Future Generation Computer Systems 92 (2019) 644–655

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

A privacy-preserving sensory data sharing scheme in Internet of Vehicles

Qinglei Kong [a], Rongxing Lu [b,*], Maode Ma [a], Haiyong Bao [c]

[a] School of Electrical and Electronics Engineering, Nanyang Technological University, 50 Nanyang Avenue, 639798, Singapore
[b] Faculty of Computer Science, University of New Brunswick, ITC Building, 550 Windsor Street Fredericton, NB, Canada E3B 5A3
[c] School of Computer Science and Information Engineering, Zhejiang Gongshang University, Hangzhou 310018, PR China

HIGHLIGHTS

- The proposed scheme enables a vehicle to structure the multi-dimensional sensory data, which greatly saves system resources.
- The proposed scheme achieves the location privacy-preserving composite data aggregation with collusion resistance.
- The proposed scheme achieves privacy-preserving sensory data querying at the network edge.
- Numerical analysis, security analysis, and performance evaluation are conducted to identify the efficiency and effectiveness of the proposed scheme.

ARTICLE INFO

Article history:
Received 30 July 2017
Received in revised form 27 October 2017
Accepted 3 December 2017
Available online 13 December 2017

Keywords:
Internet of Vehicles (IoV)
Location privacy preservation
Data collection
Data acquisition

ABSTRACT

Internet of Vehicles (IoV), which enables information gathering and disseminating among vehicles, roadside infrastructures and surrounding environments, has received considerable attention recently. However, the flourishing of IoV still faces several challenges in terms of location privacy preservation, vehicular sensory data collection and vehicular sensory data acquisition. Aiming at these challenges, in this paper, we propose a novel efficient and location privacy-preserving data sharing scheme with collusion resistance in IoV, which enables the collection and distribution of the data captured by vehicular sensors. During the vehicular sensory data collection phase, each vehicle structures the multi-dimensional sensory data captured at different locations and exploits the modified Paillier Cryptosystem to achieve the location privacy-preserving sensory data aggregation, while the data aggregation result of multiple vehicles can be recovered by the trusted central entity. During the vehicular sensory data acquisition phase, the proposed scheme exploits the proxy re-encryption technique to achieve the location privacy-preserving data querying at the network edge, i.e., the vehicles query the RSU instead of the trusted central entity. Numerical analysis are performed to validate the effectiveness of the proposed scheme, i.e., low data querying failure probability. Security analysis and performance evaluations are also carried out to validate the security properties and show the computation and communication efficiency of the proposed scheme.

1. Introduction

As one of the most important branches of Internet of Things (IoT), Internet of Vehicles (IoV), which is an extension of the vehicular ad hoc networks (VANETs), mainly focuses on the information interaction between vehicles, roadside infrastructures, and surrounding environments without human intervention [1]. By coordinating the data captured by vehicular sensors (e.g., chemical spill detectors, vibration sensors, and acoustic detectors [2]), which can record a myriad of characteristics reflecting the physical phenomena, IoV enables the data sharing among vehicles and the deployed roadside infrastructures to further improve traffic safety and on-board experience in the intelligent transportation system (ITS); meanwhile, it can also support various value-added services, such as autonomous driving [3], fuel optimal cruising [4], and early warning [5].

Although IoV can improve the ITS performance through information interaction, there exist several challenges for the wide application of vehicular sensory data sharing in IoV, including location privacy preservation, vehicular sensory data collection, and vehicular sensory data acquisition. During the sensory data collection phase, since the captured sensory data are highly location-dependent, the location-based sensory data may disclose the moving trajectory and violate the location privacy of vehicles [6]. Recently, privacy-preserving data aggregation has received much attention in the era of IoT [7,8], which are designed to achieve the

* Corresponding author.
E-mail addresses: qlkong@ntu.edu.sg (Q. Kong), rlu1@unb.ca (R. Lu), emdma@ntu.edu.sg (M. Ma), baohy@zjgsu.edu.cn (H. Bao).

https://doi.org/10.1016/j.future.2017.12.003
0167-739X/© 2017 Elsevier B.V. All rights reserved.

data aggregation result while protecting the content of each individual data report. In [7], to achieve efficient privacy-preserving data aggregation in smart grid, each user structures the small-sized multi-dimensional sensory data into one composite data report and encrypts each composite data report with the homomorphic encryption technique, and then the gateway aggregates the encrypted composite data reports. However, in IoV, vehicles are dynamically moving and the scale of sensory data reports in each data dimension (if we take different locations as different data dimensions) are constantly fluctuating in the temporal domain. The existing schemes are not able to count the number of sensory data generated in each data dimension, and further fail to calculate the average of the sensory data in each data dimension. Thus, to fully exploit sensory data generated by vehicles, the design of an efficient and location privacy-preserving data collection scheme in IoV is essential.

During the sensory data acquisition phase, since vehicles are normally interested in acquiring the sensory data in accordance with the potential moving path towards the destination, the location-based data queries may violate the location privacy of vehicles [9,10]. Location-based data querying schemes with privacy preservation has attracted considerable attention [11,12], however, the existing schemes are mainly designed for the outsourced cloud environment with fixed cloud storage, which cannot be directly applied towards the real-time services in IoV. To preserve location privacy, the querying location needs to be hid and the data captured at the unqueried locations also needs to be protected, and the oblivious transfer (OT) [13] can be employed to achieve both during the data querying process. However, in the existing OT protocols [13,14], the communication overhead is proportional to the data dimension, which are not applicable for the IoV environment with short-termed and intermittent wireless connection. Thus, to achieve the efficient sensory data sharing in IoV, an efficient and location privacy-preserving data acquisition scheme in IoV is also essential.

In this paper, aiming at the above challenges, we propose a novel efficient and location privacy-preserving data sharing scheme in IoV with the modified Paillier Cryptosystem, which enables the road side unit (RSU) to collect and distribute the sensory data generated by vehicles. Specifically, the contributions of this paper are threefold as follows.

- First, we propose a privacy-preserving vehicular sensory data collection scheme in IoV. The proposed scheme first enables a vehicle to structure the multi-dimensional sensory data captured at different locations into one composite data report, which greatly saves the communication and computation resources. The proposed scheme can also support the RSU to perform the privacy-preserving composite data reports aggregation and count the number of data reports contained in each data dimension. Meanwhile, the sensory data aggregation result can be retrieved and re-encrypted by the trusted traffic management authority for further processing.
- Second, we propose a privacy-preserving sensory data acquisition scheme in IoV. The proposed scheme enables a data querying vehicle to retrieve the sensory data captured by other vehicles at the network edge, i.e., without the involvement of the trusted central traffic management authority. Specifically, the RSU re-encrypts the aggregated sensory data into ciphertexts that can only be fully recovered by the data querying vehicle, while without the disclosure of the data querying location.
- Third, we offer numerical analysis to compute the probability of data querying failure, and perform detailed security analysis to validate the security property of the proposed scheme in terms of location privacy preservation, data integrity protection and collusion attack resistance. Through performance evaluation, we demonstrate that the proposed scheme can significantly reduce the computation cost and communication overhead in comparison with the traditional secure data aggregation and OT protocols.

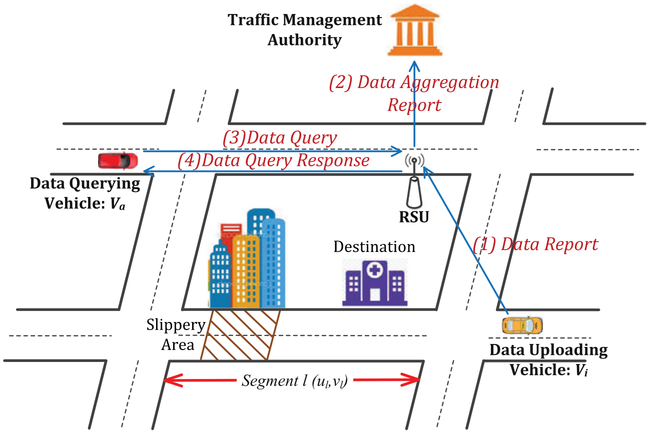

Traffic Management Authority
(2) Data Aggregation Report
(3)Data Query
(4)Data Query Response
Data Querying Vehicle: V_a
RSU
Destination
(1) Data Report
Slippery Area
Data Uploading Vehicle: V_i
Segment l (u_l, v_l)

Fig. 1. Vehicular sensory data sharing system.

The remainder of this paper is organized as follows. We describe system model, show security requirements, and identify design goals in Section 2. In Section 3, we recall the modified Paillier Cryptosystem and Chinese Remainder Theorem as the preliminaries. Then we present our proposed privacy-preserving vehicular data sharing scheme in Section 4, followed by our security analysis and performance evaluation in Section 5 and in Section 6, respectively. We show related work in Section 7, and finally conclude our work in Section 8.

2. System model, security requirements, and design goals

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

2.1. System model

In the system model, we consider a practical scenario in an urban area, in which the data querying vehicle (V_a) aims to drive towards the destination (a hospital) in an efficient and safe manner, as shown in Fig. 1. By querying the sensory information captured by other vehicles' inductive rotational-speed sensors [15] (the input data for the anti-lock braking system (ABS) system [16], which can detect the slippery area), V_a can discover the hidden road conditions. If V_a notices the existence of a slippery area (e.g., the shadow area of a building blockage after snow in winter), V_a will avoid it and take an alternative path. The proposed system consists of four roles: a traffic management authority, a road side unit (RSU), a group of data uploading vehicles (V_i, $i = 1, 2, \ldots, w$), and a data querying vehicle (V_a), as shown in Fig. 1.

- Traffic management authority is a central entity, which is responsible for the registration and key distribution of the RSU and vehicles. Traffic management authority divides a given urban area into districts (e.g., with the coverage area of approximately 3 km^2), such that each district is large enough for the vehicles to be comfortable with revealing fact that they are locating somewhere within the district. In each divided district, the sensory data generation locations are denoted in the level of segments, and a long road can also be divided into a few short segments. For segment l, it is denoted by a unique two-dimensional segment identifier (u_l, v_l), which is an approximation of the location coordinates.

