Journal of Signal Processing Systems (2021) 93:209–219
https://doi.org/10.1007/s11265-020-01559-6

TPSense: A Framework for Event-Reports Trustworthiness Evaluation in Privacy-Preserving Vehicular Crowdsensing Systems

Zhenqiang Xu[1,3,4] · Weidong Yang[3,4] · Zenggang Xiong[2] · Jiayao Wang[1] · Gang Liu[3,4]

Received: 25 April 2020 / Revised: 14 May 2020 / Accepted: 20 May 2020 / Published online: 17 June 2020

Abstract
Vehicles with abundant sensors and sophisticated communication capabilities have contributed to the emergency of vehicular crowdsensing systems. Vehicular crowdsensing is becoming a popular paradigm to collect a variety of traffic event-reports in intelligent transportation research. However, event-reports trustworthiness and drivers' privacy are under the threats of the openness of sensing paradigms. This paper proposes TPSense, a lightweight fog-assisted vehicular crowdsensing framework, which guarantees data trustworthiness and users' privacy. Firstly, we convert the data trustworthiness evaluation problem into a maximum likelihood estimation one, and solve it through expectation maximization algorithm. Secondly, blind signature technology is employed to generate a pseudonym to replace the vehicle's real identity for the sake of drivers' privacy protection. Our framework is assessed through simulations on both synthetic and real-world mobility traces. Results have shown that TPSense outshines existing schemes in event-reports trustworthiness evaluation and the reliability of vehicles.

Keywords Vehicular crowdsensing · Data trustworthiness · Privacy-preserving · Artificial intelligence · Maximum likelihood estimation · Expectation maximization

1 Introduction

Recently, the integration of sensors and embedded computing devices triggers a novel sensing paradigm, namely mobile crowdsensing, which allows individuals to acquire sensory data from their surroundings. Mobile sensing is increasingly applied to the collection of definite information as to road conditions, environmental pollution and commodity prices supervision [1–3]. Likewise, the onboard units (OBUs) installed in vehicles offer vehicular crowdsensing services in intelligent transportation. With sensing devices, vehicles can send the basic driving information, collect traffic conditions and road conditions, and upload them to server for data aggregation and publishing. Traffic management department uses these data to provide traffic and road conditions, route planning services [4–7]. This acquisition model of raw data from vehicular crowdsensing explicitly lower various economic costs. In spite of the significant conveniences given by vehicular crowdsensing, this paradigm is still confronted with sensing data trustworthiness and privacy preserving key challenges because crowdsensing is an "open" system, in which any vehicle can join the sensing activities [8–10].

In classic vehicular crowdsensing application, a lot of information such as real-time location of vehicles and event-reports are collected and analyzed; however, the utility of traffic reports depends on its correctness. Malicious vehicles will generate some event-reports that conflict with the actual situation to be uploaded to local RSU or broadcast to other vehicles. If these false reports cannot be evaluated and filtered in time, the crowdsensing system will be attacked [8], which is demonstrated in Fig. 1.

As evident from Fig. 1, a vehicle detects an accident ahead by sensor, and then it reports this accident to the local RSU. The green symbol in the figure represents the real event-report generated by honest vehicles and the red one represents the false event-report by generated by malicious vehicles (red color). If the trustworthiness of the event-report cannot be

✉ Zenggang Xiong
jkxxzg2003@163.com; xzg@hbeu.edu.cn

[1] Information Engineering University, Zhengzhou 450001, China

[2] School of Computer and Information Science, Hubei Engineering University, Xiaogan 432000, China

[3] Key Laboratory of Grain Information Processing and Control (Henan University of Technology), Ministry of Education, Zhengzhou 450001, China

[4] College of Information Science and Engineering, Henan University of Technology, Zhengzhou 450001, China

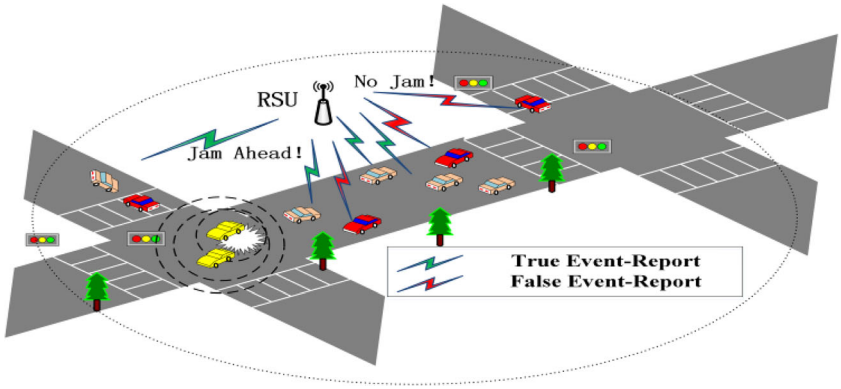

Figure 1 True event-report vs. false event-report in vehicular crowdsensing.

properly evaluated, then the RSU or traffic management department was misled and published wrong road conditions information, which ultimately led to the traffic system being hijacked by these malicious vehicles or false event-reports, which may cause serious traffic jams. Further, if these data are uploaded to the cloud server for evaluation, the best time for guiding the traffic flow may be delayed. In essence, accurate and timely assessment of the event-report is the foundation for the security of vehicular crowdsensing system; secondly, when the vehicle shares the sensing report with space-time attributes, it may leak large amount of drivers' privacy information, i.e., the user's identity, trajectory, health status, and behavior patterns, etc. Therefore, privacy protection mechanism in the vehicular crowdsensing scenario is also one of the research focuses of researchers [11, 12]. Of course, if the system provides complete anonymity, the credibility of the report is difficult to guarantee at the same time. Therefore, finding a solution that can achieve data trust and privacy protection has become a key Challenge. At present, in the field of location-based services (LBS), a large amount of research works on data collection based on privacy protection has been conducted [13–15]. This paper focuses on evaluating the truth-value of the event-reports rather than analyzing the formation of event-reports by vehicles. Vehicles gather data about the surrounding environment through the embedded sensors (such as OBU, camera, etc.) and then make the corresponding decisions. For example, an individual vehicle can determine whether there are potholes on the road surface through the changes of speed and vibration.

