A Privacy-Preserving and Verifiable Querying Scheme in Vehicular Fog Data Dissemination

Qinglei Kong, *Student Member, IEEE*, Rongxing Lu, *Senior Member, IEEE*, Maode Ma, *Senior Member, IEEE*, and Haiyong Bao

***Abstract*—Vehicular fog has attracted considerable attention recently, as the densely deployed fog devices are in proximity to vehicular end-users, and they are particularly suitable for the latency-sensitive and location-aware vehicular services. In this paper, we propose a secure querying scheme in vehicular fog data dissemination, in which the roadside units (RSUs) act as fog storage devices to cache data at network edge and disseminate data upon querying. To disrupt the association between a specific data request and its origin vehicle, the proposed scheme exploits an invertible matrix to structure multiple data requests from different vehicles, and aggregates the ciphertexts of data requests at the RSU side with the homomorphic Paillier cryptosystem. Meanwhile, given the invertible matrix and decryption result, the RSU can recover each individual data request without identifying its origin vehicle. In addition, the RSU can verify the correctness of the recovered data requests with an identity-based batch verification scheme. Through security analysis, we demonstrate that the proposed scheme can achieve the security goals of unlinkability, confidentiality, and verifiability. Performance evaluations are also conducted, in which the obtained results show that the proposed scheme can be adaptive to the fluctuating number of the data querying vehicles, and significantly reduce the computation complexity and communication overhead.**

***Index Terms*—Vehicular network, data dissemination, unlinkability, verifiability.**

I. Introduction

With the recent development of automotive industry and intelligent transportation systems (ITS), vehicular fog has been envisioned to provide location-aware and latency-sensitive services towards the vehicular end-users [1]. As an important component of vehicular fog, roadside units (RSUs), which are pervasively deployed in different areas, are perfect candidates to act as distributive fog storage nodes. Specifically, RSUs can act as fog storage devices to maintain and disseminate a plethora of traffic-related information at the network edge, which greatly reduces latency and saves transmission bandwidth [2]–[4]. That is, by receiving the uploaded data requests, the fog RSUs disseminate the queried data towards vehicles as a response at the network edge. However, there are still some challenges in aspects of security and privacy preservation in vehicular fog data querying and dissemination.

Since the data requests uploaded by vehicles are tightly associated with the personally identifiable information (PII) of people on-board. If a data request can be uniquely linked to a vehicle, some privacy-sensitive information could be disclosed (such as hospitals or churches), and those information must be strictly prohibited from unauthorized access. Even though provably good privacy in VANETs can be achieved with pseudonyms [5], [6], an adversary can still associate a data request with a particular vehicle, and further infer the driver's trajectory and behavioral pattern through observation. However, when there exist a group of vehicles sending data requests together, it is highly possible that the data request uploaded by one vehicle is indistinguishable from the rest of data requests, i.e., by disrupting the association between one vehicle and its data request.

If a group of data requests are processed separately, it could bring heavy computation and communication costs towards the RSU, especially when there are a large number of data querying vehicles. Thus, the uploaded data requests should be processed aggregately, and the proposed scheme should also be adaptive to the fluctuating number of querying vehicles in the spatio-temporal domain. To achieve privacy-preserving data aggregation, a secure multi-dimensional data aggregation scheme is proposed in [6], in which each dimension of data aggregation can still be recovered after decryption. Specifically, the proposed scheme structures the multi-dimensional data report with a super-increasing sequence, and encrypts the data with the homomorphic Paillier cryptosystem, which achieves the ciphertexts aggregation and complexity reduction. To preserve the location privacy of the participating vehicles, a privacy-preserving route reporting aggregation scheme in VANET is proposed in [7], which utilizes the homomorphic encryption technique to count the number of vehicles in each route segment. However, the above schemes do not support the scenario with high device density; meanwhile, they are only be able to

Manuscript received July 27, 2018; revised November 7, 2018; accepted December 15, 2018. Date of publication December 20, 2018; date of current version February 12, 2019. This work was supported in part by the research grant National Research Foundation of Singapore (NRF2018-SR2001-005), in part by NSERC Discovery Grants (04009), in part by the NBIF Start-Up Grant (Rif 2017-012), in part by the HMF2017 YS-04, and in part by the LMCRF-S-2018-03, in part by the Key Research and Development Program of Science and Technology Department of Zhejiang Province under Grant 2017C01015, and in part by the Natural Science Foundation of Zhejiang Province (LY17F020006). The review of this paper was coordinated by Dr. M. M. Atia. *(Corresponding author: Qinglei Kong.)*

Q. Kong and M. Ma are with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798 (e-mail: kql8904@163.com; emdma@ntu.edu.sg).

R. Lu is with the Faculty of Computer Science, University of New Brunswick, Fredericton, NB E3B 5A3, Canada (e-mail: rlu1@unb.ca).

H. Bao is with the School of Computer Science and Information Engineering, Zhejiang Gongshang University, Hangzhou 310018, China (e-mail: baohy@zjgsu.edu.cn).

Digital Object Identifier 10.1109/TVT.2018.2888854

calculate the data aggregation result, but cannot recover the content of each individual data report.

In the proposed scheme, upon the recovery of each individual data request, the correctness of the recovered requests should also be verified with privacy preservation. A verifiable data sharing scheme is proposed in [8], which exploits an identity-based signature (IBS) technique to achieve the batch verification of the involved entities. Meanwhile, an efficient verifiable data aggregation scheme with the certificateless aggregate signature scheme is also proposed in [9], which verifies the correctness of the data obtained from IoT terminals in batch, and effectively reduces the computation burden introduced to the IoT data center. Since the data request space is limited, the data contained in each individual signature could still possibly be disclosed under the brute-force attack. Therefore, how to verify the correctness of the recovered data requests without disclosing the data request content contained in each individual signature, has become an emerging challenge.

In this paper, to solve the above-mentioned challenges, we propose a privacy-preserving querying and verifiable scheme in vehicular fog data dissemination, which enables the vehicles (end-users) to acquire local data from the RSU (fog storage device). Specifically, the main contributions of the paper are three folds.

- Firstly, the proposed scheme exploits an invertible matrix to structure multiple data requests sent by different vehicles, such that each individual data request can be recovered by the recipient RSU without being associated with its origin. Meanwhile, the proposed scheme owns the advantage of scalability, such that it can be adaptive to the situation with a large number of querying vehicles.
- Secondly, the proposed scheme achieves efficient and privacy-preserving batch verification of the recovered data requests. By exploiting an identity-based signature scheme, the RSU can verify the correctness of the recovered data requests in batch, without learning the content corresponds to each individual signature.
- Thirdly, we demonstrate the security properties of the proposed data querying scheme, in terms of unlinkability, confidentiality and verifiability. Meanwhile, we conduct performance evaluations to show the efficiency of the proposed scheme, in terms of computation complexity and communication overhead. In addition, we demonstrate the feasibility of the proposed scheme in both high-speed and low-speed scenarios.

The rest of this paper is organized as follows. We describe our system model, identify out security requirements, and show our design goals in Section II. We propose our secure vehicular data dissemination scheme in Section IV, followed by security analysis and performance evaluation in Section V and Section VI, respectively. Related work is discussed in VII. Finally, we draw our conclusion in Section VIII.

II. Model, Requirements and Design Goal

In this section, we describe our system model, identify our security requirements, and show our design goals.

A. System Model

In the system model, we consider a data querying paradigm in vehicular fog data dissemination, in which a group of vehicles send data requests towards the covered RSU (fog storage device) together, for both the dynamically moving and static scenarios. After receiving the data requests, the RSU disseminates the queried data towards the querying vehicles as a response. Specifically, the proposed system model consists of three parties: data querying vehicles, an RSU and an information server, as shown in Fig. 1.

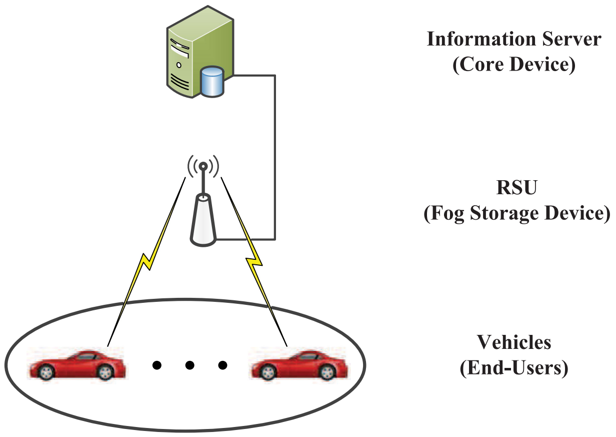


Fig. 1. Vehicular data querying architecture under consideration.

Each data querying vehicle acts as an end-user, which is equipped with adequate processing ability and storage space, so as to perform the data processing and store data item identifiers. Each vehicle generates data requests with data identifiers, receives data responses from RSUs and communicates with other entities, i.e., RSUs and other data querying vehicles.

