SCIENCE CHINA
Information Sciences

December 2020, Vol. 63 220305:1–220305:17
https://doi.org/10.1007/s11432-019-2835-4

• RESEARCH PAPER •
Special Focus on Challenges and New Insights for Network Security in 5G-and-Beyond

An authentication and plausibility model for big data analytic under LOS and NLOS conditions in 5G-VANET

S. A. SOLEYMANI[1], M. H. ANISI[2*], A. Hanan ABDULLAH[1], M. Asri NGADI[1], Sh. GOUDARZI[3*], M. Khurram KHAN[4] & M. Nazri KAMA[5]

[1]*School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia (UTM), Johor* 81310*, Malaysia;*
[2]*School of Computer Science and Electronic Engineering, University of Essex, Colchester* CO4 3SQ*, UK;*
[3]*Centre of Artificial Intelligence, National University of Malaysia (UKM), Selangor* 43600*, Malaysia;*
[4]*Center of Excellence in Information Assurance (CoEIA), King Saud University, Riyadh* 11653*, Saudi Arabia;*
[5]*Advanced Informatics School, Menara Razak, Universiti Teknologi Malaysia (UTM), Kuala Lumpur* 54100*, Malaysia*

Received 1 November 2019/Accepted 16 March 2020/Published online 12 November 2020

Abstract The exchange of correct and reliable data among legitimate nodes is one of the most important challenges in vehicular ad hoc networks (VANETs). Malicious nodes and obstacles, by generating inaccurate information, have a negative impact on the security of 5G-VANET. The big data generated in the vehicular network is also an issue in the security of VANET. To this end, a security model based on authentication and plausibility is proposed to improve the safety of network named 'AFPM'. In the first layer, an authentication mechanism using edge nodes along with 5G is proposed to deal with the illegitimate nodes who enter the network and broadcast wrong information. In the authentication mechanism, because of the growth of the connected vehicles to the edge nodes that lead to generating big data and hence the inappropriateness of the traditional data structures, cuckoo filter, as a space-efficient probabilistic data structure, is used. In the second layer, a plausibility model by performing fuzzy logic is presented to cope with inaccurate information. The plausibility model is based on detection of inconsistent data involved in the event message. The plausibility model not only tackles with inaccurate, incomplete, and inaccuracy data but also deals with misbehaviour nodes under both line-of-sight (LOS) and non-line-of-sight (NLOS) conditions. All obtained results are validated through well-known evaluation measures such as F-measure and communication overhead. The results presented in this paper demonstrate that the proposed security model possesses a better performance in comparison with the existing studies.

Keywords authentication, plausibility, fuzzy logic, cuckoo filter, 5G-VANET, big data

Citation Soleymani S A, Anisi M H, Abdullah A H, et al. An authentication and plausibility model for big data analytic under LOS and NLOS conditions in 5G-VANET. Sci China Inf Sci, 2020, 63(12): 220305, https://doi.org/10.1007/s11432-019-2835-4

1 Introduction

As the key component of smart transportation systems, VANETs, as a sub class of mobile ad hoc network (MANET) [1], are mobile networks including infrastructures and vehicles. VANETs are used for communications either between an infrastructure and a vehicle (V2I) or between vehicle and vehicle (V2V). VANETs are capable of facilitating various beneficial uses including road safety improvement, vehicular mobile data services, self-driving assistance, and traffic management [2, 3].

* Corresponding author (email: m.anisi@essex.ac.uk, shidrokh@ukm.edu.my)

In this network, issues like multi-hop connectivity, nonexistence of centrality, infrastructure-less nature, and lack of clear defence line make this network unstable [4]. In addition to these issues, the data passed on this network are very crucial and sensitive because any attack or damage on them can result in huge disasters to human lives [5]. Hence, security is one of the most imperative concerns in VANET.

In the vehicular network, one of the major challenges of security is the exchange of reliable and correct data among legitimate nodes. Because the exchanged data have a great impact on the safety and comfort of passengers, dealing with malicious and faulty nodes that create inaccurate data is an important issue. Obstacles, such as existing buildings in vehicular environments, also have a negative impact on the accuracy of the data. These objects by restricting direct communication between nodes create wrong information in the network. Plausibility checking, as an element of security model, is a way to deal with inaccurate and unreliable information in highly distributed and dynamic scenarios such as VANET. Authors in [6] stated that plausibility-check ensures the reliability of data. Authentication is also a security requirement for accepting safety messages from the VANETs legitimate users [7]. Authentication can be performed in two levels including message and node authentication. Message authentication ensures the integrity of event message; whereas node authentication certifies the legitimacy of vehicle node. An authentication outline can simple categorize non-legitimate nodes while providing security in the VANET. Motivated by this observation, this paper seeks for developing a proper security modal based on plausibility and authentication. The proposed model should be able to deal with inaccurate data created by malicious, faulty and unauthorized nodes under line-of-sight (LOS) and non-line-of-sight (NLOS) condition.

Additionally, owing to the big data created in the vehicular network, because of the ever-incrementing demand of mobile services, the security model needs to support five dimensions of large data as volume, velocity, variety, veracity and value [8]. For this purpose, usage of edge computing in the network can be effective. This is mainly because the edge nodes have much better processing compared to roadside units (RSUs) in the vehicular network. Moreover, given the shortcoming of IEEE 802.11p-based networks, fifth generation (5G) technology is impressive to improve the abilities of computation and communication in the vehicular network. Therefore, to efficiently support big data and in addition improve network connectivity for providing secure information transmission, in this paper RSUs are substituted with edge nodes to communicate intermediately along with 5G network.

The key concentrations of this study are summarized as follows:

(i) We provide a two-layer security model using edge nodes in 5G-VANET. In the first layer, we develop an authentication mechanism to deal with any illegitimate node entered the network. In the second layer, we propose a plausibility model to cope with inaccurate, incomplete and uncertainty data.

(ii) We analyse the proposed security model under LOS and NLOS circumstances in 5G-VANET. A NLOS condition occurs where the two lines intersect. This condition results in a message drop by a total signal block.

(iii) We evaluate the proposed security model over density, velocity and different percentage of malicious nodes. F-measure and communication overhead are also utilized to assess the performance of the provided model.

The structure of the other parts of the article is as follows. Section 2 discusses the related work on authentication and plausibility model in VANET. Section 3 provides the proposed security model, technically. Section 4 provides the simulation environment and performance evaluation metrics. The observations and results validating the effectiveness of the proposed model are drawn in Section 5. Ultimately, Section 6 concludes the paper.

2 Related work

Recently, many studies have paid more attention to improve passengers safety in VANET. A security model tries to prepare the network to be protected against malicious and unauthorized nodes. Authentication and plausibility, as requirements of security model [9], have vital role to improve safety of the

vehicular network.

To empower vehicular environment, it is significant to authenticate vehicle nodes that transmit through the vehicular network. Authentication is a mechanism to avoid access to the system by illegal nodes who are able to communicate whereas they do not belong to the network. According to [9], a powerful authentication model provides legal proofs using external mechanisms to find out illegitimate nodes. Plausibility check is also a way to recognize correct/incorrect messages by detecting inconsistencies in data [10]. It is similar to intrusion detection systems in traditional networks in which vehicles correlate the received information with the information already known from previous interaction or predefined thresholds such as speed limits. Based on this perspective, in this section, we separately rough out and discuss the existing authentication and plausibility models in the vehicular networks.