- Each sensory data uploading vehicle $(V_i, i = 1, 2, \dots, w)$ is equipped with the required inductive rotational-speed sensors, such that it can periodically formulate a data report with the captured data and upload the structured data report towards the RSU, as shown in Fig. 1 (*(1) Data Report*).
- The RSU collects and aggregates the on-board sensory data reports uploaded by vehicles, and delivers the data aggregation result towards the traffic management authority, as shown in Fig. 1 (*(2) Data Aggregation Report*). Meanwhile, the RSU shares the aggregated on-board sensory data report with other vehicles when it is requested, as shown in Fig. 1 (*(4) Data Query Response*). We set the transmission distance of the RSU to be 1 *km*, such that it can provide signal coverage of the whole divided district [17].
- When the data querying vehicle (V_a) travels in vicinity to the potential slippery area, i.e., several segments ahead, it sends a data query towards the RSU for the sensory data captured at the slippery area, as shown in Fig. 1 (*(3) Data Query*).

Communication Model. The wireless connections between the RSU and vehicles are realized through the IEEE 802.11p standard [18], which is a short to medium range communication technology operating at 5.9 GHz band designed for wireless access in vehicular environments (WAVE). The connection between the RSU and the traffic management authority is realized through either wired links or any other links with high bandwidth and low transmission delay.

2.2. Security requirements

In the security model, we consider the traffic management authority is fully trusted. The RSU and vehicles are honest-but-curious (i.e., they will follow the defined protocols, but they are curious about the location information of other vehicles). Meanwhile, we assume the vehicular sensory data are correctly generated and delivered towards the RSU. Moreover, we assume there is a compromised RSU, which may collude with a malicious vehicle, i.e., through mutually sharing their secret keys. We also make an assumption that there exists an adversary $\mathcal{A}$ located on the road, which could launch some active attacks to modify the data transmission and threat the data integrity. Therefore, the following security requirements should be met.

Location Privacy Preservation. In the proposed scheme, since the locations of vehicles are denoted with segments, preserving location privacy means preventing the segments from being uncovered. During the sensory data collection phase, preserving the location privacy of data uploading vehicles indicates that, the sensory data generation segments contained in an individual data report should be protected, i.e., even if the RSU obtains all the possible data reports generated by a data uploading vehicle, it still cannot identify the data generation segments contained in an individual data report. During the sensory data acquisition phase, since a data querying vehicle only interests in sensory data in accordance with the potential trajectory towards its destination, the segment identifier contained in a data query should also be protected. In addition, the location privacy preservation requirement also should include that a data querying vehicle can only obtain the sensory data captured in the segment it is querying about.

Collusion resistance. The proposed scheme should resist the collusion between a compromised RSU and a malicious vehicle by mutually sharing their secrets. Given the secret key of the compromised RSU, the malicious vehicle should not derive the content contained in any individual sensory data report during the data collection phase and should not obtain the location identifier contained in any data query during the data acquisition phase. Meanwhile, when the malicious vehicle receives a data query response from an uncompromised RSU, it can only learn the sensory data generated at the segment it is querying about. Given the session key of the malicious vehicle, the compromised RSU should also not obtain the content contained in any individual data report and the location identifiers contained in any individual data query.

Data Integrity. The proposed scheme should guarantee that any sensory data report, data query and data query response, should not be modified during the data transmission process, i.e., even if the adversary $\mathcal{A}$ alters the data transmission, the malicious operation should be detected. In this way, the security requirement of data integrity could be achieved.

2.3. Design goals

Under the aforementioned system model and security requirements, our design goal is to develop an efficient and privacy-preserving sensory data sharing scheme, which enables vehicles to upload the captured sensory data towards the RSU and allows vehicles to retrieve the aggregated sensory data captured by other vehicles. Specifically, the following three requirements should be achieved.

The proposed scheme should achieve the defined security requirements. If the security requirements are not taken into consideration, the location privacy of the involved vehicles could be violated, and the data transmission between the RSU and vehicles could be modified. Then vehicles are not willing to join in the sensory data sharing process, and the collected sensory data could not be properly shared and fully exploited.

The proposed scheme should achieve high effectiveness during the sensory data sharing process. Since the data captured by the on-board sensors are critical to safety-related applications, the accuracy of the collected on-board sensory data is mandatory. Meanwhile, in order to achieve the effective sensory data acquisition under stringent delay requirement, the RSU should be able to provide near real-time sensory data with low probability of data querying failure.

The proposed scheme should achieve high efficiency in terms of computation complexity and communication overhead. The dynamically moving characteristic of vehicles and the scarce transmission bandwidth in vehicular network result in the requirement of short transmission interval, low processing delay and low communication overhead. Thus, the proposed scheme should achieve high computation and communication efficiency.

3. Preliminaries

In this section, we briefly review the modified Paillier Cryptosystem [19] and the Chinese Remainder Theorem [20], which will serve as the basis of the proposed vehicular sensory data sharing scheme.

3.1. Modified Paillier cryptosystem

The modified Paillier cryptosystem is exploited due to its homomorphic addition property and homomorphic multiplication property between one ciphertext and one plaintext [19], whose variations have been applied to various privacy-preserving schemes [21,22]. Specifically, the proposed scheme consists of four components: key generation, encryption, decryption, and proxy re-encryption.

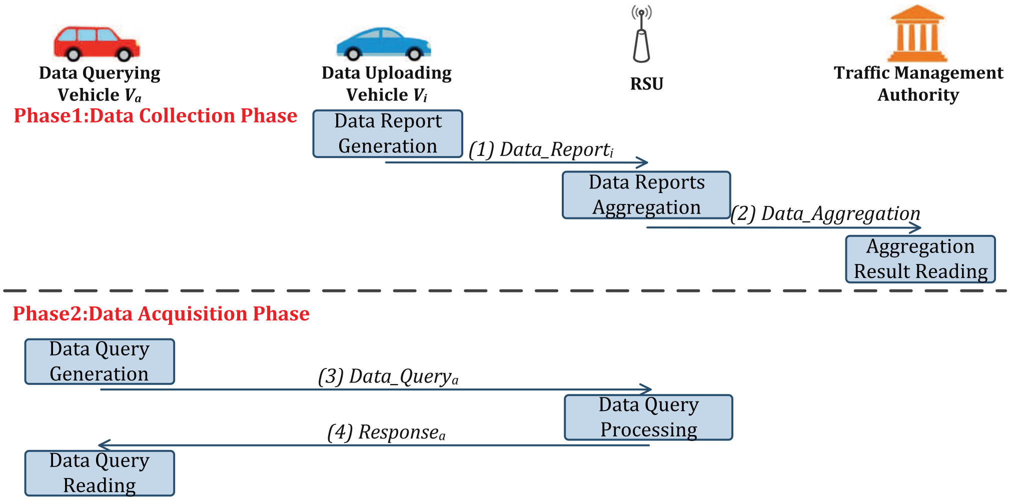


Fig. 2. Message flow during data sharing.

- *Key Generation:* Given security parameter κ, two large safe prime numbers p_1, p_2 in the form of $p_1 = 2p_1' + 1$ and $p_2 = 2p_2' + 1$ are first chosen, where $|p_1| = |p_2| = \kappa$, and p_1', p_2' are also primes. Then, the RSA modulus $n = p_1p_2$ and $\lambda = lcm(p_1 - 1, p_2 - 1)$ are computed. We consider $\mathbb{G} = QR_{n^2}$ the cyclic group of quadratic residues modulo n^2, whose order is $ord(\mathbb{G}) = \lambda(n^2)/2 = p_1p_1'p_2p_2' = n\lambda/2$, and the maximal order of an element in this group is $n\lambda/2$. Choose a random number $\mu \in \mathbb{Z}^*_{n^2}$ and a random value $x \in [1, ord(\mathbb{G})]$, then set $g = \mu^2 \bmod n^2$ and $h = g^x \bmod n^2$. The public key is represented as $pk = (n, g, h = g^x)$, and the corresponding secret key is x.
- *Encryption:* Given message $m \in \mathbb{Z}_n$, choose a random number $r \in \mathbb{Z}_{n^2}$ and generate the ciphertext pair (c_1, c_2) as $c_1 = g^r \bmod n^2$ and $c_2 = h^r(1 + mn) \bmod n^2$.
- *Decryption:* Given ciphertext pair (c_1, c_2), message m can be decrypted as $m = \frac{c_2/(c_1)^x - 1 \bmod n^2}{n}$.
- *Proxy Re-encryption:* Randomly splits the secret key x into two shares x_1 and x_2, such that $x = x_1 + x_2$. The modified Paillier Cryptosystem enables an encrypted message (c_1, c_2) to be partially decrypted to a ciphertext pair as $(\tilde{c_1}, \tilde{c_2})$ using x_1 as $\tilde{c_1} = c_1$ and $\tilde{c_2} = c_2/(c_1)^{x_1} \bmod n^2$. Then $(\tilde{c_1}, \tilde{c_2})$ can be decrypted with x_2 to recover the original message m.

3.2. Chinese remainder theorem

The Chinese Remainder Theorem can solve any pair of congruences with relatively prime moduli [20], which can be exploited to structure the multi-dimensional data into one composite data, and it is described as follows.

Theorem 3.1 (*Chinese Remainder Theorem*). *Suppose that $q_1, q_2, \dots, q_k$ are relatively prime integers, and let $d_1, d_2, \dots, d_k$ be integers. Then, the system of congruences $s \equiv d_l \bmod q_l$ for $1 \le l \le k$, has a unique solution modulo $Q = q_1 \times q_2 \times \cdots q_k$, which is*

$$s \equiv d_1Q_1y_1 + d_2Q_2y_2 + \cdots + d_kQ_ky_k \bmod Q \tag{1}$$

where $Q_l = \frac{Q}{q_l}$ and $y_l \equiv \frac{1}{Q_l} \bmod q_l$, for $1 \le l \le k$.

4. Proposed privacy-preserving sensory data sharing scheme

In this section, we present our privacy-preserving vehicular sensory data sharing scheme, which mainly consists of three phases: system initialization, sensory data collection, and sensory data acquisition. Note that in this paper we only take one district into consideration, and the data flows are shown in Fig. 2.