Each RSU under consideration acts as a fog storage device, which stores a catalog of data items indexed by data identifiers (such as local emergency contacts, local hospitals, public libraries, bank services, and so on). Upon requesting, it distributes the local information towards the vehicles under its coverage. While in high mobility scenario with a large number of querying vehicles, the RSU could also transmit the received data identifiers towards the neighboring RSUs to disseminate cooperatively. In addition, the data items kept in the local storage of the RSU are periodically updated by the backbone information server.

The information server is a trustable central entity, which initializes the whole system and is responsible for the registration of RSUs and vehicles. It also collects the local information and periodically updates the information items stored in RSUs.

Communication Model: The vehicle-to-vehicle (V2V) and the vehicle-to-infrastructure (V2I) communications are realized through the IEEE 802.11p Wireless Access for Vehicular Environment (WAVE) standard [10], which is a short to medium range communication technology operating at 5.9 GHz band. The connection between the RSU and the information server is realized through secure wired link with high bandwidth and low transmission delay.

B. Security Requirements

In our security model, we assume the RSU is honest-but-curious, i.e., it will correctly execute the operations defined in the protocol and disseminate the queried data as responses, but it may try to violate the privacy of a data querying vehicle through associating a specific recovered data request with its origin; meanwhile, we assume neither the RSU nor the data querying vehicles will collude with each other in the proposed scheme. Moreover, we assume there exists an adversary, which may eavesdrop the data transmission and launch some active attacks to threat data integrity. In addition, some free-riding vehicles may submit data requests towards the RSU without registration. Therefore, to preserve the privacy of data querying vehicles, to protect the confidentiality of the uploaded data requests, and to achieve the batch verification, the following security requirements should be satisfied in the proposed scheme.

Unlinkability: At the RSU side, the RSU can only learn the content of the recovered data requests, but it cannot link a specific data request with the origin vehicle. Meanwhile, each involved vehicle should not be able to reveal the data request sent by any other vehicle. In other words, the proposed scheme should not disclose any information about the association between a vehicle and the data request it sent [11]. The unlinkability requirement arises primarily from the need to protect a vehicle's anonymity, i.e., preserving a vehicle's privacy by removing the vehicle's identity from its data request [12].

Confidentiality: In the proposed scheme, the content of the data requests and data dissemination response should be protected from the adversary. Even if the adversary eavesdrops the data transmission, it cannot obtain the data requests and data dissemination. Moreover, the confidentiality requirement should also include that the data dissemination content can only be learnt by vehicles which have already uploaded data requests.

Verifiability: The RSU should be able to verify that the recovered data requests have not been modified during data processing and transmission, and authenticate that each received data request is actually generated and signed by a registered vehicle. Meanwhile, the coordinating vehicle should also be able to verify that the received data request has not been modified during data transmission.

C. Design Goals

Under the aforementioned system model and security requirements, our design goal is to develop a privacy-preserving and verifiable querying scheme in vehicular fog data dissemination. Specifically, the following three objectives should be achieved.

The proposed scheme should achieve the above security requirements: If the proposed scheme does not take the defined security requirements into consideration, the privacy of data querying vehicles could be disclosed, the confidentiality of data requests could be threatened, and the recovered data requests could not be properly verified. Then vehicles may not be willing to join in the data querying process, and the entire vehicular data dissemination system cannot function properly.

The proposed scheme should achieve the goal of scalability: Since the number of data querying vehicles on the parking lot varies in the spatio-temporal domain, the proposed scheme should be adaptive to the fluctuating number of vehicles, especially the situation when there involve a large number of data querying vehicles.

The proposed scheme should achieve high efficiency in terms of computation and computation overhead: Even though we assume the RSU (which acts as a fog storage device) owns relatively high processing capability, the proposed scheme should also take the computation overhead into consideration, especially with a large number of vehicles. Due to scarce and limited transmission bandwidth in VANETs, the proposed scheme should also take the communication overhead into consideration.

III. Preliminaries

In this section, we briefly review the security techniques of Paillier cryptosystem and bilinear maps, which function as the basis of the proposed data dissemination scheme.

A. Paillier Cryptosystem

In our proposed scheme, the Paillier Cryptosystem is the building block [13], which has been widely exploited due to its homomorphic additive properties. Specifically, the Paillier Cryptosystem consists of three algorithms: key generation, encryption and decryption.

- *Key Generation:* Given the security parameter κ, two large prime numbers p_1, q_1 are first chosen, where $|p_1| = |q_1| = \kappa$; meanwhile, the RSA modulus $n = p_1 q_1$ and $\lambda = lcm(p_1 - 1, q_1 - 1)$ are also calculated. After choosing a generator $g \in \mathbb{Z}^*_{n^2}$, a function $L(u) = \frac{u-1}{n}$ is defined, and $\mu = (L(g^\lambda \bmod n^2))^{-1} \bmod n$ is further computed. Then, the public key $pk = (n, g)$, and the corresponding private key $sk = (\lambda, \mu)$ are derived.
- *Encryption:* Given a message $m \in \mathbb{Z}_n$, choose a random number $r \in \mathbb{Z}^*_n$, and the ciphertext can be generated as $c = E(m) = g^m \cdot r^n \bmod n^2$.
- *Decryption:* Given the ciphertext $c \in \mathbb{Z}^*_{n^2}$, the corresponding message can be recovered as $m = D(c) = L(c^\lambda \bmod n^2) \cdot \mu \bmod n$.

B. Bilinear Maps

Let $\mathbb{G}_1$ and $\mathbb{G}_2$ be two cyclic groups with the same prime order q, i.e., $|\mathbb{G}_1| = |\mathbb{G}_2| = q$. A bilinear map $e \colon \mathbb{G}_1 \times \mathbb{G}_1 \to \mathbb{G}_2$ has the following properties:

1) Bilinearity: $\forall P, Q \in \mathbb{G}_1$, and $\forall a, b \in \mathbb{Z}^*_q$, we can derive $e(aP, bQ) = e(P, Q)^{ab}$;
2) Nondegeneracy: There exist $P \in \mathbb{G}_1$, which satisfies the condition that $e(P, Q) \neq 1_{\mathbb{G}_2}$.
3) Computable: $\forall P, Q \in \mathbb{G}_1$, there is an efficient algorithm to compute $e(P, Q)$.

Definition 1: A bilinear parameter generator $\mathcal{G}en$ denotes a probabilistic algorithm that takes a parameter κ_1 as input, and outputs a 5-tuple $(q, P, \mathbb{G}_1, \mathbb{G}_2, e)$ as the output, where q is κ_1-bit prime, $\mathbb{G}_1$ and $\mathbb{G}_2$ are two cyclic groups with order q, $P \in \mathbb{G}$ is a generator, and $e \colon \mathbb{G}_1 \times \mathbb{G}_1 \to \mathbb{G}_2$ is non-degenerated and computable bilinear map.

IV. Proposed Privacy-Preserving Data Dissemination Scheme

In this section, we propose a privacy-preserving and verifiable data querying scheme in vehicular fog data dissemination, which consists of five parts: system initialization, querying group

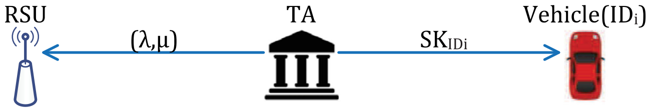


Fig. 2. Key distribution.

formulation, data requests generation, data requests aggregation and recovery, and data requests verification.

A. System Initialization

For the vehicular fog data querying and dissemination system under consideration, the information server, which also acts as the trusted authority (TA), will bootstrap the entire system. Given a security parameter κ, the trusted authority (TA) selects two large prime numbers p_1 and q_1, where $|p_1| = |q_1| = \kappa$. Then the TA generates the Paillier Cryptosystem's public key $(n = p_1 q_1, g)$, and the corresponding private key (λ, μ). The TA also selects a large prime number α with $|\alpha| = \kappa$, and chooses a constant number d with $|d| < \kappa$.

Meanwhile, given another security parameter κ_1, the TA generates the bilinear parameters $(q, P, \mathbb{G}_1, \mathbb{G}_2, e(\cdot,\cdot))$ by running $\mathcal{G}en(\kappa_1)$, and calculates the value of $h = e(P, P)$. The TA selects a random number $x \in \mathbb{Z}_q^*$ as the secret key, and computes the public key $P_0 = x \cdot P$. Moreover, the TA selects a secure encryption algorithm $\mathcal{E}nc(\cdot)$, and chooses two secure cryptographic hash functions $H(\cdot), H_1(\cdot)$, where $H : \{0,1\}^* \rightarrow \mathbb{Z}_n^*, H_1 : \{0,1\}^* \rightarrow \mathbb{Z}_q^*$. In addition, the TA publishes the system parameter as $params = \{n, g, \alpha, d, q, P, \mathbb{G}_1, \mathbb{G}_2, e(\cdot,\cdot), h, P_0, \mathcal{E}nc(\cdot), H(\cdot), H_1(\cdot), \}$.

As shown in Fig. 2, during the registration of the RSU, the TA delivers the private key (λ, μ) towards it. Meanwhile, during the registration of a vehicle with identity $ID_i \in \{0,1\}^*$, the TA computes the secret key $SK_{ID_i} = \frac{1}{H_1(ID_i)+x} \cdot P$, and securely distributes the secret key SK_{ID_i} towards it.