2.1 Authentication model

In [8], a security model is proposed to deal with unauthorized nodes using edge computing. The proposed authentication model is based on the quotient filter at both the vehicle and edge node's layer. The quotient filter is a probabilistic data structure that is utilized for query of the dataset. The main objectives of this model are to detect illegal nodes and any attacks initiated in the network.

Authors in [11] categorized the authentication schemes into four groups including huge number of anonymous keys (HAB), group signature based schemes (GSB), road side unit based schemes (RSUB) and tamper-proof device based schemes (TPDB). Comparison with other authentication schemes, they mentioned that RSUB schemes [12, 13] are more efficient. In the RSUB scheme, the computation and verification done by RSU are much more than other schemes and hence it completely depends on the infrastructure. However, the V2V communication is not supported by this scheme. Tangade et al. [14] proposed a node authentication model through RSU. In this study, a V2I pre-authenticated step is developed. Based on this, before starting communication between two vehicle nodes (V2V), the pre-authenticated phase will be performed. However, because of the large number of vehicles connected to the vehicular network, the method used in this model is not suitable.

2.2 Plausibility model

In the vehicular network, plausibility-check compares the received data with the data of the internal sensor or evaluates the messages from various sources regarding an individual occurrence. For example, the location of an adjacent vehicle is proved through received mobility data from cooperative awareness messages (CAMs) sent by the target node and other neighbours, and information from vehicle-local sensors.

In [15], a beacon-oriented trust model is provided to improve privacy in VANET. In this model, trust of data is measured by cross-checking the likelihood of event message and beacon message. For verification of the plausibility of the event-message and maintaining the trustworthiness, it computes the composite direct event trust. It considers two plausibility measurements in addition to the trustworthiness value as maximum transmission delay verifying and maximum transmission distance verifying. In case the bigger distance between the message transmitter and the message receiver than the maximum transmission distance or when the time interruption between the event message time-stamp and the current time-stamp of the receiver is higher than maximum event message postponement, then the event message's trustworthiness will be adjusted to 0. Followed by passing the plausibility verification, direct event trust will be allocated by the composite value of the cosine similarity and Tanimoto similarity with a weight value.

Bismeyer et al. [6] proposed a model to assess the trustworthiness of node using data plausibility check. They mentioned that not only a cryptographic solution is required to increase safety but also the data plausibility check is an important mechanism. They stated that every node in the VANET runs individually a plausibility checker to find the ghost vehicles. It can find abnormal happenings like overlaps, unpredicted position jumps or quickly appearing nodes.

Lo and Tsai [16] presented a new attack in vehicular environment called illusion attack. Based on this attack, using sensors mounted on the vehicle produces wrong information and broadcasts to neighbour nodes. Then, they proposed a plausibility validation network (PVN) to deal with the illusion attack. It contains rule database and plausibility network (PN) module. Each value in the PN module, in an element field of the considered message is cross confirmed by the values of other correlated element fields referring the rules defined previously in rule set.

Incorrect position information can cause problems such as increased fuel consumption, reduced passenger comfort, and in some cases even accidents. Authors in [17] developed Vouch, a secure proof-of-location scheme tailored for VANETs. The scheme leverages the node positioning capability of 5G wireless network roadside units. The key idea of Vouch is to disseminate periodic proofs of location, combined with plausibility checking of movement between proofs.

Based on available knowledge, few models of security have focused on the impact of obstacles and unauthorized vehicles on correctness of data in 5G-VANET. Both static and moving obstacles are an inseparable part of the urban vehicular network. In the NLOS condition, direct communication between two nodes restricts by obstacles. Obviously, these restrictions are able to effect the reliability, integrity, and availability of the event message. Despite the existing security models in the literature, there is lack of a security model that works correctly in both LOS and NLOS cases. Moreover, owing to the big data created in the 5G-VANET, there is lack of proper security model that not only evaluates the correctness of data but also needs to support five dimensions of large data.

3 Authentication and fuzzy plausibility model (AFPM)

In this study, we proposed a two-layer security model using edge nodes along with 5G-VANET to deal with unauthorized vehicle nodes as well as to cope with inaccurate information generated by malicious, faulty nodes and in addition obstacles. In the first layer, we develop a mechanism based on probabilistic data structure to detect illegitimate nodes entered the network. In the second layer we propose a plausibility model based on fuzzy logic to tackle uncertainty and inaccurate information (see Figure 1). In the following, we describe each layer of security model in details.

3.1 Authentication mechanism

According to [2, 18], authentication, as an element of security systems, is a way to ensure integrity and accuracy of event message exchanged among nodes. Unauthorized nodes by creating inaccurate data threat security of network. To this end, we propose a lightweight mechanism to detect unauthorized nodes entered the network. In the proposed mechanism, it is assumed that RSUs are substituted by edge nodes. This is mainly because the edge nodes contain much better processing power than RSUs to reduce latency, increase throughput and enhance security. It is supposed that each edge node has a list of registered vehicle nodes who are within its defined communication range. To assess the authentication of the vehicle node V_i by V_j, a query from the relevant edge node by V_j is enough, which is explained in detail next.

On the other hand, owing to the growth of connected vehicles to the edge nodes, which results in generating a large amount of data in the edge nodes, using the traditional data structures cannot be suitable. This is because of the much memory and high latency of processing queries in traditional data structures. Hence, it is assumed that the probabilistic data structure, as a group of data structure, is used in the proposed authentication model. This kind of data structure is extremely useful for big data because it reduces latency and analytical process [19]. Bloom filter (BF), cuckoo filter (CF), and quotient filter (QF) are three different of space-efficient probabilistic data structures that used to check whether an element is member of massive dataset or not. According to [20], QF has fast and efficient querying of the elements even in secondary memory than BF. Authors in [21] stated that CF has practically better performance than BF and QF. CF is easier to implement than BF and QF. In terms of space efficiency, CF also uses less space than BF in different applications by the false positive rate of less than

