Hindawi
Mobile Information Systems
Volume 2017, Article ID 9050787, 16 pages
https://doi.org/10.1155/2017/9050787

Research Article

Trustworthy Event-Information Dissemination in Vehicular Ad Hoc Networks

Rakesh Shrestha and Seung Yeob Nam

Department of Information and Communication Engineering, Yeungnam University, 280 Daehak-Ro, Gyeongsan-si, Gyeongsangbuk-do 712-749, Republic of Korea

Correspondence should be addressed to Seung Yeob Nam; synam@ynu.ac.kr

Received 28 May 2017; Revised 1 October 2017; Accepted 17 October 2017; Published 12 November 2017

Academic Editor: Francesco Gringoli

In vehicular networks, trustworthiness of exchanged messages is very important since a fake message might incur catastrophic accidents on the road. In this paper, we propose a new scheme to disseminate trustworthy event information while mitigating message modification attack and fake message generation attack. Our scheme attempts to suppress those attacks by exchanging the trust level information of adjacent vehicles and using a two-step procedure. In the first step, each vehicle attempts to determine the trust level, which is referred to as truth-telling probability, of adjacent vehicles. The truth-telling probability is estimated based on the average of opinions of adjacent vehicles, and we apply a new clustering technique to mitigate the effect of malicious vehicles on this estimation by removing their opinions as outliers. Once the truth-telling probability is determined, the trustworthiness of a given message is determined in the second step by applying a modified threshold random walk (TRW) to the opinions of the majority group obtained in the first step. We compare our scheme with other schemes using simulation for several scenarios. The simulation results show that our proposed scheme has a low false decision probability and can efficiently disseminate trustworthy event information to neighboring vehicles in VANET.

1. Introduction

Vehicular networks are expected to be used for traffic control, accident avoidance, parking management, and so on [1]. Communication security between vehicles needs to be addressed carefully due to the safety requirements of vehicular network applications [2]. There is a lot of ongoing research on security topics, which aims to provide secure communications and verification of data to thwart malicious attackers. One of the major issues in vehicular ad hoc network (VANET) is message trust, which can be used to secure VANET communications. It is essential to periodically evaluate the trustworthiness of event information based on trust metrics. Generally, trust computation in a static network is relatively simple, because the trust level can be calculated based on the behavior of the nodes with sufficient observations [3]. However, message trust computation in VANET is challenging due to the ephemeral nature of the network topology.

The wireless access in vehicular environments (WAVE) protocol is based on the IEEE 802.11p standard and provides the basic radio standard for dedicated short range communication (DSRC) operating in the 5.9 GHz frequency band [4]. Vehicular communications can be achieved in the infrastructure domain for vehicle-to-infrastructure (V2I) communications or in the ad hoc domain for vehicle-to-vehicle (V2V) communications. We mainly focus on V2V communications because road side units (RSUs) [1] may not be available in some parts of the country during the initial stages of deployment of the vehicular communications infrastructure. Vehicles communicate with other vehicles through on-board units (OBUs) forming mobile ad hoc networks that allow communications in a completely distributed manner [5]. We note that some event information (e.g., accident reports) needs to be disseminated quickly and accurately, with minimum delay. Failure in timely and accurate dissemination of such time-critical information might lead to collateral damage to neighboring vehicles.

Some of the issues in vehicular networks include simple routing problems and application-oriented problems like Sybil attacks and false data dissemination [6]. The traditional reputation systems may not work efficiently in vehicular networks [7]. Public key infrastructure (PKI) may not be available everywhere during the initial stages of vehicular network deployment around a country, because some regions may not be covered due to deployment costs or budget issues. Generally, cryptography-based verification of message trustworthiness is computationally expensive. It can protect against a few types of attack from external nodes. However, it will not protect against malicious nodes in the network, which already have the required cryptographic keys, and may not be suitable for V2V ephemeral network communications. Our scheme does not use cryptography and centralized servers, and, thus, it does not have a single point of failure. Most VANET models assume that the system is up and running, where all vehicles have a certain trust score. However, it is not easy to know the trustworthiness of vehicles without having had any interaction with those vehicles. In highly distributed vehicular networks, vehicles can join and leave a network frequently [8, 9]. When a new vehicle joins the network for the first time, there is no information about it. One of the challenges faced by VANET is that the trust model of the VANET should consider the requirement for anonymity of vehicles. The trust model should have minimal overhead in terms of computation complexity, as well as storage. The trust model should be robust to data-centric attacks and be able to detect those attacks [10–12]. VANET security frameworks should be light, scalable, reliable, and secure.

Our proposed scheme investigates the trustworthiness of event information received from adjacent vehicles, which serves as multiple pieces of evidence. We use truth-telling probability as a measure for the trustworthiness of a vehicle. The vehicles communicate through safety messages to report events, such as accident information, safety warnings, information on traffic jams, weather reports, and reports of ice on the road. In our proposed scheme, all vehicles are assumed to have a pseudo identity (PID), which is independent of the node identity. Each vehicle broadcasts an event message to adjacent vehicles from the time it collects information about that event. Every vehicle maintains the trust level of its neighbors in a distributed manner to cope with the propagation of false information. We introduce an enhanced K-means clustering technique to minimize the effect of malicious nodes on trust level calculation. We use a modified threshold random walk algorithm with a single threshold to make a final decision about the occurrence of an event, while supporting real-time decision. We focus on determining the trustworthiness of the event information in the received messages by considering reports from neighboring vehicles differently with a truth-telling probability.

The main contributions of our work can be summarized as follows:

(i) Our proposed scheme can contribute to dissemination of trustworthy information since each vehicle makes a decision on the trustworthiness of information in the received message individually, while dropping packets containing fake information.

(ii) Since all the decisions are made based on the information received from the neighbor vehicles, our proposed scheme can work in an infrastructure-less environment as well.

(iii) Our proposed scheme can make a better decision on the trustworthiness of a given message compared to a simple voting mechanism, since the modified threshold random walk (TRW) can give a higher weight on the opinion of a vehicle, which makes more true statements than false statements.

The remainder of this paper is organized as follows. In Section 2, we discuss the related work. In Section 3, we propose a trustworthy event-information dissemination scheme for VANET. In Section 4, we evaluate the performance of the proposed scheme using simulation. Section 5 concludes the paper with future work.

2. Related Work

Several trust management systems have been proposed for VANET [13–15]. Trust management systems evaluate the trust values of the neighbor nodes to prevent them from interacting with the malicious nodes. The authors in [16] provide a quantitative and systematic review of existing trust management schemes for VANET. They address comprehensive trust model concepts, problems, and solutions related to VANET trust management. There are several works on trust management scheme based on infrastructure framework and cryptography techniques. Trust management schemes can be divided into four categories based on the use of infrastructure and cryptographic measures such as public key infrastructure (PKI) as shown in Table 1. The first category represents the trust management techniques based on infrastructure such as RSU and PKI. In the second category, the nodes rely on infrastructure for trust management without using PKI. In the third category, each node handles the issue of message trustworthiness based on PKI without using any infrastructure. In the fourth category, the nodes are fully decentralized and operate in infrastructure-less environment, and they do not depend on PKI.

In the first category, trust management systems are based on infrastructure such as RSU and PKI and can be effective in identifying malicious nodes with some accuracy [17–23]. However, trust management schemes in this category may not work if infrastructure is not available. The trust management based on PKI is computationally expensive and cannot secure VANET against insider attack, where the malicious nodes have already acquired the cryptographic keys [5, 23]. Some researchers used group signatures (GS) techniques [17–21] to authenticate message sender and guarantee message integrity such as Identity-Based Group Signatures (IBGS) in [18], GSIS in [19], and Identity-based Batch Verification (IBV) in [20]. However, GS schemes are usually based on PKI, and message sender authentication cannot prevent legitimate nodes from sending malicious messages.

Table 1: Trust management category based on infrastructure and PKI.

	PKI based	Non-PKI based
Infrastructure based	DTE [5], ESA [17], IBGS [18], GSIS [19], IBV [20], EGSS [21], iTrust [22], PMBP [23]	STRM [13], TRIP [24], RaBTM [25]
Non-infrastructure based	ETM [26], MTM [27], LSOT [28], BTM [29]	CTID [14], ITM [15], RMCV [30], ERM [31], VARS [32], TMEP [33], CRMS [34]

Trust management schemes in the second category require infrastructure including roadside unit or central authority without using PKI [13, 24, 25]. Schemes such as Trust and Reputation Infrastructure-based Proposal (TRIP) [24] and road side unit (RSU) and Beacon-Based Trust Management (RaBTM) [25] may not work if infrastructure is not available. In the third category, the issue of message trustworthiness was investigated based on PKI in an infrastructure-less environment [26–29]. In [27], the authors proposed a multidimensional approach for trust management in decentralized VANET environments using four different types of roles for different types of vehicles. The sender needs to authenticate itself to the receiver node using PKI for verifying each other's role implemented in a distributed manner. However, VANET faces several issues while deploying the PKI scheme due to key distribution in a real world.

In order to overcome the limitation of existing approaches, some researchers investigated trust management without using PKI in an infrastructure-less environment, which corresponds to the fourth category [14, 15, 30–34]. Trust management schemes in this category such as Vehicle Ad Hoc Reputation System (VARS) [32] are more suitable for distributed VANET architecture. Our proposed scheme also belongs to this category, since the decision on the trustworthiness of event information received from neighbor nodes is made in an infrastructure-less environment without using PKI.

The existing trust management systems are established on specific application domain implementing different trust-based models to enhance intervehicular communication. The trust-based models can be classified into three main categories. They are entity based, data-centric based, and hybrid trust models [35]. Entity based trust model deals with the trustworthiness of each node considering the opinions of the peer nodes [26–28, 36]. In [24], the authors proposed a fuzzy approach for the verification of the trustworthiness of the nodes by using feedback from their neighbors. However, the trustworthiness of a message may not always agree with the trustworthiness of the node itself. Thus, this model cannot resolve the issue of message trustworthiness properly.