B. Querying Group Formulation

When a vehicle V_a (end-user) intends to query the covering RSU (fog storage device) for local information, it acts as a coordinating vehicle. Different from the most of the ad hoc networks, in which entities are deployed fixedly (such as wireless sensor networks), the network topologies in VANETs are constantly changing, and the querying group needs to be formulated for each individual data querying task. V_a first broadcasts a collaboration request towards all the neighboring vehicles under the coverage of an identical RSU to send data requests together, so as to prevent itself from being associated with its own data request. After receiving the collaboration request, if one vehicle is interested in joining in the collaborative data querying, it accepts the collaboration request, and establishes a secure session key with V_a. We assume the vehicles will voluntarily collaborate with V_a to participate in the data querying process, as the vehicles can also obtain a piece of local information. If all the vehicles are reluctant to join in the process, an incentive mechanism can be exploited to stimulate the participation [14]. Thus, a collaborative data querying group with k vehicles can be formulated.

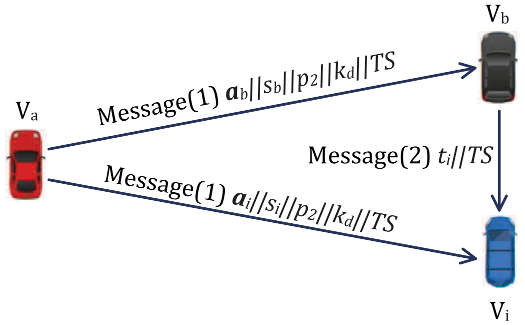

V_b
V_a
Message(1) $\boldsymbol{a}_b||s_b||p_2||k_d||TS$
Message(2) $t_i||TS$
Message(1) $\boldsymbol{a}_i||s_i||p_2||k_d||TS$
V_i

Fig. 3. Secret distribution among vehicles.

V_a selects a large prime p_2 (which satisfies the condition that $k \cdot p_2 < d$), and a random invertible matrix $\mathbf{A}$ with the size of $k \times k$, in which each element $a_{i,j} \in GF(p_2), i = 1, \ldots, k, j = 1, \ldots, k$. V_a also selects a session key k_d for the disseminated data protection, and performs the secret distribution process, which is illustrated in Fig. 3. V_a assigns a row vector $\mathbf{a_i}$ $(i \neq a)$ towards a collaborating vehicle V_i, and selects a random number $s_i \in \mathbb{Z}_n^*$ for V_i. In the proposed scheme, one row vector can only be assigned towards one vehicle. Meanwhile, $\mathbf{a_i}$ only denotes the row vector assigned to V_i, and it may or may not be the ith row vector of $\mathbf{A}$. Then V_a securely delivers $\mathbf{a_i}||s_i||p_2||k_d||TS$ towards V_i, as shown in Fig. 3 (*Message (1)*), where TS is the current timestamp.

V_a also randomly selects a subcoordinating vehicle within the collaborative data querying group, which is denoted as V_b, and acknowledges the rest of vehicles to establish a secure session key with V_b. During the secure key establishment, V_b chooses a random number $t_i \in \mathbb{Z}_n^*$ $(i \neq a, b)$ for each vehicle V_i, and securely delivers $t_i||TS$ towards V_i, as shown in Fig. 3 (*Message (2)*).

C. Data Request Generation

1) Data Request Generation of V_i: for each vehicle V_i $(i \neq a)$, it first chooses the data request identifier Q_i ($|Q_i| = l_q$ bits) from the predefined data request space, in which one data request identifier corresponds to one queried data item. Then V_i performs the following steps to generate a data request.

Step-1: Since the data request space is limited and the length of each data request identifier is short, V_i selects a random number x_i with $|x_i| = l_x$ bits and $l_x + l_q < |p_2|$, and concatenates x_i with Q_i to generate $m_i = x_i||Q_i$. V_i multiplies m_i with each element in the received row vector $\mathbf{a_i} = (a_{i,1}, a_{i,2}, \ldots, a_{i,k})$, and generates the row vector $\mathbf{d_i} = (d_{i,1}, d_{i,2}, \ldots, d_{i,k})$, where $d_{i,j} = m_i \cdot a_{i,j} \bmod p_2, j = 1, \ldots, k$.

Step-2: V_i chooses a random number $r_{i,1} \in \mathbb{Z}_n^*$, and generates the ciphertext with the public key of the RSU (n, g)

$$C_{i,1} = g^{d_{i,1}+\alpha \cdot d_{i,2}} \cdot r_{i,1}^n \cdot H(TS)^{s_i} \bmod n^2, \tag{1}$$

which prevents the RSU from directly recovering the value of $d_{i,1} + \alpha \cdot d_{i,2}$. Then V_i generates the message authentication code $MAC_{i,1} = H(C_{i,1}||TS)$ corresponds to $C_{i,1}$. Meanwhile, V_i $(i \neq a, b)$

generates a signature $\sigma_i = (\sigma_{i,1}, \sigma_{i,2})$ of m_i by choosing a random number $r_{i,2} \in \mathbb{Z}_q^*$, and calculates

$$\begin{cases}\sigma_{i,1} = h^{r_{i,2}}, \\ \sigma_{i,2} = \frac{H_1(m_i||TS)+H_1(s_i||TS)+H_1(t_i||TS)+r_{i,2}}{x+H_1(ID_i)} \cdot P\end{cases} \tag{2}$$

For V_b, it also selects a random number $r_{b,2} \in \mathbb{Z}_q^*$, and generates the signature $\sigma_b = (\sigma_{b,1}, \sigma_{b,2})$, which is

$$\begin{cases}\sigma_{b,1} = h^{r_{b,2}}, \\ \sigma_{b,2} = \\ \frac{H_1(m_b||TS)+H_1(s_b||TS)-\sum_{i\neq a,b} H_1(t_i||TS)+r_{b,2}}{x+H_1(ID_b)} \cdot P\end{cases} \tag{3}$$

Finally, V_i $(i \neq a)$ formulates the message $ID_i||C_{i,1}||MAC_{i,1}||\sigma_i||TS$, and delivers it towards the RSU, which is shown in Fig. 4 (*Message (1)*).

Step-3: For V_i $(i \neq a, b)$, it calculates the value $e_{i,j} = d_{i,j} + H_1(t_i||j||TS) \bmod p_2, j = 3, \ldots, k$, and generates a row vector $\vec{e_i} = (e_{i,3}, \ldots, e_{i,k})$. Moreover, V_i encrypts $\vec{e_i}$ with s_i to generate the ciphertext $C_{i,2}$,

$$C_{i,2} = \mathcal{E}nc_{H(s_i||TS)}(e_{i,3}||\cdots||e_{i,k}), \tag{4}$$

and transmits $C_{i,2}$ towards V_a, which prevents V_a from recovering the value of $d_{i,j}$. To protect data integrity, V_i generates the message authentication code of $C_{i,2}$, which is

$$MAC_{i,2} = H(C_{i,2}||TS) \tag{5}$$

and transmits $C_{i,2}||MAC_{i,2}||TS$ towards V_a, as shown in Fig. 4 (*Message (2)*).

For V_b, it also calculates the value $e_{b,j} = d_{b,j} - \sum_{i=1,i\neq a,b}^{k} H_1(t_i||j||TS) \bmod p_2, j = 3, \ldots, k$ and generates a vector $\vec{e_b} = (e_{b,3}, \ldots, e_{b,k})$. Then it generates the ciphertext $C_{b,2}$ with s_b, which is

$$C_{b,2} = \mathcal{E}nc_{H(s_b||TS)}(e_{b,3}||\ldots||e_{b,k}) \tag{6}$$

V_b also generates the message authentication code of $C_{b,2}$, which is

$$MAC_{b,2} = H(C_{b,2}||TS), \tag{7}$$

and transmits $C_{b,2}||MAC_{b,2}||TS$ towards V_a, as shown in Fig. 4 (*Message (2)*).

2) Data Request Generation of V_a: for V_a, it formulates the data query $m_a = x_a||Q_a$ by selecting a random number x_a with $|x_a| = l_x$ bits, and generates a row vector $\mathbf{d_a} = (d_{a,1}, d_{a,2}, \ldots, d_{a,k})$ with Step-1 defined in Section IV-C1, and performs the following steps.