4.1. System initialization

Given security parameter κ, the traffic management authority, which functions as a trusted authority (TA), selects two large safe prime numbers p_1 and p_2, where $|p_1| = |p_2| = \kappa$, then it generates the modified Paillier Cryptosystem's public key $pk = (n = p_1 \cdot p_2, g, h = g^x \bmod n^2)$ and the private key $sk = x$. Then TA randomly splits the secret key x into three shares x_0, x_1 and x_2, such that $x = x_0 + x_1 + x_2$, and it also calculates the value of $h_1 = g^{x_1+x_2} \bmod n^2$. Meanwhile, TA keeps the secret share x_0 to itself. During the registration of each RSU, TA securely delivers the secret value x_1 towards the RSU; meanwhile, during the registration of each vehicle, TA securely distributes the secret value x_2 towards the vehicle.

For a district with k segments, TA selects k equal-length prime numbers: $(q_1, q_2, \dots, q_k)$, where $|q_1| = |q_2| = \cdots = |q_k| = \kappa_1$, and it satisfies the condition that $k \cdot \kappa_1 < |n|$. For segment l with the segment identifier (u_l, v_l), TA assigns a prime number q_l towards it. Then TA computes

$$\begin{cases} Q = \prod_{l=1}^{k} q_l \\ Q_l = \dfrac{Q}{q_l}, y_l \equiv \dfrac{1}{Q_l} \bmod q_l \\ \alpha_l = Q_l \cdot y_l. \end{cases} \tag{2}$$

In addition, TA selects a constant number d with $|d| = \kappa_1 - 2$. Meanwhile, it chooses two secure cryptographic functions, $H_0 : \{0, 1\}^* \rightarrow \mathbb{Z}_n$ and $H : \{0, 1\}^* \rightarrow \{0, 1\}^{\kappa_1 - 2}$, such that $|H(\cdot) + d| < \kappa_1$. Finally, TA publishes the parameters: $params = \{pk, h_1, H_0, H, q_l : l = 1, 2, \dots, k\}$.

4.2. Sensory data collection

We describe the data report collection process as below and illustrate the process in Fig. 2 (*Phase1*). For vehicle V_i, we let $d_{i,l}$ denote the data captured by the inductive rotational-speed sensor at segment l, where $|d_{i,1}| = |d_{i,2}| = \cdots = |d_{i,k}|$ and $d_{i,l} < d$, $l = 1, 2, \dots, k$. Meanwhile, $e_{i,l}$ represents whether there is sensory data captured at segment l: (i) if there is sensory data captured at segment l, i.e., $d_{i,l} > 0$, then $e_{i,l} = 1$; (ii) if there is no sensory data captured at segment l, i.e., $d_{i,l} = 0$, then $e_{i,l} = 0$.

In order to achieve the near real-time sensory data uploading, each vehicle V_i gathers the sensory data captured at all the segments within a district during a short period T, e.g., every ten minutes, and performs the following steps to generate a sensory data report:

- V_i structures $(d_{i,1}, d_{i,2}, \ldots, d_{i,k})$ and $(e_{i,1}, e_{i,2}, \ldots, e_{i,k})$ into two composite values $s_{i,d}$ and $s_{i,e}$, i.e., $s_{i,d} = \sum_{l=1}^{k} d_{i,l} \cdot \alpha_l$ and $s_{i,e} = \sum_{l=1}^{k} e_{i,l} \cdot \alpha_l$.
- V_i selects two random numbers $(r_{i,d}, r_{i,e}) \in \mathbb{Z}_{n^2}$, and generates the following ciphertexts,

$$\begin{cases} C_{i,1} = g^{r_{i,d}} \bmod n^2, C_{i,2} = h^{r_{i,d}} \cdot (1 + s_{i,d} \cdot n) \bmod n^2, \\ C_{i,3} = g^{r_{i,e}} \bmod n^2, C_{i,4} = h^{r_{i,e}} \cdot (1 + s_{i,e} \cdot n) \bmod n^2. \end{cases} \tag{3}$$

  where $(C_{i,1}, C_{i,2})$ and $(C_{i,3}, C_{i,4})$ are the ciphertext pairs of $s_{i,d}$ and $s_{i,e}$, respectively.
- To protect data integrity, V_i also calculates the message authentication code of the ciphertext,

$$MAC_i = H_0(C_{i,1} \| C_{i,2} \| C_{i,3} \| C_{i,4} \| TS), \tag{4}$$

  where TS is the current timestamp. Note that the identity-based signature scheme can also be exploited here to protect the data integrity [23], and authenticate the origin of each sensory data report if necessary.

Thus, V_i transmits $Data\_Report_i = V_i \| C_{i,1} \| C_{i,2} \| C_{i,3} \| C_{i,4} \| MAC_i \| TS$ towards the RSU, as shown in Fig. 2 (*Message (1)*).

After receiving the data report $Data\_Report_i$, the RSU first verifies its correctness by checking $MAC_i \overset{?}{=} H_0(C_{i,1} \| C_{i,2} \| C_{i,3} \| C_{i,4} \| TS)$. If it does hold, the RSU performs the following steps to aggregate the data reports received from w data uploading vehicles, where the number of data uploading vehicles w satisfies the condition that $\lceil log_2 w \rceil + |d_{i,l}| < |d|, l = 1, 2, \ldots, k$.

$$\begin{cases} C_{d,1} = \prod_{i=1}^{w} C_{i,1} \bmod n^2 = \prod_{i=1}^{w} g^{r_{i,d}} \bmod n^2 \\ C_{d,2} = \prod_{i=1}^{w} C_{i,2} \bmod n^2 = (\prod_{i=1}^{w} h^{r_{i,d}}) \cdot (1 + \sum_{i=1}^{w} s_{i,d} \cdot n) \bmod n^2 \\ C_{e,1} = \prod_{i=1}^{w} C_{i,3} \bmod n^2 = \prod_{i=1}^{w} g^{r_{i,e}} \bmod n^2 \\ C_{e,2} = \prod_{i=1}^{w} C_{i,4} \bmod n^2 = (\prod_{i=1}^{w} h^{r_{i,e}}) \cdot (1 + \sum_{i=1}^{w} s_{i,e} \cdot n) \bmod n^2 \end{cases} \tag{5}$$

where $(C_{d,1}, C_{d,2})$ and $(C_{e,1}, C_{e,2})$ are the ciphertext pairs of the multi-dimensional sensory data aggregation $\sum_{i=1}^{w} s_{i,d}$, and the corresponding scale $\sum_{i=1}^{w} s_{i,e}$. Then the RSU calculates the message authentication code of the aggregated ciphertext pairs $MAC_{ag} = H_0(C_{d,1} \| C_{d,2} \| C_{e,1} \| C_{e,2} \| TS)$, formulates the sensory data aggregation message $Data\_Aggregation = C_{d,1} \| C_{d,2} \| C_{e,1} \| C_{e,2} \| MAC_{ag} \| TS$, and delivers it towards the traffic management authority, as shown in Fig. 2 (*Message (2)*).

After receiving *Data_Aggregation*, the traffic management authority verifies its correctness by checking whether $MAC_{ag} \overset{?}{=} H_0(C_{d,1} \| C_{d,2} \| C_{e,1} \| C_{e,2} \| TS)$, and re-encrypts the ciphertext $(C_{d,1}, C_{d,2})$ and $(C_{e,1}, C_{e,2})$ with the secret key x_0, which is shown as following

$$\begin{cases} C_{t,1} = \dfrac{C_{d,2}}{(C_{d,1})^{x_0}} \bmod n^2 \\ C_{t,2} = \dfrac{C_{e,2}}{(C_{e,1})^{x_0}} \bmod n^2 \end{cases} \tag{6}$$

Then the traffic management authority generates the message authentication code $MAC_t = H_0(C_{t,1} \| C_{t,2} \| TS)$, and delivers $C_{t,1} \| C_{t,2} \| MAC_t \| TS$ towards the RSU. In addition, the traffic management authority decrypts the received ciphertext pairs $(C_{d,1}, C_{d,2})$ and $(C_{e,1}, C_{e,2})$ with the secret key x, which is

$$\begin{cases} \sum_{i=1}^{w} s_{i,d} = \dfrac{C_{d,2}/(C_{d,1})^x - 1 \bmod n^2}{n}, \\ \sum_{i=1}^{w} s_{i,e} = \dfrac{C_{e,2}/(C_{e,1})^x - 1 \bmod n^2}{n}. \end{cases} \tag{7}$$

Thus, the traffic management authority can achieve the average of the vehicular sensory data $\bar{d}_l$ captured at segment l. According to the Chinese Remainder Theorem, $\bar{d}_l$ can be computed as

$$\bar{d}_l = \frac{\sum_{i=1}^{w} s_{i,d} \bmod q_l}{\sum_{i=1}^{w} s_{i,e} \bmod q_l} = \frac{\sum_{i=1}^{w} d_{i,l} \bmod q_l}{\sum_{i=1}^{w} e_{i,l} \bmod q_l}. \tag{8}$$

4.3. Sensory data acquisition

The data querying vehicle V_a intends to acquire the data captured by the inductive rotational-speed sensors at segment c with segment identifier (u_c, v_c), and we summarize the data acquisition process as below and illustrate the process in Fig. 2 (*Phase 2*). V_a selects three random numbers $(s_a, r_{a,1}, r_{a,2}) \in \mathbb{Z}_{n^2}$, and performs the following steps

- V_a calculates the encrypted data query

$$\begin{cases} C_{a,1} = g^{r_{a,1}} \bmod n^2, \\ C_{a,2} = h_1^{r_{a,1}} \cdot g^{s_a \cdot r_{a,1}} \cdot (1 + u_c \cdot n) \bmod n^2, \\ C_{a,3} = g^{r_{a,2}} \bmod n^2, \\ C_{a,4} = h_1^{r_{a,2}} \cdot g^{s_a \cdot r_{a,2}} \cdot (1 + v_c \cdot n) \bmod n^2. \end{cases} \tag{9}$$

  where $(C_{a,1}, C_{a,2})$ and $(C_{a,3}, C_{a,4})$ are the ciphertext pairs of u_c and v_c respectively.
- To protect the data integrity, V_a generates the corresponding message authentication code MAC_a, which is

$$MAC_a = H_0(C_{a,1} \| C_{a,2} \| C_{a,3} \| C_{a,4} \| TS). \tag{10}$$

Then V_a transmits the data query towards the RSU, which is $Data\_Query_a = V_a \| C_{a,1} \| C_{a,2} \| C_{a,3} \| C_{a,4} \| MAC_a \| TS$, as shown in Fig. 2 *(Message (3))*.

