MDPI

Article

TripSense: A Trust-based Vehicular Platoon Crowdsensing Scheme with Privacy Preservation in VANETs

Hao Hu [1], Rongxing Lu [1,*], Cheng Huang [1] and Zonghua Zhang [2]

[1] School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore; hhu002@e.ntu.edu.sg; huangcheng@ntu.edu.sg
[2] IMT/TELECOM Lille, CNRS UMR 5157 SAMOVAR Lab, France; zonghua.zhang@telecom-lille.fr
* Correspondence: rxlu@ntu.edu.sg; Tel.: +65 6790-4519

Academic Editor: name
Version May 22, 2016 submitted to Entropy; Typeset by LaTeX using class file mdpi.cls

Abstract: In this paper, we propose a trust-based vehicular platoon crowdsensing scheme, named as TripSense, in VANET. The proposed TripSense scheme introduces a trust-based system to evaluate vehicles' sensing abilities and then selects those more capable vehicles in order to improve sensing results accuracy. Besides, the sensing tasks are accomplished by platoon member vehicles and preprocessed by platoon head vehicles before the data are uploaded to server. Hence it is less time-consuming and more efficient compared with the way where the data are submitted by individual platoon member vehicles. Hence it is more suitable in ephemeral network like VANET. Moreover, our proposed TripSense scheme integrates unlinkable pseudo-ID technique to achieve PM vehicle identity privacy, and employs a privacy-preserving sensing vehicle selection scheme without involving the PM vehicle's trust score to keep its location privacy. Detailed security analysis shows that our proposed TripSense scheme not only achieves desirable privacy requirements but also resists against attacks launched by adversaries. In addition, extensive simulations are conducted to show the correctness and effectiveness of our proposed scheme.

Keywords: Trust; Privacy; Platoon; Crowdsensing; Vehicular Ad Hoc Networks

1. Introduction

Envisioned as one of the most promising applications to implement intelligent transportation systems (ITS), vehicular platooning [1,2] has the potential to enhance road safety, improve traffic efficiency and reduce energy consumption due to air drag reduction [3]. At the same time, with the increasing popularity of mobile devices and sensing technologies, a new sensing paradigm, mobile crowdsensing, attracts attention from both academia and industry [5]. Different from traditional sensor networks, this new sensing paradigm leverages the power of crowds for large scale sensing tasks and fuels the evolution of the Internet of Things [4]. Many factories are built in remote areas where the sensor resources are limited, if the authority needs to inspect those factories, it can hardly collect information with the existing traditional sensor networks. However, given the fact that many highways go through the remote areas. One solution to this problem is to invite vehicles passing by those areas to take part in the crowdsensing tasks and utilize their sensed data (e.g., temperature, humidity, noise level, air pollution level, etc.).

However, due to the inherent openness of this platform, it is easy for vehicles to contribute corrupted data [6]. As a result, several research efforts have been made on ensuring the trustworthiness of the sensed data [7,8]. One possible solution is to establish the reputation system for evaluating the trustworthiness of volunteer contributions in participatory sensing applications [6].

Furthermore, this new data aggregation way may also bring in privacy concerns into the networks. For example, the sensed data could reveal the capacity of a vehicle's sensor and hence reveal the personal

information of the vehicle. Another factor that has always been a concern is the location privacy, since the locations of the vehicles are closely related to the drivers of those vehicles [9]. To achieve location privacy, one approach is to use unlinkable pseudonyms that are periodically changed when broadcasting messages [10–12]. However, pseudonyms do not always ensure privacy, as an example shown in Fig. 1, a platoon head vehicle is asking its platoon member vehicle V_i for participating in the sensing task. When V_i responds by sending its own reputation score ts_i at Day-1 and Day-2 respectively, the platoon head vehicle can still associate V_i in different days by associating its trust scores even when its pseudonym has been changed. The unchanged trust score of a vehicle reveals its location privacy and the platoon head vehicle can even derive the driving pattern of the platoon member vehicle. Therefore, it is compelling for us to build a trust system through which vehicles take its advantages without sacrificing their privacies and a data aggregation mechanism to ensure data privacy.

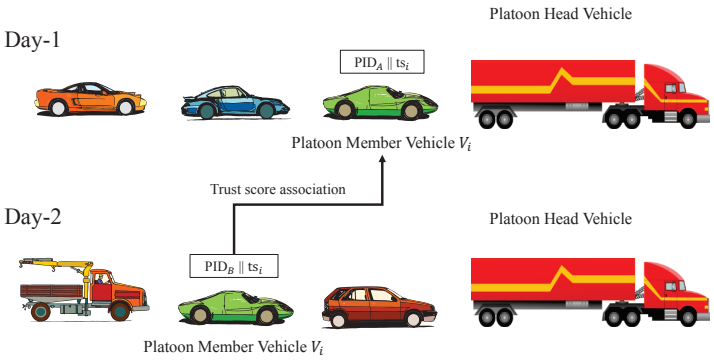

Figure 1. platoon head vehicle associates a platoon member vehicle's trust scores even when its pseudonym has been changed

Based on the observations above, we propose a trust-based vehicular platoon crowdsensing scheme, called TripSense, to improve sensing accuracy while achieving location and data privacy. This scheme is based on vehicular platooning technique to collect and aggregate data. At the same time, by establishing a trust model to measure the accuracy of a vehicle's sensed data, the service provider (SP) efficiently detects and then excludes the malicious or selfish vehicles who submit corrupted sensed data. Meanwhile, the proposed scheme is characterized by its ability to preserve the location privacy and data privacy of sensing vehicles. With assistance of platoon head vehicles, the communication overhead and computational cost can be greatly reduced. Specifically, our work features the followings:

- First, we establish the trust system based on Dirichlet distribution to evaluate the sensing accuracy of all sensing vehicles in our proposed TripSense scheme. The historical sensed data will be evaluated and finally form a reputation score. Therefore, the sensing accuracy will be improved greatly when the data are always collected from those high reputation sensing vehicles.
- Second, we propose the TripSense scheme by taking advantages of the unique features of vehicular platooning. In this scheme, platoon head vehicles firstly authenticate all sensing vehicles inside the platoon and then select some of them according to their trust values. Later, the sensed data from sensing vehicles will be collected and aggregated by platoon head vehicles before they are finally uploaded to the server. Compared with previous works, our proposed scheme reduces the communication overhead and hence is more suitable for the dynamic and ephemeral vehicular ad hoc network.
- Third, we design a privacy-preserving sensing vehicle selection scheme based on our trust system and a privacy-preserving data aggregation scheme based on the efficient commitment scheme in [13] such that platoon head vehicles can collect the data without leaking sensing vehicles' privacy.

The remainder of this paper is organized as follows. In Section 2, we formalize the system model, trust model and threat model considered in our work, and identify our design goals. In Section 3, we briefly recall the bilinear pairing and the Dirichlet distribution which have been applied in the trust and reputation system. In Section 4, the TripSense scheme is presented in details, together with the rationale how it can help the requesting

vehicles to choose a highly reliable relay vehicle without knowing its reputation score. Security analysis is then presented in Section 5, and the performance analysis is given in Section 6. Finally, we present the related work in Section 7 and draw conclusions in Section 8.

2. Problem Statement

In this section, we define the problem by formalizing the system model, security model and design goal.

