Hindawi
Security and Communication Networks
Volume 2023, Article ID 1357886, 10 pages
https://doi.org/10.1155/2023/1357886

Research Article

PSI-CA-Based Vehicle Selection Scheme for Data Sharing in Internet of Vehicles

Zhengtao Jiang, Ting Yu, Ye Chen, Huiqiang Li, and Xiaoxuan Guo

School of Computer and Cyber Science, Communication University of China, Beijing 100024, China

Correspondence should be addressed to Ting Yu; chloeyt0803@163.com

Received 18 April 2023; Revised 14 July 2023; Accepted 4 August 2023; Published 19 August 2023

Academic Editor: Salvatore D'Antonio

In recent years, the development of the Internet of Vehicles (IoV) has led to an increase in the demand for data sharing services for the IoV. In the era of big data, safe and convenient data sharing services for IoV are inseparable from the support of reliable data. With the increase in the number of smart cars, how to filter the data provided by vehicles while protecting privacy to improve the quality of data sharing services has attracted the attention of scholars. However, in the existing data sharing scheme for IoV, the vehicles that provide data are selected based on the reputation mechanism or the voting mechanism. Our analysis shows that these two mechanisms lack objectivity and are vulnerable to cooperative attack. Therefore, based on spatiotemporal matching, a new trusted relationship establishment method, this paper proposes an objective and cooperative attack-resistant vehicle selection scheme for data sharing in IoV. Bloom filters and an exponential ElGamal encryption scheme are used to implement private set intersection cardinality (PSI-CA) technology for evaluating spatiotemporal matching level. In this paper, the scheme proposed is of great significance to improve existing vehicle services and improve driving safety.

1. Introduction

In the era of the Internet of Everything (IoE), the Internet of Things (IoT) is regarded as another wave of information technology revolution after the Internet. IoT refers to the connection of any object with the network through information sensing devices in accordance with the agreed protocol, so that the objects can exchange information and communicate through the information transmission medium. IoT allows people to anticipate and control their surroundings to make more advantageous decisions. IoT will be the trend of the future society development. The world of IoE has great potential.

The concept of IoV stems from IoT. The application of private set intersection (PSI) technology in the field of IoV is one of the research hotspots in academia and industry in recent years, and it is also a typical example of the effective combination of theoretical technology and engineering application. PSI technology is researched to solve many problems in the field of IoV, such as realizing IoV social based on vehicle attribute matching [1, 2], cloud-assisted computing services and data sharing between IoV vehicles [1, 3], providing location-based services in IoV [4, 5], and privacy protection proximity testing in IoV [5]. In our vehicle selection scheme, PSI technology is used to evaluate the level of spatiotemporal matching. The vehicle with the highest level of spatiotemporal matching will be selected as the vehicle that provides the data for data sharing in IoV. Currently, vehicles are equipped with a variety of wireless communication components and an increasing number of sensors, which enable vehicles to generate large amounts of data while driving [6]. In IoV, data are shared between vehicles and road side units (RSUs) to improve vehicle-related services and driving safety.

In the process of sharing data, the accuracy and reliability of the data must be guaranteed. Unreliable data can lead drivers to make incorrect judgments. In the large-scale IoV, malicious vehicles spreading false information will threaten the life safety and property safety of drivers and even endanger public safety [7]. Therefore, designing a secure and reliable IoV data sharing scheme is a great challenge and has become a research hotspot.

The reputation mechanism and the voting mechanism used in recent research [3, 8, 9] are actually subjective behaviors and cannot truly objectively guarantee the reliability of the data. To ensure the reliability of shared data, the reliability of the vehicles from which the data originates should first be objectively guaranteed. At the same time, it is also necessary to consider the situation that multiple vehicles in IoV submit similar information at the same period in real-world scenarios. If the vehicles uploading sharing data are not objectively selected, the sharing data scheme in IoV will be messy, inefficient, and insecure. Therefore, efficient and objective selection of reliable vehicles that provide data is of great significance in the data sharing scheme of IoV.

The explosive growth of mobile-connected and location-aware devices makes it possible to have a new way of establishing trust relationships, which Sun et al. [10] defined as spatiotemporal matching. In particular, a vehicle could very easily maintain its spatiotemporal profile by recording its continuous whereabouts in time, and the level of its spatiotemporal profile matching that of the RSU can be translated into the level of trust they two have in each other. In this paper, "spatiotemporal profile matching" is also referred to as "spatiotemporal matching." As an example, if a RSU discovers through spatiotemporal matching that a vehicle is frequently present in its area, then it is natural for the RSU to trust the vehicle more than a vehicle that has only appeared once in this area. In other words, vehicles that appear more often in the RSU area provide more reliable data. The popular understanding is that this vehicle often appears in this area, which means that the vehicle knows more about this area. At the same time, the spatiotemporal profile used to assist in judging the spatiotemporal matching level of vehicles and RSUs is objective data and is not affected by subjective factors.

In this paper, spatiotemporal matching technology is used to select vehicles that provide reliable data, and the level of spatiotemporal profile matching is judged by PSI-CA technology. A spatiotemporal profile consists of private information such as unique identifiers of vehicles and RSUs as well as records of interactions between them. The position of the vehicles in a certain period can be inferred from the spatiotemporal profile. Therefore, the spatiotemporal profile is highly sensitive private information. In order to protect the privacy and security of the spatiotemporal profile of vehicles and RSUs, this paper computes PSI-CA by employing the Bloom filters to determine the existence of elements and exponential ElGamal to encrypt Bloom filters. In order to reduce the local computing burden of vehicles and RSUs, the main computation in the spatiotemporal profile matching process is outsourced to the cloud. Overall, the contributions of this paper can be summarized as follows:

(1) This paper effectively combines spatiotemporal matching and IoV. Spatiotemporal matching is a new way to establish trust relationships, and this paper applies the way to practice, which expands the application scenarios of spatiotemporal matching in IoV.

(2) This paper presents a reliable and secure vehicle selection scheme for data sharing in IoV based on spatiotemporal matching, Bloom filters, and exponential ElGamal. In this scheme, spatiotemporal matching is used to objectively select vehicles that provide data for data sharing in IoV, Bloom filters are used to achieve private set size hiding, and exponential ElGamal is used to encrypt data and prevent collusion attacks.

(3) Existing IoV data sharing schemes based on reputation mechanism or voting mechanism cannot objectively ensure the reliability of shared data and are vulnerable to cooperative attack. In order to solve these problems, the scheme proposed in this paper uses PSI-CA to achieve spatiotemporal profile matching to select the vehicle providing data, which can objectively ensure the reliability of the data and resist cooperative attack.