After receiving $Data\_Query_a$, the RSU verifies its correctness by checking $MAC_a \overset{?}{=} H_0(C_{a,1} \| C_{a,2} \| C_{a,3} \| C_{a,4} \| TS)$. If it does hold, the RSU selects two random numbers $(t_1, t_2) \in \mathbb{Z}_{n^2}$, and performs the following steps:

- The RSU calculates a secret key $\beta_l = H(m_l), l = 1, 2, \ldots, k$, for each segment l with location identifier (u_l, v_l), where $m_l = u_l \cdot t_1 + v_l \cdot t_2 \bmod n$. By computing $s_k = \sum_{l=1}^{k} \beta_l \cdot \alpha_l$, the RSU structures the multi-dimensional secret key $(\beta_1, \beta_2, \ldots, \beta_k)$ into one composite secret key s_k.

- The RSU generates the ciphertext pair of $m_c = u_c \cdot t_1 + v_c \cdot t_2 \bmod n$ without learning the value of (u_c, v_c), which is

$$\begin{cases} C_{r,1} = (C_{a,1})^{t_1} \cdot (C_{a,3})^{t_2} \bmod n^2, \\ C_{r,2} = \dfrac{(C_{a,2})^{t_1} \cdot (C_{a,4})^{t_2}}{(C_{r,1})^{x_1}} \bmod n^2, \end{cases} \quad (11)$$

such that m_c can only be recovered by V_a with the shared secret key x_2 and the secret value s_a.
- The RSU also generates the ciphertext pairs of $\sum_{i=1}^{w} s_{i,d} + s_k$ and $\sum_{i=1}^{w} s_{i,e} + s_k$ respectively, and re-encrypts the derived ciphertext pairs with the secret key x_1, such that they can be decrypted by V_a,

$$\begin{cases} C_{r,3} = C_{d,1} \\ C_{r,4} = \dfrac{C_{t,1} \cdot (1 + s_k \cdot n)}{(C_{d,1})^{x_1}} \bmod n^2 \\ C_{r,5} = C_{e,1} \\ C_{r,6} = \dfrac{C_{t,2} \cdot (1 + s_k \cdot n)}{(C_{e,1})^{x_1}} \bmod n^2 \end{cases} \quad (12)$$

- To protect the data integrity, the RSU generates the corresponding message authentication code MAC_r, which is

$$MAC_r = H_0(C_{r,1} \| C_{r,2} \| C_{r,3} \| C_{r,4} \| C_{r,5} \| C_{r,6} \| TS). \quad (13)$$

Then the RSU transmits the data query response towards V_a, which is denoted as $Response_a = C_{r,1} \parallel C_{r,2} \parallel C_{r,3} \parallel C_{r,4} \parallel C_{r,5} \parallel C_{r,6} \parallel MAC_r \parallel TS$, as shown in Fig. 2 *(Message (4))*.

After receiving the data query response $Response_a$, V_a checks its validity by checking $MAC_r \overset{?}{=} H_0(C_{r,1} \| C_{r,2} \| C_{r,3} \| C_{r,4} \| C_{r,5} \| C_{r,6} \| TS)$. If it does hold, V_a performs the following steps to recover the queried vehicular sensory data:

- V_a recovers the value of m_c with the secret key x_2 and the secret value s_a, which is

$$m_c = u_c \cdot t_1 + v_c \cdot t_2 \bmod n = \frac{\frac{C_{r,2}}{(C_{r,1})^{x_2+s_a}} - 1 \bmod n^2}{n} \quad (14)$$

Then V_a computes the corresponding secret key $\beta_c = H(m_c)$.
- V_a also exploits the secret key x_2 to decrypt $(C_{r,3}, C_{r,4})$ and $(C_{r,5}, C_{r,6})$, and recovers the value of $\sum_{i=1}^{w} s_{i,d} + s_k$ and $\sum_{i=1}^{w} s_{i,e} + s_k$, respectively.
- V_a recovers the sum of the sensory data captured at segment c, i.e., $sum_{c,d}$ and the number of sensory data reports captured at segment c, i.e., $sum_{c,e}$, which is computed as $sum_{c,d} = \sum_{i=1}^{w} s_{i,d} + s_k \bmod q_c = \sum_{i=1}^{w} d_{i,c} + \beta_c$ and $sum_{c,e} = \sum_{i=1}^{w} s_{i,e} + s_k \bmod q_c = \sum_{i=1}^{w} e_{i,c} + \beta_c$, respectively. Then V_a derives the average of sensory data $\bar{d}_c$ captured at segment c, by computing $\bar{d}_c = \frac{sum_{c,d} - \beta_c}{sum_{c,e} - \beta_c}$.

4.4. Discussion

In this subsection, we analyse the probability when a data querying vehicle fails to retrieve the sensory data from the RSU, i.e., the probability of data querying failure. Here we define the probability of data querying failure of segment l as the probability of the scenario when the number of data reports generated at segment l is no higher than a pre-defined threshold th, and the number of data querying vehicles is above zero. Meanwhile, the data uploading period is defined as T.

Let $X_{d,l}(T)$ denote the number of data reports which contains segment l within a finite time period $[0, T]$ and $X_{q,l}(T)$ be the number of data queries of segment l during $[T, 2T]$, which follows the Poisson $(\lambda_{d,l} \cdot T)$ and the Poisson $(\lambda_{q,l} \cdot T)$ distribution [24]. Then we obtain the probability of $X_{d,l}(T) = x_d$ data uploading vehicles and the probability of $X_{q,l}(T) = x_q$ data querying vehicles of segment l respectively, which are

$$\begin{cases} P\{X_{d,l}(T) = x_d\} = e^{-\lambda_{d,l} \cdot T} \cdot \dfrac{(\lambda_{d,l} \cdot T)^{x_d}}{x_d!} \\ P\{X_{q,l}(T) = x_q\} = e^{-\lambda_{q,l} \cdot T} \cdot \dfrac{(\lambda_{q,l} \cdot T)^{x_q}}{x_q!} \end{cases} \quad (15)$$

Let N_d and N_q be the average number of data reports and data queries in the given district, respectively. Meanwhile, $p_{d,l}$ and $p_{q,l}$ denote the probability of data uploading and the probability of data querying on segment l, respectively, which satisfy the condition that $\sum_{l=1}^{k} p_{d,l} = 1$ and $\sum_{l=1}^{k} p_{q,l} = 1$. Suppose all the vehicles are moving in the district with a constant speed of V_0, and all the street segments are with the length of L m. Then the value of $\lambda_{d,l}$ and $\lambda_{q,l}$ can be calculated as

$$\lambda_{d,l} = \frac{N_d \cdot p_{d,l}}{L/V_0}, \; \lambda_{q,l} = \frac{N_q \cdot p_{q,l}}{L/V_0}. \quad (16)$$

The probability of data querying failure in segment l can be calculated as following

$$\begin{aligned} P\{X_{d,l}(T) \le th \,|X_{q,l}(T) > 0\} &= \frac{P\{X_{d,l}(T) \le th\}}{P\{X_{q,l}(T) > 0\}} \\ &= \frac{\sum_{x_d=0}^{th} P\{X_{d,l}(T) = x_d\}}{1 - P\{X_{q,l}(T) = 0\}} \\ &= \frac{\sum_{x_d=0}^{th} e^{-\lambda_{d,l} \cdot T} \cdot \frac{(\lambda_{d,l} \cdot T)^{x_d}}{x_d!}}{1 - e^{-\lambda_{q,l} \cdot T}}. \end{aligned} \quad (17)$$

Then we calculate the probability of data querying failure in the given district, which is

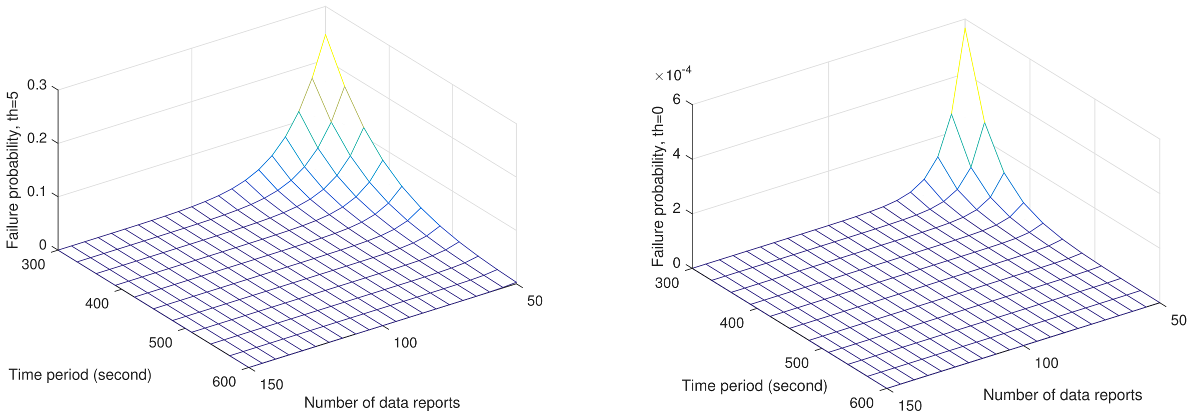
$$P_f = \sum_{l=1}^{k} p_{q,l} \cdot P\{X_{d,l}(T) \le th \,|X_{q,l}(T) > 0\}. \quad (18)$$

In Figs. 3(a) and 3(b), we show the probability of data querying failure in terms of the average number of data reports N_d and the data uploading period T, when the threshold are set to be $th = 5$ and $th = 0$ respectively. In Figs. 3(a) and 3(b), the number of data reports N_d ranges from 50 to 150, and the data uploading period T varies from 300 to 600 seconds. Meanwhile, we set the average number of the data querying vehicles to be $N_q = 50$, the length of each street segment is set to be $L = 500$ m, and the velocity of each vehicle is defined to be $V_0 = 10$ m/s. In addition, we assume that the number of segments in the given district are $k = 40$, and all the segments in the given district have an equal data uploading and data querying probability of $p_{d,l} = p_{q,l} = 0.025, l = 1, 2, \ldots, k$. As shown in Figs. 3(a) and 3(b), the probability of data querying failure decreases with respect to the increase of the number of data reports N_d and the increase of the data uploading period T.

