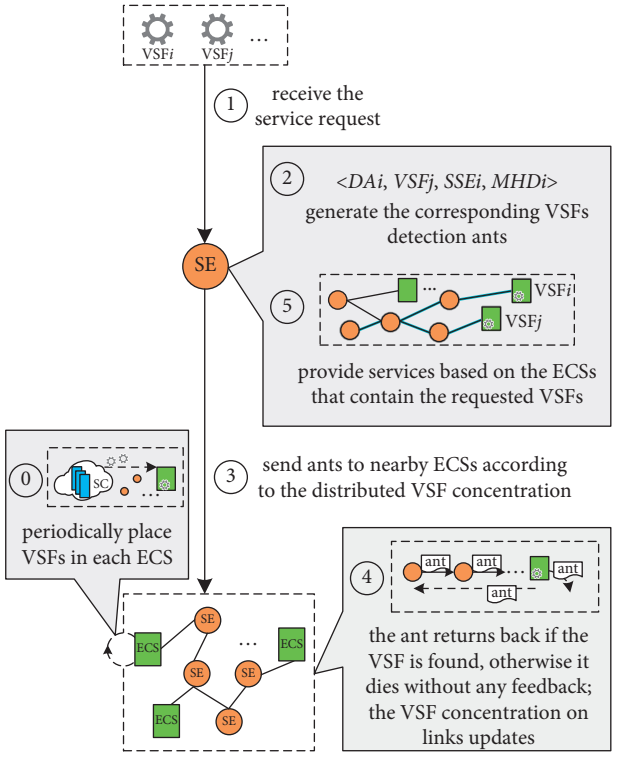


Figure 2: The overall workflow of the system.

3. VSFs Detection

In order to minimize the service provision delay for edge equipment, most of services should be provided by the near ECSs rather than mainly by the SC that is far away from ENDs. It is a key issue to find the ECSs that have already been placed in the requested VSFs as soon as possible. In this section, the VSF detection ant and the VSF concentration are defined in our proposed scheme. The VSF detection ants are sent by the source SE that receives the service request to search for the locations (i.e., ECS) of the requested VSF according to this VSF concentration on links. The information on detecting the VSF is laid over the ants' trails, and the accumulated amount of information is regarded as the VSF concentration that is recorded and updated in the VDR.

In practice, the VSF concentration on links decreases with time. It also increases according to the ants' feedback if these ants have found the locations (i.e., ECSs) of the requested VSF. For instance, if the requested VSF is found by any of the VSF detection ants within the ant's living period, the ant returns back to the source SE, and the VSF concentration on the corresponding links increases. Comparing to the conventional ACO method, our presented VSF concentration updating does not depend on multiple iterations of the same group of ants. In the actual network environment, because of a large number of ongoing service requests for edge equipment, different VSF detection ants can be generated by SEs at any time, which enables different VSF concentrations continuously and efficiently to change in practice combined with the devised detection ant's feedback method.

3.1. The VSF Detection Ant. In this paper, a VSF detection ant is defined as a four-tuple $\langle DA_i, VSF_k, SSE_i, HDA_i \rangle$. Here, DA_i is the unique identifier of a detection ant; VSF_k indicates the VSF that is detected by DA_i; SSE_i is the set of SEs that DA_i has passed; HDA_i denotes the number of survival hops of DA_i.

In detail, DA_i is one of the ants generated by the source SE that receives the service request and is forwarded to search for VSF_k according to the VSF_k concentration on links. The SEs passed by DA_i are orderly recorded in SSE_i to avoid loopback. Meanwhile, once DA_i finds VSF_k, it returns back to the source SE according to SSE_i. In addition, an ant is not allowed to detect the ECSs that are far away from the

source SE due to the delay limitation. HDA_i is used to constrain the survival time of DA_i, and DA_i dies without any feedback if it has been forwarded exceeds HDA_i.

3.2. The VSF Detection Concentration. We assume two factors will influence the VSF concentration on links, which are time and ants' feedback. $VCL_{SE_u,SE_v}^{VSF_k}(t)$ is defined as the VSF_k concentration on the link from the SE SE_u to the SE SE_v at time t, and its value changes with time. At time $t+1$, its value is shown as follows:

$$VCL_{SE_u,SE_v}^{VSF_k}(t+1) = RVCL_{SE_u,SE_v}^{VSF_k}(t,t+1) + AVCL_{SE_u,SE_v}^{VSF_k}(t,t+1). \tag{1}$$

Here, $RVCL_{SE_u,SE_v}^{NSF_k}(t,t+1)$ and $AVCL_{SE_u,SE_v}^{NSF_k}(t,t+1)$ are the remaining concentration after volatilizing and the additive concentration after ants' feedback from t to $t+1$, respectively. They are defined as follows:

$$\begin{aligned} RVCL_{SE_u,SE_v}^{VSF_k}(t,t+1) &= \int_t^{t+1} VCL_{SE_u,SE_v}^{VSF_k}(t)\cdot e^{-\gamma\cdot t}dt, \\ AVCL_{SE_u,SE_v}^{VSF_k}(t,t+1) &= \sum_{w=1}^{m} \Delta VCL_{SE_u,SE_v}^{VSF_k}(t,t+1)\cdot x_w. \end{aligned} \tag{2}$$

Different from the conventional ACO method, the VSF remaining concentration volatilizes faster and faster with time, because the frequent changing service requests from edge equipment may lead to the frequent migrating of VSFs. $RVCL_{SE_u,SE_v}^{VSF_k}(t,t+1)$ is designed to be continuously differentiable, and γ is a positive constant. The VSF additive concentration can only be brought by the survival ants that have found the requested VSF. The closer the link to the ECS that contains the requested VSF, the more this VSF concentration on the link increases. $AVCL_{SE_u,SE_v}^{VSF_k}(t,t+1)$ is designed to promote the searching efficiency for the following ants. m is the number of the VSF_k detection ants that have returned back between t and $t+1$. And x_w is related to the sequential position of SE_u in SSE_i, and it is defined as follows:

$$x_w = \begin{cases} \dfrac{SSE_i[SE_u]}{|SSE_i|}, & SE_u \in SSE_i, \\ 0, & \text{otherwise}. \end{cases} \tag{3}$$

Here, $|SSE_i|$ is the element number of SSE_i, and $SSE_i[SE_u]$ is the position of SE_u in SSE_i.

3.3. The Forwarding Probability. SE_v is defined as an adjacent SE of SE_u if a detection ant can be forwarded by SE_u to SE_v by only one hop. Not all adjacent SEs of SE_u can receive the detection ants from SE_u due to the limited number of the ants generated by the source SE. According to $\langle DA_i, VSF_k, SSE_i, HDA_i\rangle$, we define $SASE_u$ is the set of adjacent SEs of SE_u, and $SASE_u^{VSF_k}$ is the set of adjacent SEs that can receive the VSF_k detection ant (e.g., DA_i) at the next hop from SE_u, which are shown as follows:

$$SASE_u^{VSF_k} = SASE_u - SSE_i \cap SASE_u. \tag{4}$$

When searching for VSF_k, DA_i stops searching without any feedback if $SASE_u^{VSF_k}$ is empty. We assume $FP_{SE_u,SE_v}^{\forall VSF_k}(t+1)$ is the forwarding probability that DA_i is forwarded from SE_u to SE_v, $SE_v \in SASE_u$. Here, the adjacent SE with higher $FP_{SE_u,SE_v}^{VSF_k}(t+1)$ only means that it can be forwarded more VSF_k detection ants rather than be forwarded all VSF_k detection ants. The $FP_{SE_u,SE_v}^{VSF_k}(t+1)$ is defined as follows:

$$FP_{SE_u,SE_v}^{VSF_k}(t+1) = \frac{VCL_{SE_u,SE_v}^{VSF_k}(t+1)}{\sum_{SE_z \in SASE_u^{VSF_k}} VCL_{SE_u,SE_z}^{VSF_k}(t+1)}. \tag{5}$$

Especially, $FP_{SE_u,SE_v}^{VSF_k}(0) = 1/|SASE_u^{VSF_k}|$.

The working process of searching for the VSF_k is shown as Algorithm 1. In detail, firstly, when an SE receives the service request, as the source SE, it generates a certain number of detection ants to search for the requested VSF that is VSF_k here (line 2). Secondly, the same kinds of ants are divided into several groups in each arriving SE, with SSE_i of each ant being considered to avoid the loopback, the numbers of ants in different groups are determined according to the current concentrations of VSF_k on different links. Then, different groups of ants are forwarded to adjacent nodes of the current node (lines 3–5). Thirdly, if an VSF_k detection ant finds VSF_k, it adds the current node into its SSE_i and returns back to the source SE. Meanwhile, the VSF_k concentrations on the corresponding searching links update and the current node (i.e., ECS) is added into ECS^{VSF_k} (lines 6–11). Else if HDA_i of the ant is not zero, the ant adds this node into its SSE_i and joins to the next-hop searching (lines 12–14). Otherwise, the ant dies without any feedback (lines 15–16). Finally, the set of ECSs that are found currently contain VSF_k (i.e., ECS^{VSF_k}) is obtained (line 20).

4. VSFs Placement

In order to optimize the service delivery delay, the ECSs locating near to the ENDs where the service requests generate should have been deployed in or migrated to the requested VSFs. In this paper, two kinds of VSFs that should have been placed in each ECS are considered, which are deployed VSFs (DVSFs) and migrated VSFs (MVSFs). DVSFs are the VSFs that have already been deployed in the ECS and are still being massively requested recently. MVSFs are the VSFs that should be migrated to the ECS due to the rapidly growing requests for them recently. Therefore, due to the ECS limited computing capacity, the VSFs that have already been deployed but rarely be requested recently in this ECS should be replaced by the MVSFs.

The performed frequencies of the already deployed VSFs in an ECS can be estimated according to these VSFs' current concentrations on the links around this ECS. Assume that $SASE_q$ is defined as the set of adjacent SEs of the ECS, ECS_q. The ratio of performed frequency of VSF_k to the performed frequencies of all already deployed VSFs in ECS_q from t to $t+1$ is defined as $RPF_q^{VSF_k}(t,t+1)$, shown as follows:

$$\mathrm{RPF}_q^{\mathrm{VSF}_k}(t,t+1)=\frac{\sum_{\mathrm{SE}_z\in\mathrm{SASE}_q}\mathrm{VCL}_{\mathrm{SE}_z,\mathrm{ECS}_q}^{\mathrm{VSF}_k}(t+1)}{\sum_{\mathrm{VSF}_e\in\mathrm{SVSF}_q(t,t+1)}\sum_{\mathrm{SE}_z\in\mathrm{SASE}_q}\mathrm{VCL}_{\mathrm{SE}_z,\mathrm{ECS}_q}^{\mathrm{VSF}_e}(t+1)}. \tag{6}$$

Here, $\mathrm{SVSF}_q(t,t+1)$ is the set of VSFs that have been performed in ECS_q from t to $t+1$.