Furthermore, the abundant sensing data exerts a negative influence on data transmission and processing. It is worth mentioning that vehicle-generated data boasts local relevance in vehicle crowdsensing application. Local relevance indicates that the sensing data exhibit spatial-temporal features. The information of traffic jams may be valid only within half an hour and be useful to the vehicles nearby. Therefore, sending all reports to a remote cloud server for processing will cause a waste the bandwidth of network and response delay. With the emergence of fog computing, network edge devices are employed to perform storing and communication in large quantities. Hence, the necessity to upload all relevant data to the cloud server. The sensing data can be collected, stored and analyzed through fog nodes for local services. In addition, the data are processed locally. In brief, fog nodes not only save unnecessary communication bandwidth, but also support location-aware data management [8, 16]. In order to reduce the burden of data transmission and computation, fog networking is introduced into TPSense framework. Fog networking is a new architecture that provides storage, communication and other functions between terminal devices and the network. Computing functions are further applied to the edge of network for reducing the delay of data transmission. In the fog network, large-scale mobile terminals realize self-organized communication and collaboration through fog nodes. Data storage and calculation can be completed not in a large data center but on the fog nodes near the terminal. Therefore, RSUs are used as a fog node, which act as an intermediate route between the vehicles and the server, aggregates and analyze the data sensed by vehicles. Meanwhile, it can provide corresponding services for vehicles. In short, the introduction of fog nodes can not only reduce the communication bandwidth, but also contribute to the localization of data processing.

Confronted with the challenges mentioned above, we design a novel fog-assisted vehicular crowdsensing framework with event-report trustworthiness evaluation and privacy-preserving (TPSense), which treats RSU as a fog node. The RSU has powerful computing and storage capabilities, and it collect event-reports generated by vehicles, trustworthiness assessment, data aggregation, and broadcasts the results to surrounding vehicles and uploads them to the cloud server. Meanwhile, TPSense uses blind signature to achieve vehicle privacy protection. The following are the major contributions of the paper.

1) We mathematically formulate the event-reports trustworthiness evaluation problem (TEP), which is converted to a maximum likelihood estimation problem. Finally, TE-EM algorithm based Expectation-Maximization efficiently solves the TEP.

2) We propose a novel fog-assisted vehicular crowdsensing framework, which has two goal: event-report trustworthiness evaluation and vehicle privacy preserving.
3) We run a set of simulation experiments on synthetic data and real data to evaluate the effectiveness of our proposals and then make a comparison of the performance.

This paper is developed as follows. Section 2 analyzes related work. Section 3 introduces the TPSense framework, threat model and threat model. Section 4 presents the event-information trustworthiness scheme. Section 5 completes the experiments and analyzes results, while Section 6 reaches conclusions and discusses future work.

2 Related Work

2.1 Data Trustworthiness and Privacy Preservation in VANETs

As mentioned, data trustworthiness and privacy preservation is two important challenges in VANETs. Researchers proposed different data trustworthiness evaluation schemes for communication in both Vehicle-to-Vehicle and Vehicle-to-Infrastructure [17]. Most of these solutions are based on the reputation model proposed in [18]. An event report with the vehicle's reputation score is sent to a receiving entity (vehicles, RSU or cloud servers), and the latter confirms the true value of the report. The reputation of vehicle is continuously updated. A reputation-based announcement scheme is devise by Li et al. [19] to assess reliability of information, where vehicles behaviors can be collected. Accordingly, the scheme uses the accumulated feedbacks to testify the reliability of a vehicle. Similarly, in [20] a lasting reputation system in accordance with the vehicular daily behavior patterns is raised. A beacon-based trust management system is advocated by Chen et al. [17] to guard against internal attacks from sending malicious messages sent by internal malicious nodes proposed. Additionally, location privacy of VANETs is increased. However, it is proposed by Raya et al. [21] that data-oriented trust may better suit VANETs. The trust of the nodes are regarded as one parameter for data trustworthiness in this this scheme. The data trustworthiness varies in accordance with environment. However, this scheme does not explore privacy preservation. A dynamic approach was put forward to construct trusted vehicles groups by Tamper et al. [22].

Many solutions in VANETs require trusted third parties (TTP) updates the certificate revocation list (CRL) frequently and offer to the co-located RSUs [23]; The false messages are not to get circulated in VANETs. It is, however, demanding for CRL-based authentication practices as in processing the list as its size increases. In addition, encryption, and signature algorithms are combined into various schemes to achieve privacy preservation. Nevertheless, more efforts are necessary to meet efficient computing requirements.

2.2 Data Trustworthiness and Privacy Preservation in Participatory Sensing

To enhance the trustworthiness of sensing data in participatory sensing (PS) paradigm, researchers have attached great importance to the evaluation of sensing data and management of the reputation of participants. Huang et al. [24] and Yu et al. [25] proposed a reputation system by utilizing Gomeprtz function to calculate the participant's reputation score. Gomeprtz function model applies participants' past cooperation levels to reputation scores. The cooperation level is set by a module which performs an outlier detection algorithm. The main drawbacks of those models lie in the fact that the uncertainty factor in the trust assessment is not considered [26].

Fuzzy inference-based reputation model is used to calculate trust scores by way of the acquired evidence. Each participant is assumed to belong to a social network, and a social graph is able to describe the relationships of the participants.

Trustworthiness of a participant's contribution data results from such factors as their expertise, location, socializing. A trust of participant comes from those factors. However, participants' privacy (such as location, friendship relations) in social networks will lose protection. This is also not accepted by participants. In addition, several works in the literature [27, 28] have employed feedback-based reputation model to calculate the participants' trust scores in PS. Currently, some well-known crowdsensing applications such as Foursquare and Waze also use a rating feedback mechanism to allow service consumers to give positive, negative, or neutral evaluation information on products. Its advantages include simplicity, fast, and less expensive, which is the essence of PS paradigm.

In the above scheme, privacy protection of participants is less considered. Huang [29] further considers privacy needs of participants based on literature [24]. It assigns each participant a couple of pseudonyms, and relies on a trusted third party to pseudonyms change. Similarly, Christin et al. [30] used blind signatures and cloaking techniques to protect privacy. Wang et al. [31] proposes ARTSense to tackle the issue of trust under no identity. To solve the issue of time delay, Ma et al. [32] put forward two reputation maintenance schemes concerning privacy protection.