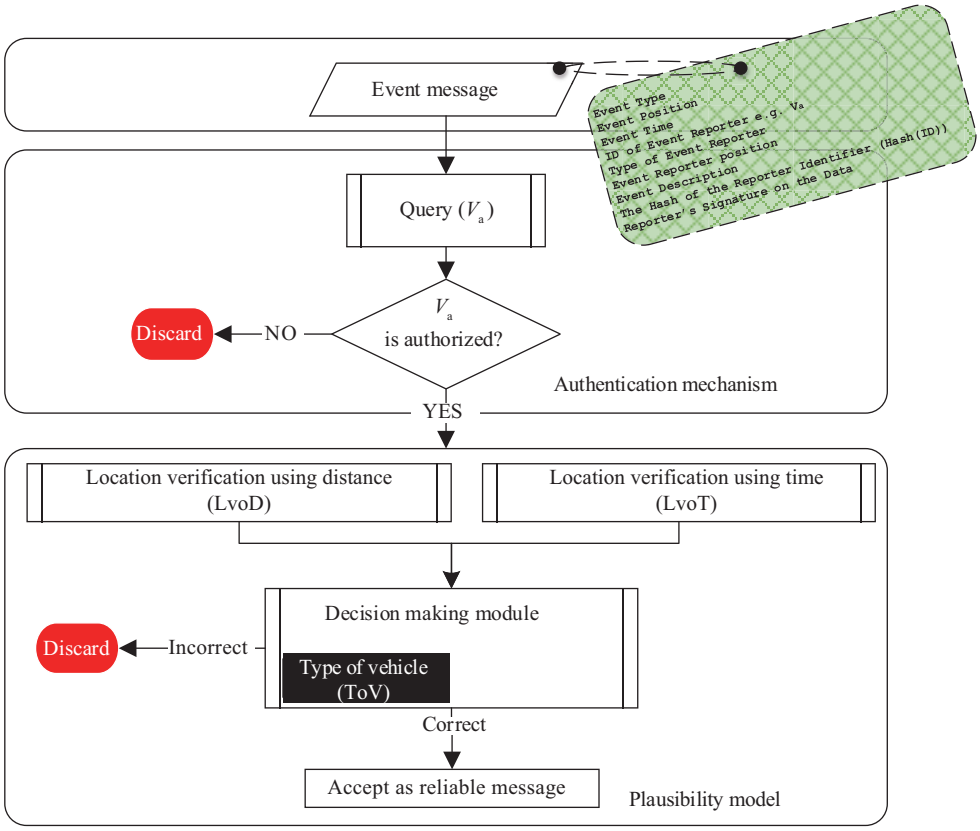

Figure 1 (Color online) Modular framework of AFPM.

3%. It also outperforms QF and BF in lookup performance. Totally, compared to BF and QF, a cuckoo filter has better throughput. Hence, we use CF to query from the edge node's dataset in the proposed authentication model.

3.1.1 *Cuckoo filter*

A cuckoo filter is a compact variant of a cuckoo hash table storing only fingerprints rather than key-value sets. Fingerprint is a bit string resultant from the element utilizing a hash function, for each inserted item. The cuckoo filter uses a hash table based on cuckoo hashing to store the fingerprints of items. Hash table is a structure of data storing the information in an associative mode. Within a hash table, the data are stored in an array format with individual index value for the data value. Cuckoo hashing [22] is an alternative open-addressing solution which ensures constant lookup in the worst case. In cuckoo filter two potential buckets in the table for a given item x are required by cuckoo hashing that are calculated by the following hash functions:

$$f = \text{fingerprint}(x), \tag{1}$$

$$h_1(x) = \text{hash}(x), \tag{2}$$

$$h_2(x) = h_1(x) \oplus \text{hash}(f). \tag{3}$$

Based on partial-key cuckoo hashing, the hash table can achieve both highly-utilization and compact because only fingerprints are stored. Lookup and delete operations of cuckoo filter are straightforward. There is a maximum of two locations to check by $h_1(x)$ and $h_2(x)$. If found, the suitable lookup or delete operation are conducted in $O(1)$ time.

The cuckoo filter is compactly occupied with fingerprints (for example 95% entries filled) conferring high space effectiveness. The hash table is easily searched by a set membership query for item x for the fingerprint of x, and returns true by finding an equal fingerprint. Using a multi-way associative cuckoo

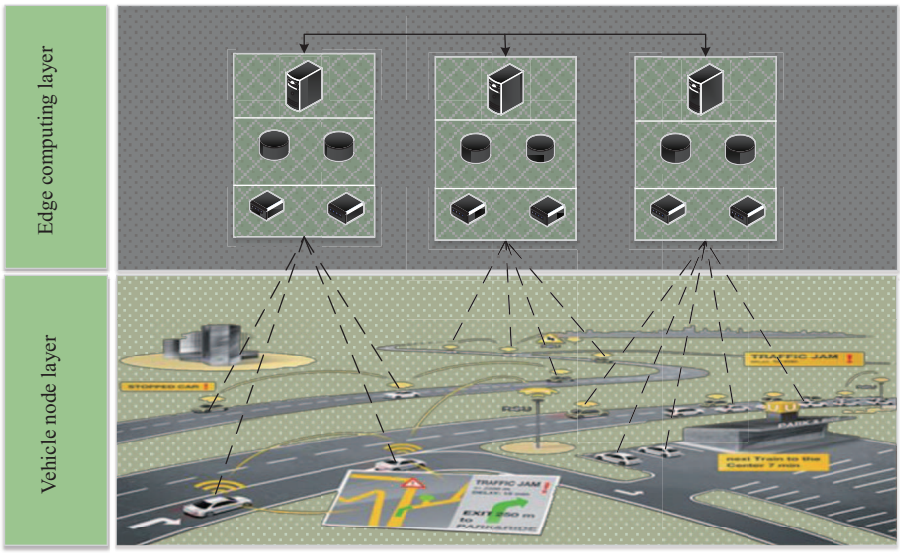

Figure 2 (Color online) Architecture of proposed model.

hash table, the cuckoo filters are greatly made space effective offering high table occupancy and high-speed lookup; for reducing the hash table size further, at first each element is hashed into a fingerprint with constant size prior to inserting into this hash table.

3.1.2 *Vehicle to edge communication*

Consider a set of authorized vehicle nodes that have been registered in the network $\Re = \{V_1, V_2, \ldots, V_N\}$. In addition, there is a set of edge nodes, instead of RSUs, with a defined communication range in the network that is $\Im = \{E_1, E_2, \ldots, E_M\}$. As shown in Figure 2, each edge node ($E_k \in \Im$) has a subset of authorized vehicle nodes that are under its communication range. When a vehicle entered the range of the edge node, the last existing list will be upgraded by the edge node. In CF, in order to add a new vehicle V_{id} to the buckets of the relevant edge node, two hash functions $h1$ and $h2$ and an array B with n buckets where the i-th bucket will be named $B[i]$ are needed. Also, a vehicle V_{id} exists that entered within the edge node's communication range.

Each vehicle such as $V_i \in \Re$ can receive data from another vehicle V_j, as long as it is within the senders transmission range. In order to check authentication of V_j, a query is performed by V_i on the relevant edge node that is under it at that time. Returning the edge node in TRUE, it means the sender is authorized; otherwise, it is highlighted as an intruder within the network while sending an alert representing that an unauthorized vehicle node has entered the network. Figure 3 shows the sequence diagram of V2E communication to check authentication of sender of event message.

3.2 Fuzzy plausibility model

As stated earlier, the main idea of this study is to develop a security model to deal with inaccurate data generated in the vehicular network. Obviously, these data have negative impact on the performance of the network. Plausibility checking is a way to cope with inaccurate and unreliable information. Plausibility-check ensures the reliability of data [6]. It is assumed that each vehicle shares some useful information such as its location and velocity with other vehicles using beacon message in the network, automatically [23]. Building on this, location verification can be a way to assess the senders plausibility to determine the true location of the sender or not [10].