Let $\mathrm{CS}_q^{\mathrm{VSF}_k}(t+1-m,t+1)$ be the concentration stability of VSF_k on the links around ECS_q during the recent m time periods, shown as follows:

$$\mathrm{CS}_q^{\mathrm{VSF}_k}(t+1-m,t+1)=\sqrt{\frac{1}{m}\sum_{x=t+1-m}^{t+1}\left(\sum_{\mathrm{SE}_z\in\mathrm{SASE}_q}\mathrm{VCL}_{\mathrm{SE}_z,\mathrm{ECS}_q}^{\mathrm{VSF}_k}(x)-\frac{\sum_{y=t+1-m}^{t+1}\sum_{\mathrm{SE}_z\in\mathrm{SASE}_q}\mathrm{VCL}_{\mathrm{SE}_z,\mathrm{ECS}_q}^{\mathrm{VSF}_k}(y)}{m}\right)^2},\quad t+1\geq m. \tag{7}$$

Assume RTV and CTV are the threshold values of the VSF's ratio of performed frequency and concentration stability, respectively. The current set of DVSFs that can be retained in ECS_q is defined as CSD_q, obviously, $\mathrm{CSD}_q\subseteq\mathrm{SVSF}_q(t,t+1)$. The VSFs in CSD_q should satisfy the following conditions:

$$\mathrm{RPF}_q^{\mathrm{VSF}_k}(t,t+1)\geq\mathrm{RTV}, \tag{8}$$

$$\mathrm{CS}_q^{\mathrm{VSF}_k}(t+1-m,t+1)\leq\mathrm{CTV},\quad t+1\geq m, \tag{9}$$

or

$$\mathrm{RPF}_q^{\mathrm{VSF}_k}(t,t+1)<\mathrm{RTV}, \tag{10}$$

$$\sum_{\mathrm{SE}_z\in\mathrm{SASE}_q}\mathrm{VCL}_{\mathrm{SE}_z,\mathrm{ECS}_q}^{\mathrm{VSF}_k}(h)\geq\sum_{\mathrm{SE}_z\in\mathrm{SASE}_q}\mathrm{VCL}_{\mathrm{SE}_z,\mathrm{ECS}_q}^{\mathrm{VSF}_k}(h-1),\quad t+1-m<h\leq t+1. \tag{11}$$

Let $\mathrm{SCV}_q(t+1)$ be the set of candidate VSFs that have been requested but not yet been deployed in ECS_q during the recent m time periods, these VSFs may replace the VSFs in $\mathrm{SVSF}_q(t,t+1)-\mathrm{CSD}_q$ with $\mathrm{SCV}_q(t+1)\cap\mathrm{SVSF}_q(t,t+1)=\varnothing$ satisfied. Let CSM_q be the current set of MVSFs that should be migrated to ECS_q, obviously, $\mathrm{CSM}_q\subseteq\mathrm{SCV}_q(t+1)$. For the VSF $\mathrm{VSF}_d\mathrm{VSF}_d\in\mathrm{SCV}_q(t+1)$, its total requested number in ECS_q during the recent m time periods is defined as $\mathrm{TRN}_q^{\mathrm{VSF}_d}$, and its incremental requested number in ECS_q in each recent time period is denoted as follows:

$$\mathrm{IRN}_q^{\mathrm{VSF}_d}(h)=\mathrm{TRN}_q^{\mathrm{VSF}_d}-\mathrm{IRN}_q^{\mathrm{VSF}_d}(h),\quad t+1-m\leq h\leq t+1. \tag{12}$$

Furthermore, the requested number growth rate of VSF_d in ECS_q in the $(h)th$ day is denoted as follows:

$$\mathrm{RNGR}_q^{\mathrm{VSF}_d}(h)=\frac{\mathrm{IRN}_q^{\mathrm{VSF}_d}(h)}{\mathrm{TRN}_q^{\mathrm{VSF}_d}},\quad t+1-m\leq h\leq t+1. \tag{13}$$

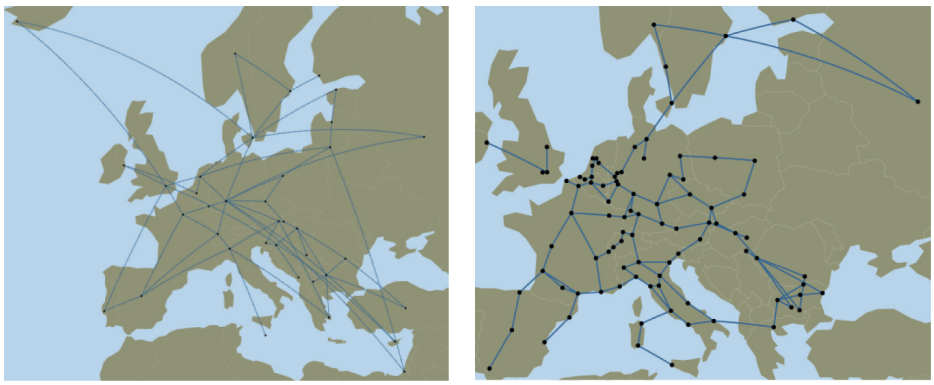
In this approach, the VSFs in CSM_q can be selected from $\mathrm{SCV}_q(t+1)$ by comparing the VSF's requested number growth rate from high to low with the ECS's available computing capacity considered. And these VSFs can be migrated from the SC in the cloud computing center or other ECSs. In addition, a VSF in CSM_q may become one element of CSD_q if it satisfies the conditions of equations (8) and (9), or equations (10) and (11) in the following time periods. Therefore, the VSFs that should be placed in ECS_q before the next time period are the ones belonging to $\mathrm{CSD}_q\cup\mathrm{CSM}_q$.

5. Simulation and Results

5.1. The Simulation Setup. In the simulation, the proposed schemes are implemented in Python and all experiments are performed on a computer with one Intel(R) Core(TM) i7-6700 CPU @ 3.40 GHz and 16 GB of RAM. We evaluate the approaches in two typical and real network topologies (e.g., ISP and backbone networks) with different numbers of nodes and links, called Geant and Interroute, which can be obtained from the Internet Topology Zoo [17]. Specifically, Geant is a middle-scale network topology with 41 nodes and 65 links, and Interroute is a large-scale network topology with 110 nodes and 148 links. They are shown in Figure 3.

Input: SE_u/ * the source switching equipment * /, VSF_k/ * the service function that is requested * /
Output: ECS^{VSF_k}/ * the set of ECSs that contain VSF_k, * /
(1) **Begin**
(2) SE_u generates detection ants, each of which is defined as $\langle DA_i, VSF_k, SSE_i, HDA_i \rangle$;
(3) **while** the set of VSF_k detection ants is not empty **do**
(4) Divide ants into several groups according to the SSE_i of each ant and the distributed VSF_k concentration
(5) Forward different groups to different adjacent nodes of the current node according to equation (5);
(6) **for** each DA_i**do**
(7) **if**VSF_k is found **then**
(8) Add this node into SSE_i;
(9) Return back to SE_u according to SSE_i;
(10) Update VSF_k concentrations on links according to equation (2);
(11) Add this node (i.e., ECS) into ECS^{VSF_k};
(12) **else if**$HDA_i \geq 1$**then**
(13) Add this node into SSE_i;
(14) Join the next-hop searching;
(15) **else if**$HDA_i = 0$**then**
(16) Die without any feedback;
(17) **end if**
(18) **end for**
(19) **end while**
(20) **return**ECS^{VSF_k}
(21) **End**

Algorithm 1: The search for a VSF.



(a) (b)

Figure 3: Network topologies. (a) Geant topology. (b) Interroute topology.

We assume that the network topology is divided into 6 edge network domains, and one ECS is placed among 10 nodes in the simulation. A variety of VSFs are simulated by ClickOS [18], and it can create small virtual machines, each of which is able to host a VSF. We assume that the computing capacity of each ECS follows a uniform distribution between 100 and 200 units. The number of VSFs requested by each request follows a uniform distribution between 2 and 4, and the type of each VSF is random. The ECS computing capacity needed to support each VSF follows the uniform distribution between 5 and 10 units. The simulation parameters and the corresponding distribution model are motivated by the literature [19, 20] that studies the network function provision problem. We compare the proposed AVDP with the scheme of AI-enabled mobile multimedia service instance placement (AMSP) according to [13]. We use the following performance metrics: the service provision delay (SPD), the computing capacity utilization (CCU), and the service access success ratio (SASR).

5.2. The Simulation Results. We compare the SPDs of the two approaches of AVDP and AMSP under Geant and Interroute. The SPD is defined as the time interval from receiving the service request to successfully providing the service. The results are shown in Figures 4 and 5.

As shown in Figures 4 and 5, the SPD of AVDP is always lower than that of AMSP with the number of service requests increasing (i.e., Figure 4). In addition, we also compare the SPDs of the two approaches with the time period increasing under the network load of 10000 service requests (i.e., Figure 5). In more detail, when the number of service

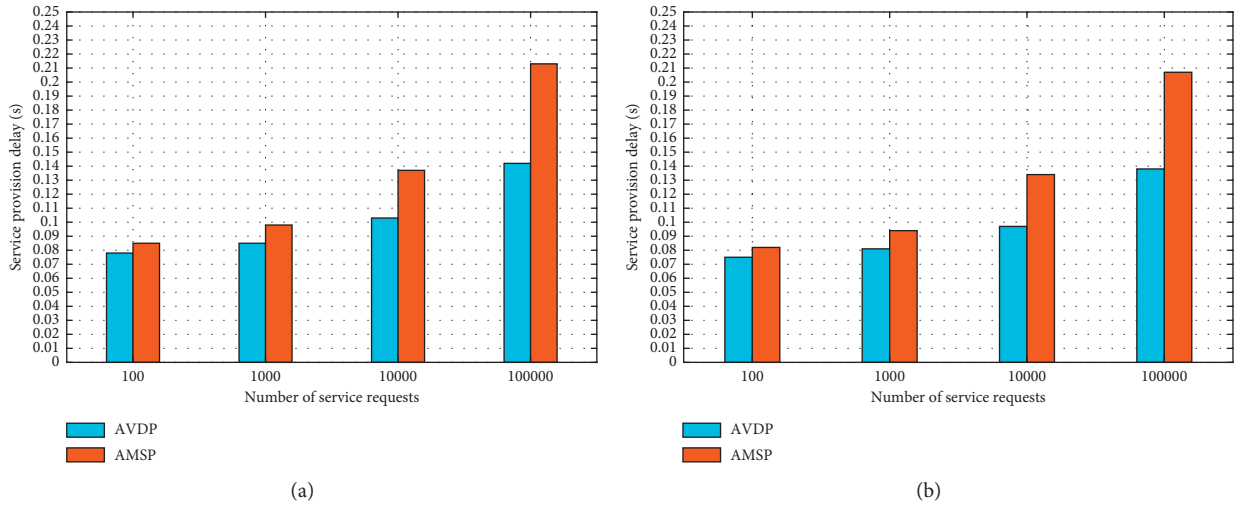


Figure 4: Service provision delay over different network loads. (a) Under Geant. (b) Under Interroute.