Different from the data trustworthiness evaluation based on the node reputation model mentioned above, this research assumes that the reliability of vehicles in participating sensing does not have any prior knowledge, and that the given research does not require a stable network topology, nor the reputation value of the vehicles.

3 TPSense Framework

Unlike previous research works, there are three differences. First, we focus on estimating the binary value of sensing reports; second, the report trustworthiness evaluation algorithm is executed on RSU, and does not need Manage and continuously update the reputation score of all vehicles. Third, in our framework, pseudonyms are used to replace the real identities of vehicles, to protect privacy of vehicles. Network model, attack model and corresponding hypotheses of the framework will be described in this part.

3.1 System Model

TPSense is composed of five entities as given in Fig. 2.

Trusted Authority (TA): To ensure system security and vehicles' privacy, we consider usually an authoritative traffic management department of government as TA. Initial parameters and cryptographic keys set by TA can help vehicles and RSUs to realize stronger privacy protection. Service providers (SP) and cloud servers (CS) are merged together to provide end users with various services including data storage, data processing, and data publishing (e.g., traffic queries from different vehicles,). Roadside Units (RSUs) are subordinated by the service providers and placed on the roadside. RSUs are viewed as a fog node with computation capabilities. Equipped with wireless devices, RSUs collect driving reports from crowd vehicles, authenticates vehicles' identity, verifies data trustworthiness and filters, and then uploads local traffic conditions to service providers and cloud servers. The benefit of data processing on RSU is to shorten the data collection process. RSUs feedback traffic conditions to vehicles faster. Therefore, it is not necessary to upload all sensing reports to the cloud server for processing. In addition, RSU is also capable of responding to the driver's query of road conditions or broadcasting the correct road condition information. As one major component of the TPSense framework, RSUs can fulfill data trustworthiness evaluation and privacy protection, and do not exist in mobile crowdsensing. Vehicle nodes (VNs) are equipped with onboard units (OBU), which enable direct communication with other vehicles and RSUs through DSRC, or 5G. A vehicle may periodically broadcast its driving information or occasionally send its traffic report to the local RSUs.

3.2 Threat Model

External and internal attackers may threaten the security of vehicular crowdsensing system. We focus on internal attack about data trustworthiness in this paper. Specifically, the malicious nodes in system may generate forged data and submit them to RSU or the server for their own benefit (for example, gaining credits for contributing to a crowdsensing task). Internal attacks mainly include malicious node or malicious conflict behavior attacks. The attacker may tamper, counterfeit or modify (the adversary may include intercepting the normal data transmission, forging or modifying data) data in conflict with the real traffic scene.

The paper proposes two assumptions.

1) RSU, SP and cloud server are trusted and impossible to be compromised. The reports generated by vehicles do not need to be transmitted to cloud server. In other words, the filtering of traffic event reports is done on the RSU.
2) Most of vehicles are honest and able to generate event-reports faithfully. The scope of this paper does not include

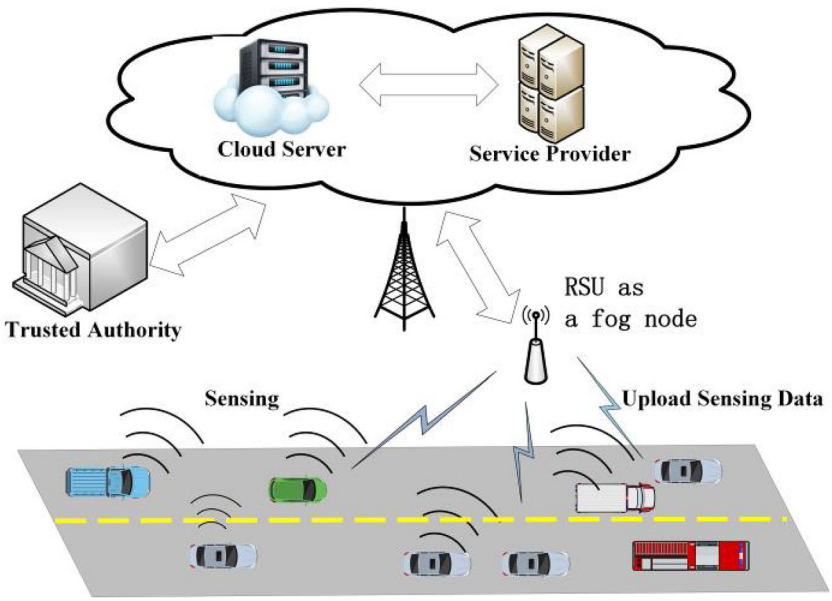

Figure 2 System architecture.

the following: Method for vehicle to generate data report and collusion attacks implemented by several malicious nodes.

Privacy threats in this system cannot be ignored. The RSUs and CS are honest-but-curious. However, they may get drivers' information ranging from drivers' identity, trajectories, and driving behavior modeling. Therefore, while ensuring the trustworthiness of the vehicle's report, users' privacy protection should also be considered. That is, the identity of vehicles should remain anonymous to other vehicles and RSUs.

4 Enabling Event-Information Trustworthiness Scheme for Vehicle Crowdsensing

Recently, several trust schemes as to crowd-sensing [31] have also been devised. Most of these schemes evaluate the trust of nodes and sensing data based on nodes' continuously updated reputation score. This section focuses on the implementation of sensing report trustworthiness evaluation model (SrTEM).

4.1 Sensing Report Trustworthiness Evaluation Module

As a core part of the whole vehicular sensing system, SrTEM provides a basis to reach our final objective. The goal of SrTEM is to judge the trustworthiness of event-reports. The content of the sensing data in the report depends on the application itself; it can also be traffic conditions, perception of air quality or noise, etc. Taking the vehicle crowdsensing traffic application as an example, RSU collected the traffic and road condition information (such as vehicle collision or road surface condition reports) which uploaded by the vehicle under its communication within a specified period. A trustworthiness evaluation algorithm evaluate false report from some selfish or malicious vehicles, and filter those false reports on RSU.

A. Problem Definition and Formalization

We take an event e_j ($e_j \in E$) as the object under monitor and a vehicle v_i ($v_i \in V$) generates a sensing report (R_j) about an event in vehicle crowdsensing system. This can be a traffic event, or potholes on a specific road, and etc.