Step-1: V_a selects a random number $r_{a,1} \in \mathbb{Z}_n^*$, and generates the ciphertext

$$C_{a,1} = g^{d_{a,1}+\alpha\cdot d_{a,2}} \cdot r_{a,1}^n \cdot H(TS)^{s_a} \bmod n^2. \tag{8}$$

where $s_a = -\sum_{i\neq a} s_i \bmod n$. Meanwhile, V_a also chooses a random number $r_{a,2} \in \mathbb{Z}_q^*$, and generates a signature $\sigma_a = (\sigma_{a,1}, \sigma_{a,2})$ corresponding to m_a, which is

$$\begin{cases}\sigma_{a,1} = h^{r_{a,2}}, \\ \sigma_{a,2} = \frac{H_1(m_a||TS)-\sum_{i\neq a} H_1(s_i||TS)+r_{a,2}}{x+H_1(ID_a)} \cdot P.\end{cases} \tag{9}$$

Step-2: After receiving $C_{i,2}$ $(i \neq a)$ from all the collaborating vehicles, V_a checks the correctness of each received message authentication codes, i.e., verifing whether $MAC_{i,2} \overset{?}{=} H(C_{i,2}||TS)$. If the message authentication codes are verified to be correct, V_a decrypts each ciphertext $C_{i,2}$ $(i \neq a)$ with $H(s_i||TS)$, and obtains the value of $e_{i,j}, j = 3, \ldots, k$. Then V_a aggregates all the recovered $e_{i,j}$ to obtain the value $\sum_{i=1,i\neq a}^{k} e_{i,j} = \sum_{i=1,i\neq a}^{k} d_{i,j}, j = 3, \ldots, k$, and further updates the matrix elements $a_{a,j}, j = 3, \ldots, k$, which is

$$a_{a,j} = \left(-\sum_{i=1,i\neq a}^{k} d_{i,j}\right) \cdot {m_a}^{-1} \bmod p_2. \tag{10}$$

Step-3: V_a encrypts the session key k_d (for data dissemination protection) with the public key (n, g), which is denoted as $C_k = g^{k_d} \cdot r_k^n \bmod n^2$ and $r_k \in \mathbb{Z}_n^*$ is a randomly chosen number. Then V_a generates the corresponding message authentication code, which is $MAC_a = H(C_{a,1}||C_k||\mathbf{A}||TS)$. Finally, V_a transmits the message $ID_a||C_{a,1}||\sigma_a||\mathbf{A}||C_k||MAC_a||TS$ towards the RSU, as shown in Fig. 4 (*Message (3)*).

D. Data Requests Aggregation and Reading

After receiving $ID_a||C_{a,1}||\sigma_a||\mathbf{A}||C_k||MAC_a||TS$, the RSU first checks the correctness of MAC_a by computing $MAC_a \overset{?}{=} H(C_{a,1}||C_k||\mathbf{A}||TS)$. Meanwhile, the RSU verifies the correctness of each $MAC_{i,1}(i \neq a)$ by computing $MAC_{i,1} \overset{?}{=} H(C_{i,1}||TS)$. If all the received message authentication codes are verified to be correct, the RSU aggregates all the encrypted data requests $C_{i,1}, i = 1, \ldots, k$ and obtains the data aggregation result C_1, which is

$$\begin{aligned} C_1 &= \prod_{i=1}^{k} C_{i,1} \bmod n^2 \\ &= \prod_{i=1}^{k} g^{d_{i,1}+\alpha\cdot d_{i,2}} \cdot r_{i,1}^n \cdot H(TS)^{s_i} \bmod n^2 \\ &= g^{\sum_{i=1}^{k} d_{i,1}+\alpha\cdot d_{i,2}} \cdot \left(\prod_{i=1}^{k} r_{i,1}\right)^n \cdot H(TS)^{\sum_{i=1}^{k} s_i} \bmod n^2 \\ &= g^{\sum_{i=1}^{k} d_{i,1}+\alpha\cdot d_{i,2}} \cdot \left(\prod_{i=1}^{k} r_{i,1}\right)^n \bmod n^2 \end{aligned} \tag{11}$$

By taking $M = \sum_{i=1}^{k} d_{i,1} + \alpha \cdot d_{i,2}$ and $R = \prod_{i=1}^{k} r_{i,1}$, the aggregated ciphertext $C = g^M \cdot R^n \bmod n^2$ is still a ciphertext of the Paillier Cryptosystem, and the RSU uses the private key (λ, μ) to recover $M = \sum_{i=1}^{k} d_{i,1} + \alpha \cdot d_{i,2}$. Then the RSU

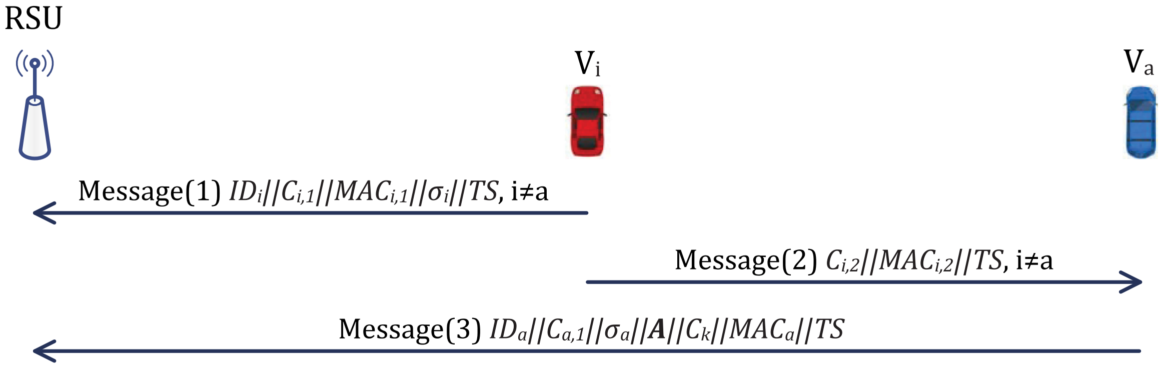

RSU
Vi
Va
Message(1) IDi//Ci,1//MACi,1//σi//TS, i≠a
Message(2) Ci,2//MACi,2//TS, i≠a
Message(3) IDa//Ca,1//σa//A//Ck//MACa//TS

Fig. 4. Vehicular data querying message flow.

generates a row vector $\mathbf{y} = (y_1, y_2, \ldots, y_k)$, which is

$$\begin{cases} y_1 = (M \bmod \alpha) \bmod p_2 = (\sum_{i=1}^{k} d_{i,1}) \bmod p_2, \\ y_2 = (\frac{M - y_1}{\alpha}) \bmod p_2 = (\sum_{i=1}^{k} d_{i,2}) \bmod p_2. \end{cases} \tag{12}$$

Meanwhile, the RSU sets $y_3 = \cdots = y_k = 0$, computes the inverse of $\mathbf{A}$ in the field $GF(p_2)$ (denoted as $\mathbf{A}^{-1}$), and calculates $\mathbf{y} \cdot \mathbf{A}^{-1}$ to recover the data request vector $\mathbf{m} = (m_1, \ldots, m_k)$, without disclosing the actual association between each vehicle and its data request. Furthermore, the RSU abandons the first l_x bits in each data request m_i to obtain the data request identifiers $(Q_1, Q_2, \ldots, Q_k)$.

E. Data Requests Verification

To verify the correctness of the recovered data request vector $\mathbf{m} = (m_1, \ldots, m_k)$, the RSU verifies the correctness of the received signatures $(\sigma_1, \ldots, \sigma_k)$, by performing the following batch verification process.

$$\begin{aligned} & h^{\sum_{i=1}^{k} H_1(m_i||TS)} \cdot \prod_{i=1}^{k} \sigma_{i,1} \\ & = e(P,P)^{\sum_{i=1}^{k} H_1(m_i||TS) + r_{i,2}} \\ & = \prod_{i=1}^{k} e\left(P, \frac{H_1(m_i||TS) + r_{i,2}}{x + H_1(ID_i)} \cdot (x + H_1(ID_i))\right) \\ & = \prod_{i=1}^{k} e(x \cdot P, \sigma_{i,2}) \cdot \prod_{i=1}^{k} e(P, H_1(ID_i) \cdot \sigma_{i,2}) \\ & = e\left(P_0, \sum_{i=1}^{k} \sigma_{i,2}\right) \cdot e\left(P, \sum_{i=1}^{k} H_1(ID_i) \cdot \sigma_{i,2}\right) \end{aligned} \tag{13}$$

If Eq. (13) verifies to be correct, the recovered data requests $\mathbf{m} = (m_1, \ldots, m_k)$ can be verified to be correctly processed and transmitted. Meanwhile, the RSU decrypts the ciphertext C_k with its private key (λ, μ) and derives the session key k_d, which is shared among the collaborative vehicles for the data transmission protection. Then the RSU identifies the data dissemination content (which is denoted as m) queried by vehicles, formulates a ciphertext $C_m = \mathcal{E}nc_{H(k_d||TS)}(m)$, and generates the corresponding message authentication code $MAC_m = H(C_m||TS)$. Moreover, the RSU broadcasts the data query response $C_m||MAC_m||TS$ towards all the data querying vehicles. For the high speed scenario, the RSU can also deliver the data item identifiers towards the neighboring RSUs for collaborative data dissemination. After receiving $C_m||MAC_m||TS$, each vehicle verifies the correctness of MAC_m by checking whether $MAC_m \overset{?}{=} H(C_m||TS)$. If it verifies to be correct, each vehicle recovers the value of m with the session key k_d. Based on the data query response m, each vehicle can identify the information queried by itself.