On the other hand, in data-centric trust model, the trustworthiness of the reported events from the neighbor vehicles is evaluated rather than the trust of the entities or the node itself [5, 30, 31, 35]. In [5], the authors used a Bayesian inference decision module to evaluate the received event reports. But, the inference module uses the prior probability, which is not easy to obtain due to dynamic topology of VANET. In [28], the authors proposed a trust model called a Lightweight Self-Organized Trust (LSOT), which contains trust certificate-based and recommendation-based trust evaluations. However, it did not distinguish between the trust value of a node and that of the reported message. The trustworthiness of the nodes does not guarantee the trustworthiness of the message as the trustworthy nodes can send fake or faulty messages, if attackers compromise them. In [30], the authors proposed Real-time Message Content Validation (RMCV) scheme in an infrastructure-less mode. This scheme assigns a trust score to a received message based on three metrics, that is, message content similarity, content conflict, and message routing path similarity. The message trustworthiness is based on the maximum value of final trust scores collected from the neighbor nodes. However, this scheme does not consider high mobility of the vehicles and its time complexity is high.

Hence, a hybrid trust model is introduced that combines the entity based and data-centric trust models to evaluate the trustworthiness of a message [32–34]. The authors in [29] proposed a hybrid trust management mechanism called Beacon-based Trust Management (BTM) system, which constructs entity trust from beacon messages and computes data trust from crosschecking the plausibility of event messages and beacon messages. However, their trust model is based on PKI and digital signature, which incurs overhead while signing and authenticating each beacon message before broadcasting.

Thus, we attempt to overcome the limitation of the existing schemes by improving the hybrid trust model for message trustworthiness. As a first step, we use initialization-step enhanced K-means clustering algorithm (IEKA) for clustering of the vehicles into normal and malicious vehicle groups to determine the trustworthiness of each neighbor node. As a second step, we use a modified threshold random walk (TRW) algorithm to decide the trustworthiness of a given message. Thus, our scheme is based on a hybrid trust model. Although RMCV is based on data-centric trust model, it belongs to the fourth category, that is, trust management scheme that requires neither PKI nor infrastructure, according to the classification in Table 1. Thus, we compare our proposed scheme with the RMCV scheme. The detailed comparison and performance evaluation is discussed in Section 4.

3. System Model

Each vehicle collects sufficient information to assess the validity and correctness of a message. Notations explains the parameters and variables used in this paper.

When an event occurs on the road, a vehicle that is near that event sends the safety event message, M_E, to neighboring

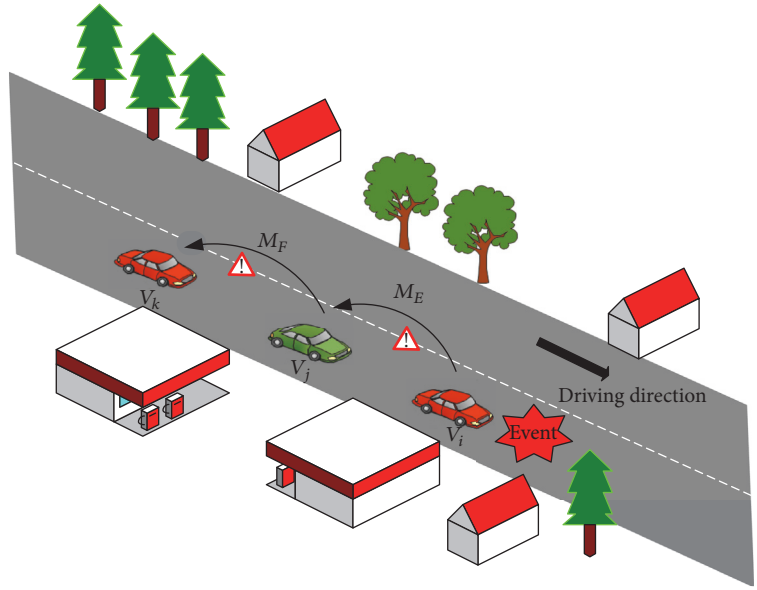


Figure 1: Trustworthy message dissemination scheme.

vehicles. Let us suppose that vehicle V_j wants to know the true information about the event reported by vehicle V_i in Figure 1.

The vehicle V_j manages an information pair $(\mathbf{p_i}, \boldsymbol{\theta_i})$ for each neighbor vehicle V_i, where $\mathbf{p_i}$ is the pseudo identity of the ith neighbor vehicle and $\boldsymbol{\theta_i}$ is the trust level, that is, truth-telling probability, of vehicle V_i. We assume that the transportation authority preloads the pseudo ID of the vehicles during vehicle registration, and it should be renewed periodically. To maintain the privacy in VANET, the pseudonym should change over time to achieve unlinkability that protects the vehicle from location tracking. Only privileged authorities are allowed to trace or resolve a pseudonym of the vehicle to a real identity under specific condition [37]. The truth-telling probability $(\boldsymbol{\theta_i})$ is the ratio of the number of true event reports propagated by vehicle V_i to the total number of event reports sent by vehicle V_i over a specific time period.

3.1. Proposed Trustworthy Information Dissemination Scheme. An outline of the proposed scheme to determine the trustworthiness of event information in the received message is shown in Algorithm 1. The vehicle parameters such as pseudo ID (PID) and default trust level are initialized at the beginning. All the vehicles periodically broadcast beacon messages, along with status information such as speed and location to neighboring vehicles. If there is no event triggered, then the vehicles will gather information from the neighboring vehicles. If a vehicle encounters any event by itself, then it broadcasts a safety message along with the trust levels of neighboring vehicles that it knows. Each vehicle accumulates the trust levels of the neighboring vehicles based on the collected safety messages. V_j creates a trust matrix based on the trust level opinions from other vehicles. Thus, the trust matrix manages the trust levels of each neighboring vehicle. Sometimes, vehicles misbehave by sending false information due to selfish motives like getting easier and faster access to the road, or due to faults. To prevent such false information that can corrupt the trust level of legitimate vehicles, we use a clustering algorithm. Our proposed clustering algorithm attempts to separate the trust level opinions of normal vehicles from the trust level opinions of malicious vehicles. The vehicle will calculate the aggregated trust level of adjacent vehicles belonging to the majority group of normal vehicles from the trust matrix. It will update the trust matrix using the average of trust levels. Then, a modified TRW is applied to know whether the event has actually occurred or not. The modified TRW can provide better decision on the trustworthiness of an event information by giving higher weights on the true event messages. After the trustworthiness of the event information has been verified, the event message is disseminated to other neighboring vehicles along with the updated trust levels. If the event information contained in the message turns out to be untrustworthy, then the message is dropped.

When new vehicles join the VANET, they are not likely to have enough information to infer the trust levels of neighboring vehicles at the beginning. We need a trust level bootstrapping procedure to assign a default trust level for this situation [38]. The trust level, that is, the truth-telling probability, ranges from 0 to 1. If vehicle A does not have any information on vehicle B, then the truth-telling probability of vehicle B is set to 0.5 at vehicle A. We assume that each vehicle sets the truth-telling probability for itself to 1 by default.

We mainly deal with two types of messages: beacon messages and safety messages. The vehicles use beacons to periodically broadcast and advertise status information to neighboring vehicles at intervals of 100 ms. The sender reports its speed, position, and so on to neighboring vehicles with beacon messages via one-hop communications [39]. On the other hand, safety messages support vehicles on the road

//The process is executed by a receiver vehicle upon receiving safety message
//p_{ij}: Pseudo ID of ith neighbor vehicle of V_j
//θ_j: truth-telling probability of V_j
//θ_{ij}: estimator of θ_i by V_j
//$\boldsymbol{\Theta}_j$: trust level opinion generated by V_j
//$\widehat{\theta}_i$: estimator for truth-telling probability of V_i
Input: $\mathbf{Y} = \{\boldsymbol{\Theta}_i\}$ $(i = 1, 2, \ldots, n)$
Output: $\{\widehat{\theta}_i\}$, updated trust matrix

(1) Information gathering from neighbor vehicles
(2) If event is triggered then goto step (5)
(3) Else goto step (2).
(4) If event source is the V_j itself then goto step (13).
(5) Else V_j accumulates the trust levels opinions of neighbor vehicles
$$\boldsymbol{\Theta}_j = ((p_{1j}, \theta_{1j}), (p_{2j}, \theta_{2j}), \ldots, (p_{jj}, \theta_{jj}), \ldots, (p_{nj}, \theta_{nj}))$$
(6) V_j generates a trust matrix based on the trust level opinions.
(7) Use modified clustering algorithm to separate trust level opinions of normal from malicious vehicles.
(8) Calculate aggregated trust level of adjacent vehicles belonging to majority group from trust matrix
$$\widehat{\theta}_i = (1/n) \sum_{j=1}^{n} \theta_{ij},$$
(9) Update the trust matrix
(10) Use modified TRW to know if the event has actually occurred or not.
(11) If we decide that the event message is trustworthy, then goto step (13).
(12) Else drop the message.
(13) Broadcast safety message and trust level to neighboring vehicles.

Algorithm 1: Determining trustworthiness of event information in the received message.

by delivering time-critical information so that proper action can be taken to prevent accidents and to save people from life-threatening situations. Safety messages include different types of events, E_x, such as road accidents, traffic jams, slippery roads, road constructions, poor visibility due to fog, and emergency vehicle warnings. Vehicles broadcast a safety message to neighboring vehicles when they encounter events on the road [1]. The message payload includes information about the vehicle's position, message sending time, direction, speed, and road events [19]. Each vehicle gathers information about the neighboring vehicles within its communication range.

One advantage of our proposed message dissemination scheme is to avoid a central trusted third party for trust accumulation in a distributed vehicular networking environment. We consider VANET without infrastructure such as RSUs. Vehicles communicate with each other in V2V mode using DSRC [40]. This allows fast data transmission for critical safety applications within a short range of 250 m. A basic safety application contains vehicle safety-related information, such as speed, location, and other parameters, and this information is broadcast to neighboring vehicles [41–43]. Let us consider two vehicles: V_i and V_j. The truth-telling probability of V_i depends on whether vehicle V_i is truthful when relaying event information. According to Velloso et al. [44], the more positive experiences vehicle V_j has with vehicle V_i, the higher the trust vehicle V_j will have towards vehicle V_i.