The remainder of this paper is organized as follows. Section 2 gives a brief overview of the related work. Section 3 describes preliminaries and notations. Section 4 presents the detailed process of our proposed scheme. Section 5 provides a theoretical analysis and performance analysis of the proposed scheme. Section 6 concludes this paper.

2. Related Work

The rapid development of the Internet has enabled data to be shared between different subjects, thereby reducing the workload and cost of data collection and enhancing the meaning and value of data. Data sharing has a wide range of applications in politics, economics, and healthcare. In addition, data sharing is also used in the field of IoV to assist traffic management, improve traffic safety, promote the prosperity of in-vehicle applications, and promote the construction of smart cities. Scholars have carried out a lot of research studies around data sharing in IoV.

Lai et al. [2] proposed to realize IoV social by matching the attribute data of the shared vehicles. Lai et al. [1] proposed to provide privacy preservices by analyzing the data shared in IoV, such as real-time updates of road traffic conditions. To prevent unauthorized access to plaintext data shared in the IoV, Ma et al. [11] proposed an attribute-based encryption algorithm using blockchain. But the authors neglected to judge the accuracy and reliability of the data.

In order to prevent the adverse consequences of sharing malicious data in the IoV, Kang et al. [8], Xie et al. [9], and Wang et al. [3] proposed different data sharing schemes for the IoV based on the reputation mechanism and voting mechanism, respectively. However, their schemes are vulnerable to cooperative attacks and lack objective judgment of the reliability of shared data. The basic idea of these schemes is to select the vehicles that provide the data through different mechanisms. In order to effectively prevent the spread of malicious data, Kang et al. [8] and Xie et al. [9] proposed sharing schemes for sharing data and video based on the reputation mechanism. Kang et al. [8] proposed quantifying the reputation value of vehicles based on the interactions

between vehicles. The vehicle selects the best data provider based on the reputation value. The calculation of the reputation value is based on the three-weight subjective logic model. If a positive shared interaction occurs between two vehicles, that is, the data shared between the vehicles are considered relevant and useful, the relationship between the vehicles is enhanced, and the reputation value of the vehicle is also increased. In order to prevent the malicious vehicle from interfering with traffic by publishing false or inaccurate data, Xie et al. [9] proposed that the authenticity of the traffic data broadcast by the vehicle is scored by the vehicles in its vicinity. RSUs calculate the reputation value of information based on the distance between the scoring vehicle and the vehicle providing the information. However, the use of reputation mechanism can be subject to cooperative attack. Some vehicles negotiate into a small group and together they give a high rating to a certain vehicle in the group or interact positively with it multiple times, which gives the vehicle a high reputation value and therefore the opportunity to share malicious data.

Wang et al. [3] proposed a data sharing scheme in a 6G vehicular ad hoc network, which can not only ensure the reliability and security of shared data but also protect the privacy of participants. In the scheme proposed in the literature [3], a cloud-assisted PSI protocol is designed as a blockchain voting consensus mechanism, which allows vehicles with the same characteristics to make subjective judgments about the accuracy of data to improve the accuracy of data evaluation. However, voting on data accuracy by filtering vehicles with the same attributes relies on the subjective judgments of users, and the final result obtained is neither objective nor reliable. This approach is also vulnerable to cooperative attacks. Even if the mechanism of adding malicious vehicles to the blacklist is introduced to restrict the behavior of malicious vehicles, it does not prevent new vehicles from acting maliciously in the first few voting processes, which can affect the final result.

When Sun et al. [10] first proposed the concept of spatiotemporal matching, they pointed out that one of its application scenarios is participatory sensing. Participatory sensing is similar to data sharing in IoV in that data are collected through devices with sensors, and after uploading the data, knowledge is obtained from the data and shared. Wang et al. [12] also studied the application of spatiotemporal matching in IoV. In order to solve the problems of location cheating and privacy disclosure in vehicular crowdsensing, they proposed a secure participant recruitment scheme based on location authentication for vehicular crowdsensing. In their scheme, spatiotemporal matching is used to implement location authentication. This is different from our scheme because spatiotemporal matching is recognized as a new trusted relationship establishment method in our scheme. However, at present, spatiotemporal matching is mainly used in contact tracing, which is used in combination with PSI-CA for the prevention and treatment of infectious diseases [13].

Under the condition of protecting the privacy of the sets of parties, the common method used to measure the spatiotemporal matching level of the two parties is PSI-CA, a variant of PSI [14]. The PSI-CA protocol allows a party to obtain the cardinality of the intersection of the private sets with another party without revealing information other than the intersection cardinality [15].

In order to reduce the computing burden of parties, this paper chooses to outsource PSI-CA computing to the cloud. Tajima et al. [16] presented two different two-party outsourced PSI-CA protocols based on the Bloom filters and a BGV-style fully homomorphic encryption (FHE) scheme. However, due to the use of FHE, the communication and computational overhead of the protocols in their literature are extremely large. Jolfaei et al. [17] designed an efficient outsourced PSI-CA protocol in multiparty scenario. The protocol calculates PSI-CA by using Bloom filters technique and the exponential ElGamal encryption over Bloom filters.

3. Preliminaries and Notations

This section summarizes notations, scheme model diagram, and building blocks used in our scheme and lists four potential attacks against our proposed scheme.

3.1. Notations. The notations used in this paper are shown in Table 1.

3.2. Scheme Model Diagram. In order to more intuitively show the entity structure of our proposed vehicle selection scheme for data sharing in IoV, the scheme model diagram is given in this subsection, as shown in Figure 1. In addition, this subsection gives descriptions of the entities or terms involved in our scheme.

3.2.1. Vehicles. Vehicles are equipped with a large number of sensors as well as storage and wireless communication modules. Vehicles can generate road condition-related announcement messages while driving, and these announcement messages can be passed to RSU and filtered with the help of the cloud. Then, the filtered announcement messages can be shared by RSU to all vehicles in the area. In our scheme, the vehicles are not trusted. It is possible for vehicles to launch cooperative attacks, forged interaction record attacks against RSUs, or collusion attacks against RSUs in the cloud.

3.2.2. RSUs. RSUs are equipped with processing, storage, and wireless communication modules. The main functions of RSUs are to collect the current road conditions, traffic conditions, and other information, communicate with vehicles, traffic lights, electronic signs, and other terminals through the C-V2X, realize vehicle-road interconnection, data sharing, assist drivers to drive, and ensure the safety of drivers and vehicles in the entire transportation field. In our scheme, RSUs are not trusted. RSUs may launch identity forgery attacks or conspire with the cloud to obtain private data from vehicles.

3.2.3. C-V2X. C-V2X is a new IoV communication standard, which has been proposed internationally. C-V2X is a cellular network-based wireless communication

Table 1: The notations used in this paper.