Report R_j is presumed to be binary in this paper. $R_j \in \{T, F\}$ where T stands for True (e.g., "There is a traffic jam ahead at the specific location"), and F stands for False (e.g., "There is no traffic jam at the specific location"). We are concerned with the binary attributes of statements in vehicle crowdsensing.

In vehicle crowdsensing, A matrix VR is used to account for the reports from all vehicles V about events E, which is named the vehicle-report matrix. The element $VR_{i,\,j} = v$ indicates that the vehicle V_i declares that the value of R_j is v. It is also possible that a vehicle does not generate report, in which the corresponding element in the vehicle-report matrix VR is assigened value "Unknown" (U for short) indicating that the vehicle did not generate anything relevant about this event. Hence, each element V_iR_j in the *vehicle-report matrix VR* may have a value of T, F or U.

First, some definitions and notations are introduced. X^v denotes that X has value v. The reliability of vehicle V_i is t_i, meaning the probability that a report is true if the vehicle V_i report it. t_i is described as:

$$t_i = P\left(R_j^v | VR_{i,j}^v\right) \tag{1}$$

Let further define T_i^v to be the probability that vehicle V_i reports the value of R_j correctly. Similarly F_i^v denotes the probability of an incorrect report by V_i. To put it another way, there exists the probability that V_i reports that R_j has value $\overline{v}$ if its value is v. Here $\overline{v}$ is the complement of v. Formally, T_i^v and F_i^v are defined as follows:

$$T_i^v = P\left(\mathrm{VR}_{i,j}^v | R_j^v\right), \quad F_i^v = P\left(\mathrm{VR}_{i,j}^{\overline{v}} | R_j^v\right) \tag{2}$$

Note that a vehicle may not assert a report, $T_i^v + F_i^v \leq 1$. Therefore, we can get:

$$1 - T_i^v - F_i^v = P\left(\mathrm{VR}_{i,j}^U | R_j^v\right) \tag{3}$$

Let the probability that vehicle V_i generates the report to be of value v be p_i^v. Let $p_i^{\overline{v}}$ stand for the probability that V_i generates a report that has a value instead of v. Let d^v represent the prior probability that R_j has value v.

Applying t_i into the equation of T_i^v and F_i^v, we find the correlations between the terms based on the Bayesian theorem:

$$T_i^v = \frac{t_i \times v_i^k}{d^v} \quad F_i^v = \frac{(1-t_i) \times v_i^k}{1-d^v} \tag{4}$$

Table 1 summarizes the introduced notations.

Therefore, the trustworthiness evaluation issue of reports is treated as a maximum likelihood estimation problem: Based on vehicle-reports matrix VR, how to calculate the reliability of each vehicle as well as the trustworthiness of every report efficiently?

B. *Trustworthiness Evaluation* by Maximum Likelihood Estimation

In this part, the Expectation-Maximization (EM) algorithm is employed to address the maximum likelihood estimation

problem proposed in the preceding section. For simplicity, we assume that all vehicles independently generate reports in vehicle crowdsensing scene. Thus, the proposed algorithm is named TE-EM.

As a common algorithm in the field of machine-learning, EM is taken to acquire maximum likelihood estimates of parameters [33]. While using the EM algorithm, it is the most difficult to mathematically formulate the problem. Firstly, the likelihood function L(θ; X, Y) is defined, in which θ is the vector of unknown parameters, X stands for an acquired data set, and Y denotes the vector of latent variables. It is through EM that maximums likelihood estimate of θ and Y is acquired after iterative performing of E-step and M-step.

$$\text{E-step}: Q\left(\theta|\theta^{(n)}\right) = E_{Z|X,\theta^{(n)}}[\log L(\theta; X, Y)] \tag{5}$$

$$\text{M-step}: \theta^{(n+1)} = \text{argmax}_{\theta} Q(\theta) \tag{6}$$

The EM model applies to the given crowdsensing problem. A latent variable Y is introduced in each report to denote the event variables. The vehicle report matrix VR is known as the observed data X, and the parameter vector θ as:

$$\theta = \left\{ \left(T_i^v, F_i^v, d_k\right) \middle| \forall i \in V, v \in \{T, F\} \right\}$$

Then the likelihood function $L(\theta; X, Y)$ is acquired as follows:

$$L(\theta; X, Y) = P(X, Y|\theta) = \prod_{j=1}^{N} P\left(y_j\right) \times P\left(X_j|y_j, \theta\right) \tag{7}$$

Where N = |R| refers to the quantity of event variables, and X_j indicates all the reports from the vehicle about the j-th event.

Therefore, we can deduce the E-step as:

$$Q\left(\theta|\theta^{(n)}\right) = \sum_{j=1}^{N} \left\{ Y_1(n,j) \times \left[\sum_{i=1}^{M} \left(VR_{i,j}^1 log T_i^1 + VR_{i,j}^2 log F_i^1 + \left(1 - VR_{i,j}^1 - VR_{i,j}^2\right) \log\left(1 - T_i^1 - F_i^1\right) + log d_1 \right) \right] \right. \\ \left. + (1 - Y_1(n,j)) \times \left[\sum_{i=1}^{M} \left(VR_{i,j}^2 log T_i^2 + VR_{i,j}^1 log F_i^2 + \left(1 - VR_{i,j}^1 - VR_{i,j}^2\right) \log\left(1 - T_i^2 - F_i^2\right) + log(1 - d_1) \right) \right] \right\} \tag{8}$$

Where $Y_1(n,j) = p(z_j = 1|X_j, \theta^{(n)})$ refers to the fact of that the conditional probability of R_j has value k ($k = 1$) if the VR matrix is correlated to the j^{th}event and present estimate of θ. X_j stands for the j^{th} column of VR matrix. Note that $Y_2(n,j) = 1 - Y_1(n,j)$ and $d_2 = 1 - d_1$.