Correctness: To demonstrate the feasibility of the proposed scheme, now we prove that the matrix updated by V_a is still an invertible matrix. Let $\mathbf{B}$ be the inverse of the matrix $\mathbf{A}$ in the field $GF(p_2)$, then matrix $\mathbf{A}$ and matrix $\mathbf{B}$ satisfy the following equation:

$$\begin{bmatrix} a_{1,1} & \ldots & a_{1,k} \\ \vdots & \ddots & \vdots \\ a_{k,1} & \ldots & a_{k,k} \end{bmatrix} \begin{bmatrix} b_{1,1} & \ldots & b_{1,k} \\ \vdots & \ddots & \vdots \\ b_{k,1} & \ldots & b_{k,k} \end{bmatrix} = \begin{bmatrix} 1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & 1 \end{bmatrix} \tag{14}$$

Suppose V_a assigns the xth row vector $\mathbf{a_x}$ to itself. Then we calculate $e_j = -a_{x,j} - (\sum_{i=1, i \neq a}^{k} m_i a_{i,j}) \cdot m_a^{-1} \bmod p_2, j = 3, 4, \ldots, k$, and update $a_{x,j}$ with $a_{x,j} + e_j \bmod p_2$, which is

$$\mathbf{A}^{'} = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \cdots & a_{1,x} & \cdots & a_{1,k} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{x,1} & a_{x,2} & a_{x,3} + e_3 & \cdots & a_{x,x} + e_x & \cdots & a_{a,k} + e_k \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{k,1} & a_{k,2} & a_{k,3} & \cdots & a_{k,x} & \cdots & a_{k,k} \end{bmatrix} \tag{15}$$

Then we multiply the updated matrix $\mathbf{A}^{'}$ with $\mathbf{B}$, and obtains

$$\mathbf{A}^{'} \times \mathbf{B} = \begin{bmatrix} 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \sum_{i=3}^{k} e_i b_{i,1} & \cdots & \sum_{i=3}^{k} e_i b_{i,x} + 1 & \cdots & \sum_{i=3}^{k} e_i b_{i,k} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 1 \end{bmatrix} \tag{16}$$

As the product $\mathbf{A}' \times \mathbf{B}$ is an invertible matrix, $\mathbf{A}'$ is also an invertible matrix.

V. Security Analysis

In this section, we discuss the security properties of the proposed data querying scheme in vehicular fog data dissemination. In particular, following the security requirements discussed earlier, our analysis will focus on how the proposed scheme can prevent data requests from being associated with their origin vehicles, protect the confidentiality of data transmission, and achieve the verifiability of recovered data requests.

- *The individual vehicle's data request is unlinkable in the proposed scheme:* Since $C_{i,1}$ is formulated as $g^{d_{i,1}+\alpha\cdot d_{i,2}} \cdot r_i^n \cdot H(TS)^{s_i} \bmod n^2$ and we do not take the collusion attack into consideration, the RSU cannot directly recover the value of $d_{i,1} + \alpha \cdot d_{i,2}$ without learning the value of s_i. At the RSU side, by aggregating $C_1 = \prod_{i=1}^{k} C_{i,1} = g^{\sum_{i=1}^{k} d_{i,1}+\alpha\cdot d_{i,2}} \cdot (\prod_{i=1}^{k} r_{i,1})^n \bmod n^2$ and decrypting C_1, the RSU can derive the values of $y_1 = \sum_{i=1}^{k} d_{i,1} \bmod p_2$ and $y_2 = \sum_{i=1}^{k} d_{i,2} \bmod p_2$. Then the RSU calculates $\mathbf{m} = \mathbf{y} \cdot \mathbf{A}^{-1}$ to recover the uploaded data requests $\mathbf{m} = (m_1, \ldots, m_k)$. Since the RSU cannot achieve the information about which row vector in $\mathbf{A}$ corresponds to which data querying vehicle and we do not take the collusion attack into consideration, given the recovered data requests, it is impossible for the RSU to associate a specific data request with its origin vehicle. In addition, given σ_i and m_i, it is impossible to determine whether the data query m_i is contained in the signature σ_i by directly calculating $e(\sigma_{i,2}, P_0 + H_1(ID_i) \cdot P) \overset{?}{=} \sigma_{i,1} \cdot h^{H_1(m_i||TS)}$. This is because σ_i is protected with $H_1(s_i||TS)$ and $H_1(t_i||TS)$, which are secrets shared with V_a and V_b respectively. Thus, the security goal of unlinkability can be achieved.
- *The confidentiality of the data transmission can be achieved in the proposed scheme:* In the proposed scheme, the ciphertext of V_i's data request m_i is formulated as $(C_{i,1}, C_{i,2})$; meanwhile, since the Paillier Cryptosystem is proved to be semantic secure against the chosen plaintext attack, the data request m_i contained in $C_{i,1}$ is also semantically secure and privacy-preserving. For the data request contained in $C_{i,2}$, it is encrypted with s_i, which is a secret value shared between V_a and V_i, and we do not take the collusion attack into consideration. As a result, the adversary cannot recover the value of the data request m_i even if it eavesdrops the ciphertext $(C_{i,1}, C_{i,2})$. For V_a, after decrypting $C_{i,2}$ with $H(s_i||TS)$, it can obtain the value of $e_{i,j} = d_{i,j} + H(t_i||j||TS), j = 3, \ldots, k$. However, V_a still cannot recover the individual value of $d_{i,j}$, this is because t_i is a secret value shared between V_i and the coordinating vehicle V_b, and we do not take the collusion attack into consideration. In addition, the responded data dissemination m from the RSU is encrypted by a session key k_d, which is a session key shared among the RSU and the data querying vehicles, and the ciphertext C_m can only be decrypted by the data querying vehicles. Even if the adversary eavesdrops C_m, it still cannot learn the value of the disseminated data. Thus, the security goal of confidentiality can be achieved in the proposed vehicular fog data querying and dissemination scheme.
- *The verifiability of recovered data requests and the data integrity of data transmission can be achieved in the proposed scheme:* In the proposed scheme, the signature σ_i can be considered as a commitment of the data request m_i, i.e., σ_i preserves the correctness of m_i while hiding the real value of m_i. Meanwhile, the data request m_i is signed by the identity-based signature (IBS) scheme proposed in [15], which supports batch verification. Since the security of the IBS scheme depends on the k-collision attack algorithm (k-CAA) complexity assumption, both the source authentication of the data querying vehicle V_i and the correctness verification of recovered data request m_i can be achieved. In addition, $MAC_{i,1}$ and $MAC_{i,2}$ are also generated to guarantee data integrity of $C_{i,1}$ and $C_{i,2}$ by checking the correctness of $MAC_{i,l} \overset{?}{=} H(C_{i,l}||TS), l = 1, 2$. Thus, the security goals of verifiability and data integrity can be achieved.

Based on the above security analysis, we can conclude that the proposed data querying scheme in vehicular fog data dissemination is secure, which can successfully achieve our security goals.

VI. Performance Evaluation

In this section, we evaluate the performance of the proposed secure vehicular fog data querying and dissemination scheme, which consists of four parts, parameter setup, computational complexity, communication overhead, and the impact of the proposed scheme brought by velocity.

A. Parameter Setup

In performance evaluation, we choose $|n| = 1024$ bits for the Paillier's cryptosystem, and the size of the ciphertext in $\mathbb{Z}_{n^2}^*$ is $|n^2| = 2048$ bits. The length of the prime number α is set to be 512 bits, and the maximum allowable length of the value d is 511 bits. In addition, we set the size of the prime number p_2 (the length of each element in the matrix) to be 384 bits, and the maximum allowable number of vehicles is 2^{127}, which greatly excesses the scale of the vehicles under the coverage of one RSU.

For comparison, we consider a traditional approach, which can also achieve the security requirements of unlinkability, confidentiality, and verifiability. Specifically, for each vehicle V_i in the traditional scheme, it structures the data query m_i with the row vector $\mathbf{a}_i = (a_{i,1}, a_{i,2}, \ldots, a_{i,k})$ to generate the row vector $\mathbf{d}_i = (d_{i,1}, d_{i,2}, \ldots, d_{i,k}) = (a_{i,1}m_i, a_{i,2}m_i, \ldots, a_{i,k}m_i)$. Then V_i generates the ciphertexts of all the elements in $\mathbf{d}_i$ with the public key of the RSU (n, g), such that the ciphertexts $C_{i,j} = g^{d_{i,j}} \cdot r_{i,1}^n \cdot H(TS)^{s_i} \bmod n^2$ $(j = 1, 2, \ldots, k)$ can be obtained; meanwhile, V_i also calculates the signature σ_i, which corresponds to m_i, with Eq. (2), and generates the message authentication code $MAC_i = H(C_{i,1}|| \cdots ||C_{i,k}||TS)$. In addition, V_i formulates and transmits the message $C_{i,1}|| \cdots ||C_{i,k}||\sigma_i||MAC_i||TS$ towards the RSU.

B. Computation Complexity

In the proposed vehicular fog data querying and dissemination scheme, V_i $(i \neq a)$ needs to perform three exponentiation operations in $\mathbb{Z}^*_{n^2}$ to generate the ciphertext $C_{i,1}$, which is a ciphertext of the Paillier cryptosystem, and it takes one symmetric encryption operation to generate $C_{i,2}$. To generate the corresponding verifiable signature $\sigma_i = (\sigma_{i,1}, \sigma_{i,2})$, V_i takes one exponentiation operation in $\mathbb{G}_2$ to generate $\sigma_{i,1}$, and it spends one multiplication operation in $\mathbb{G}_1$ to generate $\sigma_{i,2}$.