2.1. System Model

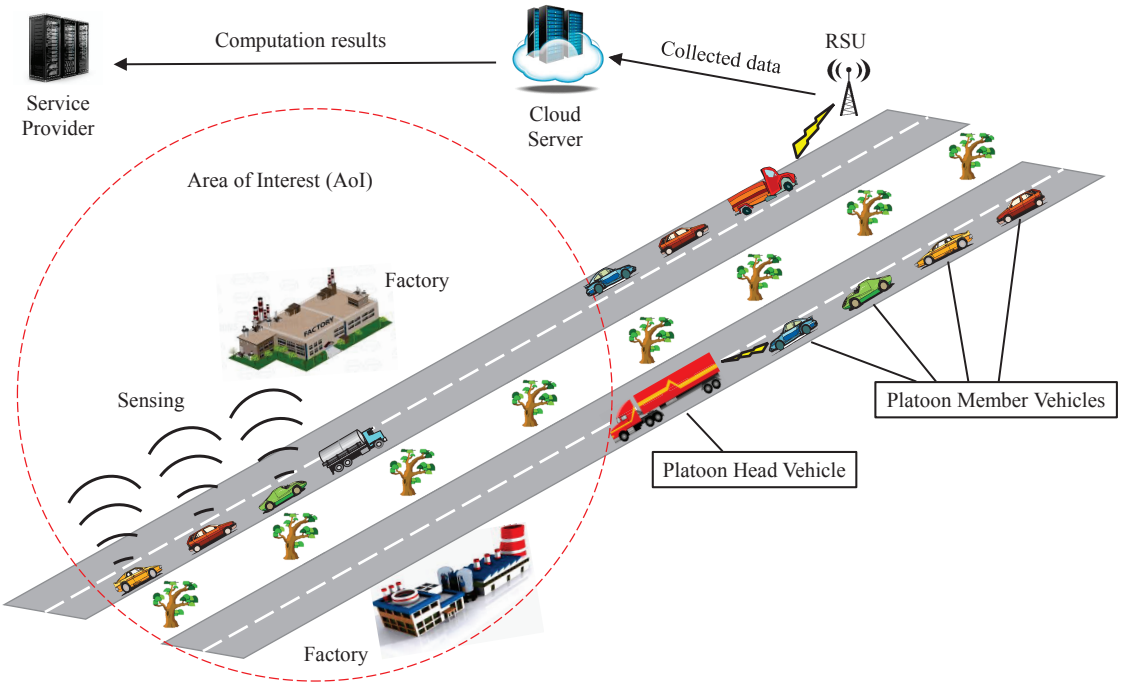

Figure 2. System model under consideration

In our model, the service provider (SP) wants to inspect an area of interest (AoI) located near a highway where many platoons pass by. As illustrated in Fig. 2, our system model consists of three roles: the service provider (SP), a cloud server (CS), the immobile roadside units (RSUs) along the highway and mobile vehicles traveling on the highway which are equipped with onboard units (OBUs) and powerful sensors.

Service Provider (SP): The SP is fully trusted because it is normally controlled by the authority who wants to inspect an area of interest by collecting the data of this AoI. The data collected is a vector of readings regarding, for example, air pollution level, noise level, temperature, humidity and so on. The duty of SP is to initialize the whole system, and distributes key materials to RSUs and vehicles. It is also responsible for storing and updating trust values for all vehicles.

Roadside Units (RSUs): The RSUs are subordinated by the SP, which are connected to the CS and SP via reliable communication channels. Equipped with wireless devices, RSUs are able to exchange data with the passing-by vehicles. However, due to the high cost of RSU installment and maintenance, especially in the early stage of VANET, RSUs are sparsely deployed along the highway. The RSUs will never disclose any internal information without permissions. However, we do not rule out the possibility that a portion of RSUs at the road side are compromised or the attackers even deploy bogus RSUs. Nevertheless, the SP can inspect all RSUs at high level: once the RSUs are compromised, they will be recovered or revoked soon by SP.

Cloud Server (CS): CS collects data from RSUs, then aggregates them in a privacy-preserving way. Besides, CS also computes the sensed data evaluations for vehicles and returns the results to SP. CS are assumed to be honest but curious about the sensed data of vehicles, which means that it follows the proposed scheme faithfully, but tends to be curious and disclose vehicles' privacy.

Vehicles: The vehicles are regarded as a group of highly mobile nodes equipped with OBUs which allow them to communicate with other vehicles or with RSUs. In the highway, vehicles follow a platoon head vehicles to form a platoon. With this driving pattern, the vehicles can be further divided into two categories:

- Platoon Head (PH) Vehicles $\mathcal{P}_0 = \{ph_1, ph_2, \cdots\}$: PH vehicles take the full control of the whole platoon when driving on the highway, they are responsible for the safety, user experience of all platoon member vehicles. Apart from that, they also claim a sensing task and submit sensed data to RSUs through V-2-I communication. PH vehicles are also honest but curious about the privacy of platoon member vehicles. In fact, PH vehicles could be malicious and provides untruthful aggregated data to server in order to subvert the system, or they may even collude with a bunch of PM vehicles with an object to victimize other PM vehicles. However, in this work, we do not consider this issue since it is not the main focus of this work.
- Platoon Member (PM) Vehicles $\mathcal{V}_0 = \{v_1, v_2, \cdots\}$: Each PM vehicle is equipped with various types of powerful sensors to meet the requirements of different tasks. Through V-2-V communication, a member vehicle authenticates itself and then submits its sensed data to the PH vehicles. Some of the PM vehicles compromised by adversaries launch attacks, other PM vehicles are all honest but curious.

Both PH vehicles and PM vehicles will be get paid by the SP for leading a platoon or contributing their sensed data.

2.2. Security Model

In our security model, we assume that all roles, except the SP, RSUs and malicious PM vehicles, are honest-but-curious, i.e., they will faithfully follow the protocol, but could also snoop into other role's privacy on account of some sensitive information available to them. In specific, we first consider the privacy requirements of PM vehicles.

Privacy requirements of platoon member vehicles: The privacy requirement of a PM vehicles include its data privacy, location privacy and identity privacy. Since the sensed data are private assets of a PM vehicle, which may reflect some sensitive information like the sensor accuracy or sensing ability, the PM vehicles will not disclose them to others. The location privacy requirement indicates that a PM vehicle will not let its PH vehicle know its past driving pattern, and the identity privacy means that the PM vehicle tries to keep his real identity secret. Meanwhile, each vehicle is also privacy-curious, i.e., it tends to disclose the privacy of other vehicles from other information available to him.

We assume that there are two kinds of adversaries according to their attacking abilities, the first kind tries to impersonate another authorized PM vehicle; the second kind is able to control a small portion of vehicles. Specifically, we list potential attacks as follows:

- *Impersonation attack:* The first kind of adversary may try to impersonate a PM vehicle to ask for a sensing task. However, this PM vehicle may not be qualified in the system. Once chosen to fulfil the task, these unqualified vehicles may submit inaccurate sensed data.
- *Malicious sensing attack:* The second kind more capable adversary is able to control a small fraction of the vehicles in the system who submit inaccurate sensed data deliberately to subvert the system. Another possible case is that the PM vehicle is selfish, so it report an arbitrary data without using sensors to save power.
- *Trust score association attack:* PH vehicles are honest but curious, in this case, if the trust score of a PM is directly given to its PH vehicle, its driving pattern will be disclosed. As described in Section 1, the reason is that since every time the PH vehicle can associate a PM vehicle in platoon according to the same trust score collected in different trips even though the pseudo-id has been changed.
- *Data analysis attack:* Due to the curious characteristics of both PH vehicles and PM vehicles, they may eavesdrop the transmission of sensed data and try to analyze the data. On the other hand, cloud server (CS) is also curious about the sensed data. If the data is not encrypted, those attacker can easily analyze the data in transmission.

2.3. Design Goal

Our design goal is to develop a trust-based privacy-preserving scheme to not only improve the sensing accuracy, but also preserve the privacy of sensing vehicles while resisting against the attacks launched by adversary. Specifically, the following desirable objectives need to be achieved.

- *Ensuring the sensed data reliability and accuracy.* According to our adversary model, the existence of selfish and malicious vehicles who submit corrupted sensed data will make the final result inaccurate and unreliable. Hence, our proposed scheme should be able to improve the sensed data accuracy by excluding those selfish and malicious vehicles' data.
- *Achieving privacy-preserving sensing vehicle selection, sensed data aggregation and evaluation.* The proposed scheme should achieve privacy requirements of PM vehicles. Particularly, i) the real identity of PM vehicles will never be disclosed; ii) when a PM vehicle replies to the PH vehicle, the PH vehicle can never know the exact trust score of this PM vehicle; iii) when a PH vehicle collects and aggregates the sensed data, it can never know what the data is; iv) the CS can never know the aggregated sensed data and the evaluations on those data.
- *Resisting against attacks launched by adversaries.* The proposed scheme should also be secure and reliable in VANET. Once an outside adversary launches some attacks, e.g., impersonation attack, data analysis attack, the proposed scheme should be able to detect them.

3. Preliminaries

3.1. Bilinear Pairing

Let $\mathbb{G}$ and $\mathbb{G}_T$ be two multiplicative cyclic groups of the same composite order n. Then, a bilinear pairing $e : \mathbb{G} \times \mathbb{G} \rightarrow \mathbb{G}_T$ will satisfy the following properties: i) Bilinear: Let $g, h \in \mathbb{G}$ and $a, b \in \mathbb{Z}_n^*$, then $e(g^a, h^b) = e(g, h)^{ab}$, ii) Non-degenerated: Let $g \in \mathbb{G}$ be a generator in $\mathbb{G}$, then $e(g, g) \neq 1_{\mathbb{G}_T}$, and iii) Computable: Let $g, h \in \mathbb{G}$, then $e(g, h)$ can be efficiently computed.

Definition 1 (Bilinear Parameter Generator)**.** A bilinear parameter generator $\mathcal{G}en$ is a probabilistic algorithm that takes a security parameter κ as its input, and outputs a 6-tuple $(p, q, g, \mathbb{G}, \mathbb{G}_T, e)$, where p, q are κ-bit prime numbers, $n = p \cdot q$, $(\mathbb{G}, \mathbb{G}_T)$ are two multiplicative groups of the same order n, $g \in \mathbb{G}$ is a generator, and $e : \mathbb{G} \times \mathbb{G} \rightarrow \mathbb{G}_T$ is a non-degenerated and efficiently computable bilinear map.