For the M-step, to get θ^* that maximizes $Q(\theta|\theta^{(n)})$, we set partial derivatives $\frac{\partial Q}{\partial T_i^k} = 0$, $\frac{\partial Q}{\partial F_i^k} = 0$ and $\frac{\partial Q}{\partial d_k} = 0$, we can get expressions of the optimal $T_{i,k}^*$, $F_{i,k}^*$ and d_k^*:

$$\begin{aligned}
T_i^{1(n+1)} &= T_i^{1*} = \frac{\sum_{j \in SJ_i^1} Y_1(n,j)}{\sum_{j=1}^{N} Y_1(n,j)} \\
F_i^{1(n+1)} &= F_i^{1*} = \frac{\sum_{j \in SJ_i^2} p\left(y_j = 1|X_j, \theta^{(n)}\right)}{\sum_{j=1}^{N} p\left(y_j = 1|X_j, \theta^{(n)}\right)} \\
T_i^{2(n+1)} &= T_i^{2*} = \frac{K_i^1 - \sum_{j \in SJ_i^1} Y_1(n,j)}{N - \sum_{j=1}^{N} Y_1(n,j)} \\
F_i^{2(n+1)} &= F_i^{2*} = \frac{K_i^2 - \sum_{j \in SJ_i^2} Y_1(n,j)}{N - \sum_{j=1}^{N} Y_1(n,j)} \\
d_1^{(n+1)} &= d_1^* = \frac{\sum_{j=1}^{N} Y_1(n,j)}{N}
\end{aligned} \tag{9}$$

Where SJ_i^1 and SJ_i^2 are the set of reports the vehicle V_i generates, each of which has true or false value respectively, and K_i^1 and K_i^2 indicate the size of reports in the two sets mentioned above.

4.2 User's Privacy Preservation

The TPSense framework utilize anonymous method to protect users' privacy. Specifically, a vehicle's real identity should not appear in the sensing data report in PS application. In this way, neither RSUs nor cloud server can link a report with a certain vehicle. Meanwhile, it is also necessary to change vehicle's pseudonym each time when a vehicle make the report submission. Otherwise, the real identity of a vehicle may still be leaked by analyzing the trajectories of the vehicle. In order to solve the above problems, our scheme utilizes Blind Signature technology [34] to generate a Blinded ID (BID) similar to a pseudonym for vehicle users. The specific process of using blind signature to generate pseudonyms for vehicles is as follows:

1) Message blinding: A vehicle randomly generates a blinding factor, and uses the signer's public key and blinding factor to process the information M. Then the vehicle can obtains the blinded information M', and sends the blinded message M' to the signer.
2) The blind message signing: When the signer receives the blinded message M' from the vehicle node, he just needs

Table 1 The set of notations.

Description	Notation
Set of vehicles	V
Set of Reports	R
Vehicle-Report matrix	VR
Vehicle Reliability	$t_i = P\left(R_j^v \middle| VR_{i,j}^v\right)$
Correctness Probability	$T_i^v = P\left(\mathrm{VR}_{i,j}^v \middle| R_j^v\right)$
Error Probability	$F_i^v = P\left(\mathrm{VR}_{i,j}^{\bar{v}} \middle| R_j^v\right)$

to prove that he receives the message without knowing the specific content. The signer encrypts M' with his private key and then gets the signature SIG (M'). Next, the signer sends it back to the vehicle node.

3) Signature recovery: After receiving the blindly signed message SIG (M'), the vehicle removes the blinding factor, and obtains the signature SIG (M) from SIG (M').
4) Pseudonym generation: The vehicle generates a pseudonym through combining the SIG (M) and the temporary public key.

5 Performance Evaluation

The effectiveness of the TPSense framework is evaluated based on the estimated error of vehicle reliability and the accuracy of event-reports evaluation including false positives and negatives rate.

In addition to providing anonymous privacy protection of vehicles, TPSense framework mainly distinguishes between true and false event-reports, which will avoid the attacks of false information from malicious nodes inside the system. We use python to develop a customized emulator to assess the efficacy of the framework. The performance of TPSense is further evaluated through synthetic and real dataset. The data contains movement trajectories of a large number of vehicles, randomly generated traffic events.

Table 2 Fundamental simulation parameters.

Parameters	Value
Simulation area (km^2)	10*10
Number of vehicles	50–200
Communication radius of RSU (km)	0.5
Nodes' Minimum Speed (km/h)	20
Nodes' Maximum Speed (km/h)	120

5.1 Evaluation of TPSense with Synthetic Data

The random waypoint model is employed to simulate the paths of vehicles. The fundamental simulation and system parameters are shown in Table 2. For the simulation dataset, several RSUs with a communication radius of 0.5 km are deployed randomly in simulation area. The coordinates of RSUs are saved into a file. We do not consider a vehicle's report unless its location attributes meet our requirements. The distribution function proposed in Barnwal et al. [35] is taken to generate random events in various simulated areas.

50–200 vehicles are taken in each experiment, and random effects are minimized through averaging the results. We assumed two types of false reports, one being submitted by those vehicles off the communication range of any RSU and the other being presented with a given probability instead of real event-reports from vehicles in the communication range.

In particular, unlike trustworthiness evaluation based on participant's reputation value in the previous schemes, TE-EM method in TPSense does not require continuous updating of the reputation value of participants. We use the classic truth-discovery algorithms in the field of data mining as the baseline. These algorithms are also commonly used to solve data fusion under information conflicts and not suitable for comparison with existing trusted models. Here, we choose the following four algorithms: Regular-EM [33], TruthFinder [36], Sums [37], and Voting. Since the existing literature has proved that simple Voting algorithm is less effective than the other three algorithms, so the experiment does not list voting algorithm. The widely accepted false positive and negative rate, estimation error are employed as the metric.

Two experiments are devised to assess the performance of TE-EM. A random number of vehicles and event-reports are produced by a simulator in Python. Each vehicle is given a random probability to represent its reliability. We assume event-reports to be binary in this paper. For each vehicle, some event-reports can be generated. It is worth mentioning that for the EM-Regular, we adapted it according to our requirements and that the report of higher probability was taken from the

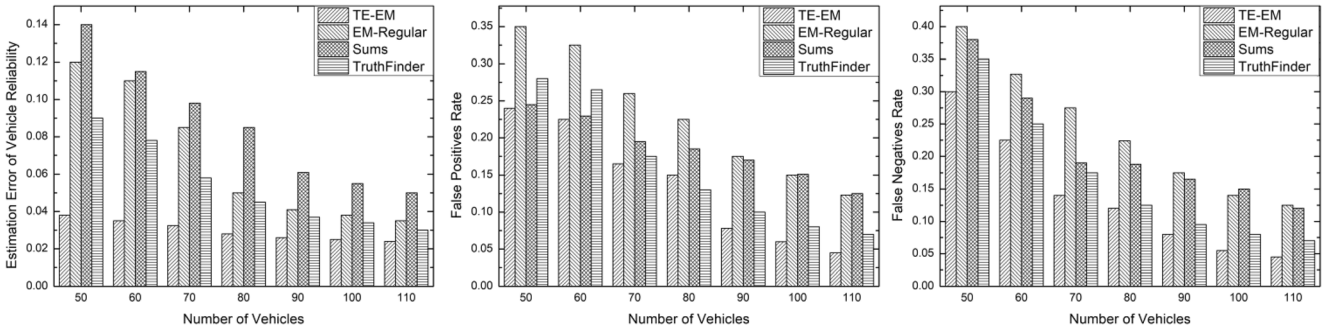

Figure 3 Impact of number of vehicles on metrics (using synthetic data).

two contrastive versions of the given event after the computation ends.

