Accepted Manuscript

PAU: Privacy Assessment method with Uncertainty consideration for cloud-based vehicular networks

Xia Feng, Liangmin Wang

PII: S0167-739X(18)32660-8
DOI: https://doi.org/10.1016/j.future.2019.02.038
Reference: FUTURE 4791

To appear in: *Future Generation Computer Systems*

Received date : 2 November 2018
Revised date : 28 January 2019
Accepted date : 20 February 2019

PAU: Privacy Assessment Method with Uncertainty Consideration for Cloud-Based Vehicular Networks

Xia Feng[a], Liangmin Wang[b]

[a]*School of Automotive and Traffic Engineering, Jiangsu University, Zhenjiang 212013, China (e-mail: u6006355@utah.edu)*
[b]*Jiangsu Key Laboratory of Security Tech. for Industrial Cyberspace, Jiangsu Univerity, Jiangsu, China;*

Abstract

With the rapid progress of wireless communication and big data, the traditional Vehicular Ad-hoc Networks (VANETs) gradually evolve into the new Heterogeneous Vehicular Networks (HetVNets). Meanwhile, with the combination of multiple forms of communication modes, it initiates the Vehicle to Everything(V2X) communication model providing more efficient services. V2X communication generates much more private data than traditional VANETs, but the concerns over privacy breaches are increasing. these big data burdens the concerns about. To protect the privacy in these cloud-based vehicular networks is remained unsolved. In this paper, we propose Privacy Assessment method with Uncertainty consideration (PAU) to estimate the nodes' capability in protecting privacy, and then choose the vehicular nodes with high priority calculated by PAU to improve the whole network's privacy protection level. PAU expands subjective logic based on two-tuple to triad and keeps uncertainty as a constituent element. It evaluates the nodes by using the historical data from the vehicular cloud and the real-time data from V2V communications. The experiments and analysis show that the improvement of privacy-preserving capability achieved when applied PAU in Mix-zone scenarios.

Keywords: Cloud-based Vehicular network, privacy, uncertainty, V2X

1. Introduction

Vehicular Ad-hoc Networks(VANETs) are envisaged to be one of the building blocks of the Internet of cognitive Things and accelerate the evolution of the Intelligent Transportation System(ITS). Based on Americans 5G white paper[1], vehicle-to-everything(V2X) communication model is mainly composed by Vehicle-to-Vehicle(V2V), Vehicle-to-Infrastructure (V2I), Vehicle-to-Network(V2N) and Vehicle-to-Pedestrian(V2P). The heterogeneous mode [2] accelerates the efficiency of information dissemination. However, it adds the concerns about privacy breaches. The long-term storage of historical data on the cloud platform adds to the worries about privacy issues. The heterogeneous vehicular networks increase the difficulties of privacy protection .

There are three main dimensions taken into account in traditional entropy-based privacy assessment methods, the specific aspects or types of privacy, the adversary and capabilities, and the privacy metric[3][4][5]. Those assessment methods are all considered to be off-line, which are quantitatively evaluated based on specific information or privacy breaches. In the cloud-based V2X network environment, on the one hand, it is challenging to evaluate every event with the high-speed of information dissemination, on the other hand, the results of the offline evaluation couldn't make up for the data leakage. In the information interaction, a node's low awareness of privacy protection will lessen the privacy protection capability of the entire communication system. To track this problem, we propose the Privacy Assessment method with Uncertainty consideration (PAU) metric based on vehicular nodes uncertainty assessment. This method focuses on evaluating the privacy protection capability of each node, and by selecting interactive nodes with high privacy awareness. Thereby achieves privacy-preserving itself and improves privacy protection level of the network. The contributions of this paper are as follows:

1)In the privacy assessment process for each vehicle, we proposed a novel method oriented subjective logic, to predict the nodes privacy breach level by analyzing the user's historical behavior under cloud-based V2X scenarios, and expand subjective logic to uncertainty to measure the undetermined records in the user's historical behavior;

2) We capture the real-time privacy capability based on real-time vehicles communication observations, therefore, present a privacy aggregation algorithm to combine the real-time and offline opinion to improve the accuracy of privacy assessment;

3)In the simulation, we design a dynamic Mix-zone construction algorithm that can efficiently coverage. The experiments and analysis show that the improvement of privacy-preserving capability achieved when applied PAU in Mix-zone scenarios.

The rest of this paper is organized as follows: Section II presents related works. Section III introduces the architecture, assumption and the formalization of our proposed method. Section IV presents the proposed PAU scheme in details. Section V shows the performance analysis based on Mix-zone. Finally, the concluding remarks and future work are given in Section VI.

2. Related work

Privacy in VANETs involves with special concern because during the communication human lives are constantly at stake. The deployment of a comprehensive security system for VANETs is very challenging[6][7] in practice, because the nature of vehicular network is highly dynamic as well as short connection duration. Most privacy issues are related to position and identifiers[8]. Many existing techniques [9] are available for the privacy protection in VANETs. Privacy metrics and privacy enhancing technologies(PETs) are proposed to measure the degree of privacy enjoyed by users. There are several[10][11][12] traditional agreement on privacy properties in VANETs. **Confidentiality** describes the possibility of an adversary obtains the privacy data. The higher the impossibility represents higher privacy level. In cloud-based Vehicular network [13], it is implemented by encryption. **Anonymity** [14][15] refers to the adversary cant distinguish the target from the anonymity set. In VANETs, the anonymity set could be a collection of vehicles at a specific location, such as an intersection. **Unlinkability** [16] indicates that the adversary couldnt establish a connection between two or more objects, actions, or locations. In VANETs, the higher privacy means that the attacker cannot link the identity to the pseudonym of the vehicle. We usually place a Mix-zone [17] at road intersections since vehicle trajectories aren't predictable. Within the Mix-zone, vehicles must change their pseudonyms and encrypt the messages. **Undetectabiltiy** [18] describes the adversary can't distinguish the information it interested from the big data generated by the communication system.