3.2. Beta Distribution

Defined on the interval of [0,1], beta distribution is a family of continuous probability distributions indexed by two parameters α and β. A random variable X beta-distributed with parameters α and β can be denoted by: $X \sim Beta(\alpha, \beta)$. Given that Gamma function is an extension of the factorial function where $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$. The probability density function (PDF) $f(x|\alpha, \beta)$ can be expressed by using gamma function Γ as: $f(x|\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$, where $0 \leq x \leq 1$, $\alpha > 0$, $\beta > 0$. The probability expectation value of the beta distribution is given by:$E(x) = \frac{\alpha}{\alpha+\beta}$.

Fig. 3 shows the PDF of beta distribution with different parameters α and β. It expresses the uncertain probability that a process will produce positive outcomes in future. Take an example, when $\alpha = 8$, $\beta = 2$, according to expectation equation, the probability expectation value of this type of beta distribution is $E(x) = 0.8$, which can be interpreted as the relative frequency of positive outcome is somewhat uncertain and that the most likely value is 0.8.

3.3. Dirichlet Distribution

The Dirichlet distribution is a family of continuous multivariate probability distributions parameterized by a priori parameter vector $\overrightarrow{\alpha}$. It is the conjugate prior distribution for the parameters of the multinomial distribution. In case of a binary state space, it is determined by the Beta distribution [14]. Generally, we can use the Dirichlet distribution to describe the probability distribution over a k-component random variable

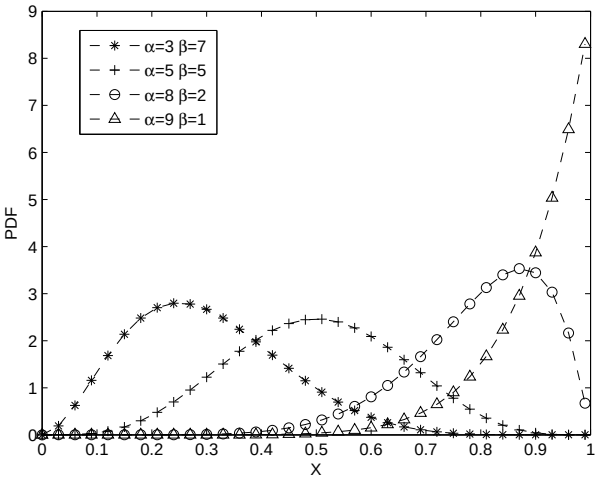

Figure 3. PDF of beta distribution with parameter α and β

$\overrightarrow{X} = \{X_1, X_2, \cdots, X_k\}$. If $\overrightarrow{p} = \{p_1, p_2, \cdots, p_k\}$ is the probability distribution vector of X, it satisfies $P\{\theta_{i-1} < X_i \leq \theta_i\} = p_i$ $(1 \leq i \leq k, \theta_i \in [0,1], \theta_{i+1} > \theta_i)$. The Dirichlet distribution captures a sequence of observations of k possible outcomes, those observations serve as the prior parameter $\overrightarrow{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_k)$, which denote the cumulative observations and initial beliefs of X. $\overrightarrow{p}$ is a k-dimensional random variable and $\overrightarrow{\alpha}$ is a k-dimensional random observation variable. The probability density function is given by:

$$f(\overrightarrow{p}\,|\,\overrightarrow{\alpha}) = \frac{\Gamma(\Sigma_{i=1}^{k}\alpha_i)}{\prod_{i=1}^{k}\Gamma(\alpha_i)}\prod_{i=1}^{k}p_i^{\alpha_i - 1} \tag{1}$$

where $0 \leq p_1, p_2, \cdots, p_k \leq 1$; $\sum_{i=1}^{k} p_i = 1$; $\alpha_1, \alpha_2, \ldots, \alpha_k > 0$. The expected value of the probability that X to be x_i given the observations vector $\overrightarrow{\alpha}$ is given by: $E(p_i|\overrightarrow{\alpha}) = \frac{\alpha_i}{\sum_{i=1}^{k}\alpha_i}$. Furthermore, if we let $\alpha_0 = \sum_{i=1}^{k}\alpha_i$, the variance of the event of X to be x_i is given by: $Var[X = x_i] = \frac{\alpha_i(\alpha_0 - \alpha_i)}{\alpha_0^2(\alpha_0+1)}$. If $i \neq j$, the covariance is: $Cov[X = x_i, X = x_j] = \frac{-\alpha_i\alpha_j}{\alpha_0^2(\alpha_0+1)}$.

4. Proposed TripSense Scheme

In this section, we propose our TripSense scheme, which consists of six parts: system initialization, trust-based privacy-preserving sensing vehicle selection, privacy-preserving sensed data aggregation, aggregated sensed data retrieval, privacy-preserving sensed data accuracy evaluation, and Dirichlet-based trust management.

4.1. System Initialization

We assume that a service provider (SP) will bootstrap the whole system. Specifically, given a security parameter κ, SP first generates the bilinear parameters $(p, q, g, \mathbb{G}, \mathbb{G}_T, e)$ by running $\mathcal{G}en(\kappa)$ and then computes by $h = g^q \in \mathbb{G}$. Next, SP chooses a secure symmetric encryption algorithm $\mathcal{E}nc()$, i.e., AES, and a collision-resistant cryptographic hash function $\mathcal{H} : \{0,1\}^* \rightarrow \mathbb{Z}_n^*$. In addition, SP chooses a random number $s \in \mathbb{Z}_n^*$ as the master key and computes $P_{pub} = g^s$, $n = pq$. Finally, SP keeps p, q secret and publishes $\{n, g, h, P_{pub}, \mathbb{G}, \mathbb{G}_T, e, \mathcal{H}, \mathcal{E}nc\}$

RSU REGISTRATION: For each RSU, SP first generates an identity RID, and then calculates its private key and public key as $(s_r; S_r)$, where s_r is randomly chosen in $\mathbb{Z}_n^*$ and $S_r = g^{s_r}$.

PH VEHICLE REGISTRATION: For any platoon head (PH) vehicle $ph_i \in \mathcal{P}_0$ which wants to participate in the sensing task, it has to register itself to SP and obtains a real identity RID_i. Then SP assigns the private key and public key to ph_i as (s_i,S_i), where s_i is randomly chosen in $\mathbb{Z}_n^*$ and $S_i = g^{s_i}$.

PM VEHICLE REGISTRATION: For each platoon member (PM) vehicle $v_j \in \mathcal{V}_0$ who wants to take part in the sensing task and contributes its sensed data, it first registers itself in the system. The following steps between SP and v_j show the registration process.

- SP first chooses a random number $k_0 \in \mathbb{Z}_n^*$ and uses $\mathcal{E}nc()$ to compute pseudo-IDs $\mathcal{PID}_j = \{PID_{j1}, PID_{j1}, \cdots\}$, where each pseudo-ID $PID_{jk} \in \mathcal{PID}_j$ is computed as $PID_{jk} = \mathcal{E}nc_{k_0}(ID_j||r_{jk})$ with a fresh random number $r_{jk} \in \mathbb{Z}_n^*$. Then, for each PID_{jk}, SP calculates its corresponding private key by $s_{jk} = \mathcal{H}(PID_{jk})^s$ and public key by $S_{jk} = g^{s_{jk}}$. Finally, SP sends $\mathcal{PID}_j$ and the corresponding public and private keys back to v_j via a secure channel.
- After receiving pseudo-IDs $\mathcal{PID}_j$ and their private keys, v_j verifies the correctness of each private key s_{jk} by checking $e(\mathcal{H}(PID_{jk}), P_{pub}) \stackrel{?}{=} e(s_{jk}, g)$.

TRUST REGISTRATION: Each registered PM vehicle v_j will be given a trust score ts_j by SP before it is able to take part in the sensing task, where $ts_j \in [0,1]$ with the precision of two decimal places. Initially, $ts_j = ts_0$. Besides, SP also defines L trust levels $\{TL_1, TL_2, \cdots, TL_L\}$ for all trust scores from 0 to 1. For instance, TL_1 is with (0,0.1], TL_2 is with (0.1,0.2], $\cdots$, TL_{10} is with (0.9,1]. Later, SP selects l random elements $\{y_1, y_2, \cdots, y_L \in \mathbb{Z}_n^*\}$ as master keys, and publishes the public keys as $\{Y_1 = g^{y_1}, Y_2 = g^{y_2}, \cdots, Y_L = g^{y_l}\}$.