Symbols	Description
V_j	Legitimate vehicle in the IoV
R_i	Legitimate RSU in the IoV
ID_{V_j}	Unique identifier issued by CA to vehicle V_j
PK_{V_j}, SK_{V_j}	Vehicle V_j's public key and private key
PK_{R_i}, SK_{R_i}	RSU R_i's public key and private key
$\mathrm{Hash}(\cdot)$	Hash function
Sig_{SK}	Signature of data with private key SK
n	The number of legal vehicles
m	The number of legal RSUs
d	The cardinality of the private set $\mathrm{SID}_{\mathrm{set}_{V_j}}$ of vehicle V_j
c	The cardinality of the private set $\mathrm{SID}_{\mathrm{set}_{Ri}}$ of RSU R_i
BF	A Bloom filter
b	Size of Bloom filter
$BF[i]$	The bit at the index i of BF
k	The number of hash functions used in Bloom filter
$h(\cdot)$	A hash function used in Bloom filter
t	The threshold for the number of vehicles to submit data
Enc_{PK}	Exponential ElGamal encryption with public key PK
Dec_{SK}	Exponential ElGamal decryption with private key SK
counter	Equals to the output of PSI-CA

technology for vehicles. The standard realizes assisted driving safety and traffic efficiency and reduces costs.

3.2.4. Certification Authority (CA). The CA is assumed to be a fully trusted authority that issues the identity identifier ID for each vehicle and distributes key information for each vehicle and each RSU for signature, encryption, or decryption. Vehicles and RSUs should register in CA before participating in the subsequent spatiotemporal profile generation and vehicles selection process.

3.2.5. Cloud. Provides cloud computing services for vehicles and RSUs. In our scheme, the cloud is not trusted. The cloud may conspire with vehicles (RSUs) to obtain private data of RSUs (vehicles).

3.3. Building Blocks

3.3.1. Bloom Filter. The Bloom filter is a probabilistic data structure that supports set membership queries. The Bloom filter has the characteristics of high efficiency and small memory consumption. The Bloom filter represents a set $S = \{s_1, s_2, \cdots, s_n\}$ by an array of m bits [18]. The key idea is to use k hash functions, $h_a(x)$, $1 \le a \le k$, to map items $x \in S$ to random numbers uniform in the range $0, 1, \cdots, m-1$. Let BF represent a Bloom filter for the set S and $BF[i]$ represent the bit at the index i. A variant of Bloom filter is described in three algorithms of setup, insertion, and membership test [19].

(1) *Setup Algorithm*. Initially all the bits in the filter are set to 1.

(2) *Insertion Algorithm*. To add an element $x \in S$ into a Bloom filter BF, x is hashed using k hash functions to get k indices as $h_1(x), h_2(x), \cdots, h_k(x)$. Then, let $BF[h_a(x)] = 0$, $1 \le a \le k$.

(3) *Membership Test Algorithm*. To check if an element x is a membership of BF or not, x is hashed with the k hash functions. Now, if at least one of $BF[h_1(x)], BF[h_2(x)], \cdots, BF[h_k(x)]$ is 1, then x is not in S; otherwise, x is probably in S.

The Bloom filter is a probability-based data structure implemented based on the hash function, which can be used to determine the existence of elements in a set. The cost of the Bloom filter being able to efficiently insert and query elements is that it can only determine that an element is absolutely not in the set, and there is a certain false positive probability for the element in the set. The reason for the false positive probability is that hash collisions cannot be completely avoided. However, the Bloom filter allows false positive probability (fpp), whereby an element that has not been inserted in the filter can pass the set membership test mistakenly. The expression for false positive probability is

$$\mathrm{fpp} = \left(1 - \left(1 - \frac{1}{m}\right)^{kn}\right)^k. \tag{1}$$

3.3.2. Additive Homomorphic Encryption. To preserve the privacy of the vehicles and RSUs input sets while cloud computing, the Bloom filters should be encrypted with an additive homomorphic encryption scheme. Hence, the cloud will be able to compute operations over the Bloom filters without leaking information of the vehicles and RSUs input sets. Our scheme has employed the exponential ElGamal, a version of ElGamal, where g^m is encrypted instead of m, as the encryption algorithm, where m is the message. The exponential ElGamal consists of the following algorithm;

(1) *Key Generation*. The public key PK_V and private key SK_V of vehicle V meet $PK_V = g^{SK_V} \bmod p$, where $SK_V \in Z_q$ is a secret share for V to decrypt a ciphertext and sign a text. The operations of RSU R are similar. Parameter $g \in F_p$ has a prime order q, where F_p is a finite filed. The value $y = PK_V \cdot PK_R \bmod p$ is the public key computed by V and R.

(2) *Encryption*. The message $m \in Z_q$ is encrypted to the ciphertext $\mathrm{Enc}_y(m) = (u, v)$ as in the following. A random number $r \in Z_q$ is chosen, and then u and v are computed as $u = g^r \bmod p$ and $v = g^m y^r \bmod p$.

(3) *Decryption*. V computes $z_V = u^{SK_V} \bmod p$ and sends it via a secure channel to R. R computes $z_R = u^{SK_R} \bmod p$ and $z = z_V \cdot z_R \bmod p$. R can decrypt ciphertext (u, v) by $Dec_z(u, v)$, where

$$\begin{aligned} Dec_z(u, v) &= \frac{v}{z} \mod p \\ &= \frac{g^m \bullet g^{(SK_V + SK_R)r}}{g^{r(SK_V + SK_R)} \bmod p} \\ &= g^m. \end{aligned} \tag{2}$$