Let us suppose that vehicle V_i has a pseudo ID p_i and broadcasts a safety warning message M_E, which is defined in (1), when event E_x, where x represents an event type, is detected. If the vehicle itself detects the event, then it broadcasts the safety message along with the trust levels of neighboring vehicles. If a vehicle receives a safety message from other vehicles, it will accumulate the safety message along with trust levels from neighboring vehicles. When vehicle V_j collects event information from vehicle V_i, it finds the type and location of event from the message. Let event message M_E be given by

$$M_E = (p_i, t, L_E, l_i), \tag{1}$$

where p_i is the pseudo ID of vehicle V_i, t is the message generation time, L_E is the location of event E_x, and l_i is the location of V_i at time t.

In addition to this, each vehicle periodically broadcasts a beacon message defined as $M_B = (p_i, t_i, l_i, s_i)$, where p_i is the pseudo ID of V_i, t_i is the beacon generation time, l_i is the location of V_i, and s_i is the speed of V_i.

Let θ_i be the trust level, that is, truth-telling probability, of vehicle V_i. Truth-telling probability θ_i is defined as the ratio of the number of true events reported by vehicle V_i divided by the total number of events reported by vehicle V_i over a specific period of time. Let m denote the total number of true events reported by V_i and let n denote the total number of events reported by the vehicle up to the current time. Then, the truth-telling probability is

$$\theta_i = \frac{m}{n}. \tag{2}$$

A value for θ_i approaching 1 indicates reliable behavior of the corresponding vehicle, whereas a value close to zero indicates a high tendency towards providing false information [45].

3.2. Calculation of Trust Level of Neighbor Vehicles. When an event occurs, the nearby vehicles broadcast safety messages with additional data, such as pseudo IDs and truth-telling probabilities of other vehicles. Based on the safety messages from the neighboring vehicles, trust matrix $[\theta_{ij}]$ can be obtained, where θ_{ij} is estimation of θ_i by vehicle V_j. The trust matrix manages the truth-telling probability of each neighboring vehicle from the viewpoint of other vehicles. We assume that each vehicle sets its own truth-telling probability to 1. If the trust matrix is constructed, the aggregated trust level, that is, truth-telling probability of vehicle V_i, is calculated from the trust matrix by

$$\widehat{\theta}_i = \frac{1}{n}\sum_{j=1}^{n}\theta_{ij}, \tag{3}$$

where $\widehat{\theta}_i$ is the estimator for the truth-telling probability of V_i.

3.2.1. Estimation of Truth-Telling Probability Based on the Correctness of Message Information. If we can decide whether specific event information received from a vehicle is correct, this information can be used to estimate the truth-telling probability of the reporting vehicle more accurately. The reliable information about a specific event might be obtained from direct observation of an event spot, or announcement from a public and reliable group.

We explain how the truth-telling probability can be estimated more accurately if we collect more evidence to decide the correctness of messages generated by a given vehicle. We can estimate the truth-telling probability θ_i, defined in (2), based on the correctness of recent N messages from V_i. We introduce a random variable X_n to estimate the number of true reports among the recent N reports from V_i on arrival of the nth report from V_i. Then, the truth-telling probability of V_i can be estimated by X_n/N. We attempt to estimate X_n from X_{n-1} using the following relation:

$$X_n = \begin{cases} 1 + \left(1 - \dfrac{1}{N}\right) X_{n-1}, & \text{when the report } n \text{ is correct,} \\ \left(1 - \dfrac{1}{N}\right) X_{n-1}, & \text{otherwise.} \end{cases} \tag{4}$$

Then, we can show that X_n/N approaches the truth-telling probability θ_i of V_i under the assumption that the correctness of one message is independent of the correctness of other messages. By taking expectation on (4), we can obtain $E[X_n]$ as

$$\begin{aligned} E[X_n] &= \Pr[\text{message } n \text{ is correct}] \\ &\quad \cdot E[X_n \mid \text{message } n \text{ is correct}] \\ &\quad + \Pr[\text{message } n \text{ is incorrect}] \\ &\quad \cdot E[X_n \mid \text{message } n \text{ is incorrect}] \\ &= \theta_i E\left[1 + \left(1 - \frac{1}{N}\right) X_{n-1}\right] + (1 - \theta_i) \\ &\quad \cdot E\left[\left(1 - \frac{1}{N}\right) X_{n-1}\right] = \theta_i + \left(1 - \frac{1}{N}\right) E[X_{n-1}]. \end{aligned} \tag{5}$$

By solving the recursive relation in (5), we can obtain

$$E[X_n] = \left(1 - \frac{1}{N}\right)^n \{E[X_0] - \theta_i N\} + \theta_i N. \tag{6}$$

Thus, regardless of the initial condition on X_0, we have $\lim_{n\to\infty} E[X_n] = \theta_i N$, and $\lim_{n\to\infty} E[X_n/N] = \theta_i$ from (6). In other words, we can say that X_n/N approaches the truth-telling probability θ_i asymptotically, and we use the estimator of X_n/N and the relation in (4) to update the truth-telling probability of some vehicle whenever we have some evidence to determine the correctness of a message from that vehicle.

3.3. Clustering Algorithm. If there is no evidence to determine the truth of a given message, then the truth-telling probability of vehicle V_i will be calculated using (3). However, malicious vehicles can modify the trust levels of neighboring vehicles to mislead vehicles in a vehicular network. Thus, we need a clustering algorithm that can separate the trust levels of normal vehicles from the trust levels of malicious vehicles. It can reduce the effect of malicious vehicles on the trust levels of normal vehicles. In this subsection, we propose a new clustering algorithm to tackle this issue.

The main goal of our modified clustering algorithm is outlier detection. Our modified clustering algorithm classifies the trust level (truth-telling probability) opinions of the vehicles into two groups, one with the trust level opinions of normal vehicles and the other with the trust level opinions of malicious vehicles. We will select the majority group and neglect the outliers corresponding to the minority group.

Let us assume that an event has occurred on the road and the vehicles near the event location send event messages along with trust level opinions to neighbor vehicles. The vehicle V_j gathers reports about a specific event from neighbor vehicles and manages the trust level opinions of other vehicles as follows. Each vehicle maintains a sorted vehicle list (SVL), which manages pseudo IDs of all the adjacent vehicles in an ascending order, and the vehicle index will be assigned based on the sequence in the sorted list as shown in Table 2. Whenever a vehicle V_j needs to disseminate its own trust level opinion to its neighbors, it sends its trust level opinion $\mathbf{\Theta}_j$ defined as

$$\mathbf{\Theta}_j = \left(\left(p_{1j}, \theta_{1j}\right), \left(p_{2j}, \theta_{2j}\right), \ldots, \left(p_{jj}, \theta_{jj}\right), \ldots, \left(p_{nj}, \theta_{nj}\right)\right), \tag{7}$$

where p_{kj} is the pseudo ID of the kth neighbor vehicle of the vehicle j and θ_{jj} is likely to be set to 1 because every node will trust itself.

If V_i receives trust level opinion $\mathbf{\Theta}_j$, then V_i updates its own SVL by adding the vehicles that are in $\mathbf{\Theta}_j$, but are not in the SVL. After updating SVL, V_i derives $\widetilde{\mathbf{\Theta}}_{\mathbf{j}'}$ from the received $\mathbf{\Theta}_j$ as

$$\widetilde{\mathbf{\Theta}}_{\mathbf{j}'} = \left(\widetilde{\theta}_{1\mathbf{j}'}, \widetilde{\theta}_{2\mathbf{j}'}, \ldots, \widetilde{\theta}_{j'\mathbf{j}'}, \ldots, \widetilde{\theta}_{n'\mathbf{j}'}\right), \tag{8}$$

Table 2: Example of sorted vehicle list (SVL).

Vehicle index	Pseudo ID
1	1135
2	2056
3	2079
4	2146
5	3012
...	...

where j' is the new index of the vehicle j according to its sequence in the updated SVL and n' is the total number of vehicles in the updated SVL. When p_{kj} in the received $\boldsymbol{\Theta}_j$ agrees with the k'th pseudo ID in the updated SVL, $\widetilde{\theta}_{k'\mathbf{j}'}$ in $\widetilde{\boldsymbol{\Theta}}_{\mathbf{j}'}$ is updated as

$$\widetilde{\theta}_{k'\mathbf{j}'} = \theta_{kj}. \tag{9}$$

In this case, n' is always larger than or equal to n since $\widetilde{\boldsymbol{\Theta}}_{\mathbf{j}'}$ accommodates all the vehicles in $\boldsymbol{\Theta}_j$. If $n' > n$, then it means that there is some pseudo ID that is in the SVL, but not in $\boldsymbol{\Theta}_j$. If an index l corresponds to such a pseudo ID, $\widetilde{\boldsymbol{\theta}}_{\mathbf{lj}'}$ will be set to 0.5 since the vehicle j' does not know the vehicle l. If $\widetilde{\boldsymbol{\Theta}}_{\mathbf{j}'}$ is derived, then the trust matrix table is updated by adding the transpose of $\widetilde{\boldsymbol{\Theta}}_{\mathbf{j}'}$ as the j'th column.

If each vehicle includes the pseudo IDs and the truth-telling probabilities of all the vehicles that it knows in the trust level opinion message defined in (7), then the traffic overhead due to this message can be excessively large. However, we can reduce the message overhead by omitting trivial information. For example, if θ_{kj} in $\boldsymbol{\Theta}_j$ is 0.5, this means that the vehicle j does not know the vehicle k since 0.5 is the default value used to initialize the truth-telling probability of a new vehicle. In this case, the vehicle j need not advertise this probability because this default value can be easily filled up by the neighbor vehicles according to the trust matrix updating rule mentioned above with (7), (8), and (9). The vehicles on the road are likely to be ignorant of each other in terms of the trust matrix table, since they need not exchange the trust level opinions if there is no event. Thus, we expect that the policy of omitting trivial information can significantly reduce traffic overhead due to trust level opinion messages.