5. Security analysis

In this section, we discuss the security properties of the proposed vehicular sensory data sharing scheme. Specifically, following the security requirements discussed earlier, our analysis will focus how the proposed scheme can achieve the security goal of location privacy preservation while taking the collusion into consideration, and achieve the security goal of the data integrity protection.

The proposed sensory data sharing scheme is location privacy-preserving during the sensory data collection phase. During the sensory data collection phase, since the locations are denoted in the level of segments, preserving location privacy of data uploading vehicles is realized through protecting the data collection segments. According to the Chinese Remainder Theorem, the sensory data



(a) Probability of data querying failure with threshold $th = 5$.

(b) Probability of data querying failure with threshold $th = 0$.

Fig. 3. Probability of data querying failure.

collection segments of V_i can be obtained from $(s_{i,d}, s_{i,e})$, which needs to be protected. In the proposed scheme, $s_{i,d}$ and $s_{i,e}$ are encrypted by the public key $(n, g, h = g^x)$ to generate the ciphertext pairs $(C_{i,1}, C_{i,2})$ and $(C_{i,3}, C_{i,4})$. Since we exploit the modified Paillier cryptosystem, which is proved to be semantically secure against the adaptive chosen-ciphertext attack [19], the messages contained in $(C_{i,1}, C_{i,2})$ and $(C_{i,3}, C_{i,4})$ are also semantic secure. In case of the collusion between a compromised RSU and a malicious vehicle, as the colluding entities can only recover the sum of the secret values $x_1 + x_2$, they cannot obtain the individual data report $s_{i,d}$ and $s_{i,e}$ without x (or x_0), which preserves the location privacy of each individual data uploading vehicle.

After receiving w data reports, the RSU aggregates the encrypted data reports by exploiting the homomorphic addition property of the modified Paillier cryptosystem, i.e., $(C_{d,1} = \prod_{i=1}^{w} C_{i,1} \bmod n^2$, $C_{d,2} = \prod_{i=1}^{w} C_{i,2} \bmod n^2)$ and $(C_{e,1} = \prod_{i=1}^{w} C_{i,3} \bmod n^2$, $C_{e,2} = \prod_{i=1}^{w} C_{i,4} \bmod n^2)$, without learning the content of the data aggregation results. Given the ciphertext pairs $(C_{d,1}, C_{d,2})$ and $(C_{e,1}, C_{e,2})$, the traffic management authority can recover the content of $\sum_{i=1}^{w} s_{i,d}$ and $\sum_{i=1}^{w} s_{i,e}$ with the private key x, and it can re-encrypt the ciphertext pairs with secret value x_0 to generate the ciphertext pairs $(C_{d,1}, C_{t,1})$ and $(C_{e,1}, C_{t,2})$. For the traffic management authority, it can only achieve the aggregated sensory data, while the individual sensory data report is kept from the traffic management authority, and the segments contained in an individual data report are protected. Thus, the location privacy of each data uploading vehicle can be preserved during the sensory data collection phase.

The proposed sensory data sharing scheme is location privacy-preserving during the sensory data acquisition phase. During the sensory data acquisition phase, a data querying vehicle V_a encrypts the segment identifier (u_c, v_c) with the public key (n, g, h_1) of the modified Paillier cryptosystem and a random number s_a, before sending towards the RSU. The segment identifier (u_c, v_c) contained in $(C_{a,1}, C_{a,2}, C_{a,3}, C_{a,4})$ is also semantic secure since $(C_{a,1}, C_{a,2})$ and $(C_{a,3}, C_{a,4})$ are also valid ciphertext pairs of the modified Paillier cryptosystem. Without the private key $x_1 + x_2$, the RSU cannot recover u_c and v_c contained in $(C_{a,1}, C_{a,2})$ and $(C_{a,3}, C_{a,4})$. In case of collusion, even though the colluding entities can obtain the value of $(x_1 + x_2)$, without the secret random value s_a, (u_c, v_c) cannot be fully recovered. Due to the homomorphic multiplicative property between one ciphertext and one plaintext, and the homomorphic additive property of the modified Paillier cryptosystem, the RSU calculates the ciphertext pair of $m_c = u_c \cdot t_1 + v_c \cdot t_2 \bmod n$, which is $((C_{a,1})^{t_1} \cdot (C_{a,3})^{t_2} \bmod n^2, (C_{a,2})^{t_1} \cdot (C_{a,4})^{t_2} \bmod n^2)$. The RSU also re-encrypts the derived ciphertext pair with the secret key x_1 and generates the ciphertext $C_{r,1} = (C_{a,1})^{t_1} \cdot (C_{a,3})^{t_2} \bmod n^2$ and $C_{r,2} = \frac{(C_{a,2})^{t_1} \cdot (C_{a,4})^{t_2}}{(C_{r,1})^{x_1}} \bmod n^2$, such that $(C_{r,1}, C_{r,2})$ can only be recovered by the data querying vehicle with x_2 and s_a. Meanwhile, the RSU generates a secret key β_l for each segment l, $l = 1, 2, \ldots, k$ through computing $\beta_l = H(u_l \cdot t_1 + v_l \cdot t_2)$, and then it structures $(\beta_1, \beta_2, \ldots, \beta_k)$ with $(\alpha_1, \alpha_2, \ldots, \alpha_k)$ to construct $s_k = \sum_{i=1}^{k} \beta_l \cdot \alpha_l$. By exploiting the homomorphic additive property of the modified Paillier cryptosystem, the RSU generates the ciphertext pairs of $\sum_{i=1}^{w} s_{i,d} + s_k$ and $\sum_{i=1}^{w} s_{i,e} + s_k$ respectively. The RSU also re-encrypts the derived ciphertext pairs with the secret key x_1 to generate $(C_{r,3}, C_{r,4})$ and $(C_{r,5}, C_{r,6})$, such that they can be decrypted with x_2.

After receiving $(C_{r,1}, C_{r,2}, C_{r,3}, C_{r,4}, C_{r,5}, C_{r,6})$, only V_a can recover the value of $\sum_{i=1}^{w} s_{i,d} + s_k$ and $\sum_{i=1}^{w} s_{i,d} + s_k$. Since $(C_{r,1}, C_{r,2})$ is a valid ciphertext pair of the modified Paillier cryptosystem, V_a can obtain the value of $m_c = u_c \cdot t_1 + v_c \cdot t_2 \bmod n$ with $x_2 + s_a$. Although V_a can recover the value of $m_c = u_c \cdot t_1 + v_c \cdot t_2 \bmod n$, it still cannot obtain the value of (t_1, t_2). Without (t_1, t_2), V_a cannot recover the secret keys of the unqueried segments, which prevents V_a from learning the sensory data aggregation of the unqueried segments. Given $(C_{r,3}, C_{r,4}, C_{r,5}, C_{r,6})$, even though all the vehicles can decrypt $(\sum_{i=1}^{w} s_{i,d} + s_k, \sum_{i=1}^{w} s_{i,e} + s_k)$ with private key x_2, without s_k, they still cannot learn the content of the sensory data in each data dimension. Even when V_a is one of the colluding entities and obtains the value of x_1, it still cannot obtain the content of the unqueried locations without (t_1, t_2), Thus, the security goal of location privacy preservation can be achieved during the sensory data acquisition phase.

The proposed data sharing scheme can achieve the security goal of data integrity protection. To protect data integrity, each data transmission flow between an RSU and a vehicle is attached with a message authentication code, which can be validated to guarantee the correctness of the data transmission. Thus, the security goal of data integrity protection can be achieved in the proposed scheme.

6. Performance evaluation

In this section, we evaluate the performance of the proposed privacy-preserving data sharing scheme, in terms of computation costs of the RSU and the vehicles, and communication overhead of the RSU and the vehicles.

6.1. Parameter setup

We evaluate the computation costs of the proposed scheme on a desktop with Intel i7-3770 3.4 GHz processor, 8 GB RAM,

Table 1
Parameters settings.

Simulation parameters	Value
κ, κ_1	$\kappa = 512, \kappa_1 = 24$
$p_1, p_2, n = p_1 p_2$	$\lvert p_1\rvert = \lvert p_2\rvert = 512, \lvert n\rvert = 1024$
$q_l, l = 1, 2, \ldots, k$	$\lvert q_l\rvert = \kappa_1 = 24$
k	$k = 40$: number of segments in a district
$d_{i,l}, i = 1, 2, \ldots, w, l = 1, 2, \ldots, k$	$\lvert d_{i,l}\rvert = 8$: length of each sensory data report (output voltage range 0–200 V [15])

and Window 7 platform. Meanwhile, we evaluate the performance of the proposed scheme with Java, and the detailed parameter settings are shown in Table 1. Given the parameter settings in Table 1, when $\kappa_1 = 24$, it has $|d| = 22$, and the maximum allowable number of data uploading vehicles in a district is 2^{13}, which fully satisfies the scalability requirement of the data uploading vehicles.

During the sensory data collection phase, we compare the proposed scheme with the traditional homomorphic data aggregation scheme, when the modified Paillier Cryptosystem is exploited. In the traditional data aggregation scheme, each data dimension is encrypted separately, i.e., during the data report generation phase, by choosing $(r_{i,d,l}, r_{i,e,l}) \in \mathbb{Z}_{n^2}$, the ciphertexts of V_i can be derived as $(C_{i,1,l}, C_{i,2,l}, C_{i,3,l}, C_{i,4,l})$, $l = 1, 2, \ldots, k$, where $C_{i,1,l} = g^{r_{i,d,l}} \bmod n^2$, $C_{i,2,l} = h^{r_{i,d,l}} \cdot (1 + d_{i,l} \cdot n) \bmod n^2$, $C_{i,3,l} = g^{r_{i,e,l}} \bmod n^2$ and $C_{i,4,l} = h^{r_{i,e,l}} \cdot (1 + e_{i,l} \cdot n) \bmod n^2$. Meanwhile, at the RSU and the traffic management authority side, each data report dimension is aggregated and decrypted independently.