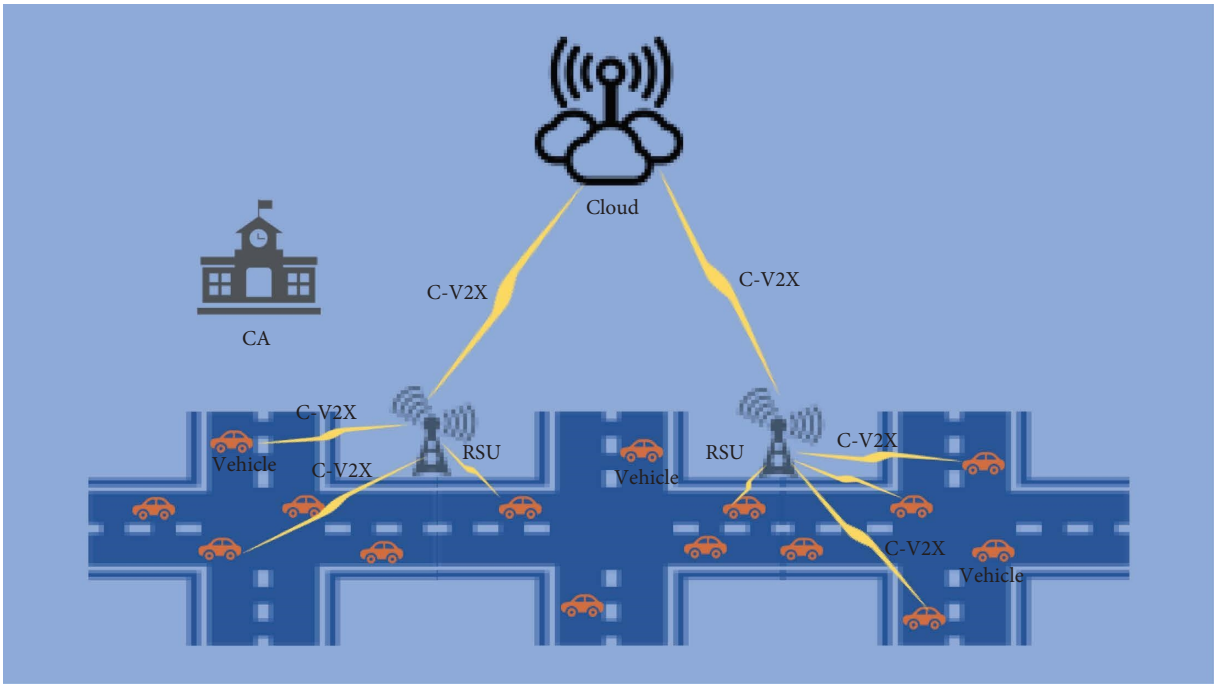

Figure 1: Scheme model diagram.

This cryptosystem has the following properties:

(1) Given two messages $m_1, m_2 \in Z_q$, an additive homomorphic property $Enc(m_1) \times Enc(m_2) = Enc(m_1 + m_2)$ holds, where $Enc(m_1) \times Enc(m_2)$ is the multiplication of two ciphertext $Enc(m_1)$ and $Enc(m_2)$ as

$$\begin{aligned} Enc(m_1) \times Enc(m_2) &= (g^{r_1}, g^{m_1} y^{r_1}) \times (g^{r_2}, g^{m_2} y^{r_2}) \\ &= (g^{r_1+r_2}, g^{m_1+m_2} y^{r_1+r_2}) \\ &= Enc(m_1 + m_2). \end{aligned} \tag{3}$$

(2) Given $m \in Z_q$, $e \in Z_q$ a scalar homomorphic property $Enc(m)^e = Enc(em)$ holds.

3.4. Potential Attacks

3.4.1. Forged Interaction Record Attack. The malicious vehicle hopes to increase the cardinality of intersection with RSU by falsifying interaction data. The vehicle is then selected as the vehicle providing the data, which in turn publishes malicious information.

3.4.2. Cooperative Attack. Multiple vehicles work together so that one of them is selected as the data provider.

3.4.3. Identity Forgery Attack. An adversary could disguise as an RSU to gain access to the vehicle's private data.

3.4.4. Collusion Attack. There are two types of possible collusion attacks against our scheme. One is when the malicious vehicle colludes with the cloud to get the plaintext of RSU's Bloom filter. Another is when the malicious RSU colludes with the cloud to get the plaintext of the vehicle's original Bloom filter.

4. Vehicle Selection Scheme

Our vehicle selection scheme is divided into two parts. The first part is the generation of a spatiotemporal profile, including the system initialization phase, interactive record generation phase, and Bloom filters construction phase. The second part is the selection of vehicles, in which RSUs select vehicles for providing shared data according to the level of matching between the vehicles and their own spatiotemporal profile. The second part contains the set outsourcing phase, the cloud computation phase, and the PSI-CA computation phase.

4.1. The Generation of Spatiotemporal Profile

4.1.1. System Initialization Phase. When a new vehicle V_j ($j \in [1, n]$) registers with the CA for the first time, the CA will issue a pair of exclusive public and private key (PK_{V_j}, SK_{V_j}) and identity identifier ID_{V_j} to V_j.

All RSUs also need to be authenticated and authorized by CA to ensure their legitimacy. When a legal RSU R_i ($i \in [1, m]$) registers with the CA, the CA will issue a exclusive public-private key pair (PK_{R_i}, SK_{R_i}) to R_i.

In this phase, the keys are generated in line with the exponential ElGamal encryption, as described in Section 3.3.2.

4.1.2. Interactive Record Generation Phase. When RSU R_i detects that vehicle V_j enters its area, R_i sends an interactive signal IS_{R_i,V_j} to V_j

$$\mathrm{IS}_{R_i,V_j} = \left\{\mathrm{SID}_{R_i,V_j}, \mathrm{RVC}_{\mathrm{SID}_{R_i,V_j}}\right\}, \tag{4}$$

where $\mathrm{SID}_{R_i,V_j} = \mathrm{Hash}(\mathrm{ID}_{V_j} \| \mathrm{timestamp})$ and $\mathrm{RVC}_{\mathrm{SID}_{R_i,V_j}} = \mathrm{Sig}_{SK_{R_i}}(\mathrm{SID}_{R_i,V_j})$ is a verifiable credential.

V_j verifies $\mathrm{RVC}_{\mathrm{SID}_{R_i,V_j}}$ after receiving the interactive signal IS_{R_i,V_j} sent by R_i. If the verification passes, V_j stores the IS_{R_i,V_j} locally and adds the SID_{R_i,V_j} to the local SID set $\mathrm{SID}_{\mathrm{set}_{V_j}} = \{\mathrm{SID}_1, \cdots, \mathrm{SID}_d\}$. $\mathrm{SID}_{\mathrm{set}_{V_j}}$ can be translated as the spatiotemporal profile of V_j. V_j signs SID_{R_i,V_j} and sends $\mathrm{VVC}_{\mathrm{SID}_{R_i,V_j}}$ to R_i, where $\mathrm{VVC}_{\mathrm{SID}_{R_i,V_j}} = Sig_{SK_{V_j}}(\mathrm{SID}_{R_i,V_j})$ is a verifiable credential.

R_i verifies the received $\mathrm{VVC}_{\mathrm{SID}_{R_i,V_j}}$ and stores the $\mathrm{IS}_{V_j,R_i} = \left\{\mathrm{SID}_{R_i,V_j}, \mathrm{VVC}_{\mathrm{SID}_{R_i,V_j}}\right\}$ locally after the verification is passed. R_i then adds the SID_{R_i,V_j} to the local SID set $\mathrm{SID}_{\mathrm{set}_{Ri}} = \{\mathrm{SID}_1, \cdots, \mathrm{SID}_c\}$. $\mathrm{SID}_{\mathrm{set}_{Ri}}$ can be translated as spatiotemporal profile of R_i.