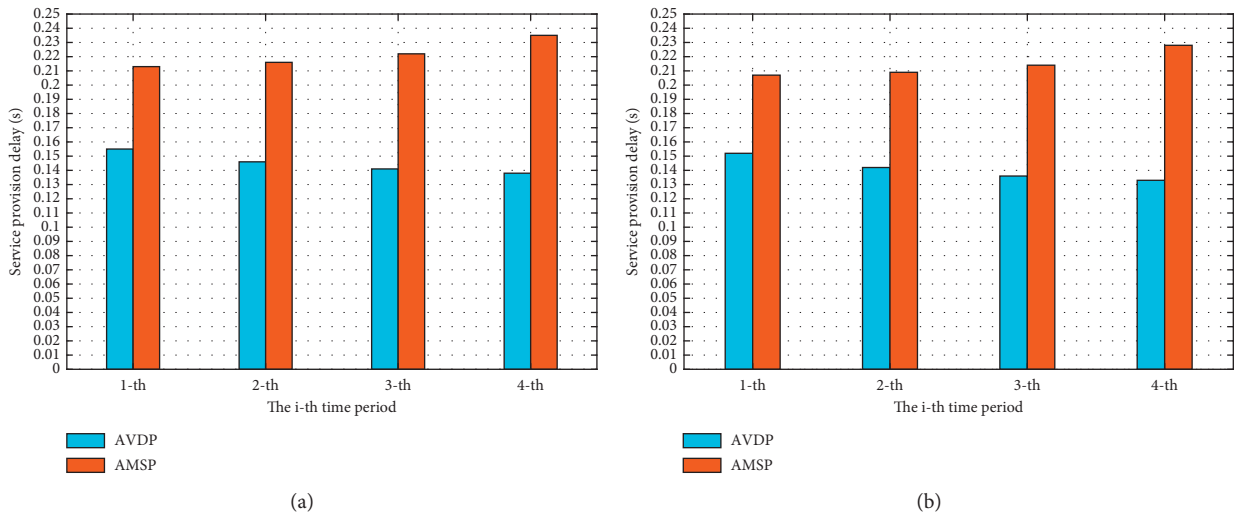


Figure 5: Service provision delay over different network loads. (a) Under Geant. (b) Under Interroute.

requests is low, the SPDs of AVDP and AMSP are 0.079 s and 0.085 s, respectively. When the number of service requests reaches the maximum, the SPD of AVDP just increases by 0.061 s, while the SPD of AMSP increases by 0.125 s. Moreover, when the network load is the heaviest, the SPD of AVDP reduces to 0.132 s with the time period increasing; that is, the SPD of AVDP is optimized with time, while that of AMSP increases to 0.235 s in the same situation. The reasons are as follows. In AVDP, the improved ACO method enables the VSF concentration on links to be updated in time by efficiently accelerating volatilization and enhancing feedback. Thus, the requested VSFs can be quickly found from the nearest ECSs, which significantly reduces the time overhead of searching for services. Furthermore, AVDP continuously optimizes the deployed locations of VSFs over time, which enables the VSF searching delay of AVDP to be reduced over time. Thus, the SDP of AVDP is further improved. However, AMSP barely changes over time, it mainly provides services on demands by instantly deploying service functions rather than leveraging already deployed functions. Real-timely deploying most of requested service functions leads to large delay for AMSP to provide new services, especially when the network load becomes heavy.

We compare the CCUs of the two approaches of AVDP and AMSP under Geant and Interroute. The CCU is defined as the ratio of the VSFs performed in an ECS to the total already deployed VSFs in this ECS. The results are shown in Figures 6 and 7.

As shown in Figures 6 and 7, the CCU of AVDP increases with the number of service requests increasing, while that of AMSP decreases when the network load becomes heavy (i.e., Figure 6). We also compare the CCUs of the two approaches with the time period increasing under the network load of 10000 service requests (i.e., Figure 7). In more detail, when the number of service requests is low, the CCUs of AVDP and AMSP are about 77% and 84%,

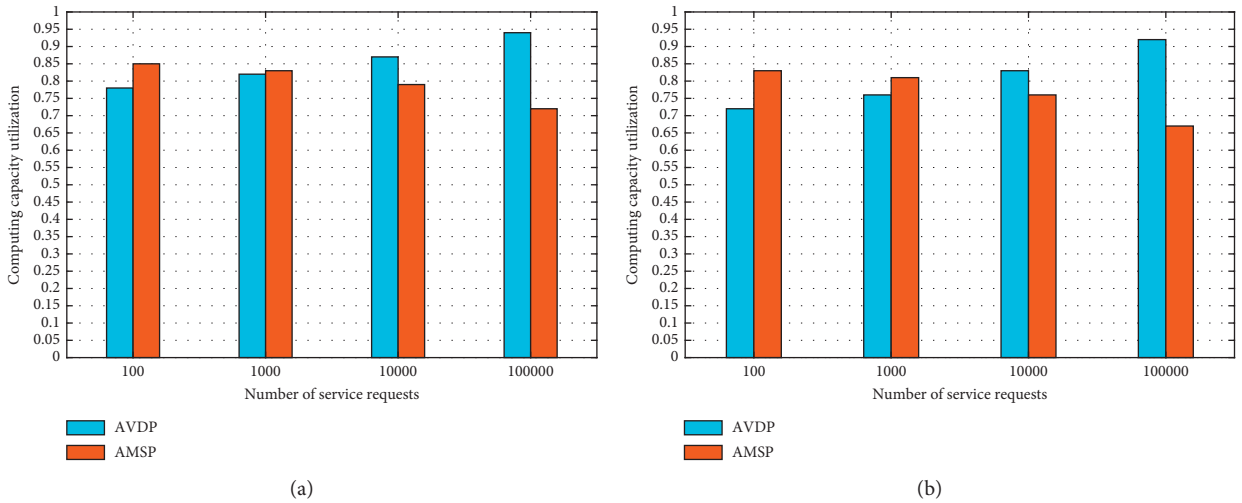


Figure 6: Computing capacity utilization over different network loads. (a) Under Geant. (b) Under Interroute.

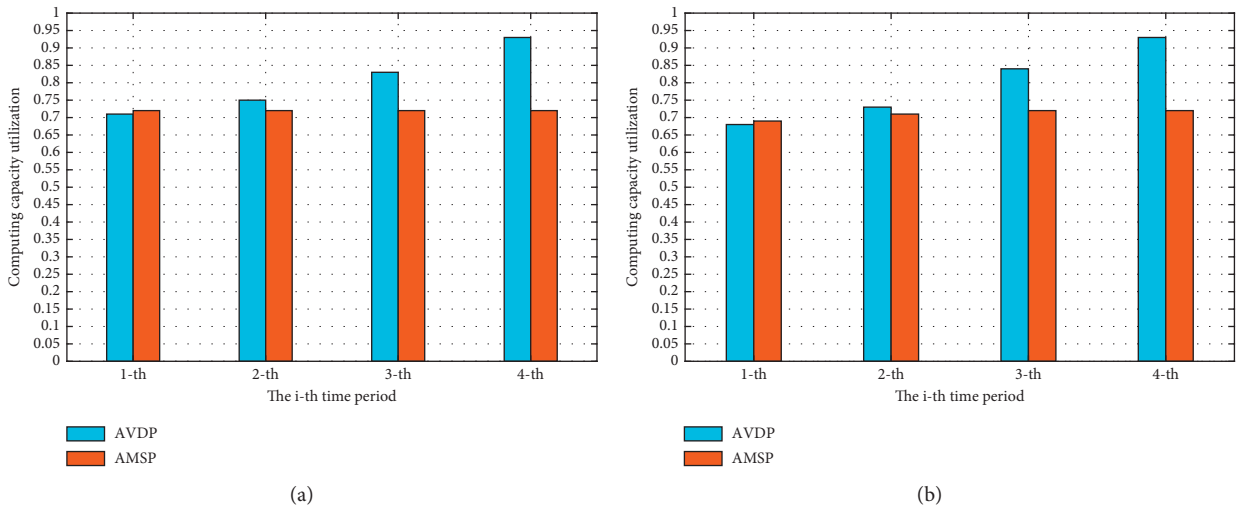


Figure 7: Computing capacity utilization over different time periods. (a) Under Geant. (b) Under Interroute.

respectively. With the network load becoming heavy, the CCU of AVDP increases and it reaches about 94% under the network load of 10000 service requests, while that of AVDP drops to 72% at the same network status. Moreover, when the network load is the heaviest, the CCU of AVDP increases from 71% to 93% with the time period increasing while that of AMSP is always keeping about at 72%. That is, the CCU of AVDP continuously optimizes while that of AMSP barely changes over time. The reasons are as follows. AVDP dynamically deploys VSFs by taking the frequencies of the VSFs being requested and performed in an ECS into account. With the number of service requests increasing, the VSFs rarely performed recently can be replaced by the ones with higher requested frequencies in an ECS; thus, the computing capacity of each ECS can be fully used. In addition, the requested VSFs can be efficiently found by AVDP, which enables the appropriate VSFs to be more frequently performed in each ECS. Thus, the CCU of AVDP is significantly optimized. However, AMSP mainly makes full use of some VSFs that are currently requested frequently in an ECS, and it does not consider replacing already deployed VSFs that are rarely requested by new popular VSFs, which cannot further optimize the ECS computing capacity utilization. In addition, AMSP does not have the ability to achieve the VSF future popularity prediction, and the CCU of AMSP cannot be improved over time.

We compare the SASRs of the two approaches of AVDP and AMSP under Geant and Interroute. The SASR is defined as the ratio of the requests that successfully obtain services to the total requests asking for services. The results are shown in Figures 8 and 9.

As shown in Figures 8 and 9, the SASR of AVDP is much higher than that of AMSP when the number of service requests increases rapidly (Figure 8). We also compare the SASRs of the two approaches with the time period increasing under the network load of 10000 service requests (Figure 9).

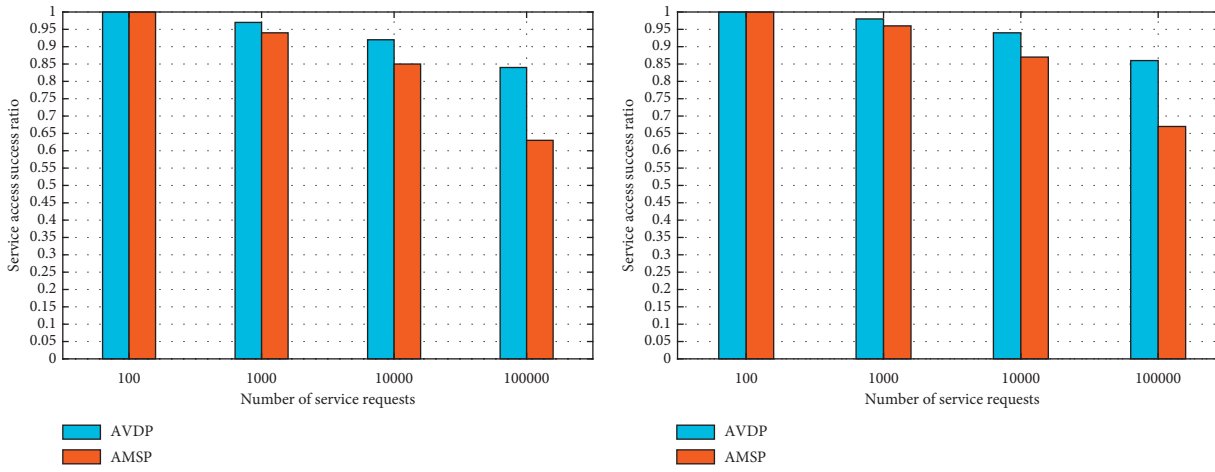


Figure 8: Service access success ratio over different network loads. (a) Under Geant. (b) Under Interroute.