During the data querying phase, we compare the proposed scheme with the traditional OT scheme, when the modified Paillier Cryptosystem is exploited. In the traditional OT scheme, the data query sent by the data requester is still in the format of $(C_{a,1}, C_{a,2}, C_{a,3}, C_{a,4})$, while the data query response is in the format of $(C_{r,1}, C_{r,2})$ and $(C_{r,3,l} = C_{t,1,l}, C_{r,4,l} = \frac{C_{t,2,l} \cdot (1+\beta_l)}{C_{t,1,l}^{x_1}} \bmod n^2, C_{r,5,l} = C_{t,3,l}, C_{r,6,l} = \frac{C_{t,4,l} \cdot (1+\beta_l)}{C_{i,3,l}^{x_1}} \bmod n^2)$, $l = 1, 2, \ldots, k$.

6.2. Computation complexity

For the proposed data sharing scheme, during the data collection phase, to generate an encrypted data report $C_{i,1} \parallel C_{i,2} \parallel C_{i,3} \parallel C_{i,4}$, V_i needs to perform 4 exponentiation operations in $\mathbb{Z}_{n^2}$ and 4 multiplication operations in $\mathbb{Z}_{n^2}$. Meanwhile, the traffic management authority performs 4 exponentiation operations in $\mathbb{Z}_{n^2}$ and 6 multiplication operations in $\mathbb{Z}_{n^2}$ for both re-encryption and decryption process. During the data acquisition phase, V_a performs 6 exponentiation operations in $\mathbb{Z}_{n^2}$ and 6 multiplication operations in $\mathbb{Z}_{n^2}$ to generate the encrypted data query $C_{a,1} \parallel C_{a,2} \parallel C_{a,3} \parallel C_{a,4}$. After receiving $C_{a,1} \parallel C_{a,2} \parallel C_{a,3} \parallel C_{a,4}$, the RSU takes 5 exponentiation operations in $\mathbb{Z}_{n^2}$ and 3 multiplication operations in $\mathbb{Z}_{n^2}$ to generate $C_{r,1} \parallel C_{r,2}$, and it takes 2 exponentiation operations in $\mathbb{Z}_{n^2}$ and 6 multiplication operations in $\mathbb{Z}_{n^2}$ to generate $C_{r,3} \parallel C_{r,4} \parallel C_{r,5} \parallel C_{r,6}$. After receiving $C_{r,1} \parallel C_{r,2} \parallel C_{r,3} \parallel C_{r,4} \parallel C_{r,5} \parallel C_{r,6}$, the vehicle takes 3 exponentiation operations in $\mathbb{Z}_{n^2}$ and 6 multiplication operations in $\mathbb{Z}_{n^2}$ to retrieve the value of m_c, $\sum_{i=1}^{w} s_{i,d} + s_k$, and $\sum_{i=1}^{w} s_{i,e} + s_k$.

We compare the proposed scheme with the traditional data aggregation scheme during the data collection phase, in which each vehicle generates a ciphertext for each data dimension during the data collection phase. Specifically, under this setting, for a k-dimension data report, it takes $4 * k$ exponentiation operations in $\mathbb{Z}_{n^2}$ and $4 * k$ multiplication operations in $\mathbb{Z}_{n^2}$ to generate $(C_{i,1,l}, C_{i,2,l}, C_{i,3,l}, C_{i,4,l})$, $l = 1, 2, \ldots, k$. Then the traffic management authority performs $4 * k$ exponentiation operations in $\mathbb{Z}_{n^2}$ and $6 * k$ multiplication operations in $\mathbb{Z}_{n^2}$ for the sensory data decryption and re-encryption. During the data acquisition phase, we compare the proposed scheme with the traditional OT protocol, in which V_a also performs 5 exponentiation operations in $\mathbb{Z}_{n^2}$ and 3 multiplication operations in $\mathbb{Z}_{n^2}$ to generate $C_{a,1} \parallel C_{a,2} \parallel C_{a,3} \parallel C_{a,4}$. After receiving $C_{a,1} \parallel C_{a,2} \parallel C_{a,3} \parallel C_{a,4}$, the RSU performs the $5+2*k$ exponentiation operations in $\mathbb{Z}_{n^2}$ and $3+6*k$ multiplication operations in $\mathbb{Z}_{n^2}$ to generate the queried data response $C_{r,1} \parallel C_{r,2}$ and $C_{r,3,l} \parallel C_{r,4,l} \parallel C_{r,5,l} \parallel C_{r,6,l}$, $l = 1, 2, \ldots, k$. Then the data querying vehicle spends 3 exponentiation operations in $\mathbb{Z}_{n^2}$ and 6 multiplication operations in $\mathbb{Z}_{n^2}$ to recover the data query results.

We run our experiments 1000 times, in the proposed scheme, during the data collection phase, the average computation cost of a vehicle is 72.83 ms and the average computation cost of the traffic management authority is 75.99 ms. In the proposed scheme, during the data acquisition phase, the average computation cost of a vehicle for data query generation is 109.65 ms and for data query response recovery is 57.10 ms; and the average computation cost of the RSU for the generation time of $C_{r,1} \parallel C_{r,2}$ is 55.14 ms and for the generation time $C_{r,3} \parallel C_{r,4} \parallel C_{r,5} \parallel C_{r,6}$ is 37.78 ms. In Fig. 4(a), we plot and compare the computation cost of the vehicle in terms of the data dimension, which ranges from 4 to 40, during the data collection phase. Meanwhile, in Fig. 4(b), we plot and compare the computation cost of the RSU in both schemes in terms of the data dimension which ranges from 4 to 40, during the data acquisition phase. The comparison results show that the proposed scheme can greatly reduce the computation complexity of the involved entities.

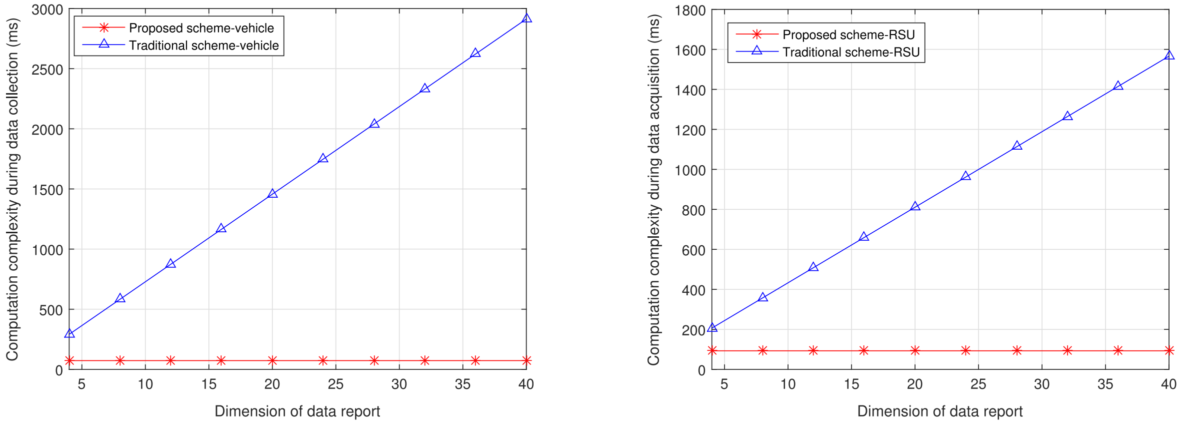
6.3. Communication overhead

In the proposed scheme, during the sensory data report collection phase, the vehicle-to-RSU communication overhead of $C_{i,1} \| C_{i,2} \| C_{i,3} \| C_{i,4} \| MAC_i$ is $2048 * 4 + 1024$ bits, if we choose 1024-bit n. If the RSU collects the w data reports from w data uploading vehicles, and the overall vehicle-to-RSU communication overhead is $2048 * 4 * w + 1024 * w$ bits. During the sensory data acquisition phase, the vehicle-to-RSU communication overhead $C_{a,1} \| C_{a,2} \| C_{a,3} \| C_{a,4} \| MAC_a$ is $2048 * 4 + 1024$ bits; meanwhile, the size of the RSU-to-vehicle communication overhead $C_{r,1} \| C_{r,2} \| C_{r,3} \| C_{r,4} \| C_{r,5} \| C_{r,6} \| MAC_r$ is $2048 * 6 + 1024$ bits.

For the traditional scheme, the overall vehicle-to-RSU communication overhead during the data report collection phase is $2048 * 4 * w * k + 1024 * w$ bits if we choose 1024-bit n. During the data acquisition phase, the vehicle-to-RSU communication overhead is $2048 * 4 + 1024$ bits, and the RSU-to-vehicle communication overhead brought by the data query is $2*2048 + 2048*4*k + 1024$ bits.

As shown in Figs. 5(a) and 5(b), during the sensory data collection phase, the communication overhead (including both data query and data query response) of the proposed scheme only increases in terms of the number of data reports, while the communication overhead of the traditional scheme increases in terms of both the number of data dimension and the number of data reports. Thus, the proposed scheme can significantly reduce the communication overhead during the sensory data collection phase.