4.1.3. Bloom Filters Construction Phase. V_j constructs its Bloom filter BF_{V_j} by setup algorithm and insertion algorithm (see Section 3.3.1 Bloom filter) as $BF_{V_j} = [BF_{V_j}[0], \cdots, BF_{V_j}[b-1]]$, where $BF_{V_j}[h_a(\mathrm{SID})] = 0$, $\mathrm{SID} \in \mathrm{SID}_{\mathrm{set}_{V_j}}$, and $1 \le a \le k$.

R_i constructs its Bloom filter in the same way as V_j. The Bloom filter corresponding to $\mathrm{SID}_{\mathrm{set}_{Ri}}$ is $BF_{R_i} = [BF_{R_i}[0], \cdots, BF_{R_i}[b-1]]$, where $BF_{R_i}[h_a(\mathrm{SID})] = 0$, $\mathrm{SID} \in \mathrm{SID}_{\mathrm{set}_{R_i}}$, and $1 \le a \le k$.

4.2. The Selection of Vehicles. When vehicles submit traffic data to R_i and the number of vehicles submitting data reaches the threshold t, R_i initiates a PSI-CA request to the cloud to achieve spatiotemporal profile matching and notifies these t vehicles to perform the following processes.

4.2.1. Set Outsourcing Phase. R_i encrypts BF_{R_i} locally to get $\mathrm{Enc}_{y_j}(BF_{R_i})$. Then, R_i sends $\mathrm{Enc}_{y_j}(BF_{R_i})$ to the cloud, where

$$y_j = PK_{V_j} \cdot PK_{R_i} \bmod p, 1 \le j \le t,$$
$$\mathrm{Enc}_{y_j}\left(BF_{R_i}\right) = \left[\mathrm{Enc}_{y_j}\left(BF_{R_i}[0]\right), \cdots, \mathrm{Enc}_{y_j}\left(BF_{R_i}[b-1]\right)\right]. \tag{5}$$

V_j encrypts BF_{V_j} locally to get $\mathrm{Enc}_{y_j}(BF_{V_j})$. Then, V_j sends $\mathrm{Enc}_{y_j}(BF_{V_j})$ to the cloud, where

$$\mathrm{Enc}_{y_j}\left(BF_{V_j}\right) = \left[\mathrm{Enc}_{y_j}\left(BF_{V_j}[0]\right), \cdots, \mathrm{Enc}_{y_j}\left(BF_{V_j}[b-1]\right)\right]. \tag{6}$$

4.2.2. Cloud Computation Phase. The cloud computes the inner product of encrypted Bloom filters denoted by

$$\begin{aligned}\mathrm{Enc}_{y_j}\left(BF_{R_i,V_j}\right) &= \mathrm{Enc}_{y_j}\left(BF_{R_i}\right) \times \mathrm{Enc}_{y_j}\left(BF_{V_j}\right) \\ &= \left[\mathrm{Enc}_{y_j}\left(BF_{R_i}[0]\right) \times \mathrm{Enc}_{y_j}\left(BF_{V_j}[0]\right), \cdots, ,\right. \\ &\quad \left.\mathrm{Enc}_{y_j}\left(BF_{R_i}[b-1]\right) \times \mathrm{Enc}_{y_j}\left(BF_{V_j}[b-1]\right)\right] \\ &= \left[\mathrm{Enc}_{y_j}\left(BF_{R_i}[0] + BF_{V_j}[0]\right), \cdots, ,\right. \\ &\quad \left.\mathrm{Enc}_{y_j}\left(BF_{R_i}[b-1] + BF_{V_j}[b-1]\right)\right].\end{aligned} \tag{7}$$

The cloud randomizes each element of $\mathrm{Enc}_{y_j}(BF_{R_i,V_j})$. The result of randomization is $\mathrm{Enc}_{y_j}(RBF_{R_i,V_j})$, where

$$\begin{aligned}\mathrm{Enc}_{y_j}\left(RBF_{R_i,V_j}\right) &= \left[\mathrm{Enc}_{y_j}\left(BF_{R_i,V_j}[0]\right)^{r_0}, \cdots, \mathrm{Enc}_{y_j}\left(BF_{R_i,V_j}[b-1]\right)^{r_{b-1}}\right] \\ &= \left[\mathrm{Enc}_{y_j}\left(\left(BF_{R_i}[0] + BF_{V_j}[0]\right) \cdot r_0\right), \cdots, ,\right. \\ &\quad \left.\mathrm{Enc}_{y_j}\left(\left(BF_{R_i}[b-1] + BF_{V_j}[b-1]\right) \cdot r_{b-1}\right)\right].\end{aligned} \tag{8}$$

It should be noted that $r_0, \cdots, r_{b-1}$ are the random elements from Z_q. Then, cloud sends $Enc_{y_j}(RBF_{R_i,V_j})$ to V_j.

4.2.3. PSI-CA Computation Phase. V_j computes L_s as follows: $\forall \mathrm{SID}_s \in \mathrm{SID}_{\mathrm{set}_{V_j}}$ and $1 \le s \le d$:

$$L_s = \prod_{a=1}^{k} \mathrm{Enc}_{y_j}\left(RBF_{R_i,V_j}\left[h_a(\mathrm{SID}_s)\right]\right). \tag{9}$$

V_j constructs vector $l = (L_1, \cdots, L_d)$. V_j adds $w - d$ dummies, i.e., $L_{d+1} = r_{d+1}, \cdots, L_w = r_w$ into the vector l to hide the number of elements of $\mathrm{SID}_{\mathrm{set}_{V_j}}$, where w is a random number selected randomly by V_j (e.g., $d < w < 2d$ can be selected).

V_j then gets $l = (L_1, \cdots, L_d, \cdots, L_w)$ and computes $z_{V_j,s} = u_s^{SK_{V_j}} \bmod p (1 \le s \le w)$. V_j shuffles l and $z_{V_j,s} (1 \le s \le w)$ and sends them to R_i.

R_i calculates $z_{R_i,s} = u_s^{SK_{R_i}} \bmod p\,(1 \le s \le w)$. R_i then calculates $z_s = z_{R_i,s} \cdot z_{V_j,s}\,(1 \le s \le w)$ for decryption.

R_i initializes $\text{counter}_j = 0$. R_i decrypts each elements of vector l with $Dec_{z_s}(L_s)\,(1 \le s \le w)$ and computes $\text{counter}_j = \text{counter}_j + 1$ if the output equals to 1.