A. Impact of Number of Vehicles on Metrics

The estimation accuracy of our algorithm and baselines are compared in the first experiment through the different numbers of vehicles in crowdsensing scenario. The size of event-reports was set at 4000, with 2000 reports being correct and 2000 being misreported. The event-reports by per vehicle was set at 50 on average and the number of vehicles ranges from 50 to 110, results of which are presented in Fig. 3. TE-EM performs the best of four algorithms in predetermined metrics. In addition, all the algorithms perform better as the number of participants increase.

B. Impact of Number of Offset O on Metrics

As mentioned above, d^v stands for the prior probability that the value of R_j is v. For example, d^1 refers to a probability that the value of a randomly chosen report is true. d^1 is set at 0.5 in the initial phase of algorithm and an offset O indicates the disparity between d^1 and ground truth. In this experiment, the value of o range from 0 to 0.45, and the results are shown as in Fig. 4.

Figure 4 indicates that all the algorithms can produce reliable results as o changes. Comparatively speaking, TE-EM algorithm outshines the other three algorithms in estimation error, false positive and negative rate. Besides, the initial value of d^v make little difference to the performance of TE-EM algorithm.

5.2 Evaluation of TPSense with Real Data

In this experiment, we employed the outdoor temperature sensing data from CRAWDAD. The data set includes about 5000 sensing items opportunistically from 300 or so taxis in Rome within a day. Temperature data with a time mark, an identity and coordinates are uploaded to the server. The data set is from the previous research on participatory sensing [38].

The city area is segregated into 9 sensing regions with each part being 56 km^2 in area, and a base station is deployed in the central area of individual regions. There are 4 time spans for temperature sensing in a day, with each one lasting 6 h. We assign a temperature value to each taxi in a time span based on Gaussian distribution.

The value of ground truth temperature come from the mean temperature. We presume a temperature range with 10% offset and if a value produced falls into this range, we regard the report as true. In addition, if the location of taxi is not in the area, but the sensing data within the normal temperature range and the fake location data are upload, we consider its contribution to be false. We varied the percentage of honest vehicles from 0.6 to 0.9. In other words, each time a different ratio of taxis are designated as malicious nodes. Experimental evaluation still uses previous metrics.

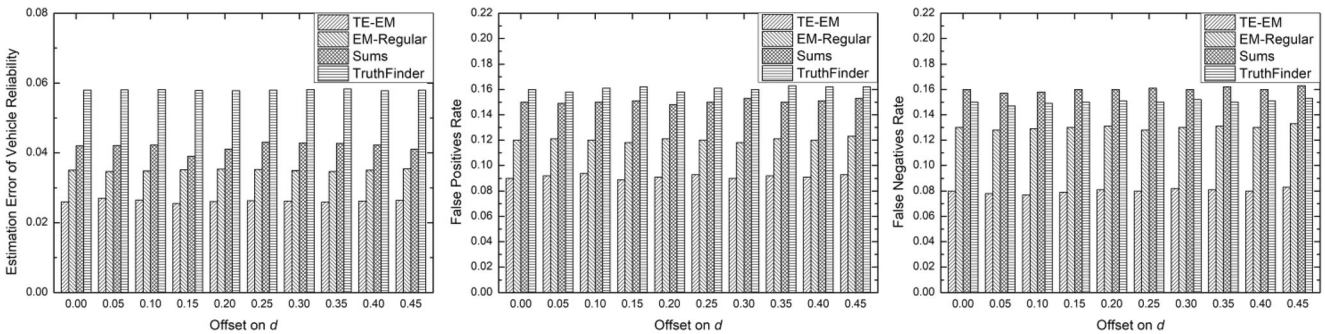

Figure 4 Impact of number of offset o on metrics (using synthetic data).

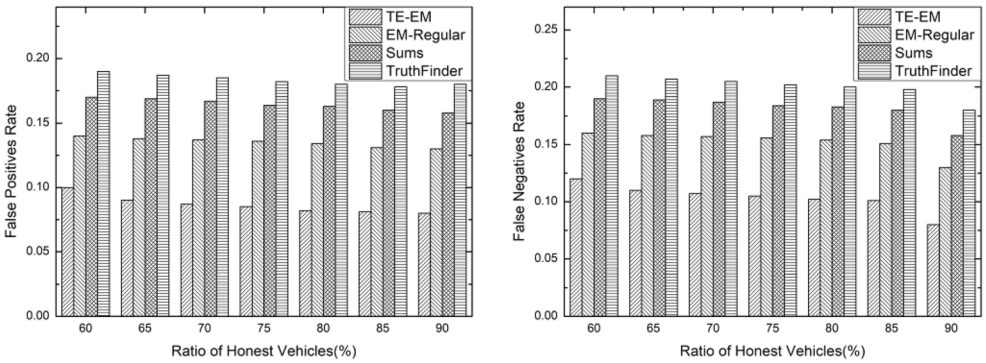

Figure 5 Impact of ratio of honest vehicles on metrics (using real data).

Figure 5 shows the changes in the performance of these four algorithms against the various percentages of honest vehicles. The increase in the percentage of honest vehicles has witnessed improvement of performance of these four algorithms to different degrees, among which TE-EM exhibits the highest accuracy. The experimental results are consistent with the results generated based on synthetic data.