- For a registered PM vehicle v_j with trust score $ts_j \in TL_x$, where $x \in [1, L]$, SP makes signatures for each of its pseudonym $PID_{jk} \in \mathcal{PID}_j$ as $A_{jk} = g^{\frac{1}{y_x + ts_j + s_{jk} + H(T_j)}}$, where T_j is the timestamp for updating the trust score of v_j.

TASK REGISTRATION: Before a task is broadcasted to PH vehicles, it should be registered by SP. First, the sensed data categories need to be decided, such as air pollution level, noise level, temperature, humidity, and so on. Second, the the format could be defined as $\overrightarrow{d} = (d^A, d^B, \cdots, d^Z)$, where each element denotes one category. Besides, SP also defines that each piece of data is with precision of two decimal places. In addition, the location of area of interest (AoI) is also included in the task. Finally, the SP will also decide a sensing vehicle trust level threshold TL_{TH} according to the accuracy requirements of the task to make sure that the sensed data only come from those more trusted vehicles.

4.2. Trust-based Privacy-preserving Sensing Vehicle Selection

We assume that there are a number of m PH vehicles in the system which would like to participate in the sensing task. They form a new set $\mathcal{P} = \{ph_1, ph_2, \cdots, ph_m\}$. For a specific PH vehicle $ph_i \in \mathcal{P}$, it needs to collect sensed data from all registered PM vehicles in its platoon and then selects those which meet the trust level requirement. Therefore, a trust-based privacy-preserving sensing vehicle selection scheme has been proposed as follows.

Step 1: When the platoon approaches the AoI, ph_i broadcasts its sensing requests $\{PID_i||H(T)^{s_i}||T\}$ to all platoon members, where T is the current timestamp.

Step 2: In ph_i's platoon, for each registered PM vehicle v_j, after receiving ph_i's requests, it first verifies whether ph_i is a registered PH vehicle by checking $e(H(T)^{s_i}, g) \stackrel{?}{=} e(S_i, H(T))$. If it holds, v_j accepts the requests, otherwise, v_j rejects ph_i. Then v_j responds to ph_i with $\{PID_{jk}||H(T')^{s_{jk}}||TL_x||\Pi||T'||T_j\}$ by calculating as follows, where T' is the current timestamp and T_j is the latest timestamp for updating v_j's trust score.

- Since the trust scores of PM vehicles are two decimal places, we need to expand them by 100 times before it can be encrypted. If v_j's trust score is with trust level TL_x where $x \in [1, L]$, v_j encrypts the expanded trust score ts_j as $C = g^{ts_j} h^r$, with a fresh random number $r \in \mathbb{Z}_n^*$.
- Similarly, the sensed data are all collected and expanded by 100 times. For one category in $\overrightarrow{d}$, v_j encrypts the expanded sensed data d_j as $D = g^{d_j} h^{r'}$, where r' is a random number in $\mathbb{Z}_n^*$.

- v_j chooses a pseudonym PID_{jk} and a random element $v \in \mathbb{Z}_n^*$ to calculate the following:

$$\begin{aligned} B &= A_{jk}^{v} = g^{\frac{v}{y_x + ts_j + s_{jk} + H(T_j)}} \\ E &= B^{-ts_j} \cdot g^{v} = g^{\frac{v(y_x + s_{jk} + H(T_j))}{y_x + ts_j + s_{jk} + H(T_j)}} \end{aligned} \tag{2}$$

- v_j randomly chooses ts_j'', r'', $v'' \in \mathbb{Z}_n^*$ and computes $C'' = g^{ts_j''} h^{r''}$, $E'' = B^{-ts_j''} g^{v''}$
- v_j calculates the proofs $\Pi = \{C, B, E, D, z_1, z_2, z_3, \phi\}$ as:

$$\begin{aligned} \phi &= H(C, B, E, D, C'', E'', H(T)^{s_j}) \\ z_1 &= ts_j'' + \phi \cdot ts_j \mod n \\ z_2 &= r'' + \phi \cdot r \mod n \\ z_3 &= v'' + \phi \cdot v \mod n \end{aligned} \tag{3}$$

Step 3: After receiving the response from v_j, ph_j first checks whether v_j is a registered PM vehicle by checking $e(g, H(T')^{s_{jk}}) \stackrel{?}{=} e(S_{jk}, H(T'))$. Then, ph_i checks whether the timestamp T' is relatively new. Next, ph_i checks whether v_j's trust score ts_j is with TL_x by checking $e(E, g) \stackrel{?}{=} e(B, Y_x \cdot S_{jk} \cdot g^{H(T_j)})$. If it does hold, ph_i calculates $\hat{C} = g^{z_1} h^{z_2} C^{-\phi}$, $\hat{E} = B^{-z_1} g^{z_3} E^{-\phi}$ and checks $\phi \stackrel{?}{=} H(C, B, E, D, \hat{C}, \hat{E}, H(T)^{s_i})$. If it holds, ph_i will finally check whether $TL_x \geq TL_{TH}$, and accept v_j's sensed data once its trust level satisfies the task requirement.

4.3. Privacy-preserving Sensed Data Aggregation

For each PH vehicle $ph_i \in \mathcal{P}$, where $i \in [1, m]$, we assume that a number of n_i PM vehicles meet the trust level threshold requirement, and those eligible PM vehicles form a set $\mathcal{V}_i = \{v_{i1}, v_{i2}, \cdots, v_{in_i}\}$. After a PH vehicle ph_i receives the sensed data from its PM vehicles, it selects those eligible data and aggregates them locally before submitting to CS for global aggregation. Therefore, a privacy-preserving data aggregation scheme has been proposed.

Local Aggregation: Take one data category, d^A, in $\overrightarrow{d}$ as an example. For simplicity, we omit the superscript and use D_{in_i} instead of $D_{in_i}^A$. As described in Section 4.2, ph_i collects encrypted sensed data from n_i PM vehicles as $D_{i1} = g^{d_{i1}} \cdot h^{r'_{i1}}, D_{i2} = g^{d_{i2}} \cdot h^{r'_{i2}}, \cdots, D_{in_i} = g^{d_{in_i}} \cdot h^{r'_{in_i}}$ together with their corresponding encrypted trust score: $C_{i1} = g^{ts_{i1}} \cdot h^{r_{i1}}, C_2 = g^{ts_{i2}} \cdot h^{r_{i2}}, \cdots, C_{in_i} = g^{ts_{in_i}} \cdot h^{r_{in_i}}$. Then ph_i aggregates the encrypted data D_{ij} and trust score C_{ij} of each PM vehicle $v_{ij} \in \mathcal{V}_i$ where $j \in [1, n_i]$ using the paring:

$$\begin{aligned} e(D_{ij}, C_{ij}) &= e(g^{d_{ij}} \cdot h^{r_{ij}}) = e(g^{d_{ij}} \cdot h^{r'_{ij}}, g^{ts_{ij}} \cdot h^{r_{ij}}) \\ &= e(g, g)^{d_{ij} ts_{ij}} \cdot e(g, h)^{d_{ij} r_{ij} + ts_{ij} r'_{ij}} \cdot e(h, h)^{r_{ij} r'_{ij}} \end{aligned} \tag{4}$$

Later ph_i aggregates all aggregated data of all PM vehicles in $\mathcal{V}_i$ together as:

$$\begin{aligned} \phi_i &= \prod_{j=1}^{m_i} e(D_{ij}, C_{ij}) = \prod_{j=1}^{m_i} e(g^{d_{ij}} \cdot h^{r_{ij}}) \\ &= \prod_{j=1}^{m_i} e(g, g)^{d_{ij} ts_{ij}} \cdot \prod_{j=1}^{m_i} e(g, h)^{d_{ij} r_{ij} + ts_{ij} r'_{ij}} \cdot \prod_{j=1}^{m_i} e(h, h)^{r_{ij} r'_{ij}} \\ &= e(g, g)^{\sum_{j=1}^{m_i} d_{ij} ts_{ij}} \cdot e(g, h)^{\sum_{j=1}^{m_i} (d_{ij} r_{ij} + ts_{ij} r'_{ij})} \cdot e(h, h)^{\sum_{j=1}^{m_i} r_{ij} r'_{ij}} \end{aligned} \tag{5}$$

Finally, when ph_i drives within the transmission range of an RSU, it submits ϕ_i together with all pseudo-IDs in $\mathcal{V}_i$ to CS via RSU.