The numerical magnitude of counter is related to the degree of spatiotemporal profile matching between the vehicle and RSU. The higher the value of counter, the higher the degree of spatiotemporal profile matching between the vehicle and the RSU, the higher the degree of mutual trust, and the more reliable the data provided by the corresponding vehicle. Finally, R_i selects the road information provided by the vehicle V_j corresponding to the maximum value counter_j in the $\{\text{counter}_1, \cdots, \text{counter}_t\}$ to broadcast.

4.3. Sequence Diagram. In order to more visually illustrate the process of our proposed vehicle selection scheme, we give a corresponding sequence diagram, as shown in Figure 2.

5. Analysis and Experimental Evaluation

5.1. Privacy Analysis. In our data sharing scheme, we provide strong privacy assurance for all vehicles and RSUs. For vehicles, vehicles first generate Bloom filters corresponding to their private sets $\text{SID}_{\text{set}_V}$ locally. The use of Bloom filters not only hides the elements of the sets but also hides the sizes of the sets. Vehicles encrypt Bloom filters locally before sending the ciphertext to the cloud, and subsequent calculations are performed on the basis of the ciphertext. When constructing the vector l, vehicles add $w - d$ random dummies, which again hides the number of sets elements. Due to the difficulty of solving the discrete logarithm problem, when vehicle transmits the private keys SK_V through $z_{V,s} = u_s^{SK_V} \bmod p$, the privacy of the private key SK_V is guaranteed.

For RSUs, RSUs first generate Bloom filters corresponding to their private sets $\text{SID}_{\text{set}_R}$ locally. Similarly, the elements and sizes of the $\text{SID}_{\text{set}_R}$ are hidden. RSUs encrypt Bloom filters locally before sending the ciphertext to the cloud, and subsequent calculations are also carried out on the basis of the ciphertext.

In the end, RSUs only get the cardinality of the intersections, and no other information is leaked. Therefore, the privacy of both vehicles and RSUs is guaranteed.

5.2. Security Analysis. We designed our data sharing scheme to be robust and resistant to the potential attacks listed in Section 3.4. Next, we will analyze the security of our scheme from four aspects corresponding to potential attacks.

5.2.1. Forged Interaction Record Attack. Vehicle wants to increase the intersection cardinality with RSU by forging $\text{SID}_{\text{set}_V}$ set elements. However, the interaction record IS of vehicle and RSU is stored locally by both parties. Vehicle cannot forge an IS stored locally in RSU. In addition, the $\text{RVC}_{\text{SID}_{R,V}}$ of IS is issued by the private key of RSU. Vehicle cannot obtain the private key of RSU, so it is impossible to forge IS. Even if vehicle uses the old $\text{RVC}_{\text{SID}_{R,V}}$, its malicious behavior can be detected through the timestamp of $\text{SID}_{R,V}$. Therefore, our scheme is resistant to a forged interaction record attack.

5.2.2. Cooperative Attack. Neither literature [9] nor literature [3] is resistant to cooperative attack, but our scheme is resistant to cooperative attack because in our scheme, vehicles only interact with RSUs and the cloud. Whether a vehicle is selected is only related to its local data and RSU's local data and is not affected by other vehicles. Therefore, our scheme is resistant to cooperative attack.

5.2.3. Identity Forgery Attack. Since each RSU is initially registered with CA to obtain its public-private key pair and prove its legal identity. When an adversary forges RSU's identity, it can be discovered by verifying its signature. Therefore, our scheme is resistant to identity forgery attack.

5.2.4. Collusion Attack. The malicious vehicle cannot collude with the cloud to obtain the plaintext of RSU's Bloom filter because neither the vehicle nor the cloud has the private key to decrypt the ciphertext of RSU's Bloom filter. The malicious RSU also cannot collude with the cloud to obtain the plaintext of the vehicle's original Bloom filter, as RSU can only get the key used to decrypt the ciphertext of the randomized Bloom filter but not the private key of the ciphertext of the vehicle's original Bloom filter. Therefore, our scheme is resistant to collusion attack.

5.3. Cost. Cost here refers to the communication cost, transmission cost, and computational cost required for an RSU to share data once within its area in our data sharing scheme. At the same time, since the system initialization phase is not necessary for each sharing process, the cost of the system initialization phase is not considered here.

In our vehicle selection scheme, before an RSU can share information, there must be threshold t vehicles to provide the information. Therefore, the RSU must at least communicate with the t vehicles.

5.3.1. Communication Cost. Suppose that two parties each send a message to the other, which is a round of communication between the two parties. The case where only one-way messages are sent is also considered a round of communication. Consider the example of communication between a vehicle, a RSU, and the cloud. In the part of the generation of the spatiotemporal profile, the RSU sends an interactive signal to the vehicle. The vehicle then sends a verifiable credential to the RSU. Therefore, the communication cost is 1 round in the part of the generation of the spatiotemporal profile. The communication cost in the part of the selection of vehicles is 3 rounds. The RSU needs to