However, referring to the cloud-based V2X communication system, little research has been done in this domain, because comparatively it is a new field. 3GPP Released 15[19] to describe the architectural enhancements for V2X services and provides more details about privacy issues. It is indispensable, but this can not resolve the concerns about insider attack. Feng[20]proposed a scheme involving with physical layer technique, in which the signatures or other unique identifiers are implanted in messages to identify legitimate nodes. Vuk[21] presented cross-layer techniques to schedule messages for the purpose of enhancing distributed security awareness.

The method we proposed focuses on the individual privacy-preserving capability. For example, if one node in the Mix-zone breaches the privacy, it not only

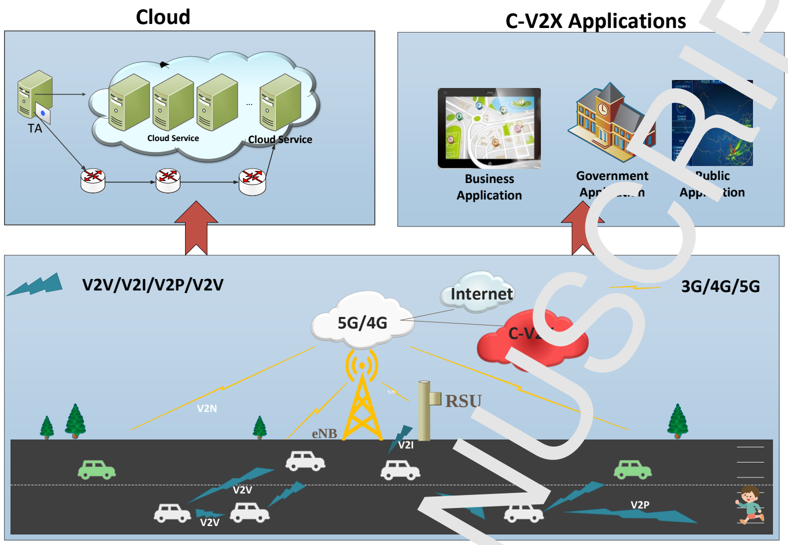


Figure 1: Architecture of the cloud-based vehicular network

reveals the identity itself but also the nodes communicate with it. Therefore, majority metrics in research cannot handle the practical situation. We take account of the historical behavior for individual node and update the assessment. And when applied the technology in the construction of Mix-zone, we would sort out the higher privacy-preserving nodes to improve the effects.

3. System model

In this section, we describe the system model in heterogeneous Vehicular network. Based on this model, we proposed our architecture of PAU, three attack models, and the research objectives. To measure the privacy, we present a formal mathematical description of the system.

3.1. System Model for cloud-based V2X

3GPP [22] and 5G Americas[1] describe the current V2X landscape, including standards and industry status with expected benefits. In this paper, as shown in Figure 1, the architecture of cloud-based V2X involves four entities and four main communication modes. The four entities are Cloud, Vehicle, Pedestrian, and Infrastructure. To connecting these entities, it specifies vehicular communications for Vehicle-to-Everything (V2X) services, which includes Vehicle-to-Vehicle (V2V), Vehicle-to-Pedestrian (V2P), Vehicle-to-Infrastructure (V2I), and

Vehicle-to-Network (V2N). In the communication systems, vehicles should be able to communicate with not only other vehicles (V2V) but also with nearby infrastructure (V2I), Internet-based networks (V2N) and even pedestrians (V2P). Collectively these use cases have become known as vehicle-to-everything (V2X) connectivity, which forms a significant part of the Intelligent Transport System(ITS). The cloud platform can store a large amount of data generated by the network, and the V2N increases the storable, queryable and usable of the data stored in the cloud. However, V2V and V2P expand the source of transport data as well as the complexity, meanwhile, raise users concerns about data privacy. After all, more individual data contains in the traffic information, and the disclosure of information, in addition to privacy threats, may bring about the threaten of life and properties.

3.2. Research goal and Threaten model

In heterogeneous VANETs, vehicles and networks are likely to undertake a variety of attacks, such as jam, eavesdrop, forge, and modify. In our paper, we focus on the assessment for the nodes privacy-preserving capability, and by selecting communication nodes based on evaluating values to reduce the risk of network privacy breaches. To this end, the design goals of this paper are different from the existing methods of privacy protection. We take advantages of the historical behaviors of users accumulated by the cloud platform to obtain the evaluation of the privacy protection capability. Meanwhile, we take account of the real-time communication data to improve the accuracy of privacy assessment. Raya et al.[23] identified four groups of security threats, in which insider attackers are considered to be the hardest part to detect, because the adversaries may propagate valid messages that cannot easily be detected using cryptographic signatures alone, pose a viable threat to information dependability[24]. The most basic security threats are summarized in [25], there are three common insider attacks can be restrained by our scheme.

1) **Bad mouth attack**: The attackers collude to give negative feedback on the victim in order to lower or destroy its reputation. This type of attack against the assessment system is common in trust evaluations[26], and it is difficult to deal with such attacks when the evaluation system does not collect enough historical data and is one of the primary sources for assessing uncertainty.

2) **Conflicting behavior attack**: In this attack, malicious nodes perform differently to different neighbor nodes, or randomly generate inconsistent privacy value for benign nodes. Conflicting opinions will decrease the assessment credibility. This kind of attack can only be detected when the network scale and running

duration is reaching the threshold.

3)**Newcomer attack**: In newcomer attack, the malicious node registers as a new user and removes its bad history, thus, this kind of attack would significantly destroy the trust management of VANETs[27]. The defense against newcomer attacks does not rely on the design of trust management, but establishes the authentication scheme to distinguish the faked or copied ID.