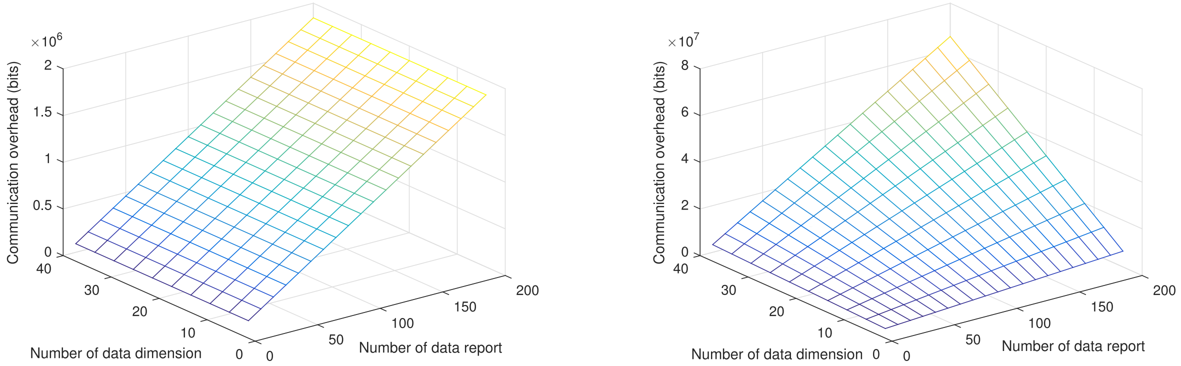
As shown in Fig. 6(a), during the sensory data acquisition phase, the communication overhead (both data query and data query response) of the proposed scheme does not increase with the number of data dimension, while the communication overhead of the



(a) Computational cost of a vehicle during data collection.

(b) Computational cost of an RSU during data acquisition.

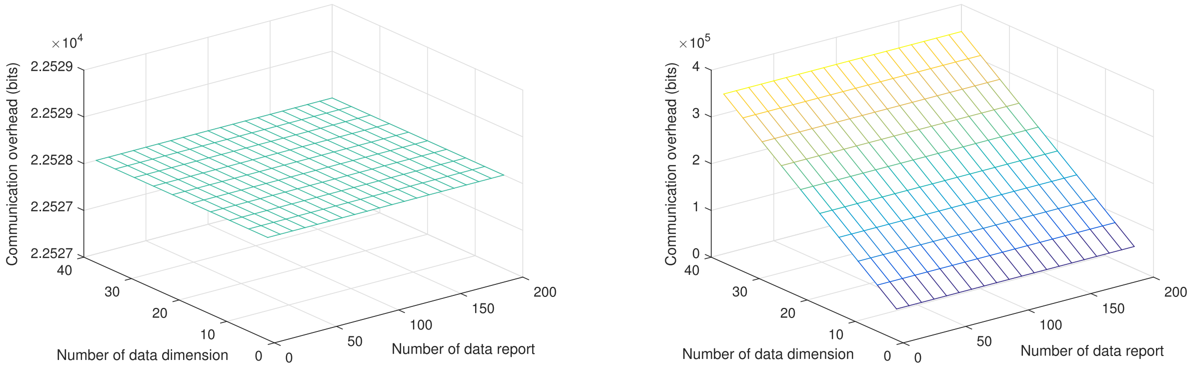
Fig. 4. Computational cost of the involved entities.



(a) Vehicle-to-RSU communication overhead during data collection of the proposed scheme.

(b) Vehicle-to-RSU communication overhead during data collection of the traditional scheme.

Fig. 5. Communication overhead during data collection.



(a) Communication overhead during data acquisition of the proposed scheme.

(b) Communication overhead during data acquisition of the traditional scheme.

Fig. 6. Communication overhead during data collection.

traditional scheme increases with the number of data dimension, as shown in Fig. 6(b). Thus, the proposed scheme can greatly reduce the communication overhead during the sensory data acquisition phase.

7. Related works

The studies of security and privacy in IoT, location privacy-preserving data aggregation and the location-based data query have gained great interests recently, and we briefly review some works closely related to this paper.

7.1. Security and privacy in IoT

A large amount of publications have been produced to solve the security and privacy problems in the era of securing IoT. With the additive homomorphic encryption technique, [25] proposed two schemes to achieve privacy preservation in trust evaluation, since the privacy leakage during the trust evidence collection is an important issue needs to be solved for the IoT devices. In the proposed scheme, two independent data servers with no collusion are introduced to achieve both privacy preservation and evaluation sharing. As defined in [26], privacy is the right of the users to remain confidentiality about their characteristics; while in the domain of IoT, personally identifiable information is widely varied with the involving applications. In [27,28], an review about the network architecture of the cloud-based IoT and the m-healthcare social networks are performed; meanwhile, the security and privacy threats are also identified, in terms of identity privacy, location privacy, node compromise attack, layer removing/adding attack, forward and backward security, data injection attack, and target-oriented compromise attack. In addition, a privacy preserving data aggregation scheme is designed without the public key homomorphic encryption. As pointed in [29], the local and private cloud providers, which are close to the end-users, can provide efficient services with low transmission delay and energy consumption, especially in the aspect of local data storage and sharing. However, for the location based services in IoV, the location privacy of the involved dynamically moving vehicles should be protected.

7.2. *Privacy-preserving data aggregation*

Due to the large volume of data generation in IoV, it is mandatory for an RSU (or a coordinating vehicle) to aggregate the sensory data reports generated by the vehicles. There exists many privacy-preserving data aggregation schemes [7,8,30,31] in the IoT domain, and we briefly review some privacy-preserving multi-dimensional data aggregation schemes, which can greatly save the scarce transmission bandwidth and are closely related to our work. In [7], a privacy-preserving multi-dimensional data aggregation scheme is proposed in smart grids, which utilized the homomorphic Paillier cryptosystem to achieve data aggregation and exploited a super-increasing sequence to compress the multi-dimensional electricity usage data into one single piece of composite data in the plaintext space. A lightweight privacy-preserving data aggregation scheme for the fog computing enhanced IoT environment with early injected false data filtering was also proposed in [8], which structures the multi-dimensional data report into one composite and filters the injected false data at the network edge.

However, the above schemes are based on the assumption that the number of data reports contained in each data dimension is pre-defined, and they are not adaptive to the scenario when the number of sensory data reports generated in each dimension is fluctuating. Our scheme enables the RSU to aggregate the different-size data reports captured at distinct locations with privacy preservation, while counting the number of sensory data reports in each dimension.

7.3. *Location privacy-preserving data query*

Location privacy preservation is regarded as one of the most important security requirements in vehicular networks. There mainly exists three categories of location privacy preservation solutions: pseudonym-based schemes [32,33], location-cloaking solutions (such as K-anonymity) [34,35], and cryptography-based schemes [11,12]. However, pseudonym-based schemes may still disclose the location information during the location-based data querying process, by associating the querying location with a particular individual, the identity of the data requester can still be inferred [36]. Meanwhile, in location-cloaking based schemes, it is hard to balance the trade-off between the privacy preservation and the accuracy of the location-based data query results [37]. Various cryptography-based location privacy-preserving data query schemes have also been proposed [11,12,14]. In [11], an efficient privacy-preserving location-based service system in the outsourced cloud is proposed, which is based on an improved homomorphic encryption technique over the composite order group. A fine-grained privacy-preserving location-based service (LBS) framework is proposed in [12], which achieves both location privacy preservation and high location-base service result accuracy without the involvement of any trusted third party. The OT technique is exploited in [14] to achieve the privacy-preserving location-based data query, in which the data server cannot learn the content of a user's query, while the user cannot obtain more than they are entitled [13]. However, if we exploit the scheme in our scenario, the communication overhead is proportional to the size of database, which cannot be directly introduced to the vehicular networks with scarce transmission bandwidth and stringent delay requirement.

However, the above schemes are designed for an outsourced cloud environment, which cannot be directly applied to our vehicular network scenario. Different from these works, our proposed scheme exploits the homomorphic modified Paillier Cryptosystem to hide the location information contained in the data query and calculate the commitment to the queried data, in which the RSU re-encrypts and shares the queried aggregated sensory data at the network edge without the involvement of the core entity. Moreover, the proposed scheme structures the multi-dimensional sensory data into one piece of composite data with the Chinese Remainder Theorem, which greatly reduces the communication overhead and computation complexity during the data acquisition phase.

8. Conclusion

In this paper, we have proposed an efficient and location privacy-preserving sensory data sharing scheme with collision resistance in IoV. By exploiting the modified Paillier Cryptosystem, the proposed scheme can achieve the location privacy-preserving multi-dimensional sensory data aggregation, while counting the number of data reports contained in each data dimension. Meanwhile, the proposed scheme also achieves the location privacy-preserving data acquisition at the network edge, i.e., the RSU re-encrypts the sensory data aggregation into ciphertexts which can only be decrypted by the data querying vehicle; meanwhile, the data querying location and the unqueried sensory data are still protected. Numerical analysis is performed to demonstrate the effectiveness of the proposed scheme in terms of the low data querying failure probability. Security analysis is conducted to validate the security properties, and experiments are also performed to show its efficiency in terms of the low computation cost and low communication overhead. In future work, we will implement the proposed scheme in our smart mobility test-bed and evaluate its performance.

Acknowledgements

This research was supported in part by the research grant S15-1105-RF-LLF URBAN from the Economic Development Board, Singapore, for the project of "Development Of NTU/NXP Smart Mobility Test-bed". This research was supported in part by Natural Sciences and Engineering Research (NSERC) Discovery Grants (no. Rgpin 04009), NBIF Start-Up Grant (Rif 2017-012), and URF Grant (no. 124419). This work was supported by Zhejiang Science Technology Department Project (No. 2016C33173), Zhejiang Education Department Project (No. Y201533448), and Natural Science Foundation of Zhejiang Province (LY17F020006).

References

[1] M. Gerla, E. Lee, G. Pau, U. Lee, Internet of vehicles: From intelligent grid to autonomous cars and vehicular clouds, in: IEEE World Forum on Internet of Things, WF-IoT 2014, Seoul, South Korea, March 6–8, 2014, pp. 241–246.

[2] U. Lee, E. Magistretti, M. Gerla, P. Bellavista, A. Corradi, Dissemination and harvesting of urban data using vehicular sensing platforms, IEEE Trans. Veh. Technol. 58 (2) (2009) 882–901.

[3] R.N. Rajaram, E. Ohn-Bar, M.M. Trivedi, Refinenet: Refining object detectors for autonomous driving, IEEE Trans. Intell. Veh. 1 (4) (2016) 358–368.

[4] S.E. Li, S. Xu, K. Deng, Q. Zhang, Cruising control of hybridized powertrain for minimized fuel consumption, in: Automotive Air Conditioning, 2016, pp. 267–289.