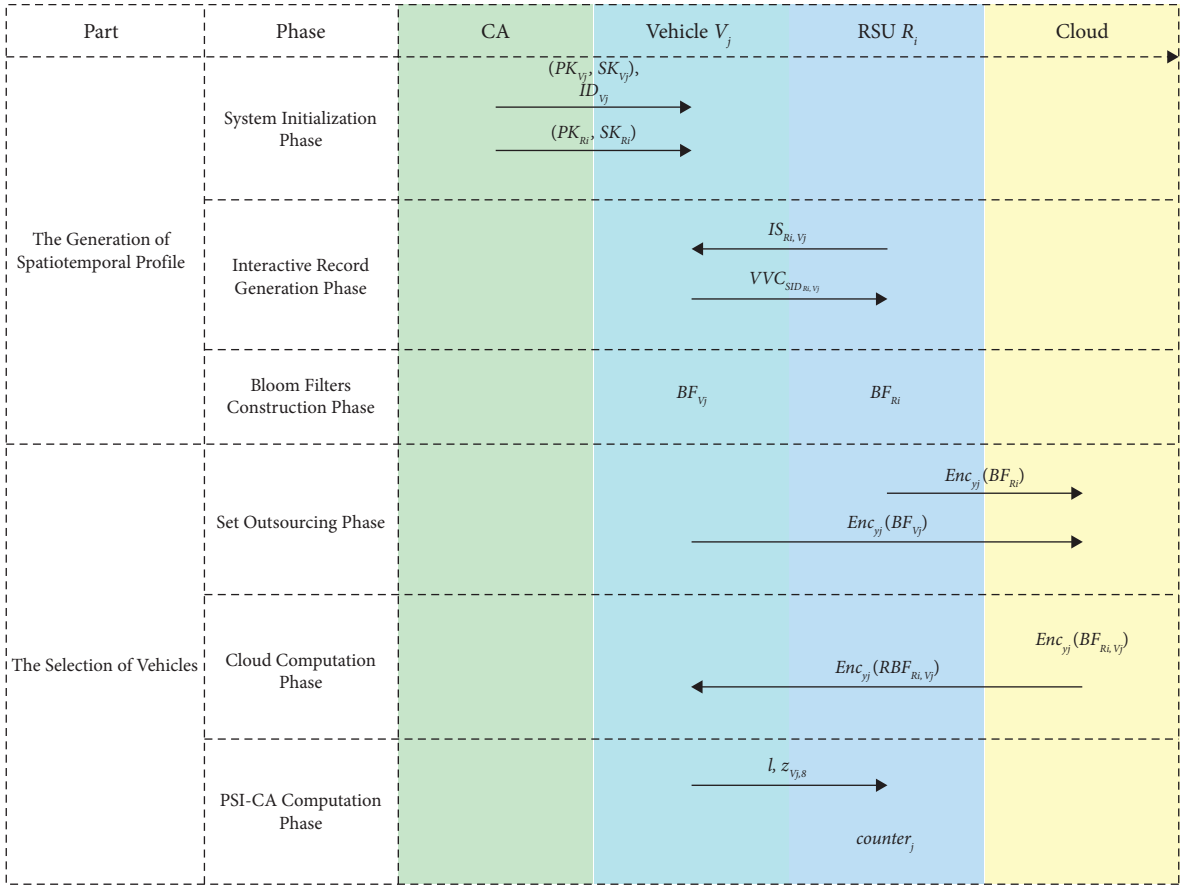

Figure 2: The sequence diagram of the proposed vehicle selection scheme.

communicate with the threshold t vehicles before each time it shares data, so the total communication cost of our scheme is 4 t rounds.

5.3.2. Transmission Cost. Consider the example of communication between a vehicle, a RSU and the cloud. The transmission cost between the vehicle and the RSU is $2(3\log_2 p - 1) + 3w\log_2 p$ bits. The transmission cost between the vehicle and the cloud is $4b\log_2 p$ bits. The transmission cost between RSU and the cloud is $2b\log_2 p$ bits. Therefore, the total transmission cost required for a RSU to share data once is $[2(3\log_2 p - 1) + 3\log_2 p\ (w + 2b)]\ t$ bits.

5.3.3. Computation Cost. Consider the example of communication between a vehicle, a RSU, and the cloud. In the part of the generation of the spatiotemporal profile, the system incurs computational overhead of $2(T_{\text{Sig}} + T_{\text{Ver}}) + (2k + 1)T_{\text{Hash}}$, where T_{Sig} is the time required to calculate a digital signature, T_{Ver} is the time required to verify a digital signature, and T_{Hash} is the time required to compute a hash function. In the selection of vehicles part, the system incurs computational overhead of $2bT_{\text{Enc}} + [2b + 2d(k-1) + w]T_{Mul} + 2(b + w)T_{\text{Exp}} + w$ T_{Dec}, where T_{Enc} is the time required to calculate an exponential ElGamal encryption, T_{Mul} is the time required to calculate an modular multiplication operation, T_{Exp} is the time required to calculate an modular exponential operation, and T_{Dec} is the time required to calculate an exponential ElGamal decryption. Table 2 shows the computation cost incurred by each entity in each phase.

5.4. Experimental Evaluation. We focus our experiment on the implementation and evaluation of Section 4.2, as this part requires high real-time performance and has a lot of time-consuming modular exponential computation. The computer configuration for the experiment is a Dell desktop computer with a 3.10 GHz CPU, 8.00 GB of RAM, and Windows 10 64 bit Home. Taking the communication between a RSU and a vehicle as an example, in the experiment, we observe the influence of different factors on the running time of the scheme. Table 3 shows the relationship between the number of Bloom Filter's bits and running time and the capacity of Bloom Filter when fpp = 0.007 and p is 512 bits.

Table 4 and Figure 3 show the relationship between the number of bits of p and the running time when fpp = 0.007, the capacity of BF is 1000, the set cardinality of RSU is 500, and the set cardinality of the vehicle is 250.

Table 5 shows the computational overhead for several key phase when fpp = 0.007, the capacity of BF is 5000, the set cardinality of RSU is 2500, and the set cardinality of vehicle is 1250. Calculating the data in Table 5 shows that the cloud undertakes about 49.53% of the computational cost.

Table 2: The computation cost incurred by each entity in each phase.

Entities	Vehicle	RSU	Cloud
The computation cost of interactive record generation phase	$T_{\text{Sig}} + T_{\text{Ver}}$	$T_{\text{Hash}} + T_{\text{Sig}} + T_{\text{Ver}}$	
The computation cost of Bloom filter construction phase	kT_{Hash}	kT_{Hash}	
The computation cost of set outsourcing phase	bT_{Enc}	bT_{Enc}	
The computation cost of cloud computation phase			$2b(T_{\text{Mul}} + T_{\text{Exp}})$
The computation cost of PSI-CA computation phase	$2d(k-1)T_{\text{Mul}} + wT_{\text{Exp}}$	$w(T_{\text{Mul}} + T_{\text{Exp}} + T_{\text{Dec}})$	

Table 3: Running time when $fpp = 0.007$ and p is 512 bits.

The capacity of the Bloom filter	The number of bits of the Bloom filter	The set cardinality of RSUs	The set cardinality of the vehicle	Running time
100	1032	50	25	3.5362
500	5163	250	125	17.4769
1000	10327	500	250	35.3105
5000	51637	2500	1250	174.9057
10000	103274	5000	2500	351.6136

Table 4: The relationship between the number of bits of p and the running time.

The number of bits of p	Running time
128	2.8033
256	9.0054
512	34.4608
1024	200.0646

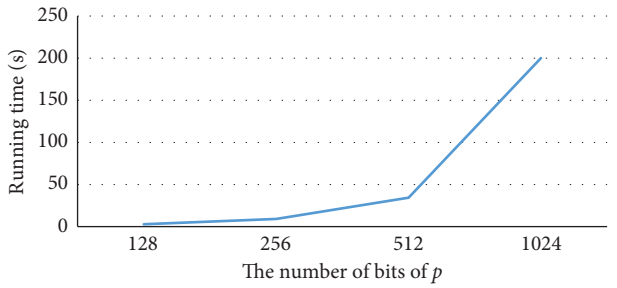

Figure 3: The relationship between the number of bits of p and the running time.

Table 5: The computational cost for several key phase.