Therefore, the scheme we designed can resist the attacks described above, and we will explain the privacy breaching level against those attackers knowledge.

4. System formalization

In VANETs, the mobility increases the difficulty of measuring the vehicle's privacy disclosure value. However, vehicle's cloud accumulates sufficient historical data, the already existing and growing data make privacy assessment possible. The privacy assessment of this article is divided into four steps to implement.

privacy calculation based on historical data: Analyzing historical data for all vehicles in the cloud-based platform, therefore, we will classify the nodes on the basis of Definition 1, and operate further calculation relying on Lemma 1.

Privacy modification adapted from real-time data: VANETs is dynamic, and in the network, the neighbor nodes keep changing. Historical data is helpful, but they may not reflect the nodes real-time surrounding environment. Therefore, the estimation of the vehicle has to consider the real-time communication status. It will described in Section 4.2.

Privacy aggregation: With the consideration of historical and modifiable data, our scheme aggregates the real-time and off-line assessment opinion to a comprehensive value. It can be applied in Algorithm 1 to decide whether or not the node is authorized to join in the network.

Privacy prediction: Based on privacy aggregation, prediction will be applied to the experiment when we discuss how privacy domains the establishment of Mix-zone in Section 5.

4.1. Privacy calculation based on historical data

The vehicle set donated as 'V', for vehicle A, we evaluate privacy-preserving capability for vehicle A, denoted as $P_{(A)}$, We define $P_{(A)}$ as follows:

Definition 1. *For any vehicle A in the vehicle set V, its privacy protection capability $P_{(A)}$ is a triplet of three attributes:*

$$P_{(A)} = (p, l, u). \tag{1}$$

where p, l, u ≤ 1 *and* $p + l \leq 1$.

In definition 1, 'p' represents the probability of vehicles in set N's belief on vehicle A depending on A's benign behaviors. Similarly, 'l' represents the probability of vehicles in set N's disbelief on vehicle A depending on A's malicious behaviors, 'u' denotes the percentage of vehicles in set N's uncertainty on vehicle A, and uncertainty means ignorance or lack of enough evidence which is a core dimension of privacy assessment.

Our paper is inspired by [28], in which the author infers the value of uncertainty by the Bayesian formula based on Beta distribution and constantly update newly evidence when observations have emerged. Combining the above ideas with the application scenario of this paper, Lemma 1 defines the privacy assessment formula for vehicle A.

Lemma 1. *'E' denotes the set of privacy events participated by vehicle A, Where '*E_1*' and '*E_2*' are two mutually disjoint subsets of E. Subset '*E_1*' and '*E_2*' contain the privacy-preserving and the privacy-leakage events respectively. The value* $P_{(A)}$ *of privacy-preserving for A can be defined as follows:*

$$p = \frac{||E_1||}{||E||} . \tag{2}$$

$$l = \frac{||E_2||}{||E||} . \tag{3}$$

$$u = \frac{12 \cdot ||E_1|| \cdot ||E_2||}{(||E_1|| + ||E_2||)^2 \cdot (||E_1|| + ||E_2|| + 1)} . \tag{4}$$

where '$||E_1||$*' indicates the cardinality of the set, that is, the number of members in the set E.*

The uncertainty 'u' in Lemma 1 is based on privacy-preserving and the privacy-leakage events. To start with, we set the prior as $Beta(1, 1)$. The value of $||E_1||$, $||E_2||$ relates to the node's privacy assessment in two important attributes. Firstly, when $||E_1|| + ||E_2||$ is higher, it implies that there is more evidence to support our assessment value. Secondly, when the evidence for privacy-preserving or privacy-leakage events dominates, uncertainty value will consequently descends. The uncertainty value will be at the peak when privacy-preserving and privacy-disclosure events have the equal value[28]. We will do some details discuss for the Formula(3) based on Lemma 1.

$$||E_1|| + ||E_2|| = ||E|| . \tag{5}$$

When compared with the physical situation of ITS, there are exceptional cases where we are unable to determine whether the participating vehicles have privacy disclosure. Therefore, the Formula (5) can be false. Studies on uncertainty are essential when the ignorance emerges. To address the problem, privacy-preserving capability of nodes is denoted as discrete interval.

Inference 1. *Denoted A's privacy-preserving capacity as $P(A) = (p, l, u)$, when A participates in a new event, the probability will be in the interval.*

$$\arg\min_{u} p \leq p \leq \arg\max_{u} p \tag{6}$$

where

$$\arg\max_{u} p = 1 - l \times (1 - u)$$

$$\arg\min_{u} p = p \times (1 - u)$$

In definition 1, Formula (7),(8),(9) satisfy the constraint: $p, l, u \leq 1$, $p + l \leq 1$, and, $p + l + u = 1$.

4.2. Privacy modification adapted from real-time data

In V2V, the node's privacy-preserving capability can be estimated through the upcoming events observed by neighbors, that is, the real-time modified privacy opinion is originated from neighbor nodes within the communication range.

Definition 2. *'N' denoted the neighbors set of vehicle A , obviously, N is a subset of V, x is an arbitrary vehicle in the set N. The real-time privacy-preserving capability of A derives from set N, the definitions of $Pr(A) = (p_r, l_r, u_r)$ are as follows:*

$$p_r = \frac{||\{x|T_x^A \geq \sigma_1, x \in N\}||}{||N||} \tag{7}$$

$$l_r = \frac{||\{x|T_x^A \leq \sigma_2, x \in N\}||}{||N||} \tag{8}$$

$$u_r = \frac{||\{x|\sigma_2 < T_x^A < \sigma_1, x \in N\}||}{||N||} \tag{9}$$

where T_x^A represents node x's opinion on node A, $\sigma_1 \geq \sigma_2$ are thresholds for the system.