For V_a, after receiving $C_{i,2}$ from the collaborative vehicles, V_a spends $k-1$ symmetric decryption operations. Meanwhile, V_a takes three exponentiation operations in $\mathbb{Z}^*_{n^2}$ to generate the Paillier cryptosystem's ciphertext $C_{a,1}$, and it spends two exponentiation operations in $\mathbb{Z}^*_{n^2}$ to generate C_k. In order to generate the verifiable signature $\sigma_a = (\sigma_{a,1}, \sigma_{a,2})$, V_a consumes one exponentiation operation in $\mathbb{G}_2$ and one multiplication operation in $\mathbb{G}_1$. In comparison with the computational cost of an exponentiation operation in $\mathbb{G}_2$ or a multiplication operation in $\mathbb{G}_1$, the computational cost of the symmetric encryption or decryption operation is negligible. Denote the computation costs of an exponentiation operation in $\mathbb{Z}^*_{n^2}$, an exponentiation operation in $\mathbb{G}_2$ and a multiplication operation in $\mathbb{G}_1$ as C_{pe}, C_{be} and C_{bm}, the total computational costs of V_i and V_a will be $Comp_1 = 3 * C_{pe} + C_{be} + C_{bm}$ and $Comp_2 = 5 * C_{pe} + C_{be} + C_{bm}$ respectively.

For the RSU, the proposed scheme enables the RSU to aggregate the data requests submitted by multiple vehicles into one compressed data request, and it takes one exponentiation operation in $\mathbb{Z}^*_{n^2}$ to decrypt $C = \prod_{i=1}^{k} C_{i,1} \bmod n^2$, which largely reduces the computational complexity of the RSU. Meanwhile, in comparison with the computational cost of an exponentiation operation in $\mathbb{Z}^*_{n^2}$, the computational cost of a multiplication operation in $\mathbb{Z}^*_{n^2}$ is negligible. After decryption, the RSU takes two bilinear pairing operations in $\mathbb{G}_1$, one exponentiation operations in $\mathbb{G}_2$ and k multiplication operations in $\mathbb{G}_1$ to verify the correctness of recovered data requests. We denote the bilinear pairing operation in $\mathbb{G}_1$ as C_{bp}, and the computational cost of the RSU is $Comp_3 = C_{pe} + 2 * C_{bp} + C_{be} + k * C_{bm}$.

Under the traditional scheme, each vehicle needs to take $3 * k$ exponentiation operations in $\mathbb{Z}^*_{n^2}$ to generate the ciphertexts $C_{i,1,l}, l = 1, \ldots, k$. After receiving the encrypted data requests, the RSU aggregates the data requests and spends k exponentiation operations in $\mathbb{Z}^*_{n^2}$ for decryption. Thus, the total computational cost of each vehicle is $3 * k * C_{pe} + C_{be} + C_{bm}$, and the computational cost of the RSU is $k * C_{pe} + 2 * C_{bp} + C_{be} + k * C_{bm}$.

We compare the computation complexity of the proposed data dissemination scheme with the traditional scheme. To test performance of the exploited Paillier cryptosystem, we conduct the experiments with the Java Paillier Library [16]. Meanwhile, to examine the performance of the exploited identity-based signature scheme, the Type-A pairing from the JPBC Library [17] is introduced. The experiments is performed on a desktop with 3.40 GHz processor and 8.00GB memory. The experimental results indicate that the cost on a single exponentiation operation in $\mathbb{Z}^*_{n^2}$ $(|n^2| = 2048)$ is $C_{pe} = 8.59$ ms, a single exponentiation operation in $\mathbb{G}_2$ is $C_{be} = 0.75$ ms, a single bilinear pairing operation in $\mathbb{G}_1$ is $C_{bp} = 5.86$ ms, and a single multiplication operation in $\mathbb{G}_1$ is $C_{bm} = 9.20$ ms. Based on the derived operation costs, we depict the computation costs of the RSU and each data querying vehicle in Fig. 5(a) and Fig. 5(b), respectively. Then we compare the proposed scheme with the traditional scheme, when the number of the data querying vehicles k varies between 5 to 100. As shown in Fig. 5, the proposed scheme outperforms the traditional scheme in terms of the computational cost of the RSU and each data querying vehicle respectively. Thus, we can conclude that the proposed vehicular data querying scheme largely reduces the computational delay of the involved entities.

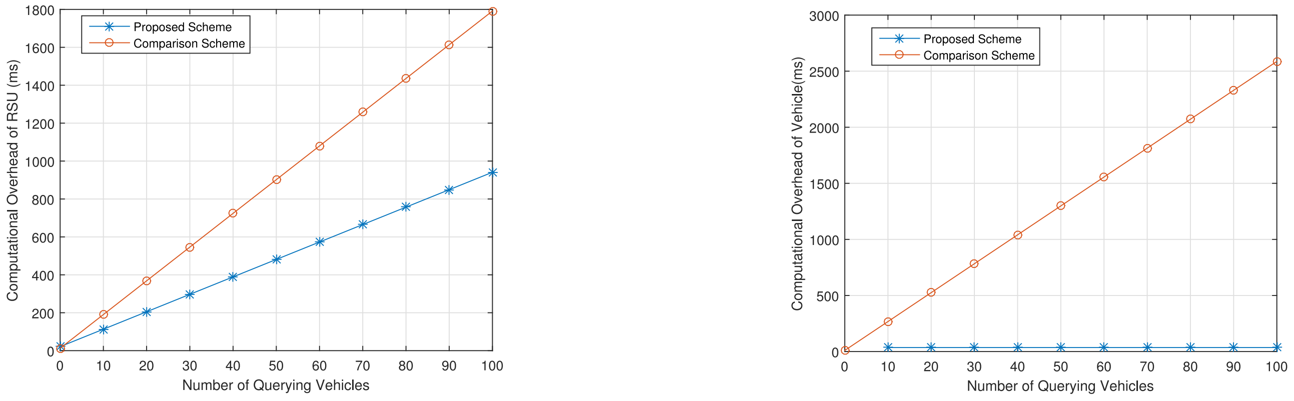
Note that one of the limitations of this algorithm is the impact brought by the matrix inverse computation. To achieve the matrix inverse computation with the Gaussian elimination method, the computational complexity is with time complexity $\mathcal{O}(n^3)$, and it is not particularly suitable for the scenario with a large number of data querying vehicles. The solution to this scenario is to divide the number of vehicles into a few subgroups (e.g., m data querying subgroups), then the time complexity of the matrix inverse in each subgroup can be reduced to $\mathcal{O}((\frac{n}{m})^3)$. While under our parameter setting in Fig. 5(a) and Fig. 5(b), the computational cost introduced by the matrix inverse computation is still negligible in comparison with the cryptographic operations.

C. Communication Overhead

The communication overhead of the proposed scheme can be divided into two parts, vehicle-to-RSU communication and vehicle-to-vehicle communication. We first consider the vehicle-to-RSU communication, for V_i $(i \neq a)$, the data request sent towards the RSU is in the format of $ID_i||C_{i,1}||MAC_{i,1}||\sigma_i||TS$, and the corresponding communication overhead should be $|ID_i| + 2048 + 1024 + 1024 * 2 + |TS|$ bits. For V_a, the data request sent towards the RSU is in the form of $ID_a||C_{a,1}||\sigma_a||\mathbf{A}||C_k||MAC_a||TS$, and the communication overhead is $|ID_a| + 2048 + 1024{*}2 + 384{*}k{*}k + 2048 + 1024 + |TS|$ bits. Thus, when we set the length of $|ID_i| + |TS|$ to be 64 bits, the overall vehicle-to-RSU communication overhead is $64 * k + 2048 * (k+1) + 1024 * 3 * k + 384 * k * k$ bits. For the traditional scheme, each vehicle V_i generates k ciphertexts $C_{i,1,l}, l = 1, \ldots, k$ with the length of 2048 bits. The overall communication overhead is $Comm_1 = 64 * k + 2048 * (k * k + 1) + 1024 * 3 * k + 384 * k * k$ bits.

Next, we examine the communication overhead of the vehicle-to-vehicle communication. The data request from V_i to V_a is in the format of $ID_i||C_{i,2}||MAC_{i,2}||TS$, and the corresponding length is $|ID_i| + 384 * (k-2) + 1024 + |TS|$ bits. Thus, the overall vehicle-to-vehicle communication overhead is $Comm_2 = 64 * (k-1) + 384 * (k-2) * (k-1) + 1024 * (k-1)$ bits. Fig. 6(a) and Fig. 6(b) depict the communication overhead between the RSU and vehicles, and the total communication overhead, respectively, when the number of data querying vehicles varies from 5 to 100. As shown in Fig. 6, we can observe that the proposed scheme can achieve lower communication overhead than the traditional scheme.