Global Aggregation: Upon receiving reports from all PH vehicles in $\mathcal{P}$, CS aggregates those data together as follows, and passes the final result Φ and all pseudo-IDs of PM vehicles to SP.

$$\begin{aligned}\Phi &= \prod_{i=1}^{m} \phi_i = e(g,g)^{\sum_{i=1}^{m}\sum_{j=1}^{n_i} d_{ij}\cdot ts_{ij}} \\ &\cdot e(g,h)^{\sum_{i=1}^{m}\sum_{j=1}^{m_i}(d_{ij}r_{ij}+ts_{ij}r'_{ij})} \cdot e(h,h)^{\sum_{i=1}^{m}\sum_{j=1}^{m_i} r_{ij}r'_{ij}}\end{aligned} \tag{6}$$

4.4. Aggregated Sensed Data Retrieval

Once SP receives the aggregated sensed data Φ from CS, it retrieves it using its secret key p:

$$\begin{aligned}\Phi^p &= e(g,g)^{p\cdot\sum_{i=1}^{m}\sum_{j=1}^{n_i} d_{ij}\cdot ts_{ij}} \\ &\cdot e(g,h)^{p\cdot\sum_{i=1}^{m}\sum_{j=1}^{m_i}(d_{ij}r_{ij}+ts_{ij}r'_{ij})} \cdot e(h,h)^{p\cdot\sum_{i=1}^{m}\sum_{j=1}^{m_i} r_{ij}r'_{ij}} \\ &= e(g,g)^{p\cdot\sum_{i=1}^{m}\sum_{j=1}^{n_i} d_{ij}\cdot ts_{ij}}\end{aligned} \tag{7}$$

Similarly, we have:

$$\begin{aligned}&e(h,h)^{p\cdot\sum_{i=1}^{m}\sum_{j=1}^{m_i} r_{ij}r'_{ij}} = e(h,h^p)^{\sum_{i=1}^{m}\sum_{j=1}^{m_i} r_{ij}r'_{ij}} \\ &= e(h,1)^{\sum_{i=1}^{m}\sum_{j=1}^{m_i} r_{ij}r'_{ij}} = 1\end{aligned} \tag{8}$$

Since the aggregated data $\sum_{i=1}^{m}\sum_{j=1}^{n_i} d_{ij}\cdot ts_{ij}$ is in small space, we can use the method of exhaustion to retrieve them. From the pseudo-IDs in $\mathcal{V}_i, i \in [1,m]$, SP is able to find their real identities and corresponding trust scores $ts_{ij}, i \in [1,m], j \in [1,n_i]$. Then SP computes the sensed data d_0 using a weighted majority method:

$$d_0 = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n_i} d_{ij}\cdot ts_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{n_i} ts_{ij}} \tag{9}$$

For each category, there will be a sensed result, therefore the sensed result vector $\overrightarrow{d_0}$ will be $\overrightarrow{d_0} = (d_0^A, d_0^B, \cdots, d_0^Z)$. After shrinking 100 times, SP will get the final sensed result.

4.5. Privacy-preserving Sensed Data Accuracy Evaluation

We assume that all PM vehicles in each PH vehicle ph_i's platoon contribute their sensed data in the task, where $ph_i \in \mathcal{P}, i \in [1,m]$. These PM vehicles form a set, denoted as $\mathcal{V} = \{v_1, v_2, \cdots, v_n\}$, where n is the total number of these PM vehicles. After the sensed result is computed, SP would like to evaluate the sensed data accuracy in this task for each PM vehicle who contributes its data. Specifically, for $v_k \in \mathcal{V}$, from Section 4.3, we learn that CS stores v_k's pseudo-ID and encrypted sensed data for sensing category A as: $D_k^A = g^{d_k^A}\cdot h^{r'^A_k}$. The evaluation score $f_k \in [0,1]$ for v_k in this task can be calculated by following the steps below:

Step 1: Given that there are many sensing categories in the sensed data vector, SP first defines a tolerance value for each sensing category in sensed result vector $\overrightarrow{d}$. We denote these tolerance values as another vector $\overrightarrow{d_t} = (d_t^A, d_t^B, \cdots, d_t^Z)$. The tolerance value can be explained in this way: if the difference between the sensed data and sensed result is larger than tolerance level, the sensed data accuracy is unacceptable and $f_k = 0$. Besides, SP also defines the weights for different sensing categories as $\omega_A, \omega_B, \cdots$ which satisfies $\omega_A + \omega_B + \cdots + \omega_Z = 1$.

Step 2: We take sensing category A as an example, SP needs to calculate the difference between sensed result and sensed data $\Delta d_k^A = |d_0^A - d_k^A|$. When there are many vehicles in the VANET, the computation cost are to large for SP so it should be done by CS in a privacy-preserving way as follows. Note that, in case d_0^A is not an integer, SP rounds it off to the nearest integer.

- SP encrypts the sensed result d_0^A as: $D_0^A = g^{d_0^A} \cdot h^{r'^A_0}$, where $r'^A_0 \in \mathbb{Z}_n^*$, and sends D_0^A to CS.
- CS pairs the encrypted sensed data D_k^A from v_k and the inverse of encrypted sensed result, $D_k^{A^{-1}}$, as: $\alpha_k^A = e(D_0^A, D_k^{A^{-1}})$. In case that $d_0 < d_k$, CS also pairs the data as: $\beta_k^A = e(D_0^{A^{-1}}, D_k^A)$. After calculation, CS passes the results together with the v_k's pseudo-ID $\{PID_k||\alpha_k^A||\beta_k^A\}$ to SP.
- Upon receiving CS's message, SP first finds the real identity of v_k according to its pseudo-ID PID_k, then retrieves $\Delta d_k^A = |d_0^A - d_k^A|$ using the same method of exhaustion proposed in Section 4.4.

Step 3: After calculating Δd_k^A, SP uses the similar way to calculate other sensing categories as: $\Delta d_k^B, \Delta d_k^C, \cdots, \Delta d_k^Z$. Finally, the evaluation score for v_k in this task is calculated by:

$$f_k = \omega_A(1 - \frac{\Delta d_k^A}{d_t^A}) + \omega_B(1 - \frac{\Delta d_k^B}{d_t^B}) + \cdots + \omega_Z(1 - \frac{\Delta d_k^Z}{d_t^Z}) \tag{10}$$

4.6. Dirichlet-based Trust Management

For a specific PM vehicle $v_k \in \mathcal{V}$, the SP would like to evaluate its trustworthiness from its evaluation scores. Since the trustworthiness of v_k reflects its performance in a long period, SP first collects v_k's evaluation scores in many tasks, denoted by a continuous random variable X ($0 \leq X \leq 1$). From these collected historic records, SP can estimate X's future distributions by using Dirichlet distribution. Since Dirichlet distribution is based on initial belief on an unknown event according to prior distribution. It provides a solid mathematical foundation for measuring the uncertainty of feedbacks based on historical data. Compared to Beta distribution which is more appropriate in a binary satisfaction level [15], Dirichlet distribution is more appropriate for multi-valued satisfaction levels [16]. In our case, the evaluation trustworthiness of user vehicles are described by continuous trust scores. Therefore, SP uses Dirichlet distribution to estimate the performance of candidate vehicles in the future and then build trust model accordingly.

In order to classify the historical and future evaluation scores, we also denote a number of l satisfaction levels of feedbacks as a set $\{\theta_1, \theta_2, \cdots, \theta_l\}$ ($\theta_i \in (0,1], i \in [1,l], \theta_i < \theta_{i+1}$). Let $\overrightarrow{p} = \{p_1, p_2, \cdots, p_l\} (\sum_{i=1}^{l} p_i = 1)$ be the probability distribution vector of X with respect to satisfaction levels, so that we have $P\{\theta_{i-1} < X_i \leq \theta_i\} = p_i (i = 1, 2, \cdots, l)$. To make it more mathematically precise, we define $\theta_0 = 0$ when $i = 1$, $X_i = 0$ is categorized into θ_1.