There is a possibility that definition 2 fails to square up the influences for the number of observed nodes. As we mentioned before, the scales of observed nodes will greatly increase the confidence of assessment. Following the analysis above, We would update the Formula (9) to Formula (10).

$$u_r = \frac{12 \cdot p \cdot l}{(p + l)^2 \cdot (\|N\| \cdot p + \|N\| \cdot l + 1)} \tag{10}$$

In definition 2, T_x^A is derived from the aggregation of first-hand and second-hand opinions. The first-hand opinion is derived from the node's privacy-disclosure during the direct V2V communication,while the second-hand opinion is a kind of trust transitivity. The two kinds of opinions will be calculated by the algorithm in [9]

4.3. Privacy aggregation algorithm

As the analysis above, vehicle A has two aspects of privacy assessments, one is $P_{(A)}$ which is calculated in accordance with historical cloud data, and the other one $Pr_{(A)}$ is obtained by the real-time neighborhood observed data. Even though the physical meaning of the parameters used in $P_{(A)}$ and $Pr_{(A)}$ are completely different, they effectively reflect the privacy-preserving capability of vehicle A. Eventually, by the aggregation algorithm, we acquire the mathematical unity of the numerical formulas.

The consistency of mathematical form for $P_{(A)}$ and $Pr_{(A)}$ can be applied to aggregate the privacy assessment value. The aim of Definition 3 is to gradually reduce the uncertainty. By integrating $P_{(A)}$ and $Pr_{(A)}$, we will obtain the comprehension privacy-preserving capability on historical cloud data and observed real-time opinions.

Theorem 1. *Let the vehicle A's historical opinion of privacy-preserving denoted as $P_{(A)} = (p, l, u)$ and the real-time opinion is $Pr(A) = (p_r, l_r, u_r)$, then the aggregation opinion will be $P_{agg(A)} = \left(p_{agg}, l_{agg}, u_{agg}\right)$. The opinions of vehicle A satisfy the piecewise functions as follows:*

$$p_{agg} \in \begin{cases} (p(1-u), 1 - l * (1-u)) \cap (p_r(1-u_r), 1 - l_r * (1-u_r)) & \Omega \neq \phi \\ (p(1-u), 1 - l * (1-u)) \cup (p_r(1-u_r), 1 - l_r * (1-u_r)) & \Omega = \phi \end{cases} \tag{11}$$

where

$$\Omega = (p(1-u), 1 - l * (1-u)) \cap (p_r(1-u_r), 1 - l_r * (1-u_r))$$

$$l_{agg} \in \begin{cases} (l(1-u), 1-p*(1-u)) \cap (l_r(1-u_r), 1-p_r*(1-u_r)) & \Psi \neq \phi \\ (l(1-u), 1-p*(1-u)) \cup (l_r(1-u_r), 1-p_r*(1-u_r)) & \Psi = \phi \end{cases} \tag{12}$$

where

$$\Psi = (l(1-u), 1-p*(1-u)) \cap (l_r(1-u_r), 1-p_r*(1-u_r))$$

Proof. According to Inference 1, $p \in (p(1-u), 1-l*(1-u))$,

$$p_r \in (p_r(1-u_r), 1-l_r*(1-u_r)),$$

p_{agg} is derived from p and p_r, therefore p_{agg} is a union set of p and p_r. In reference to the assessment for uncertainty, $p_{agg} \in (p(1-u), 1-l*(1-u)) \cup (p_r(1-u_r), 1-l_r*(1-u_r))$. Since the result of the inference 1 has fully considered its historical behavior, the value space of uncertainty equals zero. Furthermore, if $\Omega \neq \phi$, p_{agg} should be the intersection, denoted as $p_{agg} \in \Omega$. By taking account of uncertainty behaviors, we could compress the value space described by Inference 1. The uncertainty will be $(1-(1-u)*(1-u_r))$. Therefore, Formula (11) is proved. Considering the symmetry of the problem and the formula, Formula (12) can be proved, too. □

Inference 2. *Let the vehicle A's historical opinion of privacy-preserving be evaluated as $P_{(A)} = (p, l, u)$, and the real-time opinion is $Pr(A) = (p_r, l_r, u_r)$, then the aggregation opinion will be $P_{agg(A)} = \left(p_{agg}, l_{agg}, u_{agg}\right)$. When*

$$\Omega = (p(1-u), 1-l*(1-u)) \cap (p_r(1-u_r), 1-l_r*(1-u_r))$$

and

$$\Omega \neq \phi$$

the approximate privacy-preserving opinion of A is:

$$p_{agg} \approx \left[\min(1-l_r*(1-u_r), 1-l*(1-u)) + \max(p_r(1-u_r), p(1-u))\right]/2 \tag{13}$$

when

$$\Psi = (l(1-u), 1-p*(1-u)) \cap (l_r(1-u_r), 1-p_r*(1-u_r))$$

for the situation $\Psi \neq \phi$, the approximate privacy-preserving opinion of A is

$$l_{agg} \approx \left[\min(1-p*(1-u), 1-p_r*(1-u_r)) + \max(l(1-u), l_r(1-u_r))\right]/2 \tag{14}$$

the uncertainty is

$$u_{agg} = \frac{\min\left(1 - l_r * (1 - u_r), 1 - l * (1 - u)\right) - \max\left(p_r(1 - u_r), p(1 - u)\right)}{(1 - (1 - u) * (1 - u_r))} \quad (15)$$

We deduce Inference 2 by the proof of Theorem 1. The approximation opinion is assigned by the midpoint of the scope in which Theorem 1 declared, and the source of uncertainty defined by the length of values. The further discuss about $\Omega = \phi$ indicates that there is a big difference between the historical data opinion and the real-time opinion. Under the constraints $p, l, u \leq 1$ and $p + l \leq 1$, one of the forms listed below establishes.