In this subsection, a model is proposed to assess the plausibility of event message by doing fuzzy logic. In the proposed model, fuzzy logic, as an artificial intelligence, is utilized. This is because fuzzy logic performs well in decision making systems and in addition it reduces delay in computation [24]. Nevertheless, by getting an event message from adjacent vehicles, this model evaluates the plausibility of

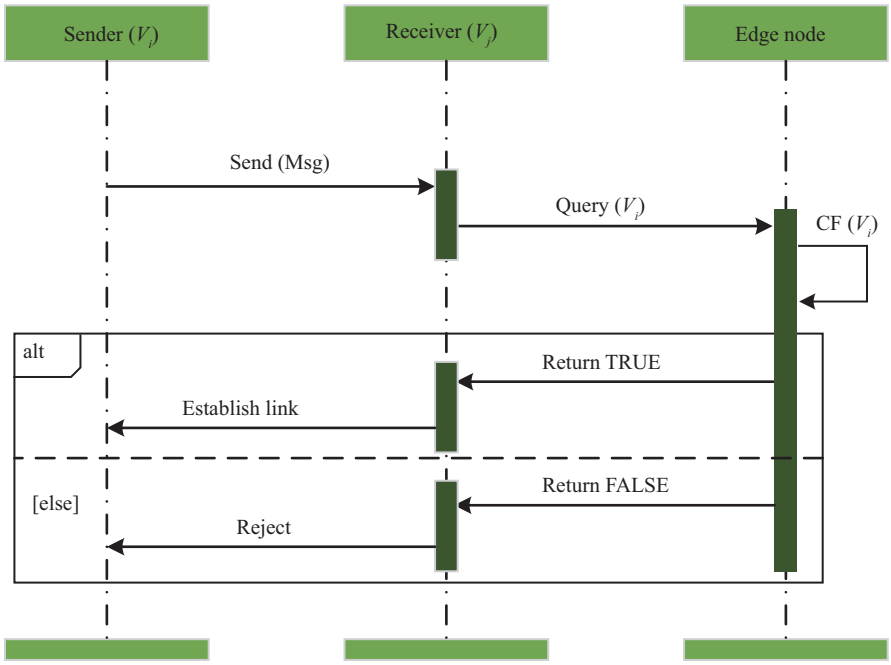

Figure 3 (Color online) A sequence diagram of vehicle to edge (V2E) communication.

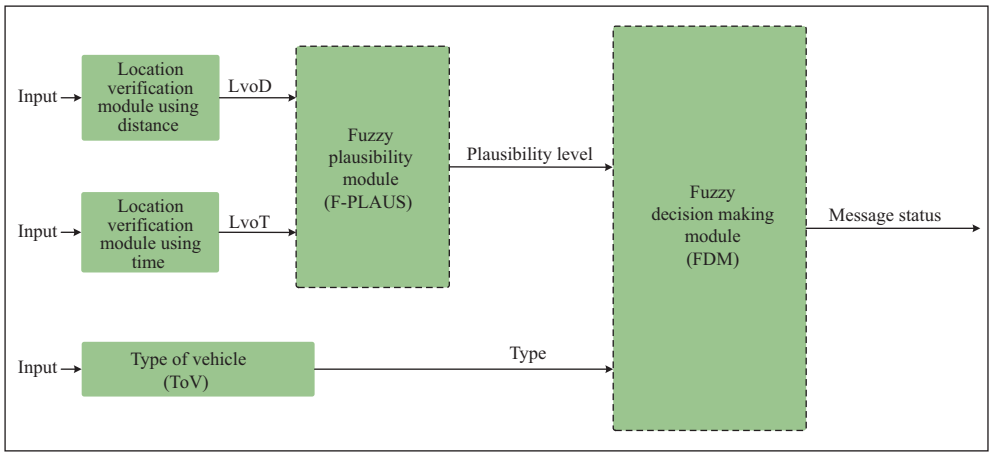

Figure 4 (Color online) Working model of plausibility model.

the event message through location verification. Then, based on type of vehicle, it decides on the received data. The work-flow of the proposed model is shown in Figure 4.

In order to verify location, two modules are presented using distance and time which we are presenting in the following.

3.2.1 *Location verifying via distance (LvoD)*

Verifying the measured distance between receiver and sender is a way to verify a claimed position. Hence, the provided outline firstly computes the distance between two vehicles in a 2D plane using GPS location information involved in the event message. It also computes the distance between two nodes using received signal strength indicator (RSSI). Finally, it evaluates the level of location verification (LvoD) based on the comparison between two measured. it is clear that obstacles have negative impact on security. This is because, obstacles decrease the transmission efficiency between two vehicles. Obstacles can prevent messages from reaching its destination. Vehicles would not receive proper beacons and are not able to directly verify one of its neighbors. Therefore, owing to the existence of obstacles in the vehicular

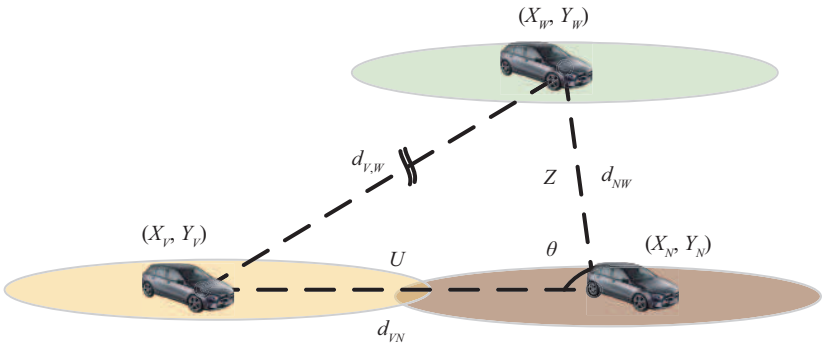

Figure 5 (Color online) Estimating the distance between two nodes using a third common neighbour node.

environment, we propose two different algorithms in LOS and NLOS statuses. On the other hand, online NLOS detection because of the obstacles, is an important issue in vehicular environment; but it is out of scope of this study. Nevertheless, in order to determine a NLOS status, we label out all NLOS conditions.

LvoD under LOS condition. In this condition, it is assumed that there is no obstacle between sender and receiver. In this condition, the proposed scheme measures the distance between V and W using GPS location information included in the event message by (4). It also calculates distance using RSSI computation. Then, it compares the announced and measured distances. Node V verifies node W if both values are a match.

$$\mathrm{Dist}_{\mathrm{gps}}(V,W) = \sqrt{|X_V - X_W|^2 + |Y_V - Y_W|^2}. \tag{4}$$

LvoD under NLOS condition. Consider node V receiving a message from node W. Under NLOS condition, owing to the presence of obstacles, to verify location W, node V sends a request to node N along with its proclaimed position (X_W, Y_W) and mobility vector. Node N is a neighbour of both V and W and under their transmission range. Node N is able to prove W location by determination of its distance via RSSI computing and comparing the proclaimed and measured values. In case both values are a match, N will send a reply back to V comprising of the distance d_{NW} and proving the location of W. By receiving, V verifies d_{VN} (utilizing the radio measurement) and computes the angle θ between vectors $\boldsymbol{U}$ and $\boldsymbol{Z}$ by

$$\theta = \mathrm{ArcCos}\left(\frac{\boldsymbol{U}\cdot\boldsymbol{Z}}{\|\boldsymbol{U}\|\cdot\|\boldsymbol{Z}\|}\right), \tag{5}$$

where $\boldsymbol{U} = \boldsymbol{N}\cdot\boldsymbol{V} = ((X_V - X_N), (Y_V - Y_N))$ and $\boldsymbol{Z} = \boldsymbol{N}\cdot\boldsymbol{W} = ((X_W - X_N), (Y_W - Y_N))$. In addition, $\boldsymbol{U}\cdot\boldsymbol{Z} = U_1Z_1 + U_2Z_2$, $\|\boldsymbol{U}\| = \sqrt{U_1^2 + U_2^2}$, and $\|\boldsymbol{Z}\| = \sqrt{Z_1^2 + Z_2^2}$. Next, node V computes the $\mathrm{Dist}_{\mathrm{rssi}}$ to node W via node N by

$$\mathrm{Dist}_{\mathrm{rssi}}(V,W) = \sqrt{d_{VN}^2 + d_{NW}^2 - 2\,d_{VN}\,d_{NW}\,\cos\theta}, \tag{6}$$

where d_{VN} is the distance between V and W, d_{NW} is the distance between N and W measured utilizing RSSI and θ is the angle between vectors $\boldsymbol{U}$ and $\boldsymbol{Z}$ (see Figure 5).