6 Conclusion

In this paper, we proposes TPSense, a lightweight fog-assisted vehicular crowdsensing framework, which addresses the evaluation of event-reports' trustworthiness and protection of users' privacy. To solve the problem of false event-report generated by malicious nodes in the crowdsensing system, we convert it into a maximum likelihood estimation problem, and handle it through the expectation maximization algorithm. Through above works, we can complete the trustworthiness evaluation of event-reports and the reliability evaluation of the vehicles, and achieve the aim of false event-reports filtering on local RSUs. Meanwhile, the blind signature technology is used to generate a pseudonym for replacing the vehicle's real identity when vehicle uploads event-reports to ensure the anonymity of the vehicle and achieve users' privacy protection. We have assessed the TPSense by means of synthetic data and real data in vehicular crowdsensing. It is shown in results that TPSense outshines previous research in increasing information reliability. In the ongoing research, we will strengthen location privacy, data privacy protection and identity privacy by using technologies such as homomorphic encryption, space-time cloak, and differential privacy.

Acknowledgements Foundation item: National Natural Science Foundation of China (61972136, 61772173,61471161); Program for the Innovative Talents of the Higher Education Institutions of Henan Province (19HASTIT027); Open fund of Key Laboratory of Grain Information Processing and Control (under Grant No. KFJJ-2018105), Department of Education Outstanding Youth Scientific Innovation Team Support Foundation under Grant T201410.

References

1. Qiu, M., Ming, Z., Li, J., Liu, S., Wang, B., & Lu, Z. (2012). Three-phase time-aware energy minimization with DVFS and unrolling for Chip multiprocessors. *Journal of Systems Architecture, 58*(10), 439–445.
2. Qiu, M., Sha, E. H.-M., Liu, M., Lin, M., Hua, S., & Yang, L. T. (2008). Energy minimization with loop fusion and multi-functional-unit scheduling for multidimensional DSP. *Journal of Parallel and Distributed Computing, 68*(4), 443–455.
3. Shao, Z., Wang, M., Chen, Y., Xue, C., Qiu, M., Yang, L. T., & Sha, E. H.-M. (2007). Real-time dynamic voltage loop scheduling for multi-Core embedded systems. *IEEE Transactions on Circuits and Systems II: Express Briefs, 54*(5), 445–449.
4. Zhang, C., Zhu, L., Xu, C., Du, X., & Guizani, M. (2019). A privacy-preserving traffic monitoring scheme via vehicular crowdsourcing. *Sensors, 19*(6), 1274.
5. Basudan, S., Lin, X., & Sankaranarayanan, K. (2017). A privacy-preserving vehicular Crowdsensing-based road surface condition monitoring system using fog computing. *IEEE Internet of Things Journal, 4*(3), 772–782.
6. Barnwal, R. P., Ghosh, N., Ghosh, S. K., & Das, S. K. (2020). Publish or drop traffic event alerts? Quality-aware decision making in participatory sensing-based vehicular CPS. *ACM Transactions on Cyber-Physical Systems, 4*(1), 1–28.
7. Li, J., Ming, Z., Qiu, M., Quan, G., Qin, X., & Chen, T. (2011). Resource allocation robustness in multi-core embedded systems with inaccurate information. *Journal of Systems Architecture, 57*(9), 840–849.
8. Ni, J., Zhang, A., Lin, X., & Shen, X. S. (2017). Security, privacy, and fairness in fog-based vehicular Crowdsensing. *IEEE Communications Magazine, 55*(6), 146–152.
9. Sun, G., Sun, S., Sun, J., Yu, H., Du, X., & Guizani, M. (2019). Security and privacy preservation in fog-based crowd sensing on the internet of vehicles. *Journal of Network and Computer Applications, 134*, 89–99.
10. Li, J., Qiu, M., Niu, J., Gao, W., Zong, Z., & Qin, X. (2010). Feedback Dynamic Algorithms for Preemptable Job Scheduling in Cloud Systems. In *2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology* (pp. 561–564). Presented at the 2010 IEEE/ACM International Conference on Web Intelligence-Intelligent Agent Technology (WI-IAT), Toronto, AB, Canada: IEEE.
11. Sun, G., Sun, S., Yu, H., & Guizani, M. (2019). Towards incentivizing fog-based privacy-preserving Mobile Crowdsensing in the internet of vehicles. *IEEE Internet of Things Journal*, 7(5), 4128-4142.
12. De Cristofaro, E., & Soriente, C. (2013). Extended capabilities for a privacy-enhanced participatory sensing infrastructure (PEPSI).