$$p(1 - u) > 1 - l_r * (1 - u_r)$$

or

$$1 - l * (1 - u) < p_r(1 - u_r)$$

according to the symmetry of formula, we will discuss

$$p(1 - u) > 1 - l_r * (1 - u_r)$$

obviously,

$$\begin{aligned} p(1 - u) &= (1 - l)(1 - u) \\ &= 1 - l * (1 - u) - u \end{aligned}$$

when it satisfies

$$1 - u - l * (1 - u) > 1 - l_r * (1 - u_r)$$

which is

$$u + l * (1 - u) < l_r * (1 - u_r)$$

In our simulation experiments, the changing trend of uncertainty 'u' and 'u_r' keeps consistent, because we derive 'u' and 'u_r' from the same vehicle. It will be safely concluded that with the descend of uncertainty, we have more confident for the assessment value. There is a tiny chance that the adversary deceives the cloud but exposes to the real-time assessment, or conversely, it exposes to the cloud. Then, there is a big gap between 'l' calculated by historical data and 'l_r' obtained by real-time observations. To theoretically implementing the tiny possibility, we make up the gap with Proposition 1.

Proposition 1. *Let the vehicle A's historical opinion of privacy preserving be evaluated as $P_{(A)} = (p, l, u)$, and the real-time opinion is $Pr(A) = (p_r, l_r, u_r)$, then the aggregation opinion will be $P_{agg(A)} = \left(p_{agg}, l_{agg}, u_{agg}\right)$ where*

$$\Omega = (p(1-u), 1 - l * (1-u)) \cap (p_r(1-u_r), 1 - l_r * (1-u_r))$$

similarly, when $\Omega = \phi$,

the approximate privacy-preserving opinion of A is:

$$p_{agg} \approx \left[\max(1 - l_r * (1-u_r), 1 - l * (1-u)) + \min(p_r(1-u_r), p(1-u))\right]/2 \tag{16}$$

if

$$\Psi = (l(1-u), 1 - p * (1-u)) \cap (l_r(1-u_r), 1 - p_r * (1-u_r))$$

and $\Psi = \phi$,

the approximate privacy-preserving opinion of A is

$$l_{agg} \approx \left[\max(1 - p * (1-u), 1 - p_r * (1-u_r)) + \min(l(1-u), l_r(1-u_r))\right]/2 \tag{17}$$

and the uncertainty is:

$$u_{agg} = \frac{\max(1 - l_r * (1-u_r), 1 - l * (1-u)) - \min(p_r(1-u_r), p(1-u))}{1 - u * u_r} \tag{18}$$

5. Performance evaluation in Mix-zone

In this section, we aim to investigate the privacy-preserving capability for proposed PAU scheme. The simulations performance on NS-2 [29]. We consider a region of 1000km^2 with 100 vehicles. Table 1 gives the definition of the basic parameters. The radio coverage radius of V2N transmission is 5 km while the V2V is 1 km, which is a typical range of the 5G protocol [1]. It is worth paying attention that the proposed scheme can be deployed in a more complex environment cause the propagation computing is operated in the cloud.

Table 1: Simulation parameters

Notation	Definition
Simulation	500s
Simulation area	1000*1000km
Total number of nodes	100
Intersections	10
V2N transmission range	5km
V2V transmission range	1km
Proportion of malicious nodes	50%
Moving Speed	0-10m/s
Packet rate	4 pkt/sec
The minimum number of vehicles joined in Mix-zone	K
τ	0.6

Table 2 is the rule of selecting nodes to join in the Mix-zone. According to threshold τ, we classify the nodes into five types, respectively tagged as N_I, N_{II}, N_{III}, N_{VI}, N_V. N_I represents that the nodes are absolute privacy[30] which can join in Mix-zone immediately. Nodes tagged N_{II} will be suspended to the network. But when the number of vehicles involved in Mix-zone is less than K, they will join in the communication. The procedure is implemented in Algorithm 1. Nodes N_{III} will be suspended and requested to verify. Nodes N_{VI} will be rejected but allowed for the second application. N_V indicates that the nodes might expose and can't join in the network.

Table 2: Regulation for vehicle nodes selection

P_{agg}	d_{agg}	u_{agg}	**Procedure**
$> \tau$	–	–	Join in Mix-zone immediately. Tag N_I
$\leq \tau$	$\leq \tau$	$\leq \tau$	Suspend the request. Tag N_{II}
$\leq \tau$	$\leq \tau$	$> \tau$	Suspend the request, request to verify. Tag N_{III}
$\leq \tau$	$> \tau$	$\leq \tau$	Reject the request, allow for the second application. Tag N_{VI}
$\leq \tau$	$> \tau$	$> \tau$	Distrust and reject the request. Tag N_V

Algorithm 1 is a general procedure for dynamic Mix-zone construction. The sampling of threshold depends on the scale of vehicles involved in the network and interaction between nodes. To meet the requirements for participation in real-time communication, we set $\tau = 0.6$ in the simulation. Thus, The baseline threshold is not suitable for all scenarios. If the minimum number of vehicles joined in Mix-zone doesn't meet expectations, We should adjust τ in comparison with the scale of the network.