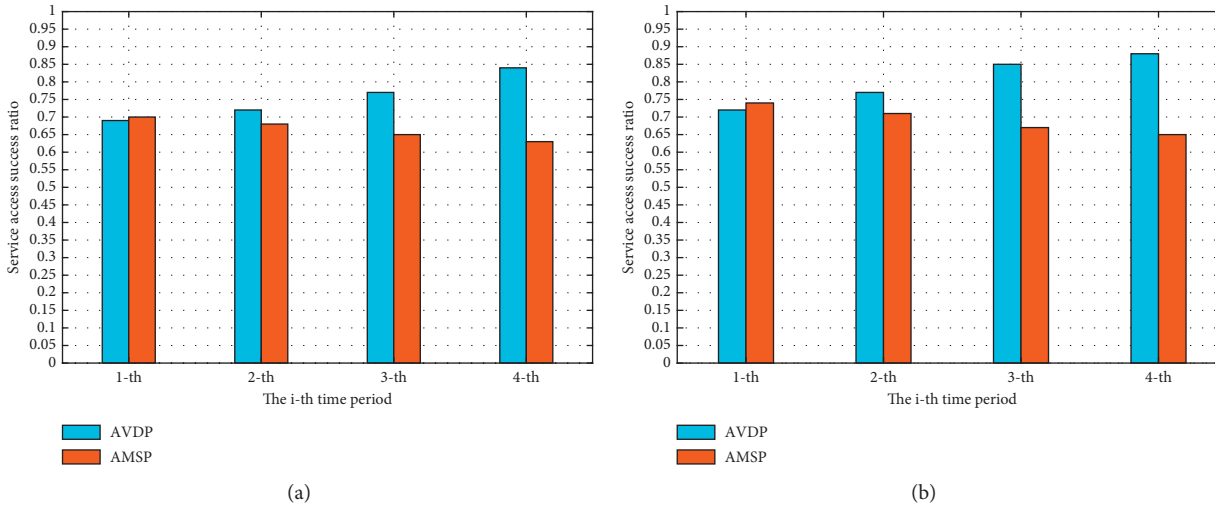


Figure 9: Service access success ratio over different time periods. (a) Under Geant. (b) Under Interroute.

In more detail, when the number of service requests is low, the SASRs of AVDP and AMSP are almost approaching 1. With the number of service requests increasing, the SASRs of AVDP and AMSP drop to about 84% and 63%, respectively. However, the SASR of AVDP still keeps beyond 84%, especially under the heaviest network load. Moreover, the SASR of AVDP increases from about 72% to about 84% with the time period increasing, while that of AMSP just improves a little and keeps at about 65% over time. Under the heaviest network load, the SASR of AVDP can achieve about 20% higher than that of AMSP. The reasons are as follows. AVDP mainly deploys most of VSFs before these VSFs are massively requested, which does not occupy much network resource to real-timely dispatch the requested VSFs. Thus, the available network resource can support as many new requests as possible, especially under the heavy network load. In addition, in AVDP, multiple ECSs that contain the requested VSFs can be found, which increases the probabilities that successfully providing corresponding services to new receiving requests when most of ECSs are busy. Thus, the SASR of AVDP is significantly optimized. However, AMSP mainly provides services by the server on the mobile path without considering balancing the working load by cooperating with multiple nearby servers. Under the heavy network load, the SASR of AMSP decreases rapidly.

6. Conclusions

In this paper, by introducing VSFs into edge computing pattern, we propose the mechanism of improved ACO-inspired VSFs detection and placement. By efficiently detecting the already deployed locations of VSFs and dynamically placing appropriate VSFs in suitable ECSs, the service provision delay to the edge equipment and the computing capacity utilization of each ECS under the edge computing scenario are significantly optimized. In this

mechanism, the approach of searching for the requested VSFs from multiple ECSs is devised, and it improves the ACO method to adapt to the edge computing scenario by defining the VSF detection ant and the VSF concentration on links. Furthermore, the approach of deploying VSFs in each ECS is devised, and it takes the frequency variations of the VSF being performed and being requested in an ECS into account, so as to select the most appropriate VSFs to place in the ECS with this ECS computing capacity utilization considered. Simulation results show that the proposed mechanism has significant improvements in the service provision delay optimization and the computing capacity utilization improvement compared with the current state of the art.

Data Availability

The data used in this study are available from the corresponding author upon request.

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 National Natural Science Foundation of China (nos. 62002261 and 62072336) and the Tianjin Municipal Education Commission Scientific Research Project (no. 2018KJ145).

References

[1] F. Alvarez, D. Breitgand, D. Griffin et al., "An edge-to-cloud virtualized multimedia service platform for 5G networks," *IEEE Transactions on Broadcasting*, vol. 65, no. 3, pp. 369–380, 2019.

[2] J. Gil Herrera and J. F. Botero, "Resource allocation in NFV: a comprehensive survey," *IEEE Transactions on Network and Service Management*, vol. 13, no. 3, pp. 518–532, 2016.

[3] K. Kaur, V. Mangat, and K. Kumar, "A comprehensive survey of service function chain provisioning approaches in SDN and NFV architecture," *Computer Science Review*, vol. 38, p. 100298, 2020.

[4] T. Taleb, K. Samdanis, B. Mada, H. Flinck, S. Dutta, and D. Sabella, "On multi-access edge computing: a survey of the emerging 5G network edge cloud architecture and orchestration," *IEEE Communications Surveys and Tutorials*, vol. 19, no. 3, pp. 1657–1681, 2017.

[5] J. Pan and J. McElhannon, "Future edge cloud and edge computing for internet of things applications," *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 439–449, 2018.

[6] M. Laroui, B. Nour, H. Moungla, M. A. Cherif, H. Afifi, and M. Guizani, "Edge and fog computing for IoT: a survey on current research activities & future directions," *Computer Communications*, vol. 180, pp. 210–231, 2021.

[7] Y. Liu, M. Peng, G. Shou, Y. Chen, and S. Chen, "Toward edge intelligence: multiaccess edge computing for 5G and internet of things," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 6722–6747, 2020.

[8] C. Li, Q. Cai, and Y. Lou, "Optimal data placement strategy considering capacity limitation and load balancing in geographically distributed cloud," *Future Generation Computer Systems*, vol. 127, pp. 142–159, 2022.

[9] A. Laghrissi and T. Taleb, "A survey on the placement of virtual resources and virtual network functions," *IEEE Communications Surveys & Tutorials*, vol. 21, pp. 1409–1434, 2019.

[10] X. Ma, S. Wang, S. Zhang, P. Yang, C. Lin, and X. Shen, "Cost-efficient resource provisioning for dynamic requests in cloud assisted mobile edge computing," *IEEE Transactions on Cloud Computing*, vol. 9, no. 3, pp. 968–980, 2021.

[11] C. Bu, X. Wang, M. Huang, and K. Li, "SDNFV-based dynamic network function deployment: model and mechanism," *IEEE Communications Letters*, vol. 22, no. 1, pp. 93–96, 2018.

[12] C. Bu and J. Wang, "Computing tasks assignment optimization among edge computing servers via SDN," *Peer-to-Peer Networking and Applications*, vol. 14, pp. 1190–1206, 2021.

[13] P. Roy, S. Sarker, M. A. Razzaque et al., "AI-enabled mobile multimedia service instance placement scheme in mobile edge computing," *Computer Networks*, vol. 182, no. 9, pp. 1–14, 2020.

[14] L. Yin, P. Li, and J. Luo, "Smart contract service migration mechanism based on container in edge computing," *Journal of Parallel and Distributed Computing*, vol. 152, pp. 157–166, 2021.

[15] L. Rui, M. Zhang, Z. Gao, X. Qiu, Z. Wang, and Ao Xiong, "Service migration in multi-access edge computing: a joint state adaptation and reinforcement learning mechanism," *Journal of Network and Computer Applications*, vol. 183-184, pp. 1–17, 2021.

[16] L. Tang, X. He, P. Zhao, G. Zhao, Yu Zhou, and Q. Chen, "Virtual network function migration based on dynamic resource requirements prediction," *IEEE Access*, vol. 7, pp. 112348–112362, 2019.

[17] The Univesity of Adelaide, "The internet topology Zoo," 2012, https://www.topology-zoo.org/.

[18] J. Martins, M. Ahmed, C. Raiciu et al., "ClickOS and the art of network function virtualization," in *Proceedings of the 11th USENIX Conference on Networked Systems Design and Implementation*, pp. 459–473, Seattle, USA, April 2014.

[19] S. Yang, F. Li, M. Shen, X. Chen, X. Fu, and Y. Wang, "Cloudlet placement and task allocation in mobile edge computing," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 5853–5863, 2019.

[20] D. Li, P. Hong, K. Xue, and J. Pei, "Virtual network function placement considering resource optimization and SFC requests in cloud datacenter," *IEEE Transactions on Parallel and Distributed Systems*, vol. 29, no. 7, pp. 1664–1677, 2018.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 3064804, 10 pages
https://doi.org/10.1155/2022/3064804

Research Article

3D Deep Heterogeneous Manifold Network for Behavior Recognition

Jinghong Chen,[1,2] **Li Zhang**,[1,2] **Zhihao Jin**,[1,2] **Chong Zhao**,[1,2] **and Qicong Wang**[1,2]

[1]*Department of Computer Science and Technology, Xiamen University, Xiamen 361005, China*
[2]*Shenzhen Research Institute, Xiamen University, Shenzhen 518057, China*

Correspondence should be addressed to Chong Zhao; zhc@xmu.edu.cn and Qicong Wang; qcwang@xmu.edu.cn

Received 1 February 2022; Accepted 26 February 2022; Published 16 March 2022

Academic Editor: Lu Liu

With the broadening of application scenarios for Internet of Things, intelligent behavior recognition task has attracted more and more attention. Since human behavior is nonrigid motion with strong spatiotemporal topological association, modeling it directly with traditional Euclidean space-based methods may destroy its underlying nonlinearity. Based on the advantages of Riemannian manifold in describing 3D motion, we propose an end-to-end 3D behavior manifold feature learning framework composed of deep heterogeneous networks. This heterogeneous architecture aims to leverage the graph construction to guide manifold backbone network to mine more discriminative nonlinear spatiotemporal features. Therefore, we first model the nonlinear spatiotemporal co-occurrence of 3D behavior in the high-dimensional Riemannian manifold space. Secondly, we implement a non-Euclidean heterogeneous architecture on the Riemannian manifold so that the backbone network can learn deep spatiotemporal features while preserving the manifold topology. Finally, an end-to-end deep graph similarity-guided learning optimization mechanism is introduced to enable the overall model to fully utilize the complex similarity relationship between manifold features. We have verified our 3D deep heterogeneous manifold network on popular skeleton behavior datasets and achieved competitive results.

1. Introduction

Behavior recognition tasks [1–3] receive much attention due to the vigorous development of artificial intelligence and the rise of computer vision. In smart security, human-computer interaction, and immersive games, behavior recognition is playing an increasingly important role. We can perform dangerous behavior warnings, provide more convenient behavior instructions for human-computer interaction, and make immersive games have a rich and exquisite game experience through behavior recognition. With the great improvement of computer and devices for capturing the movement of human skeleton, the acquisition of skeleton sequence data is more convenient, which promotes the development of skeleton-based behavior recognition [4, 5]. The skeleton-based behavior recognition method has the advantages of eliminating the influence of the background and the invariance of the perspective, which brings the ability to pay more attention to the behavior itself. For these reasons, more and more researchers are involved in skeleton-based action recognition research.