If V_j collects trust level opinions $\boldsymbol{\Theta}_i$ $(i \neq j)$ from other vehicles along with event information, then V_j can construct trust matrix Γ, defined as

$$\Gamma = \begin{bmatrix} \theta_{11} & \theta_{12} & \cdots & \theta_{1n} \\ \theta_{21} & \theta_{22} & \cdots & \theta_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{n1} & \theta_{n2} & \cdots & \theta_{nn} \end{bmatrix}. \tag{10}$$

We use a simple example to show our proposed clustering algorithm illustrated by Table 3 and Figure 2. Table 3 shows an example of trust matrix defined in (10), when $n = 3$. Three columns in Table 3 correspond to points A, B, and C in Figure 2, respectively. The three probability values in the first column of Table 3 correspond to θ_{11}, θ_{21}, and θ_{31}, respectively, and these values are estimation of θ_1, θ_2, and θ_3 by V_i. This tuple of probability values is represented as point A in the three-dimensional space of Figure 2, where each axis represents θ_1, θ_2, and θ_3, respectively. If there is no attacker, that is, all vehicles tell the truth, then all the points will be close to each other. The trust level opinions ($\boldsymbol{\Theta}_i$'s) with similar characteristics are likely to form the same cluster. If there is an attacker that tells a lie, then the corresponding point will deviate from the majority group, and this point can be distinguished as an outlier. Even if the attacker tries to change or give higher trust levels by using collusion attack, it can be detected as an outlier. Let us suppose that point C represents an attacker that tells a lie by changing the trust level, as shown in Table 3. The clustering algorithm will separate the trust level opinions into two groups, one group with normal-vehicle trust levels (i.e., A and B) and the other group with malicious vehicle trust levels (i.e., C). Figure 2 describes the outcome of one possible clustering algorithm. The final aggregated trust level is calculated based on the trust level opinions corresponding to the majority group using (3). The final aggregated trust level based on the majority group will be used to update the trust level of the vehicle itself. The resulting trust level is appended to the message during message propagation.

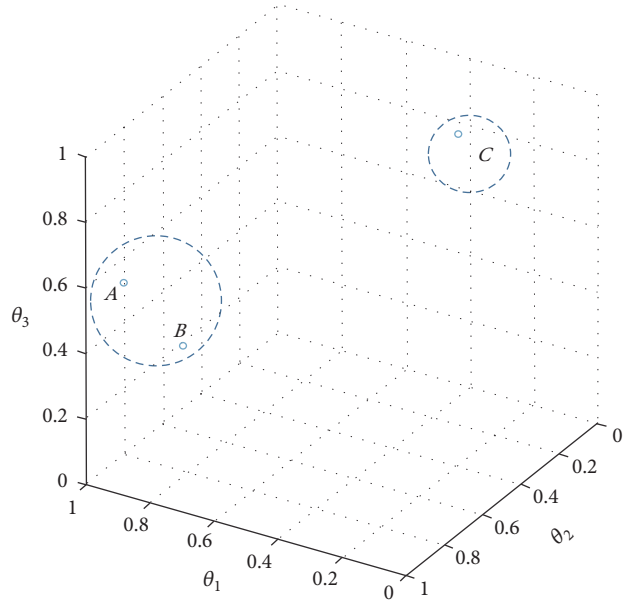


Figure 2: Clustering example for the proposed clustering algorithm.

Table 3: Trust matrix table of vehicle.

V_j	V_i 1	2	3
1	1	0.7	0.2
2	0.8	1	0.4
3	0.5	0.5	1
	A	B	C

Input: $\mathbf{Y} = \{\boldsymbol{\Theta}_i\}$ $(i = 1, 2, \ldots, n)$
Output: $\mathbf{C} = \{\mathbf{c}_j\}$ $(j = 1, 2)$
Initialize: Calculate unique centroid as initial cluster center,

$$\boldsymbol{\mu} = \frac{\sum_{i=1}^{n} \boldsymbol{\Theta}_i}{|\mathbf{Y}|}$$

(1) for each $\mathbf{c_j} \in \mathbf{C}$ do
(2) $\mathbf{c_j} \leftarrow \boldsymbol{\Theta}_i \in \mathbf{Y}$
(3) $\mathbf{c_1} = \arg\max_{\boldsymbol{\Theta}_i \in \mathbf{Y}} \|\boldsymbol{\mu} - \boldsymbol{\Theta}_i\|$
(4) $\mathbf{c_2} = \arg\max_{\boldsymbol{\Theta}_i \in \mathbf{Y}} \|\mathbf{c_1} - \boldsymbol{\Theta}_i\|$
(5) end
(6) While two centroids are not converged, do
(7) for each $\boldsymbol{\Theta}_i \in \mathbf{Y}$ do
(8) Assign $\boldsymbol{\Theta}_i$ to nearest centroid,
(9) $\mathbf{c_j} = \arg\min_j \|\mathbf{c_j} - \boldsymbol{\Theta}_i\|^2$
(10) end
(11) Update cluster centroid;
(12) Calculate new centroid $\mathbf{c_j}$ as
(13) for each $\mathbf{c_j} \in \mathbf{C}$ do
(14) $\mathbf{c}_j = (1/|\mathbf{c_j}|) \sum^{i}_{\boldsymbol{\Theta}_i \in c_j} \boldsymbol{\Theta}_i$,
(15) end
(16) end

Algorithm 2: Proposed modified clustering algorithm.

We propose a modified K-means clustering algorithm. The main problem with a K-means algorithm lies in the initialization step, so we introduce an enhanced K-means clustering technique by modifying the initialization step, which is called initialization-step enhanced K-means clustering algorithm (IEKA). We use the IEKA to cluster the trust level opinions, while reducing the effect of malicious vehicles on trust levels for other vehicles. Our proposed clustering algorithm can be described in more detail as follows.

After generating a trust matrix, IEKA partitions the trust level opinions into $K(\leq n)$ groups $\mathbf{C} = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_k\}$. We designate $\mathbf{Y}$ to be a set of the $\boldsymbol{\Theta_i}$ vectors; that is, $\mathbf{Y} = \{\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2, \ldots, \boldsymbol{\Theta}_n\}$. We consider only two clusters for our scheme. Initially, we take the mean of all the data points in $\mathbf{Y}$ to find a unique centroid, that is, $\boldsymbol{\mu}$.

$$\boldsymbol{\mu} = \frac{\sum_{i=1}^{n} \boldsymbol{\Theta}_i}{|\mathbf{Y}|}. \tag{11}$$

We calculate the Euclidean distance between $\boldsymbol{\mu}$ and each vector in $\mathbf{Y}$. Choose the point that has the maximum distance from the unique centroid; that is, the selected point is at the farthest distance from the unique centroid. We consider this point as the first centroid, $\mathbf{c}_1$, for the first cluster:

$$\mathbf{c_1} = \arg\max_{\boldsymbol{\Theta}_i \in \mathbf{Y}} \left\| \boldsymbol{\mu} - \boldsymbol{\Theta}_i \right\|. \tag{12}$$

Similarly, we compute the Euclidean distance between first centroid $\mathbf{c_1}$ and the remaining points in $\mathbf{Y}$ and select the point with the maximum distance from $\mathbf{c_1}$. Then, this point becomes the second centroid, $\mathbf{c_2}$:

$$\mathbf{c_2} = \arg\max_{\boldsymbol{\Theta}_i \in \mathbf{Y}} \left\| \mathbf{c_1} - \boldsymbol{\Theta}_i \right\|. \tag{13}$$

As a next step, we run the conventional K-means clustering algorithm, with $\mathbf{c_1}$ and $\mathbf{c_2}$ being the centroids of two separate groups. Update centroids $\mathbf{c_1}$ and $\mathbf{c_2}$ by calculating the mean value for each group. This gives new centroids $\mathbf{c_1}$ and $\mathbf{c_2}$ and then reassigns each data point to the cluster to which it is closest. We will repeat this process until those two centroids converge. The proposed modified clustering algorithm is given in Algorithm 2.

After clustering using Algorithm 2, which is based on (11), (12), and (13), the aggregated trust level of each neighbor vehicle is calculated based on the trust level opinions belonging to the majority group. We assume that the number of malicious vehicles is less than that of normal vehicles. The aggregated trust level is used for TRW calculation. In the next subsection, we discuss the decision on the event based on hypothesis testing using TRW.

3.4. Event Decision Based on Threshold Random Walk (TRW). Sequential hypothesis testing is usually used to determine if a specific hypothesis is true or not based on sequential observations [46]. Among the sequential hypothesis testing schemes, threshold random walk has been used to detect scanners with a minimal number of packet observations, while guaranteeing false positives and false negatives [47]. Since we are interested in determining whether a given message is true or not, if true message constitutes one of

Table 4: Aggregated trust level table.

Neighbor PID	Trust level (truth-telling prob.)	Event observations (X_i)
p_1	$\widehat{\theta}_1$	$X_1 = E_1$
p_2	$\widehat{\theta}_2$	$X_2 = \overline{E}_1$
p_3	$\widehat{\theta}_3$	$X_3 = \overline{E}_1$
...	...	...
p_n	$\widehat{\theta}_n$	$X_n = E_1$

the two hypotheses, then threshold random walk might be applied to this problem. The threshold random walk scheme in [47] uses two thresholds, that is, one upper bound and one lower bound, and the decision is made when a likelihood ratio reaches either threshold. However, in this threshold random walk scheme, we cannot know the number of samples required to reach either threshold in advance. This means that real-time decisions may not be possible if we cannot collect a sufficient number of samples in a short interval. In this paper, we use a modified threshold random walk scheme to determine the validity of a given event, while resolving the issue of real-time decision. We resolve this issue by applying threshold random walk with a single threshold instead of two thresholds. Hereafter, we explain the threshold random walk (TRW) scheme applied to our problem in more detail.

E_1 represents one of the events that can happen on a road. After clustering trust level opinions of neighbor vehicles using IEKA, each vehicle determines the occurrence of event E_1 based on the aggregated trust level table. The aggregated trust level table consists of vehicle PIDs, aggregated trust levels, and event observations, as shown in Table 4.