Algorithm 1 Algorithm 1 Dynamic Mix-zone Construction

Require: Input: A vehicles set V_1
Ensure: Output: Mix-zone(a set of V_{MZ})
for $V_i \in V_1$ **do**
  if $p_{agg} > \tau$; **then**
    $V_{MZ} \leftarrow V_1 + V_i$;
    $num \leftarrow num + 1$;
    $V_i \leftarrow N_I$;
  end if
end for
while $Num \leq K$ **do**
  for $V_i \in V_1$ **do**
    if $p_{agg} \leq \tau$; $l_{agg} \leq \tau$; $u_{agg} \leq \tau$ **then**
      $V_i \leftarrow N_{II}$;
      $V_{MZ} \leftarrow V_1 + V_i$;
      $num \leftarrow num + 1$;
    end if
    if $p_{agg} \leq \tau$;$l_{agg} \leq \tau$; $u_{agg} > \tau$ **then**
      $V_i \leftarrow N_{III}$;
    end if
    if $p_{agg} \leq \tau$; $l_{agg} > \tau$; $u_{agg} \leq \tau$ **then**
      $V_i \leftarrow N_{VI}$;
    end if
    if $p_{agg} \leq \tau$; $l_{agg} > \tau$; $u_{agg} > \tau$ **then**
      $V_i \leftarrow N_V$;
    end if
  end for
end while

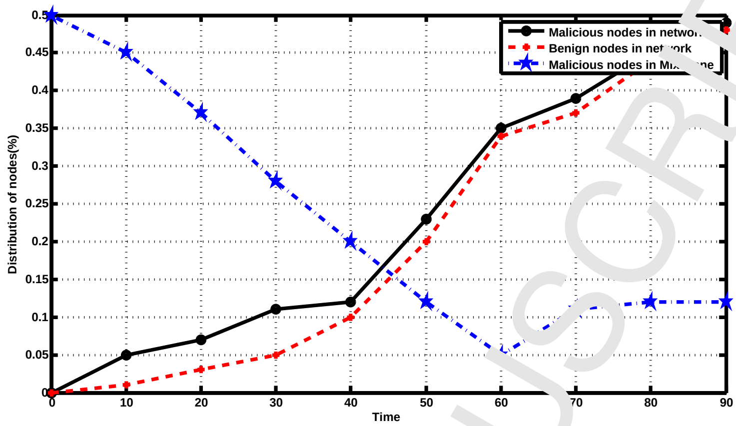


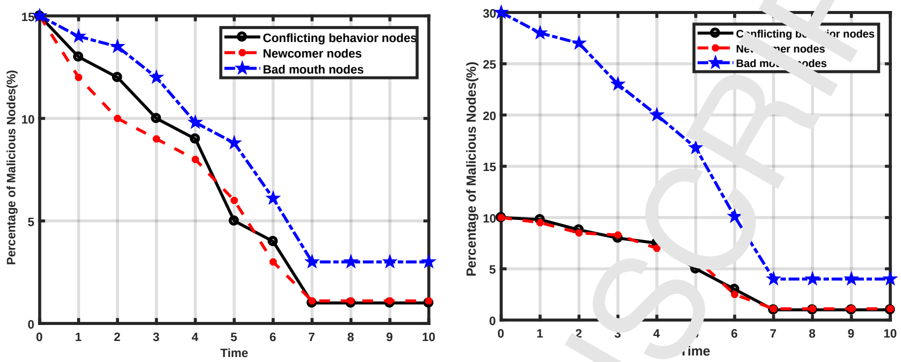
Figure 2: Simulated analysis of distribution of nodes

5.1. *Analysis of simulation*

In the cloud-based V2X scenario, we acquire historical data from V2N, and real-time privacy opinion from V2V. In the simulation, the historical data need to accumulate in the measurable time snippets. For example, in the time-line, $t = 0$ means for the first snippet there is no historical data in the cloud, while $t = 100$, the cloud server has accumulated data for 99-time snippets. In Figure 2, during the time snippets from 0 to 100, and the vertical axis represents this proportion of low privacy-preserving nodes joined in the Mix-zone. Thus we can conclude that, with the time and historical data increasing, the malicious nodes are gradually exposed, meanwhile, the percentage of the malicious nodes involved in the mix-zone decreases steadily. When the proportion of malicious nodes is more than 30% , the reduction results from the alliance of the bad mouth attack and the conflicting behavior attack. That is because, in our simulation setup, the bad mouth attack gives positive or negative evaluations for nodes with 50% probability.

5.2. *Analysis of different attacks*

Figure 3 shows the comparison for different attacks. To start with, we set the percentage for conflicting behavior, sybil and bad mouth attack is 15% separately. From Figure 3(a), we could conclude that with historical data accumulated in the cloud, the behavior and sybil attacks are efficiently restrained. However, the bad mouth nodes are still involved in the communication with a higher proportion in Mix-zone. To track the bad mouth attack, we design another comparative test and add the percentage for bad mouth attack to 30% , meanwhile decrease the



(a) Attacks with 15% bad mouth nodes (b) Attacks with 30% bad mouth nodes

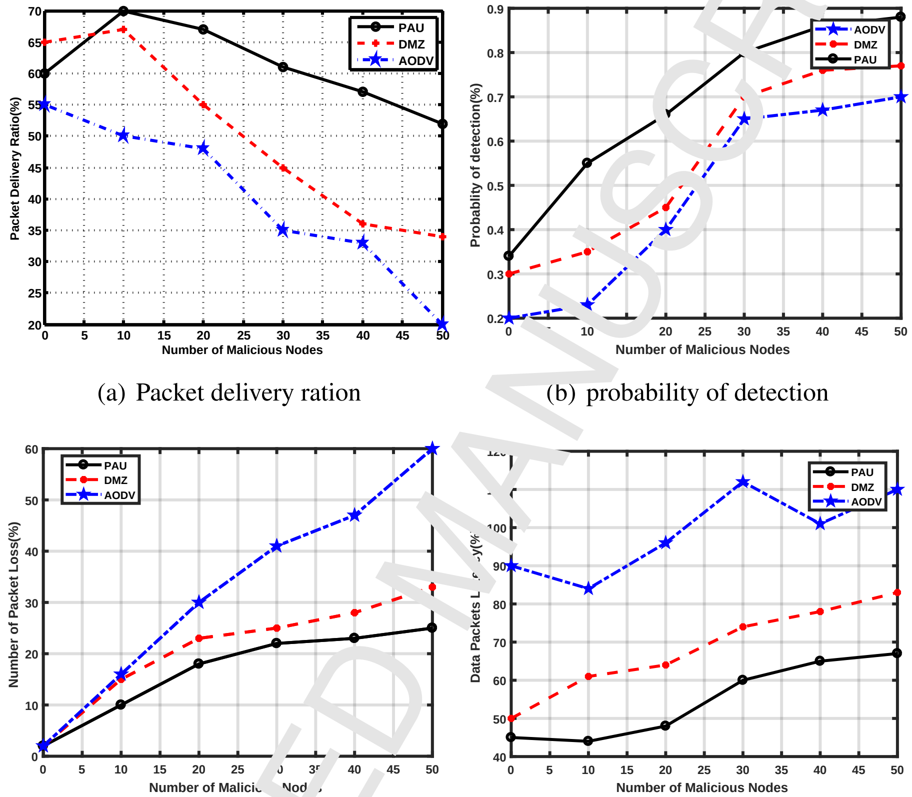
Figure 3: Performance comparison under different attack pattern