[5] K.M. Alam, M.K. Saini, A. El-Saddik, Toward social internet of vehicles: Concept, architecture, and applications, IEEE Access 3 (2015) 343–357.

[6] C. Xu, R. Lu, H. Wang, L. Zhu, C. Huang, PAVS: A new privacy-preserving data aggregation scheme for vehicle sensing systems, Sensors 17 (3) (2017) 500.

[7] R. Lu, X. Liang, X. Li, X. Lin, X. Shen, EPPA: An efficient and privacy-preserving aggregation scheme for secure smart grid communications, IEEE Trans. Parallel Distrib. Syst. 23 (9) (2012) 1621–1631.

[8] R. Lu, K. Heung, A.H. Lashkari, A.A. Ghorbani, A lightweight privacy-preserving data aggregation scheme for fog computing-enhanced iot, IEEE Access 5 (2017) 3302–3312.

[9] H. Zhu, F. Liu, H. Li, Efficient and privacy-preserving polygons spatial query framework for location-based services, IEEE Internet Things J. 4 (2) (2017) 536–545.

[10] L. Li, R. Lu, C. Huang, EPLQ: Efficient privacy-preserving location-based query over outsourced encrypted data, IEEE Internet Things J. 3 (2) (2016) 206–218.

[11] H. Zhu, R. Lu, C. Huang, L. Chen, H. Li, An efficient privacy-preserving location-based services query scheme in outsourced cloud, IEEE Trans. Veh. Technol. 65 (9) (2016) 7729–7739.

[12] J. Shao, R. Lu, X. Lin, FINE: A fine-grained privacy-preserving location-based service framework for mobile devices, in: 2014 IEEE Conference on Computer Communications, INFOCOM 2014, Toronto, Canada, April 27 - May 2, 2014, pp. 244–252.

[13] M. Naor, B. Pinkas, Oblivious transfer with adaptive queries, in: Advances in Cryptology - CRYPTO '99, 19th Annual International Cryptology Conference, Santa Barbara, California, USA, August 15–19, 1999, Proceedings, pp. 573–590.

[14] R. Paulet, M.G. Kaosar, X. Yi, E. Bertino, Privacy-preserving and content-protecting location based queries, IEEE Trans. Knowl. Data Eng. 26 (5) (2014) 1200–1210.

[15] Sensors, URL http://jenniskens.livedsl.nl/technical/tips/files/bosch.

[16] Anti-lock braking system, URL https://en.wikipedia.org/wiki/Anti-lockbrakingsystem.

[17] J. Yin, T.A. ElBatt, G. Yeung, B. Ryu, S. Habermas, H. Krishnan, T. Talty, Performance evaluation of safety applications over DSRC vehicular ad hoc networks, in: Proceedings of the First International Workshop on Vehicular Ad Hoc Networks, 2004, Philadelphia, PA, USA, October 1, 2004, 1–9.

[18] I.S. Association, et al., 802.11 p-2010-IEEE standard for information technology-local and metropolitan area networks-specific requirements-part 11: Wireless lan medium access control (mac) and physical layer (phy) specifications amendment 6: Wireless access in vehicular environments, 2010, URL http://standards.ieee.org/findstds/standard/802.11p-2010.html.

[19] E. Bresson, D. Catalano, D. Pointcheval, A simple public-key cryptosystem with a double trapdoor decryption mechanism and its applications, in: Advances in Cryptology - ASIACRYPT 2003, 9th International Conference on the Theory and Application of Cryptology and Information Security, Taipei, Taiwan, November 30 - December 4, 2003, Proceedings, pp. 37–54.

[20] C. Ding, D. Pei, A. Salomaa, Chinese Remainder Theorem: Applications in Computing, Coding, Cryptography, World Scientific, 1996.

[21] E. Ayday, J.L. Raisaro, P.J. McLaren, J. Fellay, J. Hubaux, Privacy-preserving computation of disease risk by using genomic, clinical, and environmental data, in: 2013 USENIX Workshop on Health Information Technologies, HealthTech '13, Washington, D.C., August 12, 2013.

[22] X. Liu, R.H. Deng, K.R. Choo, J. Weng, An efficient privacy-preserving outsourced calculation toolkit with multiple keys, IEEE Trans. Inf. Forensics Secur. 11 (11) (2016) 2401–2414.

[23] Z. Jin, H. Zuo, H. Du, Q. Wen, An efficient and provably-secure identity-based signcryption scheme for multiple pkgs, IACR Cryptol. ePrint Arch. 2008 (2008) 195.

[24] T.L. Saaty, Elements of queueing theory: with applications, 1961.

[25] Z. Yan, W. Ding, V. Niemi, A.V. Vasilakos, Two schemes of privacy-preserving trust evaluation, Future Gener. Comput. Syst. 62 (2016) 175–189.

[26] P. Porambage, M. Ylianttila, C. Schmitt, P. Kumar, A.V. Gurtov, A.V. Vasilakos, The quest for privacy in the internet of things, IEEE Cloud Comput. 3 (2) (2016) 36–45.

[27] J. Zhou, Z. Cao, X. Dong, A.V. Vasilakos, Security and privacy for cloud-based iot: Challenges, IEEE Commun. Mag. 55 (1) (2017) 26–33.

[28] J. Zhou, Z. Cao, X. Dong, X. Lin, A.V. Vasilakos, Securing m-healthcare social networks: challenges, countermeasures and future directions, IEEE Wirel. Commun. 20 (4) (2013).

[29] M.R. Rahimi, J. Ren, C.H. Liu, A.V. Vasilakos, N. Venkatasubramanian, Mobile cloud computing: A survey, state of art and future directions, MONET 19 (2) (2014) 133–143.

[30] C. Castelluccia, A.C. Chan, E. Mykletun, G. Tsudik, Efficient and provably secure aggregation of encrypted data in wireless sensor networks, TOSN 5 (3) (2009) 20:1–20:36.

[31] L. Chen, R. Lu, Z. Cao, K. Alharbi, X. Lin, Muda: Multifunctional data aggregation in privacy-preserving smart grid communications, Peer-to-Peer Netw. Appl. 8 (5) (2015) 777–792.

[32] R. Lu, X. Lin, T.H. Luan, X. Liang, X.S. Shen, Pseudonym changing at social spots: An effective strategy for location privacy in vanets, IEEE Trans. Veh. Technol. 61 (1) (2012) 86–96.

[33] K. Mano, K. Minami, H. Maruyama, Privacy-preserving publishing of pseudonym-based trajectory location data set, in: 2013 International Conference on Availability, Reliability and Security, ARES 2013, Regensburg, Germany, September 2–6, 2013, pp. 615–624.

[34] W. Ni, M. Gu, X. Chen, Location privacy-preserving k nearest neighbor query under user's preference, Knowl.-Based Syst. 103 (2016) 19–27.

[35] K. Vu, R. Zheng, J. Gao, Efficient algorithms for K-anonymous location privacy in participatory sensing, in: Proceedings of the IEEE INFOCOM 2012, Orlando, FL, USA, March 25–30, 2012, pp. 2399–2407.

[36] G. Ghinita, P. Kalnis, A. Khoshgozaran, C. Shahabi, K. Tan, Private queries in location based services: anonymizers are not necessary, in: Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2008, Vancouver, BC, Canada, June 10–12, 2008, pp. 121–132.

[37] M. Ghaffari, N. Ghadiri, M.H. Manshaei, M.S. Lahijani, P4QS: A peer to peer privacy preserving query service for location-based mobile applications, 2016, CoRR abs/1606.02373.

Qinglei Kong received the M.Eng. degree in electronic and information engineering from Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen, China, in 2015, and the B.Eng. degree in communication engineering from Harbin Institute of Technology, Harbin, China, in 2012. She is currently pursuing Ph.D. degree in School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, since 2015.

Rongxing Lu has been an assistant professor at the Faculty of Computer Science, University of New Brunswick (UNB), Canada, since August 2016. Before that, he worked as an assistant professor at the School of Electrical and Electronic Engineering, Nanyang Technological University (NTU), Singapore from May 2012 to August 2016. Rongxing Lu worked as a Postdoctoral Fellow at the University of Waterloo from May 2012 to April 2013. He was awarded the most prestigious "Governor Generals Gold Medal, when he received his Ph.D. degree from the Department of Electrical and Computer Engineering, University of Waterloo, Canada, in 2012; and won the 8th IEEE Communications Society (ComSoc) Asia Pacific (AP) Outstanding Young Researcher Award, in 2013. He is presently a senior member of IEEE Communications Society. His research interests include applied cryptography, privacy enhancing technologies, and IoT-Big Data security and privacy. He has published extensively in his areas of expertise (with more than 7500 citations from Google Scholar), and was the recipient (with his students and colleagues) of the Student Best Paper Award, ITS Summit Singapore 2015, the IEEE IES Student Best Paper Award 2014, the Best Paper Awards of TSINGHUA Science and Technology Journal 2014, IEEE ICCC 2015, IEEE WCNC 2013, BodyNets 2010, and IEEE ICCCN 2009. He was/is on the editorial boards of several international referred journals, e.g., IEEE NETWORK, and currently serves the technical symposium co-chair of IEEE GLOBECOM16, and many technical programme committees of IEEE and others international conferences, including IEEE INFOCOM and ICC. In addition, he is currently organizing a special issue on "security and privacy issues in fog computing in Elsevier Future Generation Computer Systems and a special issue on "big security challenges in big data era in IEEE INTERNET OF THINGS JOURNAL. Dr. Lu currently serves as the Secretary of IEEE ComSoc CIS-TC (Communications and Information Security Technical Committee).

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 co-authored about 200 international academic publications, including over 80 journal papers, over 140 conference papers and/or book chapters, and three academic books. His research interests are wireless networking and wireless network security. Dr. Ma is a member of a few technical committees in the IEEE Communication Society. He has been a member of the technical programme committees for over 100 international conferences. He has been a General Chair, Technical Symposium Chair, Tutorial Chair, Publication Chair, Publicity Chair, and Session Chair for over 50 international conferences. He serves as 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.