In Table 4, X_i is the report received from the ith neighbor vehicle about event E_1. Event E_1 represents the occurrence of an event, and $\overline{E}_1$ represents nonoccurrence of event E_1. We need a rule to make a decision about the occurrence of the event. We assume that X_i's are independent of each other among different vehicles. For a given event, suppose random variable Y_i can take only two values (0 and 1); that is,

$$Y_i = \begin{cases} 0; & \text{if } X_i = E_1 \\ 1; & \text{if } X_i = \overline{E}_1. \end{cases} \tag{14}$$

After collecting a sufficient number of reports, we wish to determine whether the event (E_1) has really occurred using sequential analysis [46].

Let us consider two hypotheses: one is null and the other is an alternate hypothesis (i.e., H_0 and H_1), where H_0 is the hypothesis that event E_1 has occurred and H_1 is the hypothesis that event E_1 has not occurred, that is, $\overline{E}_1$. We also assume that conditionals on the hypothesis $Y \mid H_j$ where $j = 0, 1$ are independent. From the definition of the truth-telling probability and (14), we obtain

$$\begin{aligned} \Pr(Y_i = 0 \mid H_0) &= \theta_i, \\ \Pr(Y_i = 1 \mid H_0) &= 1 - \theta_i, \\ \Pr(Y_i = 0 \mid H_1) &= 1 - \theta_i, \\ \Pr(Y_i = 1 \mid H_1) &= \theta_i, \end{aligned} \tag{15}$$

where $\Pr(Y = k \mid H_j)$ is the conditional probability that the observation of Y, given hypothesis H_j, is k. Then, $\Pr(Y_i = 0 \mid H_0) = \theta_i$ becomes the truth-telling probability, and $\Pr(Y_i = 1 \mid H_0) = 1 - \theta_i$ becomes the lying probability. In order to make a timely decision, we collect report samples from neighbor vehicles during an interval of fixed duration T. Let N denote the number of report samples collected during this interval. Following the approach of Wald [46], we use collected report samples to calculate the likelihood ratio by

$$\Lambda = \prod_{i=1}^{N} \frac{\Pr(Y_i \mid H_1)}{\Pr(Y_i \mid H_0)}. \tag{16}$$

Although the TRW scheme in [47] makes a decision based on two thresholds, the upper and lower bounds, we use a single threshold to make a decision without the issue of long waiting time. When the threshold is η, the decision rule is as follows:

If $\Lambda \geq \eta$, then accept hypothesis H_1.

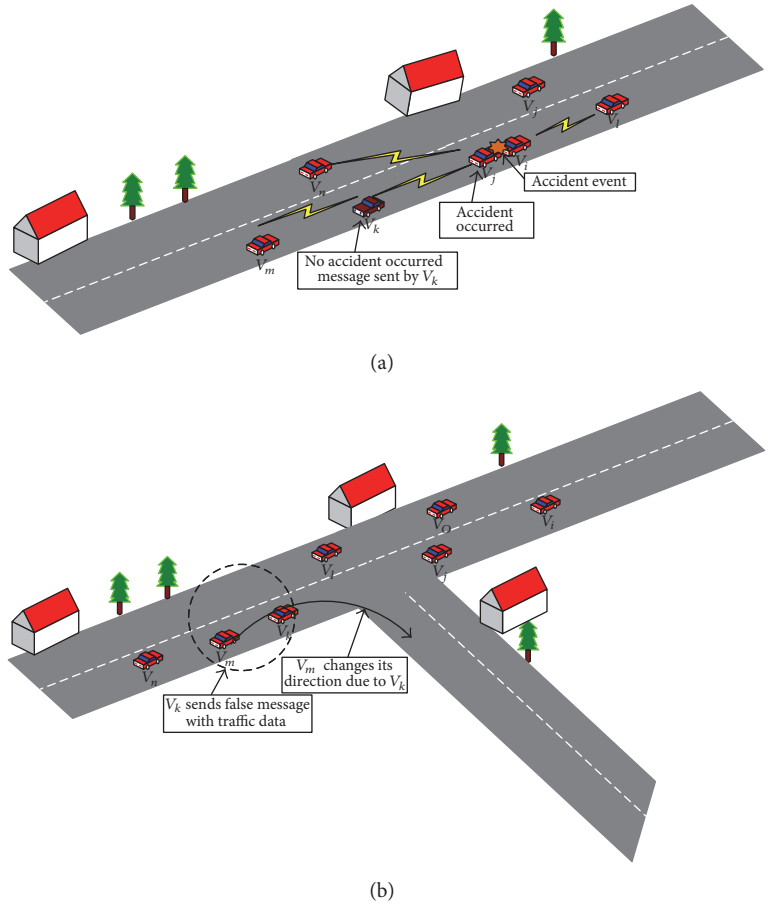
If $\Lambda < \eta$, then accept hypothesis H_0.

In this paper, the threshold η will be set to 1, and the truth-telling probability θ_i of an unknown vehicle i will be set to 0.5. When a vehicle receives N report messages, if the Nth report has come from a vehicle with no information on the truth-telling probability, $\Pr(Y_N \mid H_1) = \Pr(Y_N \mid H_0)$ since $\theta_i = 1 - \theta_i$. Thus, the report from the unknown vehicle will not affect the likelihood ratio by (15) and (16). Furthermore, if all the report messages are from the vehicles with no history information, then the likelihood ratio in (16) becomes 1, and, thus, it is fair to put $\eta = 1$, since it is not easy to make a decision in this case.

The advantage of our threshold random walk compared to a simple voting scheme can be described with a simple example as follows. Let us consider a case where an event E_1 is true, and a vehicle receives 5 report messages. Among them, only two report that E_1 is true, and the other three claim that E_1 did not happen. If we make a decision based on a simple voting, then the decision will be $\overline{E}_1$. However, if we apply threshold random walk considering the truth-telling probability of each node, the decision can be different as follows. If the truth-telling probability of the two nodes claiming E_1 is 0.8 and the truth-telling probability of the three nodes claiming $\overline{E}_1$ is 0.6, then the likelihood ratio defined in (16) becomes

$$\begin{aligned} \Lambda(X) &= \frac{0.2}{0.8} \times \frac{0.2}{0.8} \times \frac{0.6}{0.4} \times \frac{0.6}{0.4} \times \frac{0.6}{0.4} = 0.21 \\ &< 1 (= \eta). \end{aligned} \tag{17}$$

Thus, we will select the hypothesis H_0 according to the decision rule mentioned above, since the likelihood ratio calculated in (17) is less than the threshold η. This means



(a)

(b)

Figure 3: Two types of attack patterns considered in this paper: (a) message modification attack and (b) fake message generation attack.

the correct decision of E_1 is made by the proposed threshold random walk. This advantage comes from the fact that the likelihood ratio in (16) gives a higher weight to the opinion of vehicles with a high truth-telling probability.

After the decision on the actual occurrence of the event is made, vehicle j will forward the received message to its neighboring vehicles (with aggregated trust levels) within radio range, which is denoted by

$$M_F = \left(p_j, t, M_E, \mathbf{\Theta}_j\right), \tag{18}$$

where p_j is the PID of vehicle j, which forwards the message, t is the time at which M_F was sent, and $\mathbf{\Theta}_j$ denotes the trust level opinion of vehicle j defined in (7).

3.5. Attack Model. We consider two types of attacks: message modification attack and fake message generation attack in a VANET environment. Figure 3 shows an example of both message modification and fake message generation attack. A malicious vehicle might modify warning messages, either with malicious intent or due to an error in the communications system. In the message modification attack, malicious vehicles can modify message information at any time and falsify the parameters.

In Figure 3(a), when an accident event occurs on the road, the vehicles in an accident or the vehicles which are close to that accident broadcast the accident event message. After vehicle V_j sends an accident report to other vehicles, a malicious vehicle V_k modifies the message and sends the modified no-accident message as M_F, defined in (18), with the intent to affect decisions taken by other vehicles. Similarly, in a fake message generation attack, malicious vehicles generate a false warning message. For example, in Figure 3(b) [48], a malicious vehicle might send an accident message to neighboring vehicles, even when there is no such event on the road, to clear the route it wants to take. In this case, the malicious vehicle wants to convince other vehicles that an event has occurred. In this scenario, the attacker may have already compromised one or more vehicles and launches attacks by generating a fake message for neighboring vehicles. We assume that the number of malicious vehicles is less than the number of normal vehicles [12]. In simulation, we vary the number of malicious vehicles from 5% to 50% of overall vehicles to evaluate the performance of our proposed scheme in an adversarial environment.

4. Performance Evaluation

4.1. Simulation Setup. The performance of our proposed scheme was evaluated through simulation. We used the Vehicles in Network Simulation (VEINS) framework version

Table 5: Simulation parameters.

Parameters	Value
Network simulation package	OMNET++
Vehicular traffic generation tool	SUMO
Wireless protocol	802.11p
Simulation time	300 s
Scenario	Urban/highway
Transmission range	250 m

4a2 [49], which is based on both OMNeT++ version 4.6 [50], a discrete event-driven network simulator, and Simulation of Urban Mobility (SUMO) version 22 for road traffic simulation [51]. VEINS connects OMNeT++ and SUMO through Traffic Control Interface (TraCI). VEINS provides realistic models for IEEE 802.11p networks. It provides OMNeT with a set of application programming interfaces to connect the SUMO platform and to dynamically access information about SUMO simulated objects. SUMO allows the creation of scenarios that include realistic mobility patterns, such as vehicle movement and overtaking, as well as lane changing.

We use the default map of Erlangen, Germany, from the VEINS framework with the map size of 2500 m × 2500 m for our simulation. We evaluated our scheme under different traffic densities to consider diverse situations. When the vehicles reach the edge of the road, the vehicles reroute their path and can meet other vehicles multiple times during simulation. The number of vehicles increases linearly with time from 0 s to 300 s. The average vehicle speed changes from 40 km/h in an urban scenario to 110 km/h for highway scenarios. The key parameters considered in our simulation are summarized in Table 5.