There are three main methods of existing behavior recognition: methods based on spatial features of skeleton coordinates, methods based on temporal information of skeleton sequence, and methods based on spatiotemporal features. In the method based on spatial features of skeleton coordinates, the covariance matrix of the joint position trajectory is calculated to build the temporal model of skeleton sequence [2]. In [3], the paired relative positions of joints are also used to describe the posture and joint changes of the skeleton sequence, and the principal component analysis is applied to normalize features to obtain the representation of the principal features. In [4], the rotation and translation between body parts are used as features, and the Fourier temporal pyramid (FTP) is utilized to model the temporal dynamics. These methods pay more attention to

the spatial relationship of the joints in the skeleton behavior, which weakens the attention to the temporal features to a certain extent.

For the temporal information, Wang et al. [1] calculate relative positions of each joint and other joints to represent each frame of the skeleton sequence and then model temporal information. In [6], the histogram of the 3D joint position is calculated to represent each frame of the skeleton sequence, and HMMs are used to model the temporal dynamics. Kim and Reiter [7] propose to use temporal convolutional neural network (TCN) for 3D human behavior recognition. Compared with the popular LSTM-based recurrent neural network model, the TCN-based model is more intuitive and interpretable [7]. These methods can take the spatiotemporal features of behavior into account, but may ignore some spatial features that are globally related and cannot closely link temporal and spatial features.

In the method based on spatiotemporal features, Yan et al. [8] design skeleton sequence graph containing temporal information and use the spatiotemporal graph convolution network to learn the spatiotemporal features in the behavior sequences. Ke et al. [9] use a deep convolutional neural network to obtain the temporal features of the skeleton sequence, use a multitask learning network to process all the frames of the generated fragments, and finally combine the spatial information for behavior recognition. Some scholars use graph convolutional network (GCN) combined with LSTM or dual-stream network structure [5, 10–12] to extract spatiotemporal information in behavior sequences. These methods can pay attention to the close relationship between temporal and spatial features, but since behavior features also have the temporal and spatial co-occurrence, these methods cannot accurately describe this property.

To learn more discriminative spatiotemporal manifold features by the deep model, we need to comprehensively consider the spatiotemporal co-occurrence relationships between the connected and disconnected skeleton parts. To this end, we intend to represent the spatial structure based on the transformation group for each frame of original nonrigid 3D skeleton behavior sequences and use the Riemannian manifold to construct the relative spatial transformation relationships between all pairs of skeleton parts. This spatial structure representation method can describe the relative motion relationship between all pairs of skeleton parts in a frame as a point in the high-dimensional Riemannian manifold space.

Since each action sequence consists of many frames, we employ an interpolation method based on the transformation group to integrate the points in the manifold surface space into a transformation group curve, so as to model the co-occurrence relationship of the spatiotemporal features of original 3D skeleton sequence. However, directly inputting features with manifold constraints into neural network will bring high time and space complexity. Currently, it is difficult to use the neural network to mine rich information contained in manifold input while preserving the manifold constraints. To this end, Wang et al. [13] propose a GCN-based method to solve the problem of edge prediction between nodes. Inspired by this method, we try to treat an action as node, construct similarity graph of all nodes based on its manifold trajectory, use graph convolution to predict connections, and finally achieve the classification of behaviors. With respect to this idea, the difficulty to be solved is how to construct graph of feature nodes in manifold space.

The graph construction method is currently commonly used in determining the similarity of members in social network analysis [14, 15], and the constructed graph is used for intelligent recommendation. In these applications, the multidimensional features of the task are usually data in Euclidean space, and existing methods such as KNN [16] can solve this problem. However, in the application scenario of our problem, we hope to realize the construction of behavior feature nodes on manifold space. Therefore, in this study, a graph construction method based on the Riemannian metric on manifold is proposed. This method can take full advantage of rich information of data on manifold. At the same time, the Riemannian metric method can map behavior nodes isometrically into projected space.

This study proposes a 3D behavior recognition method based on spatiotemporal trajectory graph construction, whose description of framework is shown in Figure 1. This method uses Riemannian metric to measure the spatiotemporal trajectory properties, which make similar nodes closer and dissimilar or different types of nodes far apart. The model mainly has the following stages, data preprocessing, Riemannian metric graph construction, graph convolution, and behavior classification. In the data preprocessing stage, we process the 3D coordinate data of the skeleton sequences into a behavior trajectory curve representing relative behavior relationship between any pair of bones. In order to express as much spatial information as possible to reflect rich spatiotemporal co-occurrence, we calculate the relative behavior relationship between any two bones. In the stage of Riemannian metric graph construction, we roll and expand the processed manifold spatiotemporal trajectory curve along the direction of the trajectory into a corresponding continuous rolling tangent space curve. This process tries to ensure that the distance between any two points in a tangent space curve is equivalent to the distance between two points in the original manifold, use DTW to measure the similarity between curves, and use the similarity between behavior nodes to construct a similarity graph. In the graph convolution stage, through the update between each iteration of graph convolution, similar nodes are pulled closer and different are pushes apart so that behavior nodes of the same category are gathered together. Finally, in the classification stage, the labels are spread from the central point of each cluster to achieve the classification of behaviors. The main contributions of this study are as follows:

(1) For skeleton sequences, we extract rotation and translation relationships from bone pairs and represent them as discrete trajectories in Riemannian manifold, which can describe spatiotemporal co-occurrence and global relative relationships.

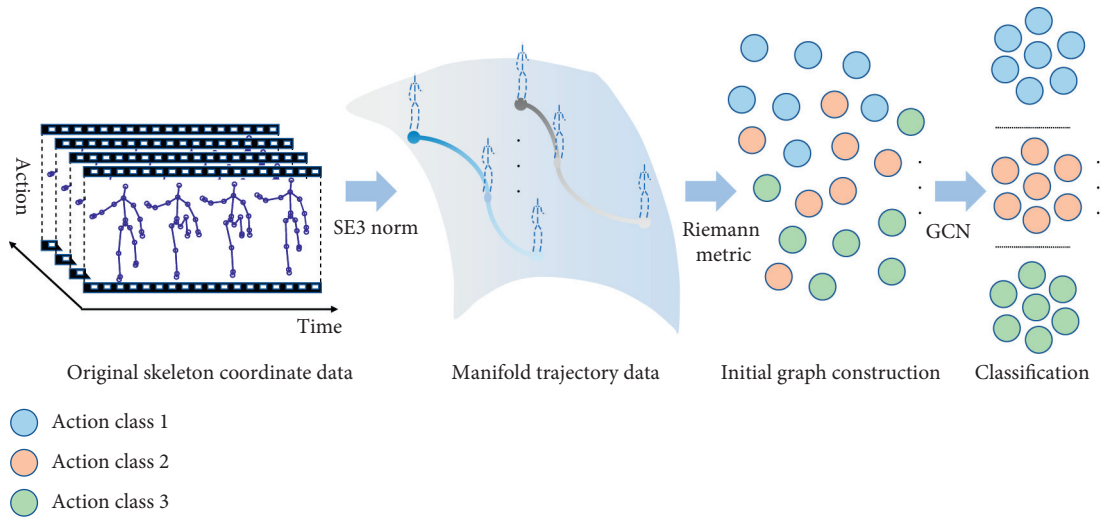


Figure 1: Framework of 3D behavior recognition network based on spatiotemporal trajectory graph construction.

(2) We propose a graph construction method based on continuous projections on Riemannian manifold, which is employed to map the spatiotemporal trajectories on the manifold isometrically to preserve more complex similarity distribution relationship between manifold features.

(3) We propose a deep heterogeneous manifold model consisting of two subnetworks with different structures. It incorporates an end-to-end optimizable manifold backbone network, which exploits the powerful representative ability of Riemannian manifold and can be guided by the subsequent graph-based subnetwork.

2. Spatiotemporal Manifold Trajectory Representation

To fully exploit the nonlinearity of behavior data, we represent them as curves in manifold space. Specifically, we represent it in the Lie manifold SE_3 in the form of Cartesian product, which can contain rich spatiotemporal co-occurrent relationships.

Given 3D coordinates of the joints of the skeleton behavior sequence, we assume that the number of frames of an behavior sequence is F, and the number of joints is N_j, so the coordinate of the nth joint in the frame f is expressed as $X_n^f = (x_n^f, y_n^f, z_n^f)$, and the 3D coordinate of a behavior sequence is represented as $\{X_n^f | n = 1, \ldots, N_j; f = 1, \ldots, F\}$. With these 3D coordinates and the body structure data given in the dataset, i.e., the above joint points are connected in the body structure, here we might as well assume that the joint i and the joint j are the two ends of the bone B_{ij} in the first frame, and this bone can be represented as $B_{ij} = X_i^1 - X_j^1 = (x_i^1 - x_j^1, y_i^1 - y_j^1, z_i^1 - z_j^1)$; in this way, a bone can intuitively be represented as a vector in 3D Euclidean space, and the set of bones $\{B_{ij}^f | 1 < i < j < N_j; f = 1, \ldots, F\}$ can also be obtained. Since the spatiotemporal graph of the body structure in the current skeleton data are all acyclic graphs, the number of bones is $N_j - 1$. In the skeleton of body, the relationship between any two different bones is $(N_j - 1) * (N_j - 2)$ pairs.

The elements in the trajectory manifold have the following constraints:

$$SE_3 = \left\{ T = \begin{bmatrix} R & d \\ 0^T & 1 \end{bmatrix} \in \Re^{4\times 4} | R \in SO_3, d \in \Re^3 \right\}, \quad (1)$$

where SE_3 is special Euclidean group and SO_3 is special orthogonal group.

The manifold trajectory using relative relationships has the following advantages:

(1) The features used to represent the rotation relationship between skeletons are scale invariant; in other words, no matter how large the scale is to represent the skeleton, the rotation relationship between the skeletons is unchanged

(2) The relative relationship of SE_3 has spatial co-occurrence, i.e., we can explore the relationship between not only any two bones but also spatially connected skeleton pairs

(3) Representing the relative relationship of the skeleton based on the trajectory curve can closely combine the spatial information and the temporal information, so different spatial features can be represented point by point to form a discrete curve on manifold space, which helps to increase the similarity of features with the similar temporal information

3. Backbone Network of Deep Heterogeneous Manifold Network

3.1. Riemannian Manifold Preservation Network. Since the input data of our deep Riemannian manifold network is the initialized high-dimensional Riemannian manifold transformation group data, it is necessary to maintain the richness and topology of their nonlinear structures during the feature learning process. The commonly used Euclidean

spatial convolution layer may destroy this property, so we employ a convolution-like Riemannian transform layer that contains transform parameters optimized for deep model learning and whose output still conforms to the Riemannian manifold constraints, which preserve the Riemannian manifold topology of the data.