As described in Section 4.5, once v_k finishes many sensing tasks,the SP is able to collect v_k's historical evaluation scores, then we let $\overrightarrow{\gamma} = \{\gamma_1, \gamma_2, \cdots, \gamma_l\}$ denote the vector of cumulative evaluation score and initial belief of X. With a posterior Dirichlet distribution, $\overrightarrow{p}$ can be modeled as:

$$f(\overrightarrow{p}|\xi) = Dir(\overrightarrow{p}|\overrightarrow{\gamma}) = \frac{\Gamma(\Sigma_{i=1}^{l}\gamma_i)}{\prod_{i=1}^{l}\Gamma(\gamma_i)} \prod_{i=1}^{l} p_i^{\gamma_i - 1} \tag{11}$$

where ξ denotes the background information represented by $\overrightarrow{\gamma}$. Let:$\gamma_0 = \sum_{i=1}^{l} \gamma_i$. The expected value of the probability of $X_i \in (\theta_{i-1}, \theta_i]$ with the historical distribution of evaluation scores is given by:

$$E(p_i|\overrightarrow{\gamma}) = \frac{\gamma_i}{\gamma_0} \tag{12}$$

Consider the time factor of historical evaluation scores, we introduce a forgetting factor η and give greater weight to more recent evaluation scores:

$$\overrightarrow{\gamma^{(n)}} = \begin{cases} \overrightarrow{S}^{(0)} & (n = 0) \\ \sum_{i=1}^{n} \eta^{t-t_i} \overrightarrow{S}^{(i)} + c_0 \overrightarrow{S}^{(0)} & (n \geq 1) \end{cases} \tag{13}$$

where n is the total number of historical evaluation scores, $\overrightarrow{S^{(0)}}$ is the initial belief vector when $n = 0$. Since no prior information is available, all elements of $\overrightarrow{S^{(0)}}$ have equal probability which makes $\overrightarrow{S^{(0)}} = (\frac{1}{l}, \frac{1}{l}, \cdots, \frac{1}{l})$. Parameter $c_0 > 0$ is a weight on the initial beliefs. In the i^{th} sensing task of $v_k (i \in [1,n])$, $\overrightarrow{S^{(i)}}$ denotes the

satisfaction level of its evaluation score, which contains only one element set to 1 corresponding to the selected satisfaction level and all the other $l-1$ elements set to 0. t_i stands for the timestamp when the i^{th} task took place and t is the moment of running the algorithm. The forgetting factor is $\eta \in [0,1]$, smaller η means that the system is easier to forget the historical records and vice versa. In order to defend against on-off attack [17], we choose an adaptive value as: $\eta = c_1 \cdot (1 - ts_k)$, where c_1 is a parameter to control the forgetting factor, the larger value of c_1 makes the system more forgettable about the historical behaviors and vice versa. From the equation we can see that when v_k has a high trust score, its forgetting factor is small, which means that those good performances will be easily forgotten. On the contrary, once v_k provides a low accuracy sensed data, its trust score gets lower and forgetting factor becomes larger. This means that all of those poor sensing performances will be memorized and it takes even longer time for v_k to build up a high trust score again.

To calculate v_k's trust score when as sensing vehicle, we first assign the weight ω_i to each satisfaction level $\theta_i (i \in [1,l])$. Let p_i denote the probability that v_k's evaluation score is categorized into the satisfaction level of θ_i. $\overrightarrow{p} = (p_1, p_2, \cdots, p_l) | \sum_{i=1}^{l} p_i = 1$. We model $\overrightarrow{p}$ using Eq. 11, Eq. 12, Eq. 13. Let Y be the random variable denoting the weighted average of the probability of each evaluation score in $\overrightarrow{p}$, the trust score ts_k of v_k is represented as:

$$ts_k = E[Y] = \sum_{i=1}^{l} \omega_i E[p_i] = \frac{1}{\gamma_0} \sum_{i=1}^{l} \omega_i \gamma_i \tag{14}$$

where γ_i is the cumulated evidence that v_k's evaluation score is with satisfaction level of θ_i.

5. Security Analysis

In this section, we discuss the security and privacy properties of the proposed TripSense scheme. In specific, following the design goals discussed early, we examine whether the proposed TripSense scheme can achieve the desirable security and privacy requirements.

5.1. The proposed TripSense scheme is privacy-preserving for PM vehicles

- *PM vehicle's identity privacy and location privacy are preserved in the proposed TripSense scheme:* In our proposed TripSense scheme, each PM vehicle $v_j \in \mathcal{V}_0$ uses pseudo-ID PID_{jk} instead of real identity in the network. Hence, the identity privacy can be achieved. In addition, to preserve the location privacy of the PM vehicle, v_j changes its unlinkable pseudo-IDs at different trips and locations to ensure that its past and future trip and location information will not be linked by pseudo-IDs.

  However, as analyzed in Section 2.2, v_j still suffers from trust score link attack. Thus, in our proposed scheme, when a PH vehicle ph_i checks whether its PM vehicle v_j's trust score satisfies the task requirement, it uses discrete trust levels TL_x in place of accurate trust scores. In other words, v_j can prove itself a highly trusted vehicle in front of ph_i without revealing its exact trust score. In addition, v_j's trust score ts_j is encrypted as $C = g^{ts_j} h^r$ and its trust level is in PS's signature $A_{jk} = g^{\frac{1}{y_x + ts_j + s_{jk} + H(T_j)}}$, which make the other PM vehicles impossible to get either v_j's trust score or trust level.
- *The sensed data privacy preservation is achieved:* Once the sensed data are aggregated by a PM vehicle v_j, they are encrypted as: $D = g^{d_j} \cdot h^{r'}$. In the whole process of local aggregation and global aggregation, those data are all aggregated without decryption until it reaches SP, where SP is able to recover with its private key p. Therefore, unless the other vehicles know the private key p, the sensed data information will never be disclosed.

5.2. The proposed TripSense scheme achieves robustness against attacks launched by adversary

- *Resilience to malicious sensing attack:* According to our proposed scheme, the selfish or malicious vehicles which submit arbitrary sensed data will get low evaluation score will get low evaluation scores in the trust system. Those low evaluation scores will be accumulated and finally lead to low trust scores if they keep behaving in that way. When their trust scores are lower than threshold TL_{TH}, their sensed data will be excluded from data aggregation or they will be given a low weight in data aggregation due to low trust scores. In both ways, the attacker will be mitigated in our proposed scheme.

- *Resilience to Trust Score Spoofing Attack:* We assume that the majority of PM vehicles follow the scheme honestly, but we don't rule out a possibility that a small fraction of PM vehicles cheat PH vehicles by using the fake trust scores. There are two possible cases: one case is that the PM vehicle v_j spoofs a higher trust score ts'_j with the hope to participate in sensing ask. However, with pseudo-ID PID_jk, v_j's trust score ts_j is signed by SP as $A_{jk} = g^{\frac{1}{y_x+ts_j+s_{jk}+H(T_j)}}$, where y_x, s_{jk}, $H(T_j)$ indicate the trust level, private key and updating timestamp respectively. Therefore, without knowing the spoofed trust score ts'_j's master key y'_x, v_j is unable to launch attack. Another case is that v_j provides a fake trust score ts'_j after encryption as $C' = g^{ts'_j}h^{r'}$, $E' = B^{ts'_j}g^{v}$. To deal with this type of attack, PH vehicle needs to check $\phi \stackrel{?}{=} H(C, B, E, D, \hat{C}, \hat{E}, H(T)^{s_i})$, it won't hold once the original C and E are changed to C' and E'.
- *Resilience to Impersonation Attack:* Both PH vehicle and PM vehicle could be impersonated by unqualified vehicles which want to take part in the sensing tasks. Specifically, for an impersonated PM vehicle, it may submit false data and escapes punishments; for a PH vehicle, it may collect sensed data without submitting to CS and renders the sensed data result incomplete. However, this attack can be thwarted by our proposed scheme. In the initialization phase in Section 4.1, both PH and PM vehicles will be given a pair of private and public keys once they are registered. In Section 4.2, a PM vehicle v_j first checks PH vehicle ph_i by checking $e(H(T)^{s_i}, g) \stackrel{?}{=} e(S_i, H(T))$ before it accepts ph_i's sensing task requirement. The timestamp $H(T)$ is used to resist against replay attack. Similarly, the PH vehicle ph_i also checks $e(g, H(T')^{s_{jk}}) \stackrel{?}{=} e(S_{jk}, H(T'))$ before accepting the sensed data from v_j. As a result, our proposed TripSense scheme is resistant against impersonation attacks.

6. Performance Evaluation

We will evaluate the performance of our proposed TripSense scheme in this section, the numerical data is generated in MATLAB. The performance metrics used in the evaluation are: i) *trust scores variations* for different PM vehicles in terms of the task number; ii) *detection ratio* defined as the ratio of the number of detected malicious vehicles with respect to the total number of malicious vehicles with the increase of task number.

6.1. Simulation Settings

We design a simulation to evaluate our proposed TripSense scheme in which only a set of key factors are considered and specified in order to validate the PM vehicles' sensing accuracy. It is worth noting that the selected factors are related to the movement of vehicles and the packets collision problems. In this case, we simulate the proposed scheme in the environment of MATLAB where there are a total number of n registered PM vehicles. To ensure the fairness, we suggest that each PM vehicle provides m times of sensing report in totally m independent tasks.

Table 1. Simulation Parameter Settings

Notation	Definition	Value
n	registered PM vehicle number	500
ρ	malicious PH vehicle proportion	20%
l_v	sensing accuracy level (SAL)	0.05;0.2;0.7;0.95
m	number of tasks for each PM vehicle	50
c_0	initial belief weight	1
c_1	forgetting factor parameter	1
c_2	variance sensitivity	10
c_3	forgetting factor parameter	1
T_0	initial trust score	0.5

6.2. Modeling the Sensing Behaviors of PM Vehicles

Due to the lack of real data, we need to model the behaviors of not only PM vehicles who take part in the sensing tasks, especially for malicious vehicles, in order to test the performance of our proposed scheme.