Based on the new coordinate of sender, node V also measures distance ($\mathrm{Dis}_{\mathrm{gps}}$) to W using

$$\mathrm{Dis}_{\mathrm{gps}}(V,W) = \sqrt{|X_V - X'_W|^2 + |Y_V - Y'_W|^2}, \tag{7}$$

where (X'_W, Y'_W) is the W location caused by mobility. Owing to mobility, the actual position has changed since the information was received. $X'_W = X_W + \triangle x$ and $Y'_W = Y_W + \triangle y$. Finally, it compares the announced and measured distances. Node V verifies node W if both values are a match.

To verify sender, the provided outline calculates difference between two measured distance $\mathrm{Dist}_{\mathrm{gps}}$ and $\mathrm{Dist}_{\mathrm{rssi}}$. This value will be normalized by scaling between 0 and 1 using

$$\Delta^{D}_{\text{gps-rssi}} = \frac{|\mathrm{Dist}_{\mathrm{gps}} - \mathrm{Dist}_{\mathrm{rssi}}|}{\max(\mathrm{Dist}_{\mathrm{gps}}, \mathrm{Dist}_{\mathrm{rssi}})}, \tag{8}$$

where max is a function to extract maximum value between Dist_{gps} and $\text{Dist}_{\text{rssi}}$. Next, it assesses the severity level of $\Delta^{D}_{\text{gps-rssi}}$ by altering this value to fuzzy data.

In order to simplify, three fuzzy sets are considered to represent LvoD including Low, Medium, and High (see Figure 6(a)). It is more trustable if LvoD falls into Low function and it is not trustable if LvoD falls into High function. The LvoD membership functions are selected in terms of experience, trial and error of the application condition, hence, the range initiates at 0 and ends at 1.

3.2.2 *Location verifying based on time (LvoT)*

Time verifying is another method to find a falsely stated position [25]. Assuming correct location information for both the receiver and sender, the anticipated received message time will be computed. Shaikh and Alzahrani [26] mentioned that the value of this time is depending on the propagation speed and the distance between two vehicles. Considering the physical medium of the link, propagation speed is within 2×10^8 (m/s) and 3×10^8 (m/s). In this study, the signal propagation speed will be $c = 3 \times 10^8$.

LvoT under LOS condition. Under LOS, it supposes that a message is sent by node W to V at t_1 and node V receives the message at time_{rec}. It is predicted that node V receives the message at time_{exp} measured via [25–27]

$$\text{time}_{\text{exp}} = t_1 + \frac{\text{Dist}(V_{t_2}, W_{t_1})}{c}, \tag{9}$$

where Dist is the distance between sender and receiver obtained from (4) and $c = 3 \times 10^8$.

LvoT under NLOS condition. Under NLOS circumstance, to verify node W, not only time_{exp} is calculated by node V but also a request is sent by it to its straight neighbours having straight communication with W (for example node N). Then, a request is sent by node N to W waiting for the response. By receipt the response from W, time_{exp} is measured by node N directly via (9) to check the validity of W via comparison of the expected time and received time as previously stated. Then, a reply will be sent by node N back to node V, if the validity of W is confirmed.

To assess the senders validity, the provided outline calculates $\Delta^{\text{T}}_{\text{exp-rec}}$ using

$$\Delta^{\text{T}}_{\text{exp-rec}} = \frac{|\text{time}_{\text{exp}} - \text{time}_{\text{rec}}|}{\max(\text{time}_{\text{exp}}, \text{time}_{\text{rec}})}, \tag{10}$$

where max is a function to find maximum value between time_{exp} and time_{rec}. Then, it investigates the severity level of this value by changing the value to fuzzy information. For this purpose, two fuzzy memberships termed not-acceptable and acceptable are regarded to provide LvoT. Node W is verified by node V, if this value is positioned in the acceptable level and is not approved otherwise. According to Figure 6(b), the LvoT range is between 0 and 1.

Fuzzy inference procedure is the 2nd stage for implementing the fuzzy logic. In this phase, the membership functions are combined with the control rules for deriving the fuzzy yield. The fuzzy inference engine is a set of rules advanced through the professional knowledge.

To assess the definite level of plausibility ($\text{PLAUS}_{\text{level}}$), a knowledge-centered rule is designed to connect the outputs and the inputs. These rules are based on the philosophy behind the vehicular environment and traffic engineering. Based on Table 1, the fuzzy inference engine is composed of six rules. The input parameters for this module are LvoD and LvoT and the output is plausibility level. Figure 6(c) shows the membership functions Low, Medium, and High to present the $\text{PLAUS}_{\text{level}}$.

3.2.3 *Decision-making module*

In this subsection, the fuzzy decision-making module is proposed to assess the event messages status. As shown in Figure 7, this module is based on the outputs of plausibility measurement module and in addition type of vehicle.

The type of vehicle depends on the level of legitimacy of the vehicle node and it will be determined in the registration step. Based on this, the registered vehicles will be classified into three groups: high ($\Re^{\text{H}}$), medium ($\Re^{\text{M}}$), and low ($\Re^{\text{L}}$) level. The high-level vehicle nodes ($\Re^{\text{H}}$) chiefly denote for the police

Table 1 Fuzzy inference engine to determine plausibility level

Rule No.	LvoD	LvoT	$\text{PLAUS}_{\text{level}}$
1	Low	Acceptable	High
2	Low	Not acceptable	Medium
3	Medium	Acceptable	Medium
4	Medium	Not acceptable	Low
5	High	Acceptable	Medium
6	High	Not acceptable	Low

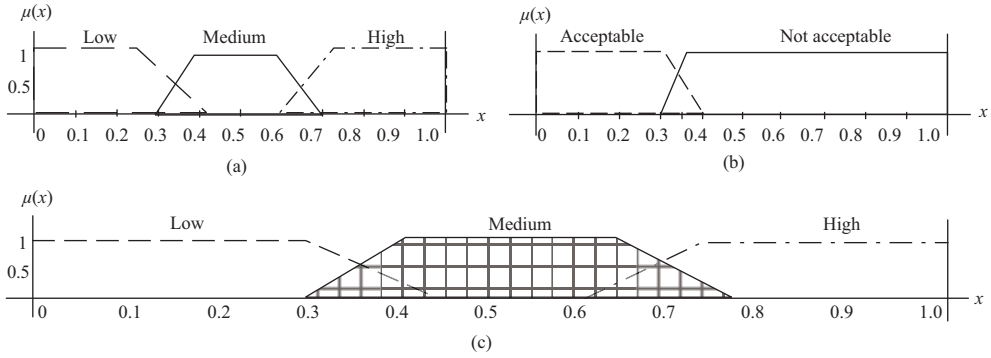

Figure 6 Membership functions. (a) LvoD membership function; (b) LvoT membership function; (c) plausibility level membership function.