conflicting behavior, sybil attacks to 10% separately. The result is shown in Figure 3(b), we can conclude that even with the increase of bad mouth nodes, attacks are controlled and our scheme can discriminate the bad mouth nodes for more than 95% .

5.3. Comparison with existing schemes

The simulation is carried out to evaluate the effectiveness of our scheme. We compare our scheme with AODV[31] and DMZ [32]. For VANETs scenario, AODV is highly dynamic in nature and reducing overhead, because packet headers are not included in routes. Therefore, AODV seems to be theoretical because nodes in VANETs don't have any safety aware scheme. DMZ increases privacy significantly because it changes Pseudonym synchronously with dynamic privacy metric as well as location-based routing protocol. Thus, AODV and DMZ can represent the efficiency and privacy respectively. As shown in Figure 4(a), the packet delivery ratio for DMZ and AODV declines greatly with the increasing number of malicious nodes. Because the characteristic of DMZ and AODV schemes cannot distinguish malicious behaviors. As a result, our scheme performs better due to its privacy aggregation algorithm as mentioned in Section 4.3. Figure 4(b) reveals the probability of detection for malicious nodes. For Figure 4(c),the packet loss ratio for DMZ, AODV, and PAU is rising fast when the malicious number increases. The reason rests on the computing time for schemes to establish a valid communication route. As expected in Figure 4(d), it confirms that packet latency for DMZ and AODV schemes is fallen steeply as the malicious nodes increasing.

Our scheme could capture the evidence of malicious nodes overhead due to the historical data stored in the cloud.



(a) Packet delivery ration (b) probability of detection

(c) Packet loss ratio (d) Packet latency

Figure 4: Performance Analysis of PAU, DMZ and AODV

In summary, the simulation results confirm that our scheme performs better due to its ability to discriminate against malicious nodes and eliminate the issues related to three attack patterns.

6. Conclusion

In this paper, we present a Privacy Assessment Method with Uncertainty Consideration(PAU) to address the privacy breach problems in cloud-based V2X communication. By taking cognitive computing in the historical data, offline assessment oriented uncertainty could be accumulated in the cloud. PAU also captures the privacy-preserving capability based on real-time vehicles communication. Further, we present a privacy aggregation algorithm to combine the real-time and off-line opinion to improve the accuracy of privacy assessment. In the simulations, we design an algorithm by selecting nodes with high privacy awareness to establish the Mix-zone. The feature of experiments could verify our scheme in different aspects. Due to the usage of historical data stored in the cloud, our scheme performs well when defends against conflicting behavior and bad mouth attack. Comparison with existing privacy-preserving scheme, our scheme achieves high privacy-preserving and improves privacy protection level for the cloud-based V2X scenario.

7. Acknowledge

The work is supported by the National Natural Science Foundation of China under Grant N1601170045 and N176426[illegible], Young Scientists Fund of the National Project Foundation of China under Grant N51605197 and the Young Scientists Fund of the National Science Foundation of Jiangsu Province of China under No.BK20160524.

References

[1] 5G Americas, 5G Americas white paper: Cellular v2x communications towards 5g, `http://www.5gamericas.org/files/9615/2096/4441/2018_5G_Americas_White_Paper_Cellular_V2X_Communications_Towards_5G__Final_for_Distribution.pdf/` 1 (1) (2018) 1–38.

[2] Q. Zheng, K. Zheng, H. Zhang, V. C. Leung, Delay-optimal virtualized radio resource scheduling in software-defined vehicular networks via stochastic learning, IEEE Transactions on Vehicular Technology 65 (10) (2016) 7857–7867.

[3] S. Clauß, S. Schiffner, Structuring anonymity metrics, in: Proceedings of the second ACM workshop on Digital identity management, ACM, 2006, pp. 55–62.

[4] S. Seys, B. Preneel, Arm: Anonymous routing protocol for mobile ad hoc networks, International Journal of Wireless and Mobile Computing 3 (3) (2009) 145–155.

[5] X. Feng, L. Wang, S2PD: A selective sharing scheme for privacy data in vehicular social networks, IEEE Access 6 (2018) 55139–55148.

[6] Y. Li, C. Luo, R. Zhu, Y. Chen, H. Zeng, Efficient spatial keyword query processing in the internet of industrial vehicles, Mobile Networks and Applications 23 (4) (2018) 864–878.

[7] R. Zhu, X. Zhang, X. Liu, W. Shu, T. Mao, B. Jalaian, Erdt: Energy-efficient reliable decision transmission for intelligent cooperative spectrum sensing in industrial iot, IEEE Access 3 (2015) 2366–2378.

[8] J. Wang, R. Zhu, S. Liu, A differentially private unscented kalman filter for streaming data in iot, IEEE Access 6 (2018) 6487–6495.

[9] V. Balakrishnan, V. Varadharajan, U. Tupakula, Subjective logic based trust model for mobile ad hoc networks, in: Proceedings of the 4th international conference on Security and privacy in communication netowrks, ACM, 2008, pp. 30–38.