Sensing accuracy level (SAL) of PM vehicles: We define a parameter as sensing accuracy level (PQL) $l_v \in [0,1]$ to describe the capability of a PM vehicle to provide high accuracy sensed data. A PM vehicle with higher l_v may submit more accurate sensing reports. Specifically, given a PM vehicle with l_v, we use the beta distribution to describe the performance quality variable X of that PM vehicle, the probability density function of beta distribution can be expressed as:

$$f(x|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \tag{15}$$

where $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx$. $f(x|\alpha,\beta)$ is the probability that a PH vehicle with PQL of l_v provides a service with the quality value of $x \in [0,1]$. Higher values of l_v imply that the PH vehicle provides a higher quality service. To achieve this goal, we define α and β as follows:

$$\begin{aligned} \alpha &= c_2 \cdot l_v \\ \beta &= c_2 \cdot (1-l_v) \end{aligned} \tag{16}$$

where c_2 is the parameter to control the variance of the distribution, when c_2 is given a larger value, the performance quality values will have a larger variance and vice versa. For a PH vehicle with SAL of l_v, the above model has the property of generating a service quality score which follows a beta distribution with the expectation $E(X) = l_v$. We define that the malicious vehicles are vehicles with SAL $l_v \leq 0.2$.

6.3. Simulation Results

6.3.1. Correctness

In this experiment, we target on comparing the trust scores between malicious and honest sensing PM vehicles with different sensing accuracy levels (SALs). For a better comparison, we choose two honest PM vehicles with FAL of $l_v = 0.7$ and $l_v = 0.95$ respectively. Besides, another malicious PM vehicles who provide corrupt sensed data are also put in the system. After "50" number of tasks, we plot their trust scores in Fig. 4.

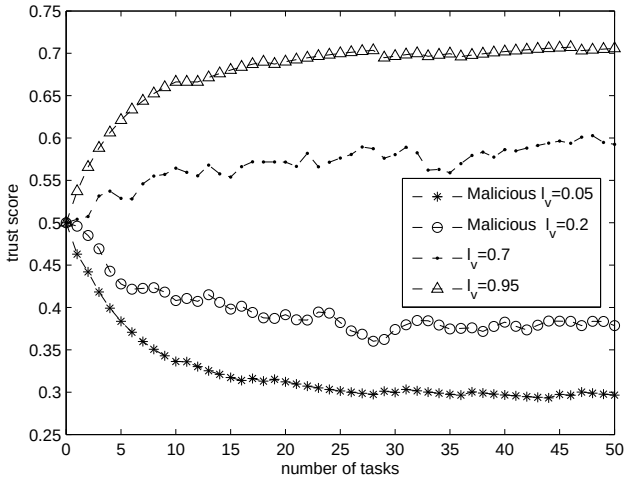

Figure 4. Trust scores comparison between honest PM vehicles with $l_v = 0.7$, $l_v = 0.95$ and malicious PM vehicles with $l_v = 0.05$, $l_v = 0.2$

We notice that the trust scores of all PM vehicles converge after "30" tasks. It is obvious that the honest PM vehicles with $l_v = 0.7$ and $l_v = 0.95$ get the highest trust scores after the experiments, on the contrary, both of the attackers get the low trust scores. We also notice that a PM vehicle with larger SAL will achieve higher trust score, which shows the correctness of our trust model to identify PM vehicles according to their actual SALs.

6.3.2. Effectiveness

To demonstrate the effectiveness of our proposed scheme in detecting malicious PM vehicles. We define a proportion of $\rho = 20\%$ number of PM vehicles with the lowest l_v as "malicious PM vehicles". After the $m = 50$ tasks, all PM vehicles will re-ranked, so the detection ratio is defined as the ratio of "malicious PM vehicles" who remain lowest 20% in the new ranking list.

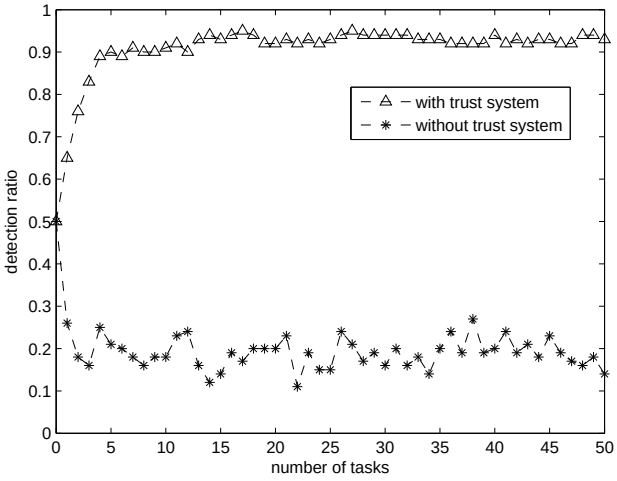

Figure 5. Detection ratio comparison between sensing system with / without trust system

Fig. 5 depicts the detection ratio between our proposed trust-based sensing system with a sensing system without trust, from the figure, we can see that our proposed system's detection ratio increases quickly with the increase of task numbers, and after around 5 tasks, it will be convergent to 92%. On the contrary, for a sensing system without a trust system, the selection of sensing PM vehicle is random and the detection ratio remains as low as 20%. Therefore, the effectiveness of our proposed scheme has been demonstrated.

7. Related Work

Recently, several research works have appeared on trust and reputation management in VANET [18–21], privacy preserving data aggregation [22–24] and crowdsensing [4,25], which are closely related to the techniques in our proposed TripSense scheme.

For trust and reputation management, Zhang et al. have done a survey for effective trust management in VANET in [18]. Specifically, it discusses challenging issues for trust management caused by the important characteristics of VANET environments, and points out that robustness should be drawn particular attention. Patwardhan et al. present a distributed reputation management scheme for VANET, which enables vehicles to quickly adapt to changing local conditions and provides a bootstrapping method for establishing trust relationships [19]. However, their scheme is not quite scalable and robust. Different from traditional entity-based trust model, Raya et al. suggest a data-oriented trust establishment framework [20]. By combining trust values of each piece of data together, their framework deals well with ephemerality and functions well in sparse areas. However, in dense urban area, due to large amount of data, their framework is less efficient.

For There has also been extensive work on data aggregation schemes in VANET [26,27]. These works share the same assumption that vehicles or servers are trusted and the communications are secure, which,

however, is not the case in real scenario. In reality, data can be eavesdropped and disclosed. Therefore, a lot of works have been done in privacy-preservation data aggregation [22–24]. Xing et al. have proposed M-PERM, a mutual privacy-preserving regression modeling approach to address the issue of keeping both participants and users data private while still utilizing them for analysis [22]. In this paper, data are aggregated at each node and each cluster, finally at the user with maximum privacy protection. He et al. present two privacy-preserving data aggregation schemes for additive aggregation functions, which bridges the gap between collaborative data collection and data privacy [23]. Bilogrevic et al. have proposed a state-of-art privacy preservation framework to preserve data utility and simultaneously provide user privacy [24]. Users in this framework only contribute encrypted and aggregated model of their files to the aggregator to tackle trust and incentive challenges.

Burke is the first to introduce the concept of participatory sensing, and describes an initial architecture to enhance data credibility, quality, privacy, and 'shareability' [25]. Ganti gives an overview of crowdsensing by introducing existing mobile crowdsensing applications and explaining their unique characteristics, illustrating various research challenges, and discussing possible solutions [4].

Combining the above privacy preserving data aggregation techniques and trust models together, our proposed TripSense scheme is focused on evaluating the platoon member vehicles' sensing ability based on the accuracy of their historical sensed data. Specifically, there are several aspects which make our proposed scheme different: first, we establish a trust system as a long-term evaluating metric. Second, we make use of platoon head vehicles for authentication local data aggregation, which greatly reduces the communication overhead between vehicles and infrastructures and hence very suitable in VANET. Third, our proposed scheme is privacy-preserving on platoon member vehicles' identities, locations and data.

8. Conclusion

In this paper, we have proposed a trust-based privacy-preserving scheme for vehicular platoon crowdsensing, called TripSense. The proposed scheme mainly focuses on establishing a trust model to improve the sensed data reliability and accuracy of the whole system, while preserving the location and data privacy of sensing vehicle in the process of sensing vehicle selection, sensed data aggregation and evaluation. Detailed security analysis shows that the proposed TripSense scheme can not only achieve vehicle's identity privacy, location privacy and data privacy, but also resists against adversary attacks on malicious sensing reports. Moreover, through extensive performance evaluation, we have demonstrated that our proposed scheme can achieve better sensing accuracy. In our future work, we will consider more scenarios in crowdsensing rather than data aggregation. Besides, we may also consider the collusion among PM and PH vehicles to launch attacks in order to victimize other vehicles.