We considered two scenarios (urban and highway) by varying parameters such as speed, vehicle density, and percentage of malicious vehicles, as shown in Table 6. The number of malicious vehicles was varied considering the mobility of vehicles in a realistic simulation environment by adjusting vehicle densities and vehicle speeds. We assume that the normal vehicles and the malicious vehicles are uniformly distributed on the roads for each ratio of malicious vehicles [52].

4.2. Simulation Results. In this section, we analyze the simulation results based on OMNet++. The traffic density increases from free-flow traffic (5 vehicles/km^2) to congested traffic (100 vehicles/km^2) where vehicles can meet multiple times. The simulation scenarios are summarized in Table 6. For performance evaluation, we have considered false decision probability and message overhead. We compared our scheme with other schemes under different scenarios. In order to evaluate our proposed scheme, we considered the message modification attack and the fake message generation attack one by one, while increasing the number of malicious vehicles from 5% to 50% in both scenarios. The positions of normal vehicles and the initial distribution of the attackers were randomly determined. We calculated the average false decision probability by averaging the simulation results for 30 simulation runs. A decision is regarded as a false decision when the decision result does not agree with the true status of the event at the time of the decision. In other words, a false decision probability is the ratio of the number of incorrect decisions to the total number of decisions.

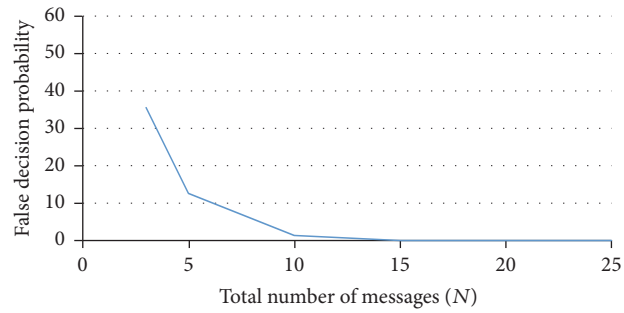


Figure 4: False decision probability versus total number of messages (N).

In order to update the truth-telling probability of vehicle V_i based on the truth of a given message according to (4), we need to decide the parameter N, that is, the number of recent messages from V_i that will be considered in this estimation. In order to decide N, we run 20 simulations under fake message attack with 30% of malicious vehicles in a highway scenario. We calculate the average false decision probability for various values of N, and Figure 4 shows the result. As the value of N increases, the false decision probability tends to decrease. The false decision probability reaches zero at $N = 15$ and does not change for larger values of N. Thus, N is fixed to 15 hereafter based on this result.

In Figure 4, the false decision probability is high for lower values of N. Let us consider an example to explain worse performance for lower values of N. Let us take an extreme case of $N = 1$. Then, this means that when V_j receives a message from V_i, it decides the truth-telling probability of V_i only based on the last message, since $N = 1$. Thus, if V_j finds that the last message from V_i was false, then V_j will think that the truth-telling probability of V_i is 0, according to the updating rule described in (4). On the other hand, if V_j finds that the last message from V_i was true, then V_j will think that the truth-telling probability of V_i is 1. Thus, the estimated truth-telling probability of each vehicle is either 0 or 1. However, if the truth-telling probability of a given vehicle is different from 0 or 1, then this updating rule (with $N = 1$) will never find the accurate truth-telling probability, since the truth-telling probability is always 0 or 1 according to the updating rule. Hence, the truth-telling probability can be significantly different from the correct truth-telling probability for lower values of N, especially when $N = 1$.

For our modified TRW scheme, we need to determine an optimal value of message collection time T to achieve a good decision accuracy. We run several simulations under fake message generation attack with 30% of malicious vehicle in highway scenario. We calculated the average false decision probability against the message collection time under the simulation parameters given in Table 5. In the sparse network, we received report message as low as five messages

Table 6: Simulation scenarios.

Scenario number	Type	Total vehicles	Malicious vehicle ratio	Average speed	Std. of speed
(1)	Urban	100	0~50%	40 km/h	4.76
(2)	Highway	100	0~50%	110 km/h	7.52

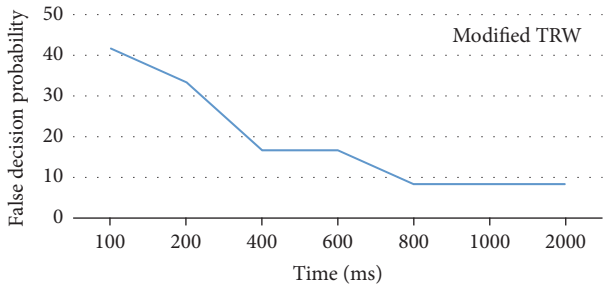


Figure 5: False decision probability for various values of message collection time (T).

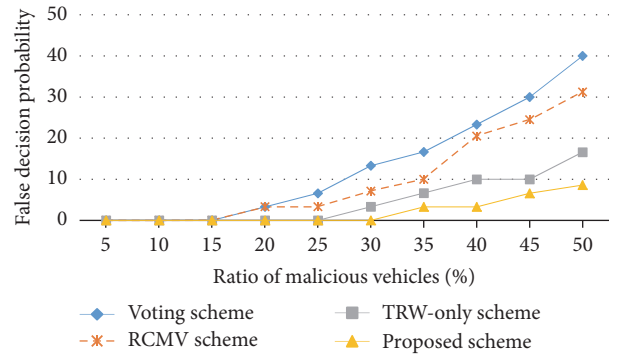


Figure 6: Comparison of the proposed scheme with other schemes under a message modification attack in a highway scenario.

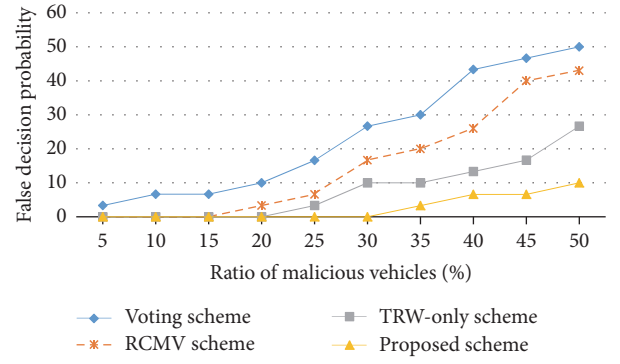


Figure 7: Comparison of the proposed scheme with other schemes under a fake message generation attack in a highway scenario.

with report collection time less than 100 ms which results in high false decision probability. The simulation result is shown in Figure 5. As the report collection time interval increases, the false decision probability decreases and, from 800 ms, the false decision probability does not decrease anymore. Based on this, we set the value of T to 1 sec, and this value will be used for T hereafter.

We now compare our proposed scheme with other schemes: RMCV scheme, a simple voting scheme, and TRW-only scheme. RMCV is an information oriented trust model and the outcome of the scheme is a trustworthiness value associated with each received message. In RMCV scheme, we consider the message trustworthiness based on the content similarity. The message trustworthiness is likely to increase as the message contents are similar among different vehicles. In the TRW-only scheme, we used the modified threshold random walk to make a decision about the event in the warning message without applying our proposed clustering algorithm. Several voting methods have been proposed to estimate the trustworthiness of each report message [53–55]. In the simple voting mechanism, each vehicle collects a fixed number of warning messages from the neighboring vehicles regarding an event and makes a decision by following the opinion of the majority group [55]. For the voting scheme, we collected 15 messages to make a decision, as this was the optimal number according to our simulation.

We compare our proposed scheme with other schemes in terms of false decision probability for various ratios of malicious vehicles under the message modification attack in a highway scenario as shown in Figure 6. We can see that our proposed scheme yields a lower false decision probability compared to the other mechanisms, even when the number of malicious vehicles increases. The simple voting mechanism performs worst among the four schemes. The performance of the RMCV scheme is close to TRW-only scheme when the malicious vehicle ratio is low. However, it degrades significantly compared to our proposed scheme as the malicious vehicle ratio increases. Our proposed scheme has a false decision probability of 0% when the ratio of malicious vehicles is 30% in highway scenario.

We compare our proposed scheme with other schemes in terms of false decision probability under a fake message attack in a highway scenario as shown in Figure 7. We consider a case where the attacker generates messages about a fake event. Our proposed scheme yields better performance compared to the RMCV, simple voting, and TRW-only schemes with a low false decision probability of less than 10%. The false decision probability of RMCV and voting scheme exceed 40% when the ratio of malicious vehicles increased to 50%. In Figure 7, the false decision probability increases as the ratio of malicious vehicles increases, with a tendency similar to Figure 6.

We compare our proposed scheme with other schemes in terms of false decision probability under a message modification attack in an urban scenario in Figure 8. Our proposed scheme exhibits better performance compared to other schemes. The false decision probability does not exceed

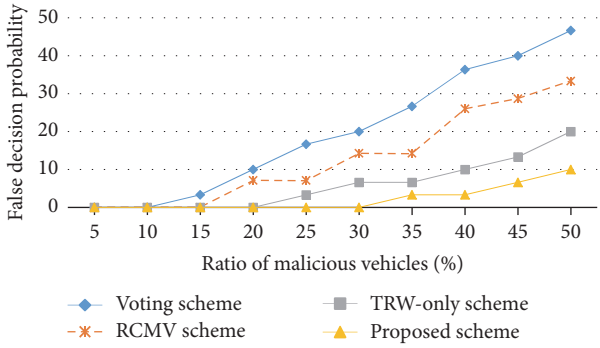


Figure 8: Comparison of the proposed scheme with other schemes under a message modification attack in an urban scenario.

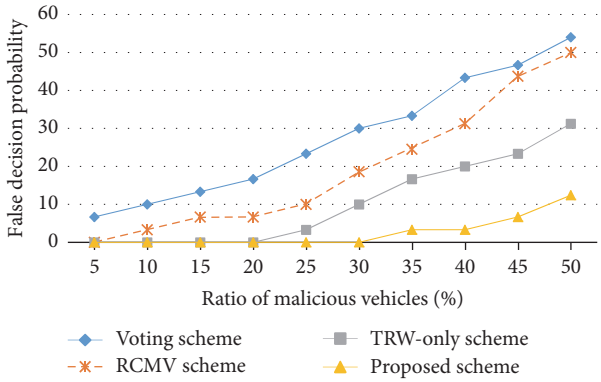


Figure 9: Comparison of the proposed scheme with other schemes under a fake message generation attack in an urban scenario.