[10] I. Wagner, D. Eckhoff, Privacy assessment in vehicular networks using simulation, in: Proceedings of the 2014 Winter Simulation Conference, IEEE Press, 2014, pp. 3155–3166.

[11] M. Deng, K. Wuyts, R. Scandariato, B. Preneel, W. Joosen, A privacy threat analysis framework: supporting the elicitation and fulfillment of privacy requirements, Requirements Engineering 16 (1) (2011) 3–32.

[12] A. Pfitzmann, M. Hansen, A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management, http://dud.inf.tu-dresden.de/literatur/Anon_Terminology_v0.34.pdf/.

[13] K. Emara, Safety-aware location privacy in vanet: Evaluation and comparison, IEEE Transactions on Vehicular Technology 66 (12) (2017) 10718–10731.

[14] D. Chaum, The dining cryptographers problem: Unconditional sender and recipient untraceability, Journal of cryptology 1 (1) (1988) 65–75.

[15] D. Kesdogan, J. Egner, R. Büschkes, Stop-and-go-mixes providing probabilistic anonymity in an open system, in: International Workshop on Information Hiding, Springer, 1998, pp. 83–98.

[16] A. Studer, E. Shi, F. Bai, A. Perrig, Tacking together efficient authentication, revocation, and privacy in vanets, in: Sensor, Mesh and Ad Hoc Communications and Networks, 2009. SECON'09. 6th Annual IEEE Communications Society Conference on, IEEE, 2009, pp. 1–9.

[17] J. Freudiger, M. Raya, M. Félegyházi, P. Papadimitratos, J.-P. Hubaux, Mix-zones for location privacy in vehicular networks, in: ACM Workshop on Wireless Networking for Intelligent Transportation Systems (WiN-ITS), no. LCA-CONF-2007-016, 2007.

[18] V. M. Vishnevsky, A. A. Larionov, R. E. Ivanov, M. Dudin, Applying graph-theoretic approach for time-frequency resource allocation in 5g mmwave backhaul network, in: Advances in Wireless and Optical Communications (RTUWO), 2016, IEEE, 2016, pp. 221–224.

[19] 3GPP, Architecture enhancements for v2x services(release 15), TS23.285 V15.0.0.

[20] Y. Feng, B. Hu, H. Hao, Y. Gao, Z. Li, J. Tan, Design of distributed cyber-physical systems for connected and automated vehicles with implementing methodologies, IEEE Transactions on Industrial Informatics.

[21] V. Marojevic, C-v2x security requirements and procedures: Survey and research directions, arXiv preprint arXiv:1807.09338.

[22] 3GPP, Security architecture and procedures for 5g system, 3GPP TS 33.501.

[23] M. Raya, J.-P. Hubaux, Securing vehicular ad hoc networks, Journal of computer security 15 (1) (2007) 39–68.

[24] S. Dietzel, R. van der Heijden, J. Petit, F. Kargl, Context-adaptive detection of insider attacks in vanet information dissemination schemes, in: Vehicular Networking Conference (VNC), 2015 IEEE, IEEE, 2015, pp. 287–294.

[25] A. Dhamgaye, N. Chavhan, Survey on security challenges in vanet, International Journal of Computer Science (2013) 88–96.

[26] G. Theodorakopoulos, J. S. Baras, On trust models and trust evaluation metrics for ad-hoc networks, IEEE Journal on selected areas in Communications 24 (LCA-ARTICLE-2007-016) (2006) 318–328.

[27] X. Feng, C.-y. Li, D.-x. Chen, J. Tang, A method for defensing against multi-source sybil attacks in vanet, Peer-to-Peer Networking and Applications 10 (2) (2017) 305–314.

[28] F. Li, J. Wu, Mobility reduces uncertainty in manets, in: INFOCOM 2007. 26th IEEE International conference on computer communications., IEEE, 2007, pp. 1946–1954.

[29] T. Issariyakul, E. Hossain, Introduction to network simulator 2 (ns2), in: Introduction to Network Simulator NS2, Springer, 2012, pp. 21–40.

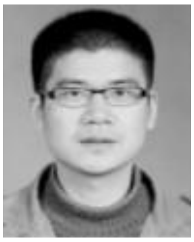
[30] M. K. Reiter, A. D. Rubin, Crowds: Anonymity for web transactions, ACM transactions on information and system security (TISSEC) 1 (1) (1998) 66–92.

[31] C. Perkins, E. Belding-Royer, S. Das, Ad hoc on-demand distance vector (aodv) routing, Tech. rep. (2003). doi:10.17487/RFC3561.

[32] F. Xia, Y. W. Liu, Dynamic mix-zone scheme with joint-entropy based metric for privacy protection in iov, Journal on Communications 39 (3) (2018) 76–85.

1) Proposing a novel method oriented subjective logic to predict the node's privacy breach level by analyzing the user's historical behavior;
2) Capturing the online privacy capability based on real-time vehicles communication observations,
3) Presenting a privacy aggregation algorithm to combine the online and offline opinion ;
4) Designing a dynamic Mix-zone construction algorithm to efficiently coverage and improve privacy protection level.

XIA FENG received the B.S. degree in Computer Science and Technology from University of Jiangsu, Jiangsu, China, in 2008, and the Ph.D degree in Computer technology from Anhui University, Anhui, China, in 2017. She is currently visiting the Computing School of University of Utah. She has published several technical papers at international journals and conferences. Her current research interests include security protocols, Vehicular and Ad-hoc network.



LIANGMIN WANG received the B.S. degree in computational mathematics from Jilin University, Changchun, China, in 1999, and the Ph.D. degree in cryptology from Xidian University, Xi an, China, in 2007. He is currently a full Professor with the School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang, China. He has published over 60 technical papers at international journals and conferences. His current research interests include security protocols and Internet of Things. He is a member of IEEE and ACM.