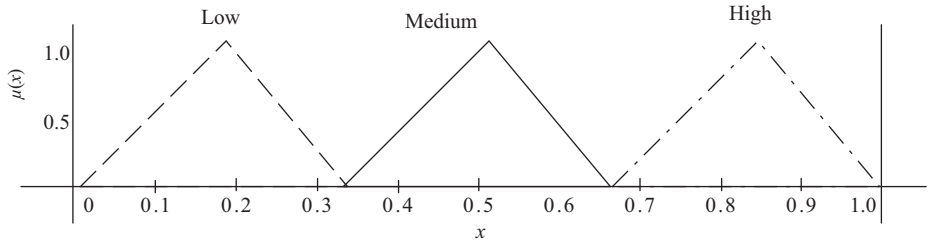

Figure 7 Membership function of vehicles type.

wagon. The police wagons authority level is obviously high (police's car $\in \Re^{\text{H}}$). The medium level nodes ($\Re^{\text{M}}$) denote for public services vehicles, like bus and ambulance. The low-level nodes denote for the private car which are managed principally by people ($\Re^{\text{L}}$). Figure 7 shows the membership functions based on the authentication level. Each edge node has the CF of authorized vehicle nodes along with type of vehicle. In order to find the type of vehicle V_j, a query by V_i needs to performs on the CF. Then, the edge node sends a reply to V_i within a certain time.

Finally, based on the outputs of previous phases, the decision-making module determines the message is acceptable or not. In the suggested model, the predefined input membership functions are used to fuzzify the input parameters gathered by the source vehicle (Figure 6(c), Figure 7). In this module, there are two input parameters as follows:

$$\text{Message}_{\text{status}} = \text{DM}(\text{PLAUS}_{\text{level}}, \text{ToV}), \tag{11}$$

where $\text{PLAUS}_{\text{level}}$ is the level of plausibility and ToV is the type of vehicle.

Then, fuzzified generated values are utilized for evaluating the rules for obtaining the status of message ($\text{Message}_{\text{status}}$). The fuzzy inference system of FDM is designed based on nine (9) of IF-THEN rules. Table 2 represents the fuzzy inference engine of decision-making module. As shown in Figure 8, the membership functions called Correct and Incorrect are utilized to characterize the level of trust ($\text{Message}_{\text{status}}$). Choosing $\text{Message}_{\text{status}}$ membership functions can be achieved in terms of plausibility level, type of vehicle and error and trial of the application necessity.

Table 2 Fuzzy inference system of decision-making module

Rule No.	ToV	$\text{PLAUS}_{\text{level}}$	$\text{Message}_{\text{status}}$
1	High	High	Correct
2	High	Medium	Correct
3	High	Low	Incorrect
4	Medium	High	Correct
5	Medium	Medium	Correct
6	Medium	Low	Incorrect
7	Low	High	Correct
8	Low	Medium	Incorrect
9	Low	Low	Incorrect

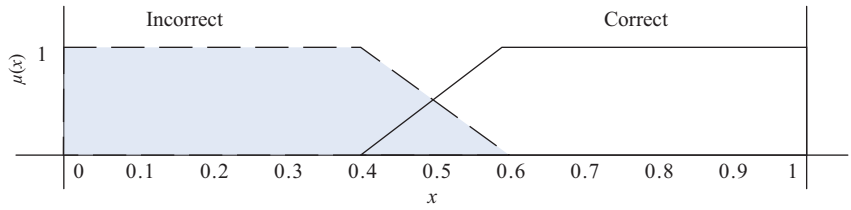

Figure 8 (Color online) Membership function of message's status.

4 Performance evaluation

In this section, simulation environment and performance evaluation metrics are described in detail.

4.1 Simulation environment

The proposed research is simulated using network simulator (NS-2). To create the simulation setting closed to the real traffic situation, a traffic network was made by simulation of urban mobility (SUMO) and mobility model generator for vehicular networks (MOVE). For making the subject more clear, the simulation setting and the associated parameters are explained in the following.

Physical layer. The wireless channel is modelled using two-ray ground reflection model as radio propagation model. Furthermore, the vehicles transmission range is adjusted at 300 m.

Mobility model and vehicular setting. The urban vehicular setup is modelled using MOVE. The vehicles maximum speed is defined at 10 km/h. The simulation area is adjusted at 2 km $\times$ 2 km with the maximum node density of 500 nodes on the simulation area and 10% of nodes are selected as malicious nodes that always provide false or bogus messages.

Media access control (MAC) and network layer. The simulation was performed using the IEEE Standard 802.11p distributed coordination function (DCF). The protocol stacks MAC layer. In our simulation, the used channel bandwidth is 6 Mbps. Using interface queue between MAC and logical link control layer (LLC) with maximum 25 packets, the packets waiting for channel access are stored.

Traffic model. The simulation traffic source is constant bit rate (CBR) with a value of 36 kbps, in terms of UDP packet generation traffic.

Simulation time. The total simulation time is 360 s in each simulation run. The setting time is set to 30 s at the start of simulation to eliminate the impact of transient behaviour on the outcomes. The overall simulation time also contains 30 s of stop to send packets from the simulation termination.

Owing to the impact of movable and immovable problems on the act of the suggested models in vehicular network, AFPM is evaluated on both LOS and NLOS conditions. In terms of NLOS condition, we label out all NLOS situations.

4.2 Evaluation metrics

In this study, state-of-the-art evaluation measures are utilized for assessing the proposed models performance, such as F-measure and communication overhead. These metrics are well-known evaluation measures to validate the obtained results [28].

4.2.1 *F-measure*

F-measure (F) is the weighted harmonic mean of both the recall and precision. It reaches to its best value at one and worst at zero [29]. F-measure is measured using the confusion matrix which is composed of four parameters including false positive (FP), true positive (TP), false negative (FN), and true negative (TN) [30]. Nevertheless, it should be stated that the true negative rate is not considered in F-measure. F-measure is computed using

$$\text{F-measure} : F = \frac{2PR}{P+R} = \frac{2\text{TP}}{2\text{TP}+\text{FP}+\text{FN}}, \tag{12}$$

where $P = \frac{\text{TP}}{\text{TP+FP}}$ and $R = \frac{\text{TP}}{\text{TP+FN}}$ are the precision and recall rate, respectively. In this study, TP is the number of nodes correctly found as malicious nodes, FN is the number of nodes incorrectly detected as non-malicious nodes and FP is the number of nodes incorrectly detected as malicious nodes.

4.2.2 *Communication overhead*

Owing to the huge size of vehicular ad hoc networks that lead to large message dissemination as well as frequent message exchange of nodes, communication overhead (CO) is a suitable metric for comparative evaluation of the proposed trust model. Villalba et al. [31] stated that communication overhead is the total number of packets to be transferred from one node to another. In this study, additional messages exchange in the network, except the event messages, is defined as communication overhead by

$$\text{CO} = \text{Total}_{\text{Msg}} - \text{Event}_{\text{Msg}} - \text{Beacon}_{\text{Msg}}, \tag{13}$$

where $\text{Total}_{\text{Msg}}$ is the all messages transferred between nodes in the vehicular environment, $\text{Event}_{\text{Msg}}$ is the total event messages and $\text{Beacon}_{\text{Msg}}$ is the total beacon messages created by nodes. Because the proposed model is a beacon-less trust model, the value of $\text{Beacon}_{\text{Msg}}$ is 0.