IEEE Transactions on Information Forensics and Security, 8(12), 2021–2033.
13. Boukoros, S., Humbert, M., Katzenbeisser, S., & Troncoso, C. (2019). On (the lack of) location privacy in crowdsourcing applications. In Proceedings of the 28th USENIX Conference on Security Symposium (pp. 1859–1876). Santa Clara, CA, USA: USENIX Association.
14. Xiao, Z., Yang, J.-J., Huang, M., Ponnambalam, L., Fu, X., & Goh, R. S. M. (2018). QLDS: A novel design scheme for trajectory privacy protection with utility guarantee in participatory sensing. *IEEE Transactions on Mobile Computing, 17*(6), 1397–1410.
15. Gao, S., Ma, J., Shi, W., Zhan, G., & Sun, C. (2013). TrPF: A trajectory privacy-preserving framework for participatory sensing. *Information Forensics and Security, IEEE Transactions on, 8*, 874–887.
16. Li, M., Zhu, L., & Lin, X. (2019). Privacy-preserving traffic monitoring with false report filtering via fog-assisted vehicular Crowdsensing. *IEEE Transactions on Services Computing*, 1–1.
17. Chen, Y.-M., & Wei, Y.-C. (2013). A beacon-based trust management system for enhancing user centric location privacy in VANETs. *Journal of Communications and Networks, 15*(2), 153–163.
18. Hubaux, J. P., Capkun, S., & Luo, J. (2004). The security and privacy of smart vehicles. *IEEE Security & Privacy Magazine, 2*(3), 49–55.
19. Li, Q., Malip, A., Martin, K. M., Ng, S.-L., & Zhang, J. (2012). A reputation-based announcement scheme for VANETs. *IEEE Transactions on Vehicular Technology, 61*(9), 4095–4108.
20. Park, S., Aslam, B., & Zou, C. C. (2011). Long-term reputation system for vehicular networking based on vehicle's daily commute routine. In 2011 IEEE Consumer Communications and Networking Conference (CCNC) (pp. 436–441). Presented at the 2011 IEEE Consumer Communications and Networking Conference (CCNC),Las Vegas, NV, USA: IEEE
21. Raya, M., Papadimitratos, P., Gligor, V. D., & Hubaux, J.-P. (2008). On data-centric trust establishment in ephemeral ad hoc networks. In *IEEE INFOCOM 2008 - The 27th Conference on Computer Communications* (pp. 1238–1246). Presented at the IEEE INFOCOM 2008 - IEEE Conference on Computer Communications, Phoenix, AZ, USA: IEEE.
22. Timpner, J., Schurmann, D., & Wolf, L. (2016). Trustworthy parking communities: Helping your neighbor to find a space. *IEEE Transactions on Dependable and Secure Computing, 13*(1), 120–132.
23. Shim, K.-A. (2012). ${\cal CPAS}$: An efficient conditional privacy-preserving authentication scheme for vehicular sensor networks. *IEEE Transactions on Vehicular Technology, 61*(4), 1874–1883.
24. Huang, K. L., Kanhere, S. S., & Hu, W. (2010). Are you contributing trustworthy data?: The case for a reputation system in participatory sensing. In *Proceedings of the 13th ACM international conference on Modeling, analysis, and simulation of wireless and mobile systems - MSWIM '10* (p. 14). Presented at the the 13th ACM international conference, Bodrum, Turkey: ACM Press.
25. Yu, R., Liu, R., Wang, X., & Cao, J. (2014). Improving data quality with an accumulated reputation model in participatory sensing systems. *Sensors, 14*(3), 5573–5594.
26. Amintoosi, H., Kanhere, S. S., & Allahbakhsh, M. (2015). Trust-based privacy-aware participant selection in social participatory sensing. *Journal of Information Security and Applications, 20*, 11–25.
27. Restuccia, F., & Das, S. K. (2014). FIDES: A trust-based framework for secure user incentivization in participatory sensing. In *Proceeding of IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks 2014* (pp. 1–10). Presented at the 2014 IEEE 15th International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM), Sydney, Australia: IEEE.
28. Xiang, Q., Zhang, J., Nevat, I., & Zhang, P. (2017). A trust-based mixture of Gaussian processes model for robust participatory sensing. In *Proceedings of the 16th conference on autonomous agents and MultiAgent systems* (pp. 1760–1762). São Paulo: International Foundation for Autonomous Agents and Multiagent Systems.
29. Huang, K. L., Kanhere, S. S., & Hu, W. (2012). A privacy-preserving reputation system for participatory sensing. In *37th Annual IEEE Conference on Local Computer Networks* (pp. 10–18). Presented at the 2012 IEEE 37th Conference on Local Computer Networks (LCN 2012), Clearwater Beach, FL, USA: IEEE.
30. Christin, D., Roßkopf, C., Hollick, M., Martucci, L. A., & Kanhere, S. S. (2013). IncogniSense: An anonymity-preserving reputation framework for participatory sensing applications. *Pervasive and Mobile Computing, 9*(3), 353–371.
31. Wang, X., Cheng, W., Mohapatra, P., & Abdelzaher, T. (2014). Enabling reputation and trust in privacy-preserving mobile sensing. *IEEE Transactions on Mobile Computing, 13*(12), 2777–2790.
32. Ma, L., Liu, X., Pei, Q., & Xiang, Y. (2019). Privacy-preserving reputation management for edge computing enhanced mobile crowdsensing. *IEEE Transactions on Services Computing, 12*(5), 786–799.
33. Wang, D., Kaplan, L., & Abdelzaher, T. F. (2014). Maximum likelihood analysis of conflicting observations in social sensing. *ACM Transactions on Sensor Networks, 10*(2), 1–27.
34. Chaum, D. (1983). Blind signatures for untraceable payments. In D. Chaum, R. L. Rivest, & A. T. Sherman (Eds.), *Advances in cryptology* (pp. 199–203). Boston: Springer US.
35. Barnwal, R. P., Ghosh, N., Ghosh, S. K., & Das, S. K. (2019). PS-Sim: A framework for scalable data simulation and incentivization in participatory sensing-based smart city applications. *Pervasive and Mobile Computing, 57*, 64–77.
36. Yin, X., Han, J., & Yu, P. S. (2008). Truth discovery with multiple conflicting information providers on the web. *IEEE Transactions on Knowledge and Data Engineering, 20*(6), 796–808.
37. Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM), 46*(5), 604–632.
38. Alswailim, M. A., Hassanein, H. S., & Zulkernine, M. (2016). A Reputation System to Evaluate Participants for Participatory Sensing. In *2016 IEEE Global Communications Conference (GLOBECOM)* (pp. 1–6). Presented at the GLOBECOM 2016–2016 IEEE Global Communications Conference, Washington, DC, USA: IEEE.

Zhenqiang Xu received the B.S degree in computer science and technology from Henan Normal University, Henan, China, in 2001, and the M.S. degree in computer technology from Huazhong University of Science and Technology, Hunan, China, in 2005, where he is currently pursuing the Ph.D. degree of surveying and mapping science and technology in Information Engineering University. His current research interests include trajectory data mining, information security and privacy perserving.

Weidong Yang received his B.S. in industrial automation, and M.S. and Ph.D. degree in Computer Science from Xidian University in China in 1999, 2005, and 2008, respectively. He is now a professor in Henan University of Technology, deputy chair of Key Laboratory of Grain Information Processing and Control (Henan University of Technology) ministry of Education. He is also a senior member of China Computer Federation (CCF). His research focuses on wireless networks security, privacy protection, and vehicular ad-hoc networks, and so on.

Zenggang Xiong received the Ph.D. degree from the University of Science and Technology Beijing, Beijing, China, in 2009. He is currently a Professor with the School of Computer and Information Science and Master Tutor, Hubei Engine erring University, Xiaogan, China. His main research interests include cloud computing, big data, pattern recognition, and computer vision.

Jiayao Wang received the B.S degree in surveying and mapping science and technology from Information Engineering University, Henan, China, in 1961. His current research interests include Cartography and Geographic Information Engineering and science and technology of surveying and mapping.

Gang Liu received the Ph.D. degree from the University of Science and Technology Beijing, Beijing, China, in 2009. He is currently Professor with the College of Information Science and Engineering, Henan University of Technology, Zhengzhou, China. His main research interests include intelligent control, optimization algorithm, and neural network.