10% for our proposed scheme. However, it reaches 20% for the TRW-only scheme. The RMCV and the simple voting schemes exhibit much higher false decision probabilities compared to our proposed scheme.

We compare our proposed scheme with other schemes in terms of false decision probability under a fake message attack in an urban scenario in Figure 9. The false decision probability of the proposed scheme increases when the density of the malicious vehicles generating the false message increases, resulting in a false decision probability slightly greater than 10%. In an urban scenario, the high density of vehicles and low speeds help the propagation of false event messages generated by attackers. Thus, the false decision probability in this case is slightly higher than that for the highway scenario. In this scenario, the RMCV and the simple voting schemes exhibit higher false decision probabilities compared to the proposed scheme, with a tendency similar to Figure 8.

We now compare our scheme with the RMCV scheme in terms of message overhead. We considered a situation where there is an actual accident without malicious vehicles. In Figure 10, we present the simulation results of the message overhead with respect to the varying density of vehicles per square kilometer in both urban and highway scenarios. The

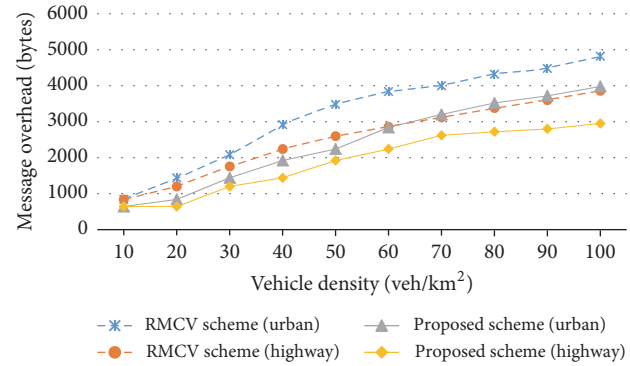


Figure 10: Message overhead for various values of vehicle densities under no malicious vehicle.

message overhead is the cost incurred due to the extra message that is exchanged with neighboring vehicles. In terms of message overhead, as the vehicle density increases, the message overhead also increases in both scenarios, as shown in Figure 10. In the beginning, when the vehicle density is low, our proposed scheme has low message overhead as the scheme does not advertise the default trust level of new neighbor vehicles; however the pseudo IDs in the trust level opinion pair cause some message overhead. We can see that the message overhead is higher for the RMCV as compared to our scheme in both scenarios because in their scheme the vehicle nodes send query messages to the neighboring vehicles and then receive response messages regarding the accident event. On the contrary, there is no query message, but only one-way report messages are sent in our scheme. The message overhead in the urban scenario for both schemes is slightly higher than the highway scenario, because the speed of the vehicles in the urban scenario is less than that of highway scenario, and, thus, each vehicle accumulates more messages, compared to the highway scenario.

We also compare our scheme with RMCV in terms of message overhead in the presence of malicious vehicles under fake message attack. The average vehicle density increases from 1 vehicle per km^2 to 100 vehicles per km^2 throughout the simulation time in urban and highway scenarios. We run several simulations by increasing the ratio of the malicious vehicles from 5% to 50% in both scenarios. Our scheme collects warning messages from neighboring vehicles to detect the trustworthiness of event information contained in the received messages. In both schemes, the message overhead increases as the ratio of the malicious vehicles increases because the vehicles accumulate more messages due to the presence of the malicious vehicles. The message overhead for different ratios of malicious vehicles is shown in Figure 11. Our scheme has a lower message overhead compared to the RMCV scheme in both scenarios.

5. Conclusion and Future Work

In this paper, we proposed a trustworthy event-information dissemination scheme in VANET. We determine and disseminate only the trustworthy event messages to neighbor

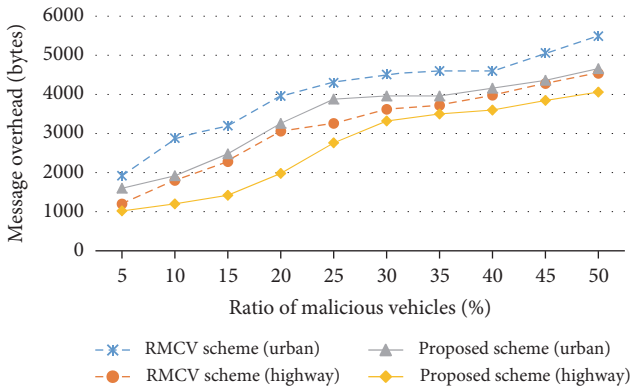


Figure 11: Message overhead for various ratios of malicious vehicles under fake message generation attack.

vehicles. We introduced a modified K-means clustering algorithm to reduce the effect of malicious vehicles on the trust levels (i.e., the truth-telling probabilities) of other vehicles. In other words, the issue of node trustworthiness is resolved through a modified K-means clustering algorithm in our proposed scheme. In the next step, the issue of message trustworthiness is resolved by applying a modified TRW to the report messages received from neighbor vehicles along with the information on node trustworthiness. We compared our proposed scheme with RMCV, simple voting, and TRW-only schemes through simulation. The simulation results show that our proposed scheme has a lower false decision probability compared to other schemes as well as low message overhead compared to the RMCV scheme. The simulation results also show that our proposed scheme can effectively cope with message modification attack and fake message generation attack as long as the number of benign vehicles is larger than the number of malicious vehicles. Our scheme has an additional advantage that the decision on the trustworthiness of a given message is made in an infrastructure-less environment without using PKI.

In this paper, we assumed that the malicious vehicles are uniformly distributed on the roads. However, this assumption may not be valid if colluding malicious vehicles move as a group to increase their influence on the nearby vehicles. Such a complicated issue will be studied in more detail in our future work.

Notations

p_i: Pseudo ID of vehicle V_i
θ_i: Trust level (truth-telling prob.) of vehicle V_i
E_x: Event of type x
M_E: Event message
M_B: Beacon message
M_F: Forwarded message
L_E: Location of event E_x
l_i: Location of vehicle V_i
s_i: Speed of vehicle V_i
θ_{ij}: Estimation of trust level θ_i for vehicle i by vehicle j.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was supported in part by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2013R1A1A2012006 and 2015R1D1A1A01058595), and by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2017-2016-0-00313) supervised by the IITP (Institute for Information & communications Technology Promotion).

References

[1] H. Hartenstein and L. Kenneth, *VANET: Vehicular Applications and Inter-Networking Technologies*, Wiley, New Jersey, NJ, USA, 1st edition, 2009.

[2] P. Papadimitratos, L. Buttyan, T. Holczer et al., "Secure vehicular communication systems: design and architecture," *IEEE Communications Magazine*, vol. 46, no. 11, pp. 100–109, 2008.

[3] K. Govindan and P. Mohapatra, "Trust computations and trust dynamics in mobile adhoc networks: a survey," *IEEE Communications Surveys & Tutorials*, vol. 14, no. 2, pp. 279–298, 2012.

[4] Y. L. Morgan, "Notes on DSRC & WAVE standards suite: its architecture, design, and characteristics," *IEEE Communications Surveys & Tutorials*, vol. 12, no. 4, pp. 504–518, 2010.

[5] M. Raya, P. Papadimitratos, V. D. Gligor, and J.-P. Hubaux, "On data-centric trust establishment in ephemeral ad hoc networks," in *Proceedings of the 27th IEEE Communications Society Conference on Computer Communications (INFOCOM '08)*, pp. 1912–1920, Arizona, Ariz, USA, April 2008.

[6] R. Shrestha, S. Djuraev, and S. Y. Nam, "Sybil attack detection in vehicular network based on received signal strength," in *Proceedings of the 3rd International Conference on Connected Vehicles and Expo, ICCVE '14*, pp. 745-746, November 2014.

[7] J. Doucer, "The Sybil Attack," in *Proceedings of the IPTPS '01 Revised Papers from the First International Workshop on Peer-to-Peer Systems*, 2002.

[8] G. Martuscelli, A. Boukerche, and P. Bellavista, "Discovering traffic congestion along routes of interest using VANETs," in *Proceedings of the 2013 IEEE Global Communications Conference, GLOBECOM '13*, IEEE, Atlanta, GA, USA, December 2013.

[9] Z. Huang, S. Ruj, M. Cavenaghi, and A. Nayak, "Limitations of trust management schemes in VANET and countermeasures," in *Proceedings of the IEEE 22nd International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '11)*, pp. 1228–1232, IEEE, Toronto, Canada, September 2011.

[10] M. Raya, P. Papadimitratos, and J.-P. Hubaux, "Securing vehicular communications," *IEEE Wireless Communications Magazine*, vol. 13, no. 5, pp. 8–15, 2006.

[11] Z. Li and C. T. Chigan, "On joint privacy and reputation assurance for vehicular ad hoc networks," *IEEE Transactions on Mobile Computing*, vol. 13, no. 10, pp. 2334–2344, 2014.

[12] F. Ishmanov, S. W. Kim, and S. Y. Nam, "A secure trust establishment scheme for wireless sensor networks," *Sensors*, vol. 14, no. 1, pp. 1877–1897, 2014.

[13] N. Yang, "A similarity based trust and reputation management framework for vanets," *International Journal of Future Generation Communication and Networking*, vol. 6, no. 2, pp. 25–34, 2013.

[14] K. Rostamzadeh, H. Nicanfar, N. Torabi, S. Gopalakrishnan, and V. C. M. Leung, "A context-aware trust-based information dissemination framework for vehicular networks," *IEEE Internet of Things Journal*, vol. 2, no. 2, pp. 121–132, 2015.

[15] R. A. Shaikh and A. S. Alzahrani, "Intrusion-aware trust model for vehicular ad hoc networks," *Security and Communication Networks*, vol. 7, no. 11, pp. 1652–1669, 2014.