5 Simulation results and discussion

In this section, the results of the simulation show the comparison of F-measure and communication overhead of AFPM with a secure plausibility scheme (Vouch) over density, velocity and different percentage of malicious nodes under both LOS and NLOS states. Vouch is a secure proof-of-location scheme using plausibility checking for VANETs [17].

To obtain some confidence in the simulation results, it is favoured to practice to launch various runs for each simulated setup. In this study, each simulation scenario has 30 runs. At the beginning of each simulation, the initial node placement is reassigned on a random basis, with various random seed, hence, all primary circumstances are essentially dissimilar to each another. It ensures a unique node placement for each run. In addition, each experimental result is the average of the 30 runs for each simulation scenario. In the following, the performance evaluation of proposed model is presented.

5.1 F-measure

As mentioned above, F-measure is the weighted harmonic mean of both the precision (P) and recall (R). It is a measure of a test's accuracy by substituting the values of P and R. In this subsection, the impact of density, velocity and presence of malicious nodes on the F-measure for both AFPM and Vouch under LOS and NLOS condition is presented.

Figure 9(a) shows the impact of node density on AFPM and Vouch. As shown in this figure, the AFPM has a higher F-measure score than the Vouch when the density of nodes varies. Moreover, when the node density is higher, both methods yield a better F-measure. This is true because it is more likely to receive true data from others when there are a higher number of well-behaved nodes.

Because movable/immovable obstacles prevent nodes to exchange true data, this is no surprise that F-measure of AFPM in NLOS condition is lower than LOS. As demonstrated by simulation results, the impact of obstacles on Vouch model is more than the proposed model. It means that the F-measure of AFPM reduces about 4% in NLOS condition while it is about 12% for baseline model (Vouch).

Figure 9(b) illustrates the comparison of the F-measure of the AFPM and Vouch when the nodes move at different velocities. In both LOS and NLOS conditions, this figure shows the values of F-measure have a decreasing trend when the vehicles are moving faster. This is true because when the vehicles are moving faster, it is generally more difficult for the information regarding the unauthorized vehicles to propagate. Figure 9(b) shows the value of F-measure of AFPM is about 88% when the velocity of vehicles is 20 km/h. The score of F-measure gradually decreases to 85%, 83%, 80% and 78% when the motion speed of vehicles is 40, 60, 80 and 100 km/h, respectively. As mentioned above, the F-measure score for Vouch scheme is lower than AFPM. For example, it is 74% when the node's velocity is 100 km/h.

Figure 9(b) also shows the F-measure score of both schemes under NLOS condition. The comparison with LOS condition shows that the F-measure of AFPM decreases about 4% in different speeds. The score for Vouch model decreases by approximately 6%. This is mainly because of the negative impact of obstacles on receive signal power. Hence, the F-measure in case of NLOS is lower than LOS condition.

It is obvious that malicious nodes avoid exchange of proper messages in the network and hence have the negative impact on F-measure. Figure 9(c) depicts the value of F-measure for the AFPM and the Vouch with different percentages of malicious nodes. Obviously, the value of F-measure will be decreased when a high percentage of malicious nodes participate in the network. As shown in this figure, the F-measure of AFPM is more than 73% when malicious nodes participate in the network. This value for Vouch reaches to 71% when 50% of nodes in the networks are malicious node.

Figure 9(c) also demonstrates that obstacles cause the F-measure of AFPM to reach 80%, 78%, 73%, 72% and 69% when malicious nodes participated in the network are 10%, 20%, 30%, 40% and 50%, respectively. While the Vouchs F-measure decreases from 79% till 63% when the number of malicious nodes is increasing. This is mainly because the obstacles and malicious nodes prevent nodes to exchange proper data.

5.2 Communication overhead

According to [32], communication overhead is the total number of packets to be transferred or transmitted from one node to another. In this paper, the messages exchange between sender and third-party node as well as receiver and third-party node is known as communication overhead. Obviously, the communication overhead will be increased under NLOS state. In the following, the communication overhead of AFPM module is evaluated over density, velocity and the different number of malicious nodes. As shown in the following figures, overhead communication is increased when the node density, velocity and number of malicious nodes are ascending.

Figure 10(a) demonstrates the impact of density on communication overhead of AFPM and Vouch under LOS and NLOS obstruction. Obviously, this metric in LOS condition is lower than NLOS. This is true because the obstacles cause the vehicles use third-party nodes to evaluation. On the other hand, the high density of nodes in the network increases the communication overhead. Figure 10(a) shows that the AFPM scheme is more cost-effective than Vouch in terms of the communication overhead. For instance, when there are 100 nodes in the network, AFPM introduces around 1% of communication overhead whereas it is about 6% for Vouch approach under LOS state. On the other hand, AFPM introduces about 11% of communication overhead when there are 500 nodes, whereas Vouch approach introduces almost 12%. As shown in this figure, AFPM is also more cost-effective than Vouch in case of NLOS. This figure shows that AFPM introduces about 13% of communication overhead when the number of nodes