According to the above description, we know that the feature is a set of points in the motion group SE_3, which is represented by the discrete curves' form on the manifold of the Lie group [17, 18]. We denoted this manifold as $\mathscr{M}$, and the set of points is $\mathbb{S}$; then, the feature of the fth frame in the kth behavior is represented as $\mathbb{S}_f^k$. Since any point on the manifold $\mathscr{M}$ has constraints: if we have any $U \in \mathscr{M}$, then $U \cdot U^T = I$ and $\det(U) = 1$, where I is the identity matrix, which is also the identity element on the manifold, and det is the operation to find the value of the determinant. So, there is

$$\mathrm{SE}_n = \left\{R \in \mathbb{R}^{n\times n} | R^\top R = I_n, \det R = 1\right\}. \tag{2}$$

If we have $V \in \mathscr{M}$, then $V \cdot U \in \mathscr{M}$.

This property can be summarized as

$$\mathrm{SE}(3) \times \mathrm{SE}(3) \longrightarrow \mathrm{SE}(3). \tag{3}$$

The SE_3 matrix has the invertible property $R^{-1} = R^\top$. Therefore, the behavior trajectory curve l is in the form of $\mathrm{SE}(3) \times \mathrm{SE}(3) \times \ldots \times \mathrm{SE}(3)$.

The initialized high-dimensional Riemannian manifold transformation group data are also a spatiotemporal co-occurrence representation of the original 3D data, thus requiring spatial and temporal pooling techniques on the Riemannian manifold. We can not only reduce the data dimension and preserve topology but also further obtain more discriminative spatiotemporal manifold features between action sequence frames.

3.2. Graph Construction Based on Manifold Trajectory. On the obtained manifold trajectory curves, we use the Riemannian similarity metric method to construct graph for the behavior features on Riemannian manifold. The distance on a manifold is obtained by measuring geodesics on the manifold. To ensure that the distance between any two points on the manifold remain constant in the constructed graph, we can map the points on the manifold isometrically to a convenient measurement space. The implementation process of the graph construction method based on Riemannian similarity metric is shown in Algorithm 1.

The dimension of the SE_3 matrix is 6, which brings high computational and space complexity to operations such as multiplication and inverse. Therefore, in this study, we do not use the method of directly calculating the distance between two points on the SE_3 manifold. We explore the use of a certain method that can isometrically map the points on the manifold to a space that is convenient for measurement. If we directly expand the projection at a point, for example, we expand at the pole, the result may be that the closer to the pole, the more similar the curve after projection is to the original curve on the manifold, and the farther away from the pole, the more distorted the curve is after projection. Inspired by methods of geodetic distance [19], we propose a method for measuring the distance of a curves on manifold based on a continuous projection.

Figure 2 shows a continuous projection of a behavior trajectory curve on the manifold along the quasi-average curve to its corresponding tangent space. In the curve l_{ABC} on the manifold, we use the continuous projection method along the average curve of the class (i.e., the dotted line in the figure) to project the points on the curve one by one into the tangent space. The lengths of line segments l_{AB}, l_{BC}, and l_{AC} on the manifold are, respectively, equal to the lengths of l_{ab}, l_{bc}, and l_{ac} of the corresponding tangent space.

Below, we explain this continuous projection process in detail. Specifically, the continuous projection mapping on the manifold is a smooth mapping h: along a smooth average curve α: $[0, T] \longrightarrow \mathscr{M}$:

$$\begin{aligned} h\colon [0,T] &\longrightarrow SE_3 = SO_3 \propto \mathbb{R}^3, \\ t &\mapsto h(t) = (R(t), s(t)). \end{aligned} \tag{4}$$

In particular, this rolling continuous mapping needs to meet the three conditions defined in [20] at any time $t \in [0, T]$, namely, rolling conditions, no-slip conditions, and no-twist conditions. The continuous projection $h(t)$ is a continuous map that satisfies the above three conditions and maps the manifold trajectory to the corresponding tangent space.

Since the area near the point on the Lie group manifold is smooth, any point in this area can be represented by a slight rotation and translation change from a point to its neighbors. Assuming that P is a point on the manifold space of SE_3, α: $[0, \tau] \longrightarrow \mathrm{SE}_3, \alpha(t) = U(t)P_0W(t)^T$ is a curve on SE_3 starting from P_0 when $t = 0$, and at any subsequent time, you can find a point on the curve corresponding to that time. We can find such a smooth curve; then, this meets the continuous projection condition. Since our calculation cannot exhaust every point on the continuous curve, in order to facilitate the calculation, in the following calculation, we will continue to project the points on the curve frame by frame. Under the three constraints of manifold described above, this mapping process can be expressed as

$$\begin{aligned} h\colon [0,\tau] &\longrightarrow G = \mathrm{SE}_3 \times \mathrm{SE}_3 \propto \mathbb{R}^{4\times 4}, \\ t &\mapsto h(t) = \left(U^\top(t), W^\top(t), X(t)\right), \end{aligned} \tag{5}$$

where $\langle i$mes denotes semidirect product symbol and $(U^\top(t), W^\top(t), X(t))$ is the solution of the motion equation in the projection process at time t.

This process is a continuous projection V along the curve $\alpha(t)$ on the Lie group manifold V: $= T_{P_0}^{\mathrm{Aff}}\mathrm{SE}_3 \cong T_{P_0}\mathrm{SE}_3$; the curve $\alpha(t)$ has the following expression:

$$\alpha(t) = U(t)P_0W(t)^\top. \tag{6}$$

$\alpha_{\mathrm{dev}}(t)$ is the expansion of the curve $\alpha(t)$ under the effect of continuous projection $h(t)$ at P_0:

$$\alpha_{\mathrm{dev}}(t) = h(t)^{\circ}\alpha(\mathrm{t}) = \mathrm{U}^\top(\mathrm{t})\alpha(\mathrm{t})\mathrm{W}(\mathrm{t}) + \mathrm{X}(\mathrm{t}) = \mathrm{P}_0 + \mathrm{X}(\mathrm{t}). \tag{7}$$

Suppose we perform continuous projection in the time interval $[0, T]$ on a certain behavior curve. Since the curve on

Input: trajectory curves of all skeletons $\mathbb{S}$; behavior sequence label in training set L; total number of behavior categories M;
(1) **for** Given behavior category $L_i \in [L_1, L_M]$ **do**
(2) Calculate the average trajectory curve of each class on the manifold;
(3) Average trajectory curve $L_i^a vr = DT\ W$ (All train behavior curves $\in L_i$);
(4) **end for**
(5) **for all** Training trajectory curve $\mathbb{S}$ with label L_i **do**
(6) Continuously project training trajectory curve $\mathbb{S}$ along the average trajectory curve $L_i^a vr$,
(7) Obtain the curve features on the tangent space S_{train} after continuous projection;
(8) **end for**;
(9) **for all** Training trajectory curve $\mathbb{S}$ **do**
(10) Given test set trajectory curve $\mathbb{S}$
(11) **for** $i = 1$; $i < M$; $i++$ **do**
(12) Continuously project test set trajectory curve $\mathbb{S}$ along the average trajectory curve $L_i^a vr$;
(13) **end for**
(14) Continuously unfold test set trajectory curve $\mathbb{S}$ along the path of M average curves, obtain a set of curves $\{S_1, S_2 \ldots S_M\}$
(15) Calculate the set of similarity scores between each curve in the curve set and the corresponding average curve Score;
(16) Obtain the features S_{test} under the score reflecting to the highest similarity;
(17) **end for**;
(18) **for all** Training trajectory curve S **do**
(19) Given a curve S_{train} feature, use DTW to calculate the most similar K trajectory curve to this curve;
(20) Get adjacency list T_{train};
(21) **end for**;
(22) **for all** Test track curves S **do**
(23) Given a curve S_{test} feature, use DTW to calculate the most similar K trajectory curve to this curve;
(24) Get adjacency list T_{test};
(25) **end for**;
Output: Curve features of training set S_{train} and test set S_{test} after continuous projection; The adjacency list obtained of the training set T_{train} and test set T_{test};

Algorithm 1: Graph construction method based on Riemannian similarity metric.

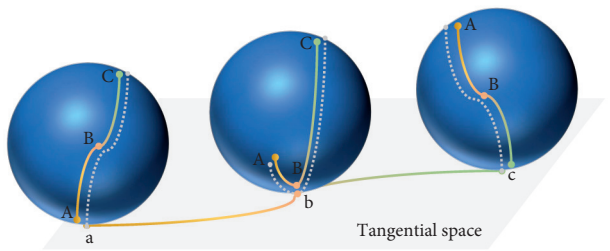


Figure 2: The behavior trajectory curve on the manifold is continuously projected along the curve to its corresponding tangent space.

the manifold we use is discrete on the time axis, we get the corresponding points in the mapping space. It is $\alpha_{\text{dev}}(t), t \in \{0, 1, 2, \ldots, T-1, T\}$.

Using the continuous projection method, the process of obtaining the similarity between the behavior curves from the manifold space is shown in Figure 3. We take the three points A, B, and C of a certain behavior curve on the manifold as an example. After continuous projection, they correspond to the three points a, b, and c in the tangent space. Our method aims to make the distances between AB, BC, and AC on the manifold are basically similar to the mapped distances ab, bc, and ac, especially to ensure that the distances between nodes of the same category are as similar as possible.

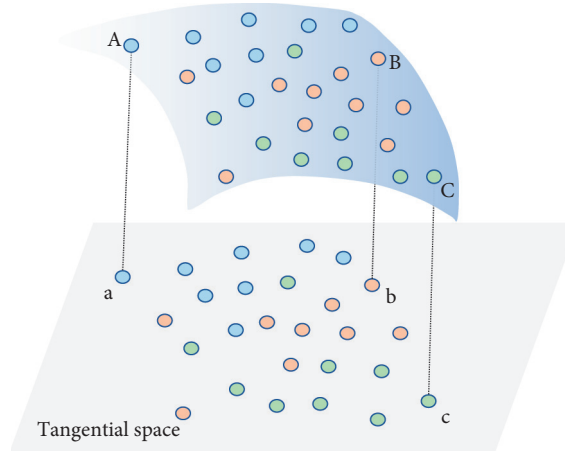


Figure 3: The distribution of nodes in manifold and in projection space by the continuous projection method.

The projection method based on the tangent space of a certain point has a problem, that is, the closer the data to the projection point, the better the retention of features and local similarities between the data. On the contrary, the farther away from the projection point is, the relative distance of the data is pulled away after being projected, which causes the local similarity of the data far from the projection point and the global similarity of the whole data to be destroyed. We keep the local similarity and global similarity between nodes

as much as possible in the projection process, avoiding the distortion of the distance between nodes that affects the subsequent node classification.

Generally, the behavior curves of a certain type on the manifold does not completely coincide with the geodesic. In particular, when this continuous projection curve satisfies certain constraints, the continuous projection curve we get degenerates into a geodesic curve. In a part of the projection of a certain point, the curve on the manifold and the two curves in the corresponding tangent space have the same geodesic curvature. That is to say, the geodesic curve is a projection curve that meets certain constraints, so the applicable range is narrow. Our continuous projection method can be applied to more manifold projection scenes; expanding average curve of a class along the behavior curve can better measure the similarity between different classes.