[16] S. A. Soleymani, A. H. Abdullah, W. H. Hassan et al., "Trust management in vehicular ad hoc network: a systematic review," *EURASIP Journal on Wireless Communications and Networking*, vol. 2015, no. 1, article 146, 2015.

[17] M. Raya, A. Aziz, and J. P. Hubaux, "Efficient secure aggregation in VANETs," in *Proceedings of the 3rd ACM International Workshop on Vehicular Ad Hoc Networks (VANET '06)*, pp. 67–75, ACM, New York, NY, USA, September 2006.

[18] B. Qin, Q. Wu, J. Domingo-Ferrer, and L. Zhang, "Preserving security and privacy in large-scale VANETs," in *Proceedings of the 13th International Conference Information and communications security (ICICS '11)*, vol. 7043 of *Lecture Notes in Computer Science*, pp. 121–135, Springer Berlin Heidelberg, Berlin, Germany, 2011.

[19] X. Lin, X. Sun, P.-H. Ho, and X. Shen, "GSIS: a secure and privacy-preserving protocol for vehicular communications," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 6 I, pp. 3442–3456, 2007.

[20] C. Zhang, R. Lu, X. Lin, P.-H. Ho, and X. Shen, "An efficient identity-based batch verification scheme for vehicular sensor networks," in *Proceedings of the 27th IEEE Communications Society Conference on Computer Communications (INFOCOM '08)*, pp. 816–824, IEEE INFOCOM, Arizona, Ariz, USA, April 2008.

[21] A. Wasef and X. Shen, "Efficient group signature scheme supporting batch verification for securing vehicular networks," in *Proceedings of the IEEE International Conference on Communications (ICC '10)*, pp. 1–5, Cape Town, South Africa, May 2010.

[22] H. Zhu, S. Du, Z. Gao, M. Dong, and Z. Cao, "A probabilistic misbehavior detection scheme toward efficient trust establishment in delay-tolerant networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 1, pp. 22–32, 2014.

[23] D. Tian, Y. Wang, H. Liu, and X. Zhang, "A trusted multi-hop broadcasting protocol for vehicular ad hoc networks," in *Proceedings of the 2012 1st International Conference on Connected Vehicles and Expo, ICCVE '12*, pp. 18–22, December 2012.

[24] F. Gómez Mármol and G. Martínez Pérez, "TRIP, a trust and reputation infrastructure-based proposal for vehicular ad hoc networks," *Journal of Network and Computer Applications*, vol. 35, no. 3, pp. 934–941, 2012.

[25] Y.-C. Wei and Y.-M. Chen, "An efficient trust management system for balancing the safety and location privacy in VANETs," in *Proceedings of the 11th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom '12)*, pp. 393–400, Liverpool, UK, June 2012.

[26] U. F. Minhas, J. Zhang, T. Tran, and R. Cohen, "Towards expanded trust management for agents in vehicular ad-hoc networks," *International Journal of Computational Intelligence: Theory and Practice*, vol. 5, no. 1, pp. 3–15, 2010.

[27] U. F. Minhas, J. Zhang, T. Tran, and R. Cohen, "A multifaceted approach to modeling agent trust for effective communication in the application of mobile Ad Hoc vehicular networks," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 41, no. 3, pp. 407–420, 2011.

[28] Z. Liu, J. Ma, Z. Jiang, H. Zhu, and Y. Miao, "LSOT: a lightweight self-organized trust model in VANETs," *Mobile Information Systems*, vol. 2016, no. 1, Article ID 7628231, pp. 1–15, 2016.

[29] Y.-M. Chen and Y.-C. Wei, "A beacon-based trust management system for enhancing user centric location privacy in VANETs," *Journal of Communications and Networks*, vol. 15, no. 2, Article ID 6512239, pp. 153–163, 2013.

[30] S. Gurung, D. Lin, A. Squicciarini, and E. Bertino, "Information-oriented trustworthiness evaluation in vehicular ad-hoc networks," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Preface*, vol. 7873, pp. 94–108, 2013.

[31] Q. Ding, X. Li, M. Jiang, and X. Zhou, "Reputation-based trust model in vehicular ad-hoc networks," in *Proceedings of the International Conference on Wireless Communications And Signal Processing (WCSP)*, 2010.

[32] F. Dötzer, L. Fischer, and P. Magiera, "VARS: a vehicle ad-hoc network reputation system," in *Proceedings of the 6th IEEE International Symposium on a World of Wireless Mobile and Multimedia Networks, WoWMoM '05*, pp. 454–456, June 2005.

[33] C. Chen, J. Zhang, R. Cohen, and P.-H. Ho, "A trust modeling framework for message propagation and evaluation in VANETs," in *Proceedings of the 2nd International Conference on Information Technology Convergence and Services (ITCS '10)*, IEEE, Cebu, Phillippines, August 2010.

[34] A. Patwardhan, A. Joshi, T. Finin, and Y. Yesha, "A data intensive reputation management scheme for vehicular ad hoc networks," in *Proceedings of the 2006 3rd Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services, MobiQuitous*, July 2006.

[35] J. Zhang, "A survey on trust management for VANETs," in *Proceedings of the 25th IEEE International Conference on Advanced Information Networking and Applications (AINA '11)*, pp. 105–112, Biopolis, Singapore, March 2011.

[36] P. Wex, J. Breuer, A. Held, T. Leinmüller, and L. Delgrossi, "Trust issues for vehicular ad hoc networks," in *Proceedings of the IEEE 67th Vehicular Technology Conference-Spring (VTC '08)*, pp. 2800–2804, Singapore, May 2008.

[37] J. Petit, F. Schaub, M. Feiri, and F. Kargl, "Pseudonym schemes in vehicular networks: a survey," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 228–255, 2015.

[38] Z. Malik and A. Bouguettaya, "Reputation bootstrapping for trust establishment among web services," *IEEE Internet Computing*, vol. 13, no. 1, pp. 40–47, 2009.

[39] W. R. J. Joo and D. S. Han, "An enhanced broadcasting scheme for IEEE 802.11p according to lane traffic density," in *Proceedings of the 20th International Conference on Software, Telecommunications and Computer Networks (SoftCOM '12)*, September 2012.

[40] W. Fehr, *Security system design for cooperative vehicleto-vehicle crash avoidance applications using 5.9 GHz Dedicated Short Range Communications (DSRC) wireless communications*, 2012.

[41] H. Xiong, K. Beznosov, Z. Qin, and M. Ripeanu, "Efficient and spontaneous privacy-preserving protocol for secure vehicular communication," in *Proceedings of the 2010 IEEE International Conference on Communications, ICC 2010*, May 2010.

[42] F. Jiménez, J. E. Naranjo, and Ó. Gómez, "Autonomous manoeuvring systems for collision avoidance on single carriageway roads," *Sensors*, vol. 12, no. 12, pp. 16498–16521, 2012.

[43] X. Tang, D. Hong, and W. Chen, "Data Acquisition Based on Stable Matching of Bipartite Graph in Cooperative Vehicle–Infrastructure Systems," *Sensors*, vol. 17, no. 6, p. 1327, 2017.

[44] P. B. Velloso, R. P. Laufer, D. D. O. O. Cunha, O. C. M. B. Duarte, and G. Pujolle, "Trust management in mobile ad hoc networks using a scalable maturity-based model," *IEEE Transactions on Network and Service Management*, vol. 7, no. 3, pp. 172–185, 2010.

[45] F. Bao, I. Chen, M. Chang, and J. Cho, "Hierarchical trust management for wireless sensor networks and its applications to trust-based routing and intrusion detection," *IEEE Transactions on Network and Service Management*, vol. 9, no. 2, pp. 169–183, 2012.

[46] A. Wald, *Sequential Analysis*, John wiley ad Sons, New York, NY, USA, 7th edition, 1965.

[47] J. Jung, V. Paxson, A. W. Berger, and H. Balakrishnan, "Fast portscan detection using sequential hypothesis testing," in *Proceedings of the IEEE Symposium on Security and Privacy*, IEEE, California, Calif, USA, 2004.

[48] G. Peter, S. Zsolt, and A. Szilard, "Highly automated Vehilce Systems," Mechatronics Engineer MSc Curriculum Development, BME MOGI, 2014.

[49] C. Sommer, R. German, and F. Dressler, "Bidirectionally coupled network and road traffic simulation for improved IVC analysis," *IEEE Transactions on Mobile Computing*, vol. 10, no. 1, pp. 3–15, 2011.

[50] A. Varga and R. Hornig, "An overview of the OMNeT++ simulation environment," in *Proceedings of the 1st International ICST Conference on Simulation Tools and Techniques for Communications, Networks and Systems (SIMUTools '08)*, March 2008.

[51] D. Krajzewicz, J. Erdmann, M. Behrisch, and L. Bieker, "Recent Development and Applications of SUMO - Simulation of Urban MObility," in *Proceedings of the Recent Development and Applications of SUMO - Simulation of Urban MObility*, vol. 5, pp. 128–138, December 2012.

[52] W. Ben Jaballah, M. Conti, M. Mosbah, and C. E. Palazzi, "The impact of malicious nodes positioning on vehicular alert messaging system," *Ad Hoc Networks*, vol. 52, pp. 3–16, 2016.

[53] A. Tajeddine, A. Kayssi, and A. Chehab, "A privacy-preserving trust model for VANETs," in *Proceedings of the 10th IEEE International Conference on Computer and Information Technology, CIT-2010, 7th IEEE International Conference on Embedded Software and Systems, ICESS-2010, 10th IEEE Int. Conf. Scalable Computing and Communications, ScalCom '10*, pp. 832–837, July 2010.

[54] J. Petit and Z. Mammeri, "Dynamic consensus for secured vehicular ad hoc networks," in *Proceedings of the 2011 IEEE 7th International Conference on Wireless and Mobile Computing, Networking and Communications, WiMob '11*, pp. 1–8, October 2011.

[55] B. Ostermaier, F. Dötzer, and M. Strassberger, "Enhancing the security of local danger warnings in VANETs - A simulative analysis of voting schemes," in *Proceedings of the 2nd International Conference on Availability, Reliability and Security, ARES '07*, pp. 422–431, April 2007.