Phase	Running time (s)
The vehicle encrypts its Bloom filter	85.1727
The RSU encrypts its Bloom filter	86.2159
Cloud calculates the inner product	0.1271
Cloud randomizes Bloom filter	86.5032
Vehicle and RSU calculate PSI-CA	2.0585
The total time required for the scheme	174.9057

6. Conclusion

Based on the new trust relationship establishment method of spatiotemporal matching and PSI-CA technology, this paper proposes a reliable and objective vehicle selection scheme for data sharing in IoV. The PSI-CA technology used in our scheme is based on Bloom filters and an exponential ElGamal encryption scheme with additive homomorphic.

The use of spatiotemporal matching, a new trust relationship establishing method, ensures the reliability and trustworthiness of data shared in the IoV. Using PSI-CA technology to measure the level of spatiotemporal matching can protect the privacy of the private sets of vehicles and RSUs, so that RSUs can only get intersection cardinality and cannot obtain other private information. Using Bloom filters to implement PSI-CA can hide the size of privacy sets of vehicles and RSUs. In addition, we have delegated some of the computing work to the cloud, reducing the 49.53% computing burden of vehicles and RSUs and improve the efficiency of the scheme.

In the existing vehicle selection schemes for data sharing in IoV, the method of filtering data is based on the reputation mechanism or voting mechanism. These methods are vulnerable to cooperative attack, and the results are not objective. However, our scheme is based on spatiotemporal matching technology, and vehicles cannot influence each other. As a result, our scheme is resistant to cooperative attack. In addition to this, our scheme is resistant to forged interaction record attack, identity forgery attack, and collusion attack and allows RSUs to obtain objective and reliable results.

Our future work will be based on ensuring the objectivity and reliability of shared data and strive to improve the security and efficiency of the scheme, reduce communication and computing costs, and realize an efficient vehicle selection scheme for data sharing in IoV.

Data Availability

The proposed scheme and its analysis need only theoretical and experimental support. There is no additional dataset to be provided in this paper.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Beijing Municipal Natural Science Foundation (M22002, 4212019), National Natural Science Foundation of China (62172005), and Fundamental Research Funds of Communication University of China (CUC22GP006).

References

[1] C. Chengzhe Lai, R. Rongxing Lu, D. Zheng, and X. Xuemin Shen, "Security and privacy challenges in 5G-enabled vehicular networks," *IEEE Network*, vol. 34, no. 2, pp. 37–45, 2020.

[2] C. Lai, Y. Du, Q. Guo, and D. Zheng, "A trust-based privacy-preserving friend matching scheme in social Internet of Vehicles," *Peer-to-Peer Networking and Applications*, vol. 14, no. 4, pp. 2011–2025, 2021.

[3] Z. Wang, Y. Xu, J. Liu et al., "An efficient data sharing scheme for privacy protection based on blockchain and edge intelligence in 6G-VANET," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 5031112, 18 pages, 2022.

[4] Q. Zhou, Z. Zeng, K. Wang, and M. Chen, "Privacy protection scheme for the Internet of vehicles based on private set intersection," *Cryptography*, vol. 6, no. 4, p. 64, 2022.

[5] L. Zhang, W. Gao, S. Chen, W. Ren, K. K. R. Choo, and N. N. Xiong, "A privacy-preserving proximity testing using private set intersection for vehicular ad-hoc networks," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 10, pp. 7373–7383, 2022.

[6] W. Xu, H. Zhou, N. Cheng et al., "Internet of vehicles in big data era," *IEEE/CAA Journal of Automatica Sinica*, vol. 5, no. 1, pp. 19–35, 2018.

[7] Z. Wang, J. Liu, C. Guo, S. Hu, Y. Wang, and X. Yang, "An efficient and secure malicious user detection scheme based on reputation mechanism for mobile crowdsensing VANET," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 5302257, 16 pages, 2021.

[8] J. Kang, R. Yu, X. Huang et al., "Blockchain for secure and efficient data sharing in vehicular edge computing and networks," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 4660–4670, 2019.

[9] L. Xie, Y. Ding, H. Yang, and X. Wang, "Blockchain-based secure and trustworthy Internet of Things in SDN-enabled 5G-VANETs," *IEEE Access*, vol. 7, pp. 56656–56666, 2019.

[10] J. Jingchao Sun, R. Zhang, and Y. Yanchao Zhang, "Privacy-preserving spatiotemporal matching," *2013 Proceedings IEEE INFOCOM. IEEE*, vol. 3, pp. 800–808, 2013.

[11] J. Ma, T. Li, J. Cui, Z. Ying, and J. Cheng, "Attribute-based secure announcement sharing among vehicles using blockchain," *IEEE Internet of Things Journal*, vol. 8, no. 13, pp. 10873–10883, 2021.

[12] D. Danxin Wang, C. Chuanhe Huang, X. Xieyang Shen, and N. Naixue Xiong, "A general location-authentication based secure participant recruitment scheme for vehicular crowdsensing," *Computer Networks*, vol. 171, 2020.

[13] D. Danxin Wang, X. Xianhao Chen, L. Zhang, Y. Yuguang Feng, and C. Chuanhe Huang, "A blockchain-based human-to-infrastructure contact tracing approach for COVID-19," *IEEE Internet of Things Journal*, vol. 9, no. 14, pp. 12836–12847, 2021.

[14] S. K. Debnath, K. Sakurai, K. Dey, and N. Kundu, "Secure outsourced private set intersection with linear complexity," in *Proceedings of the 2021 IEEE Conference on Dependable and Secure Computing (DSC)*, pp. 1–8, IEEE, Aizuwakamatsu, Japan, January 2021.

[15] M. J. Freedman, K. Nissim, and B. Pinkas, "Efficient private matching and set intersection," *Advances in Cryptology-EUROCRYPT 2004: International Conference on the Theory and Applications of Cryptographic Techniques*, Springer Berlin Heidelberg, Interlaken, Switzerlandpp. 1–19, 2004.

[16] A. Tajima, H. Sato, and H. Yamana, "Outsourced private set intersection cardinality with fully homomorphic encryption," in *Proceedings of the 2018 6th International Conference on Multimedia Computing and Systems (ICMCS)*, pp. 1–8, IEEE, Rabat, Morocco, May 2018.

[17] A. A. Jolfaei, H. Mala, M. Zarezadeh, and C. A. Eo-Psi, "Efficient outsourced private set intersection cardinality," *Journal of Information Security and Applications*, vol. 65, 2022.

[18] S. Tarkoma, C. E. Rothenberg, and E. Lagerspetz, "Theory and practice of bloom filters for distributed systems," *IEEE Communications Surveys & Tutorials*, vol. 14, no. 1, pp. 131–155, 2011.

[19] S. Tarkoma, *Overlay Networks: Toward Information Networking*, Auerbach Publications, Boca Raton, FL, USA, 2010.