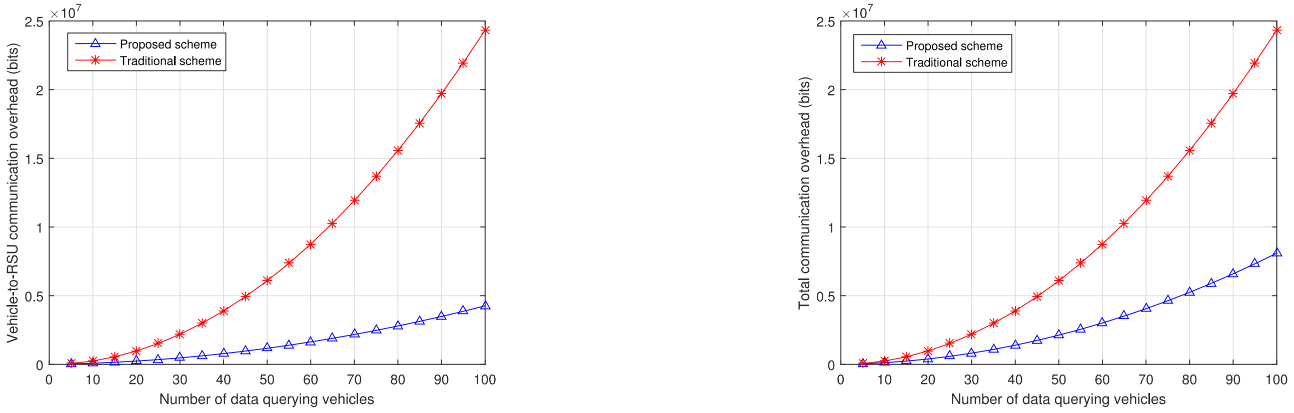
Another limitation of the proposed scheme is the communication cost introduced by the transmissions of the matrix and inverse matrix. The cost of the matrix vectors distribution



(a) Computation complexity of the RSU

(b) Computation complexity of each vehicle

Fig. 5. Computation complexity of the proposed vehicular data querying scheme.



(a) Vehicle-to-RSU Communiation Overhead

(b) Total Communication Overhead

Fig. 6. Communication overhead during the data querying.

between the data querying vehicles is $384 * (k-2) * (k-1)$ bits; meanwhile, the cost introduced by the inverse matrix transmission between the coordinating vehicle and the RSU is $384 * k * k$ bits.

D. Impact of Velocity

To demonstrate the feasibility of the proposed scheme, we also check whether the entire data querying process could be completed within the coverage of a single RSU. That is, we calculate the relative distance the vehicle V_a passes between the starting and ending locations of the entire data querying process, which is a value depends on the moving velocity and the number of vehicles. Meanwhile, the relative distance is calculated as $dist = (Comp_2 + Comp_3 + (Comm_1 + Comm_2)/Rate) * v$, where $Rate = 3$ Mbps is the data rate selected in VANET [10] and v is the moving speed of V_a.

In Fig. 7, we show the relative distance between the starting and ending locations of the entire data querying process, with respect to the number of joining vehicle and moving velocity. As shown in Fig. 7, when the moving speed ranges from 0 km/h to 160 km/h and the number of vehicles varies from 5 to 100, the maximum passing distance of V_a is 200 m, such that the entire data querying process could be completed within the coverage of one single RSU.

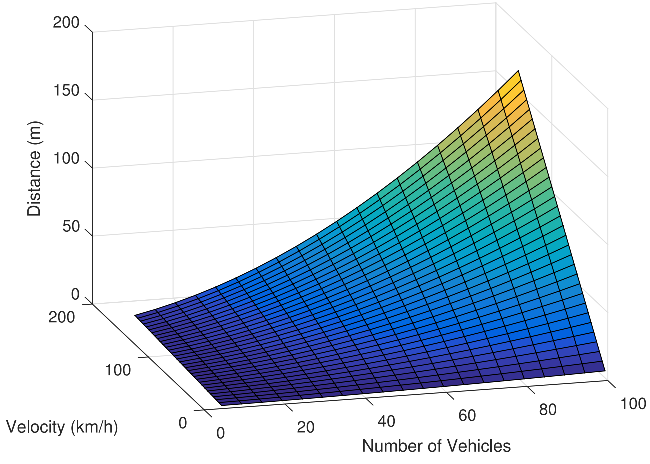


Fig. 7. Vehicular data querying architecture under consideration.

Note that, in our above analysis, we only consider the impact of velocity and data rate in VANET. However, due to the medium access control constraints of the current VANET protocol, it does not accommodate to the situation when there exists a large number of vehicles sending time-critical communication requests together. To resolve this issue, a random access control protocol in VANET with constraint fading channel is proposed in [18], which models the fading of each channel as an independent stochastic process and exploits the Markov random access protocol for channel assignment. To further solve the problem of the constraint number of access control channels, a co-designed framework for medium access scheduling and platoon control is also designed in [19], which resolves the network access conflicts with string stability for vehicular platoon in VANETs.

VII. Related Works

In this section, we briefly review some papers that are closely related to our scheme, in terms of secure data dissemination in VANETs and privacy-preserving data aggregation.

A. Secure Data Dissemination in VANETs

The fixed RSUs function as fog storage devices at the network edge, which are connected to the backbone network, can also function as a data server which store a catalog of information, and disseminate the stored data towards the vehicles under their coverage [20], in either proactive or request-response data dissemination [21], [22]. For the proactive data dissemination schemes, RSUs broadcast data towards vehicles in a proactive manner, which is particularly suitable for safety-related and urgent information distribution [23]. Instead of initiatively broadcasting data, in request-response data dissemination schemes, information is made available with a search-response paradigm. Specifically, vehicles send data queries towards the RSUs to solicit the needed content, and the RSUs disseminate data according to the collected data requests, which are particularly suitable for the value-added and infotainment data services.

A privacy-preserving navigation scheme with the anonymous credential is proposed in [24], which exploits the speed and road condition data collected by the RSU and can lead vehicles towards the intended destination in a distributive fashion. In [25], a secure and efficient message dissemination scheme is proposed in VANET, which outsources most of the decryption computation to the nearest RSU, such that the emergency messages with stringent delay requirements can be effectively disseminated towards vehicles with limited processing capability. An on-demand request-response unicast data dissemination in VANET is proposed in [26], which enables the identification of authentic vehicles in VANET and achieves data dissemination among authentic vehicles.

However, the above schemes mainly focus on the authentication of the involved vehicles, or the verification of the received messages, and they do not take the unlinkability into consideration. Even though [24] exploits the pseudonym credentials to protect the real identities of the query issuing vehicles, the content of the data query is still disclosed, and it may lead to the threat of privacy disclosure.

B. Privacy-Preserving Data Aggregation

In the proposed scheme, we achieve the security goal of unlinkability through the introduction of multiple vehicles, and each data request recovered from the data aggregation result cannot be associated with its origin. Thus, we perform a brief review on the existing schemes on secure and privacy-preserving data aggregation. A concealed data aggregation scheme is presented for wireless sensor networks in [27], which achieves the end-to-end data confidentiality and enables the data aggregation on ciphertexts with privacy homomorphism techniques. To reduce the communication overhead, the intermediate nodes aggregate the encrypted data and then upload the aggregation results towards the base station with privacy preservation. In [6], a privacy-preserving multi-dimensional data aggregation scheme is proposed with the homomorphic Paillier Cryptosystem, which can aggregate and recover the aggregation of the data reports in each dimension. In [7], a privacy-preserving route reporting aggregation scheme in VANET is proposed, which employs the data aggregation technique to calculate the routes of the vehicles in each segment. To achieve both privacy preservation in the domain of security and fault tolerance in the domain of statistic analysis, a novel secure data aggregation for smart grid communications is proposed in [28], which can flexibly adjust the malfunctioning smart meters and can also resist the differential attacks. A privacy-preserving verifiable data aggregation scheme is proposed in [29], which protects the identity privacy of each involved work, prevents the cloud from learning the content of the aggregation results, and allows the data requester to verify the correctness of the retrieved results.

However, the above schemes can only calculate the aggregation result of the data reports, the content of each individual data report cannot be recovered. In our proposed scheme, based on the data aggregation result, each data query can be still recovered by the RSU without being associated with its origin vehicle.

VIII. Conclusion

In this paper, we have proposed a privacy-preserving and verifiable querying scheme in vehicular fog data dissemination. By exploiting the homomorphic Paillier cryptosystem and structuring multiple data requests sent by different vehicles with an invertible matrix, the proposed scheme can prevent the RSU (fog storage device) from associating each recovered data request with its origin vehicle, and it can also support batch verification of the recovered data requests. Through security analysis, we demonstrate that the proposed scheme can achieve the security goals of unlinkability, confidentiality and verifiability. Through performance evaluation, we examine the performance of the proposed scheme and demonstrate its flexibility and high efficiency. For the future work, we will take the possible collusion attack into consideration, and improve the current scheme to resist such attacks.

References

[1] C. Huang, R. Lu, and K. R. Choo, "Vehicular fog computing: Architecture, use case, and security and forensic challenges," *IEEE Commun. Mag.*, vol. 55, no. 11, pp. 105–111, Nov. 2017.

[2] M. Chiang and T. Zhang, "Fog and IOT: An overview of research opportunities," *IEEE Internet Things J.*, vol. 3, no. 6, pp. 854–864, Dec. 2016.