4. End-to-End Optimizable Graph-Guided Heterogeneous Model

In the previous 3D action recognition methods based on deep learning, most methods usually use a fully connected layer at the end of the backbone network and use cross-entropy loss to complete the task. In the iterative learning process, they do not fully consider the similarities and changes between deep features of similar actions as well as the differences between deep features of different action categories. Since the output of our backbone network is still topologically preserved Riemannian manifold data, we need a construction method of nearest neighbor graph on a high-dimensional Riemannian manifold surface to model local similarities, combined with graph convolutional network to achieve deep global similarity prediction to guide the feature learning of backbone network. This can make full use of the potential local similarity relationship in the local context information of each action sequence so that the whole heterogeneous network can integrate the common features of the same category and suppress their changes and at the same time expand the differences of different categories through the aggregating capability of graph convolution.

Our deep heterogeneous manifold network consists of two subnetworks with different structures. The former is the backbone network for learning deep manifold spatiotemporal features, and the latter is the graph convolution-guided learning subnetwork, which is built on the previous trajectory curves. In the backbone network, two pooling learning submodules are added to learn more discriminative features for further promoting of the graph convolutional network. In an end-to-end manner, the latter subnetwork can guide the feature learning of the former backbone subnetwork. However, its backpropagation will be more complicated, and the whole heterogeneous model is built on the Riemannian manifold, making the optimization problem with manifold constraints. If the manifold is embedded in linear space, the dimension problem will increase, thereby increasing the complexity. It is very difficult to optimize in Euclidean space. However, in some specific Riemannian manifold, the constraints can be eliminated to become unconstrained optimization, so we consider to solve an end-to-end optimization problem directly on the Riemannian manifold.

In the first module of the trajectory curve feature learning part, we set the learning parameter R_S in a Lie group manifold and then perform a spatial pooling on the data that has undergone manifold learning so that we can select more discriminative spatial features learned by the previous layer, and it reduces the computational complexity of spatial features and facilitates the subsequent computation. Similarly, the second module also sets a learning parameter R_T in the Lie group manifold and then performs a temporal pooling on the data. In this way, on the one hand, it is possible to select more discriminative temporal features after learning from the previous layer, and on the other hand, it reduces the computational complexity of temporal features.

Given $R_S \in \mathrm{SE}_3$ and $R_T \in \mathrm{SE}_3$, we suppose that the data passed in each time are $D \in \mathrm{SE}_3$. Due to the retention of Lie group operations, there is

$$\begin{aligned} D \cdot R_S &\in \mathrm{SE}_3, \\ D \cdot R_T &\in \mathrm{SE}_3. \end{aligned} \tag{8}$$

Therefore, in this part, the network parameters' learning is constrained in the Lie group manifold. In the graph-guided convolution module, we loop all behavior nodes, put all nodes into a queue, construct a domain subgraph with each node as the central point, and predict the connection relationship between the included peripheral nodes and the central point. As a result, a set of edges whose weights are the connection probability can be obtained. In order to cluster similar nodes together, a simple method is to prune all edges whose weights are lower than a certain threshold and use breadth-first search method to propagate pseudolabels. In each iteration, the edge is updated below a certain threshold, and in the next iteration, the connected clusters are greater than the predefined maximum value. In the next iteration, the threshold for updating the edge is increased. Repeat this loop process until the queue is empty. At this time, all nodes have been marked with pseudolabels of the category. We take the label of the central node of each cluster to propagate, i.e., the classification of nodes is realized.

5. Experimental Verification

5.1. Dataset Description

5.1.1. G3D Dataset. This dataset is a skeleton-based dataset [21] collected from game data. It contains 10 participants, who perform 20 categories of game behaviors. Most behavior sequences are recorded by a specific camera in a controlled indoor environment. Participants perform basic behaviors in strict accordance with instructions, and each sequence was repeated 3 times by each subject. Nevertheless, participants are free to complete the collection of different exercise sequences according to their own exercise habits. The dataset contains manually labeled behavior category labels for all sequences.

The skeleton in this dataset consists of 20 joints, and the position of the participant's joints is expressed in X, Y, and Z

coordinates in meters. The skeleton data also includes a joint tracking state, including accurately tracked joints, imported joint coordinates, and predicted joint coordinates. In many cases, the predicted joints are accurate, but in some cases, the limbs are occluded and the predicted joints may be inaccurate. Since some joint points in the dataset are obtained through prediction, the accuracy of the final classification will be affected to a certain extent if the predicted joints are inaccurate.

5.1.2. HDM05 Dataset. The behavior sequences in this dataset are performed by 5 nonprofessional actors [22]. Most of the behavior sequences are performed multiple times by all five actors according to the specific instructions in the script. The script contains five parts, and each part is divided into several scenes. Each behavior sequence is only collected in the corresponding single scene. The skeleton in this dataset consists of 31 joints, and the 3D coordinates of the joints are represented in X, Y, and Z coordinates in centimeters.

Although the dataset is small in scale, the behavior categories are more detailed, with a total of 130 behavior categories, some of which may look similar. Therefore, this dataset is also somewhat challenging.

5.1.3. NTU-RGBD Dataset. The NTU-RGBD dataset contains 60 behavior classes and 56880 video samples [23]. This dataset contains RGB video, depth mapping sequence, 3D bone data, and infrared (IR) video for each sample. Each data is captured simultaneously by 3 Kinect V2 cameras. Here, we use three-dimensional skeleton data, and the three-dimensional coordinates of the joints are expressed in X, Y, and Z coordinates. The three-dimensional skeleton data contain the three-dimensional coordinates of 25 human body joints per frame. The original benchmark provides two evaluation methods, namely, cross-subject (CS) and cross-view (CV) evaluation. In CS evaluation, the training set contains 40,320 videos from 20 subjects, and the remaining 16,560 videos are used for testing. In CV evaluation, 37920 videos captured from No. 2 and No. 3 cameras were used for training, and the remaining 18,960 videos from No. 1 camera were used for testing.

This dataset is widely used in skeleton-based behavior recognition. It has several scene categories, including daily behaviors, medical scenes, and multiperson sports. Since it contains both single-person sequences and multiperson interaction sequences, it is quite challenging to perform recognition tasks on this dataset.

Table 1 summarizes the main data distribution characteristics of the above three datasets. It can be seen that the number of joints and the number of bones selected in the three datasets are roughly similar, and the number of frames in each behavior sequence varies widely, ranging from a few frames to a few hundred frames, i.e., it is linearly adjustable within certain limits. From this perspective, it is very important to fully dig out the temporal information to complete the task of behavior recognition. Judging from the number of behavior sequences contained, the scales of the three datasets from small to large are G3D-Gaming, HDM05, and NTU-RGBD; from the perspective of the divided behavior categories, HDM05 has the most behavior categories, indicating the classification of behavior sequences is finer, and the corresponding recognition difficulty is also greater. In addition, in order to further improve the generalization ability of recognition in the future, we have implemented a behavior recognition data acquisition system with multichannel video input. The system can be connected to the mainstream RGBD cameras on the market, and the number of channels is linearly adjustable within a certain range. The collected videos can be processed into the current major formats, for example, AVI, MPEG, and MP4. We can estimate the 3D skeleton sequences as datasets from the collected video data.

In the G3D dataset and HDM05 dataset, we follow the principle of cross-validation experiment, using half of the dataset for training and the remaining half for testing. The experimental settings of the NTU dataset adopts the commonly used cross settings, including the cross subject and cross view. In order to keep the number of frames consistent for all behavior sequences, we downsample the execution frames of the skeleton sequences so that each dataset has a fixed number of frames. The number of frames selected for the G3D dataset is 100, the HDM05 dataset is 300, and the NTU dataset is 300. For the three datasets, we apply similar normalization preprocessing to achieve the invariance of position and view changes.

5.2. Experiment and Comparative Analysis. We first test the classification result of the proposed method on the G3D dataset. The 663 sequences in the dataset are divided into the training set and the test set according to the participating objects. The behavior sequences performed by the participants 1, 3, 5, 7, and 9 are used as the training set, and the behavior sequences performed by the remaining participants are used as the test set; thus, 333 training set sequences and 330 test set sequences are obtained.

Due to the small size of the dataset, we consider that the number of neighbor nodes' set when constructing the graph is relatively small. In the update process of graph convolution, around each node, the closest node and the 11 closest nodes around it are selected. Initially, they are considered to be of the same class, and then, the edge weights are updated.

The experimental results on G3D dataset are shown in Table 2. From the data in the table, it can be seen that the proposed method has better performance than the previous methods. The reason is that the previous method directly expands the manifold data and inputs them into the network for learning. In this process, some manifold constraints are destroyed, making the latter network unable to mine the rich information originally contained on the manifold data. The proposed method continuously projects manifold curves into the corresponding projection space along the average curve of the class, which can keep the distance between the curves projected from manifold curves as consistent as possible. In this way, the subsequent graph convolution can use the similarity between the projected curves to classify.

Table 1: Datasets' summary.

Datasets	Class	Sequence	Joint	Frame	Subject
G3D-gaming	20	663	20	6–330	10
HDM05	130	2343	31	50–721	5
NTU-RGBD	60	56 880	25	50–300	40

Table 2: Performance comparison on the G3D dataset.

Methods	Accuracy (%)
RBM + HMM [24]	86.4
SE3 + FTP [4]	87.23
SO3 [25]	87.95
SO3 + deep [26]	89.10
Ours	90.69

The proposed method has an improvement of 1.59% compared with the method combining deep neural network. This is due to the fact that the spatiotemporal trajectory can mine more abundant co-occurrent features, and using these features, we can achieve better similarity construction. Graph convolution network in the following can improve the classification result through pulling similar nodes closer and pushing others far apart.

In the HDM05 dataset, we randomly select half of the behavior sequences from each class as the training set and the remaining half as the test set. There are a total of 2343 behavior sequences in the dataset and 130 detailed behavior categories. Each category has an average of less than 20 behavior sequences. After dividing the training set and the test set, the training set and test set have about 10 behavior sequences for each category. Therefore, in the update process of graph convolution, one of the closest nodes around each node and the 7 closest nodes around it are selected.

The experimental results on the HDM05 dataset are shown in Table 3. The proposed method is compared with the method that only uses the manifold learning. There is about 20% improvement. We reckon that the continuous projection method based on the manifold curve can learn the features that contain rich spatiotemporal co-occurrence from the manifold data, and the similarity graph between behavior nodes is better constructed; thus, the graph convolution method can be used for further similarity learning. In this process, the method based on continuous projection can maintain the similarity between curves, especially the similarity between curves of the same category. This step is a key step to connect the manifold data and the deep network.

Compared with some methods using deep learning, such as PB-GCN [28], our method also has a certain improvement. The reason may be that the conventional deep learning network just arranges the data according to a certain dimension. For example, the data separated into different body parts are sent to the network for learning. In this process, the local behavior information of most of the skeleton coordinates can be used, but it is difficult to learn the essential complicated features of the relative relationship of the movement in the network. Nonetheless, the proposed network can use this information by learning the features of the manifold trajectory.

Table 3: Performance comparison on HDM05 dataset.

Methods	Accuracy (%)
SPDNet [27]	61.45
SE3 + FTP [4]	70.26
SO3 [25]	71.31
SO3 + deep [26]	75.78
PB-GCN [28]	88.17
Ours	90.05

In NTU-RGBD dataset, we conduct training and testing according to the currently commonly used data division and conduct subject-cross and view-cross experiments, respectively. Due to the large number of behavior sequences for each category in the dataset, each node cannot be directly connected to its peers when constructing a graph. When constructing the graph, 200 nearest neighbor nodes of each node are selected to form the adjacency list. In the update process of graph convolution, one of the closest nodes around each node and the 20 closest neighbors around it are selected.