Acknowledgments: This research is supported 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, and ZJNSF No.LR13F020003.

Author Contributions: Hao Hu conceived of the work, designed the schemes, analyzed the experimental results and drafted the manuscript. Rongxing Lu contributed to the original ideas and scheme designing. Cheng Huang contributed to the original ideas and scheme designing. Zonghua Zhang commented on and revised the work.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. C. Bergenhem, S. Shladover, E. Coelingh, C. Englund, and S. Tsugawa, "Overview of platooning systems," in *Proceedings of the 19th ITS World Congress, Oct 22-26, Vienna, Austria (2012)*, 2012.
2. H. Hu, R. Lu, Z. Zhang, and J. Shao, "Replace: A reliable trust-based platoon service recommendation scheme in vanet," *IEEE Trans. Vehicular Technology*, to appear.
3. A. A. Alam, A. Gattami, and K. H. Johansson, "An experimental study on the fuel reduction potential of heavy duty vehicle platooning," in *13th International IEEE Conference on Intelligent Transportation Systems (ITSC), 2010*. IEEE, 2010, pp. 306–311.
4. R. K. Ganti, F. Ye, and H. Lei, "Mobile crowdsensing: current state and future challenges," *IEEE Communications Magazine*, vol. 49, no. 11, pp. 32–39, 2011. [Online]. Available: http://dx.doi.org/10.1109/MCOM.2011.6069707

5. B. Guo, Z. Yu, X. Zhou, and D. Zhang, "From participatory sensing to mobile crowd sensing," in *2014 IEEE International Conference on Pervasive Computing and Communication Workshops, PerCom 2014 Workshops, Budapest, Hungary, March 24-28, 2014*, 2014, pp. 593–598. [Online]. Available: http://dx.doi.org/10.1109/PerComW.2014.6815273
6. K. L. Huang, S. S. Kanhere, and W. Hu, "Are you contributing trustworthy data?: the case for a reputation system in participatory sensing," in *Proceedings of the 13th International Symposium on Modeling Analysis and Simulation of Wireless and Mobile Systems, MSWiM 2010, Bodrum, Turkey, October 17-21, 2010*, 2010, pp. 14–22. [Online]. Available: http://doi.acm.org/10.1145/1868521.1868526
7. X. O. Wang, W. Cheng, P. Mohapatra, and T. F. Abdelzaher, "Artsense: Anonymous reputation and trust in participatory sensing," in *Proceedings of the IEEE INFOCOM 2013, Turin, Italy, April 14-19, 2013*, 2013, pp. 2517–2525. [Online]. Available: http://dx.doi.org/10.1109/INFCOM.2013.6567058
8. A. Dua, N. Bulusu, W.-C. Feng, and W. Hu, "Towards trustworthy participatory sensing," in *Proceedings of the 4th USENIX conference on Hot topics in security*. USENIX Association, 2009, pp. 8–8.
9. R. Lu, X. Lin, T. H. Luan, X. Liang, and X. S. Shen, "Pseudonym changing at social spots: An effective strategy for location privacy in vanets," *IEEE Trans. Vehicular Technology*, vol. 61, no. 1, pp. 86–96, 2012. [Online]. Available: http://dx.doi.org/10.1109/TVT.2011.2162864
10. M. Raya and J.-P. Hubaux, "Securing vehicular ad hoc networks," *Journal of Computer Security*, vol. 15, no. 1, pp. 39–68, 2007. [Online]. Available: http://iospress.metapress.com/index/ch4d4dg8yl2qhr0w.pdf
11. L. Buttyán, T. Holczer, and I. Vajda, "On the effectiveness of changing pseudonyms to provide location privacy in vanets," in *Security and Privacy in Ad-hoc and Sensor Networks*. Springer, 2007, pp. 129–141.
12. C. Zhang, X. Lin, R. Lu, P. Ho, and X. Shen, "An efficient message authentication scheme for vehicular communications," *IEEE Trans. Vehicular Technology*, vol. 57, no. 6, pp. 3357–3368, 2008. [Online]. Available: http://dx.doi.org/10.1109/TVT.2008.928581
13. G. Arfaoui, J. Lalande, J. Traoré, N. Desmoulins, P. Berthomé, and S. Gharout, "A practical set-membership proof for privacy-preserving NFC mobile ticketing," *CoRR*, vol. abs/1505.03048, 2015. [Online]. Available: http://arxiv.org/abs/1505.03048
14. A. Jøsang and J. Haller, "Dirichlet reputation systems," in *The Second International Conference on Availability, Reliability and Security, 2007. ARES 2007.* IEEE, 2007, pp. 112–119.
15. A. Jøsang and R. Ismail, "The beta reputation system," in *Proceedings of the 15th bled electronic commerce conference*, vol. 5, 2002, pp. 2502–2511.
16. C. J. Fung, J. Zhang, I. Aib, and R. Boutaba, "Dirichlet-based trust management for effective collaborative intrusion detection networks," *IEEE Trans. Network and Service Management*, vol. 8, no. 2, pp. 79–91, 2011. [Online]. Available: http://dx.doi.org/10.1109/TNSM.2011.050311.100028
17. Y. L. Sun, Z. Han, W. Yu, and K. J. R. Liu, "A trust evaluation framework in distributed networks: Vulnerability analysis and defense against attacks," in *INFOCOM 2006. 25th IEEE International Conference on Computer Communications, Joint Conference of the IEEE Computer and Communications Societies, 23-29 April 2006, Barcelona, Catalunya, Spain*, 2006.
18. J. Zhang, "A survey on trust management for vanets," in *25th IEEE International Conference on Advanced Information Networking and Applications, AINA 2011, Biopolis, Singapore, March 22-25, 2011*, 2011, pp. 105–112. [Online]. Available: http://dx.doi.org/10.1109/AINA.2011.86
19. A. Patwardhan, A. Joshi, T. Finin, and Y. Yesha, "A data intensive reputation management scheme for vehicular ad hoc networks," in *3rd Annual International ICST Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, MOBIQUITOUS 2006, San Jose, California, USA, July 17-21, 2006*, 2006, pp. 1–8. [Online]. Available: http://dx.doi.org/10.1109/MOBIQ.2006.340422
20. M. Raya, P. Papadimitratos, V. D. Gligor, and J.-P. Hubaux, "On data-centric trust establishment in ephemeral ad hoc networks," in *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE.* IEEE, 2008.
21. H. Hu, R. Lu, and Z. Zhang, "Tpsq: Trust-based platoon service query via vehicular communications," *Peer-to-Peer Networking and Applications*, pp. 1–16, 2015.
22. K. Xing, Z. Wan, P. Hu, H. Zhu, Y. Wang, X. Chen, Y. Wang, and L. Huang, "Mutual privacy-preserving regression modeling in participatory sensing," in *Proceedings of the IEEE INFOCOM 2013, Turin, Italy, April 14-19, 2013*, 2013, pp. 3039–3047. [Online]. Available: http://dx.doi.org/10.1109/INFCOM.2013.6567116
23. W. He, X. Liu, H. Nguyen, K. Nahrstedt, and T. F. Abdelzaher, "PDA: privacy-preserving data aggregation in wireless sensor networks," in *INFOCOM 2007. 26th IEEE International Conference on Computer Communications, Joint*

Conference of the IEEE Computer and Communications Societies, 6-12 May 2007, Anchorage, Alaska, USA, 2007, pp. 2045–2053. [Online]. Available: http://dx.doi.org/10.1109/INFCOM.2007.237

24. I. Bilogrevic, J. Freudiger, E. De Cristofaro, and E. Uzun, "What's the gist? privacy-preserving aggregation of user profiles," in *Computer Security - ESORICS 2014 - 19th European Symposium on Research in Computer Security, Wroclaw, Poland, September 7-11, 2014. Proceedings, Part II*. Springer, 2014, pp. 128–145.
25. J. A. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. B. Srivastava, "Participatory sensing," *Center for Embedded Network Sensing*, 2006.
26. B. Yu, J. Gong, and C. Xu, "Catch-up: a data aggregation scheme for vanets," in *Proceedings of the Fifth International Workshop on Vehicular Ad Hoc Networks, VANET 2008, San Francisco, California, USA, September 15, 2008*, 2008, pp. 49–57. [Online]. Available: http://doi.acm.org/10.1145/1410043.1410053
27. C. Lochert, B. Scheuermann, and M. Mauve, "Probabilistic aggregation for data dissemination in vanets," in *Proceedings of the Fourth International Workshop on Vehicular Ad Hoc Networks, VANET 2007, Montréal, Québec, Canada, September 10, 2007*, 2007, pp. 1–8. [Online]. Available: http://doi.acm.org/10.1145/1287748.1287750