[3] X. Hou, Y. Li, M. Chen, D. Wu, D. Jin, and S. Chen, "Vehicular fog computing: A viewpoint of vehicles as the infrastructures," *IEEE Trans. Veh. Technol.*, vol. 65, no. 6, pp. 3860–3873, Jun. 2016.

[4] F. Al-Turjman, "Cognitive caching for the future sensors in fog networking," *Pervasive Mobile Comput.*, vol. 42, pp. 317–334, 2017.

[5] D. Huang, S. Misra, M. Verma, and G. Xue, "PACP: An efficient pseudonymous authentication-based conditional privacy protocol for vanets," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 3, pp. 736–746, Sep. 2011.

[6] R. Lu, X. Liang, X. Li, X. Lin, and X. Shen, “EPPA: An efficient and privacy-preserving aggregation scheme for secure smart grid communications,” *IEEE Trans. Parallel Distrib. Syst.*, vol. 23, no. 9, pp. 1621–1631, Sep. 2012.

[7] K. Rabieh, M. M. E. A. Mahmoud, and M. F. Younis, “Privacy-preserving route reporting schemes for traffic management systems,” *IEEE Trans. Veh. Technol.*, vol. 66, no. 3, pp. 2703–2713, Mar. 2017.

[8] R. Li, H. Asaeda, J. Li, and X. Fu, “A verifiable and flexible data sharing mechanism for information-centric IOT,” in *Proc. IEEE Int. Conf. Commun.*, Paris, France, May 21–25, 2017, pp. 1–7.

[9] J. Liu, J. Han, L. Wu, R. Sun, and X. Du, “VDAS: Verifiable data aggregation scheme for Internet of things,” in *Proc. IEEE Int. Conf. Commun.*, Paris, France, May 21–25, 2017, pp. 1–6.

[10] G. Karagiannis *et al.*, “Vehicular networking: A survey and tutorial on requirements, architectures, challenges, standards and solutions,” *IEEE Commun. Surveys Tut.*, vol. 13, no. 4, pp. 584–616, Fourth Quarter 2011.

[11] J. Bohli and A. Pashalidis, “Relations among privacy notions,” *ACM Trans. Inf. Syst. Secur.*, vol. 14, no. 1, pp. 4:1–4:24, 2011.

[12] O. Cetinkaya and A. Doganaksoy, “Pseudo-voter identity (PVID) scheme for e-voting protocols,” in *Proc. 2nd Int. Conf. Availability, Rel. Int. Dependability Conf. Bridging Theory Practice*, Vienna, Austria, April 10–13, 2007, pp. 1190–1196.

[13] P. Paillier, “Public-key cryptosystems based on composite degree residuosity classes,” in *Proc. Adv. Cryptol., Int. Conf. Theory Appl. Cryptograph. Tech.*, Prague, Czech Republic, May 2–6, 1999, pp. 223–238.

[14] Q. Kong, R. Lu, H. Zhu, A. Alamer, and X. Lin, “A secure and privacy-preserving incentive framework for vehicular cloud on the road,” in *Proc. IEEE Global Commun. Conf.*, Washington, DC, USA, Dec. 4–8, 2016, pp. 1–6.

[15] S. Cui, P. Duan, C. W. Chan, and X. Cheng, “An efficient identity-based signature scheme and its applications,” *Int. J. Netw. Secur.*, vol. 5, no. 1, pp. 89–98, 2007.

[16] K. Liu, “Paillier’s homomorphic cryptosystem (java implementation).”

[17] A. D. Caro and V. Iovino, “JPBC: Java pairing based cryptography,” in *Proc. 16th IEEE Symp. Comput. Commun.*, Kerkyra, Greece, 28 Jun.–1 Jul., 2011, pp. 850–855.

[18] G. Guo and L. Wang, “Control over medium-constrained vehicular networks with fading channels and random access protocol: A networked systems approach,” *IEEE Trans. Veh. Technol.*, vol. 64, no. 8, pp. 3347–3358, Aug. 2015.

[19] G. Guo and S. Wen, “Communication scheduling and control of a platoon of vehicles in VANETs,” *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 6, pp. 1551–1563, Jun. 2016.

[20] K. Liu and V. C. Lee, “RSU-based real-time data access in dynamic vehicular networks,” in *Proc. 13th Int. IEEE Conf. Intell. Transp. Syst.*, 2010, pp. 1051–1056.

[21] E. Lee, E. Lee, M. Gerla, and S. Oh, “Vehicular cloud networking: Architecture and design principles,” *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 148–155, Feb. 2014.

[22] F. Ye, S. Roy, and H. Wang, “Efficient data dissemination in vehicular ad hoc networks,” *IEEE J. Sel. Areas Commun.*, vol. 30, no. 4, pp. 769–779, May 2012.

[23] C. Barberis and G. Malnati, “Design and evaluation of a collaborative system for content diffusion and retrieval in vehicular networks,” *IEEE Trans. Consum. Electron.*, vol. 57, no. 1, pp. 105–112, Feb. 2011.

[24] T. W. Chim, S. Yiu, L. C. K. Hui, and V. O. K. Li, “VSPN: VANET-based secure and privacy-preserving navigation,” *IEEE Trans. Comput.*, vol. 63, no. 2, pp. 510–524, Feb. 2014.

[25] X. Liu *et al.*, “SEMD: Secure and efficient message dissemination with policy enforcement in VANET,” *J. Comput. Syst. Sci.*, vol. 82, no. 8, pp. 1316–1328, 2016.

[26] A. Mondal and S. Mitra, “Secure data dissemination in VANET—A pull based approach,” in *Proc. IEEE Int. Conf. Commun., Netw. Satell.*, 2015, pp. 60–67.

[27] S. Özdemir, “Concealed data aggregation in heterogeneous sensor networks using privacy homomorphism,” in *Proc. IEEE Int. Conf. Pervasive Serv.*, Istanbul, Turkey, 2007, 15–20 Jul., 2007, pp. 165–168.

[28] H. Bao and R. Lu, “A new differentially private data aggregation with fault tolerance for smart grid communications,” *IEEE Internet Things J.*, vol. 2, no. 3, pp. 248–258, Jun. 2015.

[29] G. Zhuo, Q. Jia, L. Guo, M. Li, and P. Li, “Privacy-preserving verifiable data aggregation and analysis for cloud-assisted mobile crowdsourcing,” in *Proc. 35th Annu. IEEE Int. Conf. Comput. Commun.*, San Francisco, CA, USA, Apr. 10–14, 2016, pp. 1–9.

Qinglei Kong (S’15) received the B.Eng. degree in communication engineering from the Harbin Institute of Technology, Harbin, China, in 2012, the M.Eng. degree in electronic and information engineering from Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen, China, in 2015, and the Ph.D degree from the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, in 2018. She was a Research Associate with the School of Electrical and Electronic Engineering, Nanyang Technological University from December 2015 to July 2018. Her research interests include applied cryptography, security and privacy protocols, VANET, and cloud computing.

Rongxing Lu (S’09–M’11–SM’15) received the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada, in 2012. He has been an Assistant Professor with the Faculty of Computer Science, University of New Brunswick (UNB), Fredericton, NB, Canada, since August 2016. Before that, he was an Assistant Professor with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, from April 2013 to August 2016. He was a Postdoctoral Fellow with the University of Waterloo from May 2012 to April 2013. He has authored or coauthored extensively in his areas of expertise. His research interests include applied cryptography, privacy enhancing technologies, and IoT-Big Data security and privacy. He was the recipient of the most prestigious Governor Generals Gold Medal and the 8th IEEE Communications Society (ComSoc) Asia Pacific (AP) Outstanding Young Researcher Award, in 2013. He is currently a senior member of the IEEE Communications Society. He was the recipient of eight Best (student) Paper awards from some reputable journals and conferences. He is currently the Vice-Chair (Publication) of IEEE ComSoc CIS-TC (Communications and Information Security Technical Committee). He is the winner of 2016–2017 Excellence in Teaching Award, FCS, UNB.

Maode Ma received the B.E. degree from Tsinghua University, Beijing, China, in 1982, the M.E. degree from Tianjin University, Tianjin, China, in 1991, and the Ph.D. degree from The Hong Kong University of Science and Technology, Hong Kong, in 1999. He is currently an Associate Professor with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. He has authored or coauthored about 200 international academic publications, including more than 80 journal papers, more than 140 conference papers, and/or book chapters, and three academic books. His research interests include wireless networking and wireless network security. He is a member of a few technical committees in the IEEE Communication Society. He was a member of the technical program committees for more than 100 international conferences. He was a General Chair, Technical Symposium Chair, Tutorial Chair, Publication Chair, Publicity Chair, and Session Chair for more than 50 international conferences. He is currently an Editor-in-Chief/Associate Editor for six international journals.

Haiyong Bao received the Ph.D. degree in computer science from Shanghai Jiao Tong University, Shanghai, China, in 2006. Since February 2011, he has been an Associate Professor with the School of Computer Science and Information Engineering, Zhejiang Gongshang University, Hangzhou, China. From May 2014 to May 2015, he was a Postdoctoral Fellow with the INFINITUS Laboratory, School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. His research interests include secure data aggregation, insider attack detection, and applied cryptography.