The experimental results on the NTU dataset are shown in Table 4. The proposed method is greatly improved compared to the method that only uses the Lie group. The reason is that, after the graph construction by continuous projection, the introduced graph convolution module can leverage backpropagation to enhance the learning ability of the Lie group. Compared with some existing deep learning methods such as Deep-LSTM [23], ST-LSTM [29], TCN [7], and GCA-LSTM [30], our method also has some advantages. When these methods are mining behavior sequences, the main focus is on one of the temporal features and spatial features, and our method can organically combine the temporal and spatial features of the behavior characteristics by means of the manifold behavior trajectory. Compared with the current mainstream behavior recognition methods HCN [31], ST-GR [32], ST-GR [32], and ST-GCN [8], our method is still comparable.

5.3. Ablation Study. In order to verify the effectiveness of the proposed method, we performed ablation experiments on HDM05 dataset to validate each module. We have done five experiments to compare the method of directly stretching the manifold data into European data (Stretch), the method of logarithmic mapping (LogMap), the method of continuous projection (Ours/G), and the continuous projection combined with graph convolution.

The results of the ablation experiments on the HDM05 dataset are shown in Table 5. It can be seen from the table that the result of directly stretching the manifold data into the Euclidean data is the worst. In this process, the constraints of manifold data are broken, so a large amount of spatiotemporal information contained is difficult to be utilized by subsequent networks. The logarithmic mapping method can retain part of the data constraints by projecting the data into the tangent space. After projection, the data can still express most of the spatiotemporal feature information. Compared with the logarithmic mapping

Table 4: Performance comparison on the NTU-RGBD dataset using cross-subject and cross-view protocol.

Methods	Accuracy	
	Xsub (%)	Xview (%)
Lie group [4]	50.1	82.8
Deep-LSTM [23]	60.7	67.3
ST-LSTM [29]	69.2	77.7
TCN [7]	74.3	83.1
GCA-LSTM [30]	74.4	82.8
HCN [31]	86.5	91.1
ST-GR [32]	86.9	92.3
ST-GCN [8]	81.5	88.3
DGNN [33]	87.5	94.3
Ours	85.3	93.8

Table 5: Comparison of ablation experiments on the HDM05 dataset.

Methods	Accuracy (%)
Stretch	69.34
Logmap	75.65
Ours/G	82.35
Ours	90.05

method, the method based on continuous projection still has a lot of improvement, which shows that the continuous projection maintains the stronger similarity of the data after the projection than the logarithmic mapping. Finally, the method of continuous projection combined with graph convolution achieves the best results, which shows that the graph convolution method used here can achieve the function of pulling similar nodes closer and pushing others far apart to improve the classification result of the algorithm.

6. Conclusion

In this study, a deep heterogeneous manifold network is proposed. It incorporates a graph construction method based on Riemannian metric, which can preserve the nonlinear constraints of the spatiotemporal trajectory to a large extent and obtain better data projection through continuous projection. The graph nodes of behavior sequences built by this method are input to graph convolutions to realize the clustering and classification, which can improve the classification result of behavior recognition. The whole architecture combines a manifold learning backbone subnetwork and a graph convolutional network. The two parts learn from each other through end-to-end optimization, and manifold-based graph construction can guide the manifold network. The proposed method has been validated on several mainstream skeleton-based datasets and achieved competitive results. In the future, we will investigate how to automatically learn features represented in Riemannian manifold from raw data, which will further improve the discriminativeness of Riemannian representations.

Data Availability

All datasets are public datasets that can be downloaded online. G3D dataset is publicly available at https://dipersec.king.ac.uk/G3D/G3D.html, NTU RGB + D dataset is publicly available at https://rose1.ntu.edu.sg/dataset/actionRecognition/, and HDM05 dataset is publicly available at https://resources.mpi-inf.mpg.de/HDM05/.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Jinghong Chen and Li Zhang are contributed equally to this work.

Acknowledgments

This work was supported by the Shenzhen Science and Technology Programs under Grant nos. JCYJ20180306-173210774 and JCYJ20200109143035495.

References

[1] J. Wang, Z. Liu, Y. Wu, and J. Yuan, "Mining actionlet ensemble for action recognition with depth cameras," in *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1290–1297, IEEE, Providence, RI, USA, 16-21 June 2012.

[2] M. E. Hussein, M. Torki, M. A. Gowayyed, and M. El-Saban, "Human action recognition using a temporal hierarchy of covariance descriptors on 3d joint locations," in *Proceedings of the Twenty-third international joint conference on artificial intelligence*, Beijing China, August 2013.

[3] X. Yang and Y. L. Tian, "Eigenjoints-based action recognition using naive-bayes-nearest-neighbor," in *Proceedings of the 2012 IEEE computer society conference on computer vision and pattern recognition workshops*, pp. 14–19, IEEE, Providence, RI, USA, 16-21 June 2012.

[4] R. Vemulapalli, F. Arrate, and R. Chellappa, "Human action recognition by representing 3d skeletons as points in a lie group," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 588–595, Columbus, OH, USA, 23-28 June 2014.

[5] A. Jain, A. R. Zamir, S. Savarese, and A. Saxena, "Structural-rnn: deep learning on spatio-temporal graphs," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5308–5317, Las Vegas, NV, USA, 27-30 June 2016.

[6] L. Xia, C.-C. Chen, and J. K. Aggarwal, "View invariant human action recognition using histograms of 3d joints," in *Proceedings of the 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp. 20–27, IEEE, Providence, RI, USA, 16-21 June 2012.

[7] T. S. Kim and A. Reiter, "Interpretable 3d human action analysis with temporal convolutional networks," in *Proceedings of the 2017 IEEE Conference On Computer Vision And Pattern Recognition Workshops (CVPRW)*, pp. 1623–1631, IEEE, Honolulu, HI, USA, 21-26 July 2017.

[8] S. Yan, Y. Xiong, and D. Lin, "Spatial temporal graph convolutional networks for skeleton-based action recognition," in

Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, Seattle WA USA, October 2018.

[9] Q. Ke, M. Bennamoun, S. An, F. Sohel, and F. Boussaid, "A new representation of skeleton sequences for 3d action recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3288–3297, Honolulu, HI, USA, 21-26 July 2017.

[10] R. Zhao, K. Wang, H. Su, and Q. Ji, "Bayesian graph convolution lstm for skeleton based action recognition," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 6882–6892, Seoul, Korea (South), 27 Oct.-2 Nov. 2019.

[11] C. Si, W. Chen, W. Wang, L. Wang, and T. Tan, "An attention enhanced graph convolutional lstm network for skeleton-based action recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1227–1236, Long Beach, CA, USA, 15-20 June 2019.

[12] L. Shi, Y. Zhang, J. Cheng, and H. Lu, "Two-stream adaptive graph convolutional networks for skeleton-based action recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 12026–12035, Long Beach, CA, USA, 15-20 June 2019.

[13] Z. Wang, L. Zheng, Y. Li, and S. Wang, "Linkage based face clustering via graph convolution network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1117–1125, Long Beach, CA, USA, 15-20 June 2019.

[14] L. A. Adamic and E. Adar, "Friends and neighbors on the web," *Social Networks*, vol. 25, no. 3, pp. 211–230, 2003.

[15] D. Liben-Nowell and J. Kleinberg, "The link-prediction problem for social networks," *Journal of the American Society for Information Science and Technology*, vol. 58, no. 7, pp. 1019–1031, 2007.

[16] I. Mani and I. Zhang, "Knn approach to unbalanced data distributions: a case study involving information extraction," in *Proceedings of the Workshop On Learning From Imbalanced Datasets*, vol. 126, Menlo Park, CA, USA, August 2003.

[17] N. Boumal and P.-A. Absil, "A discrete regression method on manifolds and its application to data on so(n)," *IFAC Proceedings Volumes*, vol. 44, no. 1, pp. 2284–2289, 2011.

[18] K. Hüper and F. Silva Leite, "On the geometry of rolling and interpolation curves on S n , SO n , and grassmann manifolds," *Journal of Dynamical and Control Systems*, vol. 13, no. 4, pp. 467–502, 2007.

[19] S. Banerjee, "On geodetic distance computations in spatial modeling," *Biometrics*, vol. 61, no. 2, pp. 617–625, 2005.

[20] R. Caseiro, P. Martins, J. F. Henriques, F. Silva Leite, and J. Batista, "Rolling riemannian manifolds to solve the multiclass classification problem," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 41–48, Portland, OR, USA, 23-28 June 2013.

[21] V. Bloom, D. Makris, and V. Argyriou, "G3d: A gaming action dataset and real time action recognition evaluation framework," in *Proceedings of the 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp. 7–12, IEEE, Providence, RI, USA, 16-21 June 2012.

[22] M. Müller, T. Röder, M. Clausen, B. Eberhardt, B. Kruger, and A. Weber, "Documentation mocap database hdm05," *Computer Graphics Technical Reports*, 2007.

[23] A. Shahroudy, J. Liu, T.-T. Ng, and G. Wang, "Ntu rgb+ d: A large scale dataset for 3d human activity analysis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1010–1019, Las Vegas, NV, USA, 27-30 June 2016.

[24] S. Nie and Q. Ji, "Capturing global and local dynamics for human action recognition," in *Proceedings of the 2014 22nd International Conference on Pattern Recognition*, pp. 1946–1951, IEEE, Stockholm, Sweden, 24-28 Aug. 2014.

[25] R. Vemulapalli and R. Chellapa, "Rolling rotations for recognizing human actions from 3d skeletal data," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4471–4479, Las Vegas, NV, USA, 27-30 June 2016.

[26] Z. Huang, C. Wan, T. Probst, and L. Van Gool, "Deep learning on lie groups for skeleton-based action recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 6099–6108, Honolulu, HI, USA, 21-26 July 2017.

[27] Z. Huang and L. Van Gool, "A riemannian network for spd matrix learning," in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, San Francisco, CA, USA, February 2017.

[28] K. Thakkar and P. J. Narayanan, "Part-based graph convolutional network for action recognition," 2018, https://arxiv.org/abs/1809.04983.

[29] J. Liu, A. Shahroudy, D. Xu, and G. Wang, "Spatio-temporal lstm with trust gates for 3d human action recognition," in *Proceedings of the European Conference on Computer Vision*, pp. 816–833, Springer, Amsterdam, The Netherlands, October 2016.

[30] J. Liu, G. Wang, P. Hu, L.-Y. Duan, and A. C. Kot, "Global context-aware attention lstm networks for 3d action recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1647–1656, Honolulu, HI, USA, 21-26 July 2017.

[31] C. Li, Q. Zhong, D. Xie, and S. Pu, "Co-occurrence feature learning from skeleton data for action recognition and detection with hierarchical aggregation," 2018, https://arxiv.org/abs/1804.06055.

[32] B. Li, X. Li, Z. Zhang, and F. Wu, "Spatio-temporal graph routing for skeleton-based action recognition," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 8561–8568, Honolulu, HA, USA, January 2019.

[33] L. Shi, Y. Zhang, J. Cheng, and H. Lu, "Skeleton-based action recognition with directed graph neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7912–7921, Long Beach, CA, USA, 15-20 June 2019.