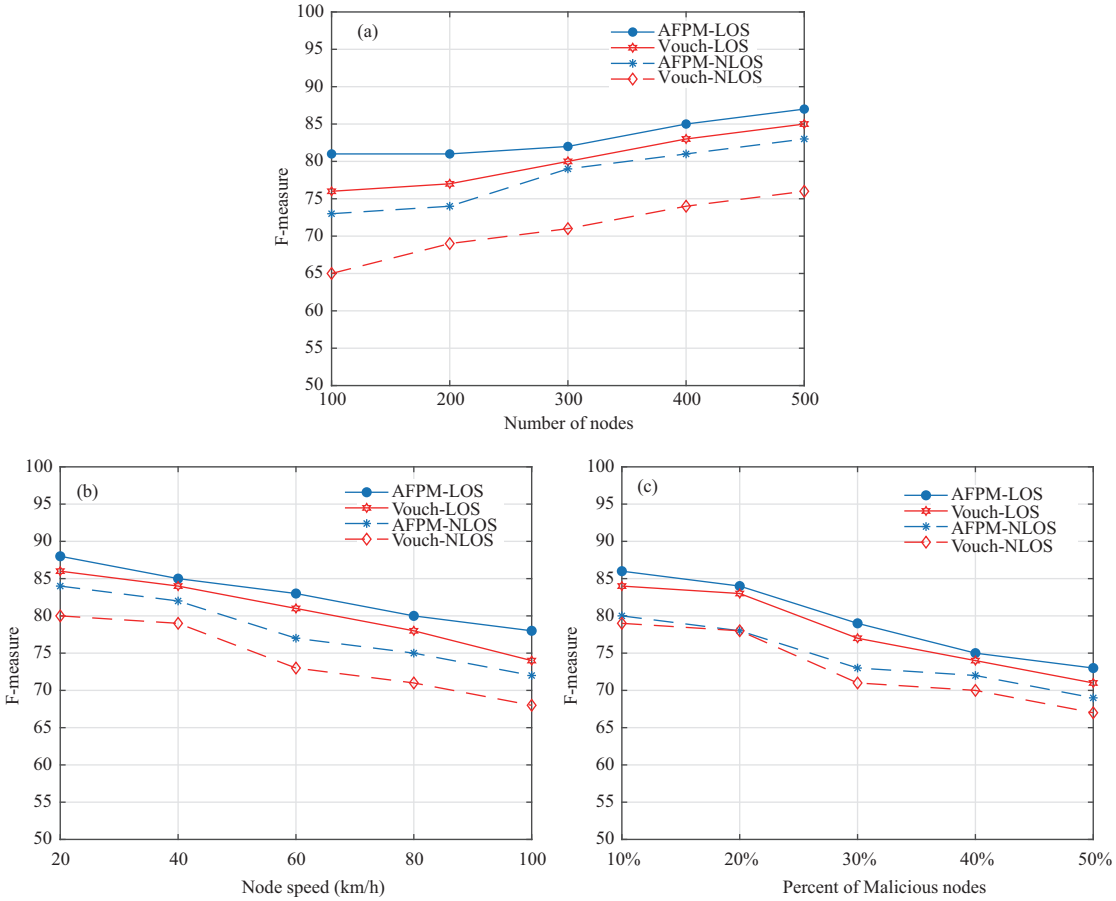

Figure 9 (Color online) F-measure of under different (a) density, (b) velocity, and (c) percent of malicious nodes on LOS and NLOS.

in the network is 500, whereas Vouch approach introduces almost 14%.

Figure 10(b) shows the impact of velocity on communication overhead. As shown in this figure, velocity also increases the communication overhead. This is because the vehicles exit each other transmission range owing to the high speed, hence, nodes need to send request to third-party nodes. This is more in case of LOS obstruction. Figure 10(b) shows that the AFPM introduces communication overhead lower than Vouch method. This metric for AFPM is almost 1% when the velocity is 20 km/h, whereas, overhead communication of Vouch is about 4.5%. When the velocity of nodes increases, the difference of communication overhead between AFPM and Vouch is more. As shown in this figure, the communication overhead of AFPM in case of NLOS is also lower than Vouch under LOS condition. Therefore, AFPM is more cost-effective than Vouch scheme in different velocity.

Figure 10(c) shows the performance of AFPM when there are different percentages of malicious nodes in the network. As shown in this figure, malicious nodes are more effective than velocity and density on increasing communication overhead. Under LOS condition, Figure 10(c) displays the overhead communication of AFPM is around 1% when 10% of existing nodes in the network are malicious nodes but it is 7% for Vouch model. In case of NLOS, it is about 10% and 14.5% for AFPM and Vouch scheme respectively, when the existing malicious nodes in the network is 50%. As shown in this figure, the communication overhead of Vouch scheme is more than AFPM module in both LOS and NLOS state. This is mainly because Vouch is beacon-based and infrastructure-based scheme.

As mentioned earlier, the aim of AFPM is to deal with inaccurate data that generated by malicious and or faulty nodes. The results obtained from the evaluation of both AFPM and Vouch over density, velocity, and different percentages of malicious nodes under LOS and NLOS conditions represented AFPM are more precise and accurate than Vouch to detect inconsistencies in different conditions. The accuracy

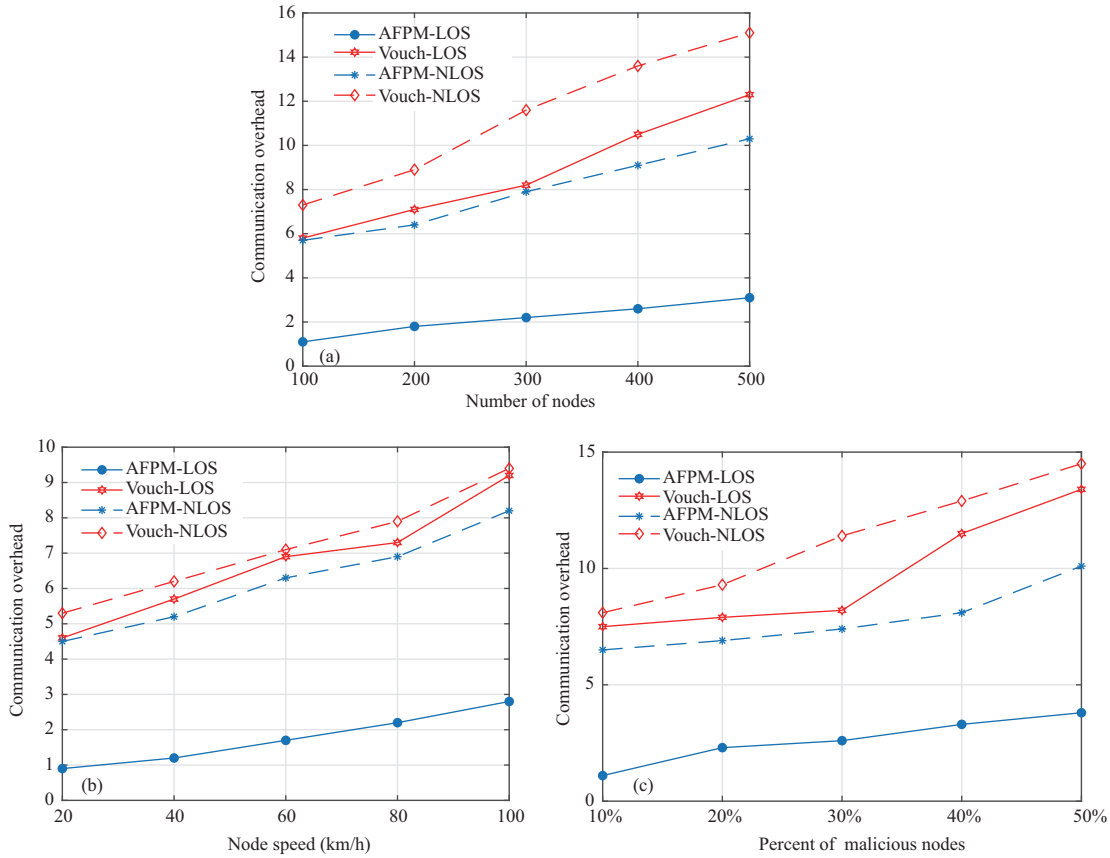

Figure 10 (Color online) Communication overhead under different (a) density, (b) velocity, and (c) percent of malicious nodes on LOS and NLOS.

of AFPM does not decrease or worsen as the size of the VANET increases. In other words, the AFPM achieves a higher accuracy score than the Vouch when the node density varies. Both methods yield a better F-measure, when the node density is higher. This is true because it is more likely to receive true data from others when there are a higher number of well-behaved nodes.

In addition, the results represents that the AFPM always outperforms the Vouch when the vehicles are moving faster. In the worst case of velocity (i.e., velocity of nodes is 100 km/h), AFPM and Vouch can be succeed about 78%, and 74% of the times, respectively. In the worst realistic scenario, when 50% of nodes behave improperly, AFPM is also able to be succeed about 73% of the times. It is 2% more than Vouch. In terms of communication overhead, it is also more cost-effective than Vouch. This is because Vouch scheme is based on beacon and infrastructure and hence creates more messages to proof of location of event reporter.

6 Conclusion

In this paper, a security model, namely AFPM, is proposed to assess the accuracy and integrity of event message under LOS and NLOS situation in 5G-VANET. The proposed security model is based on node's authentication and data plausibility level. Authentication model is developed to deal with illegitimate nodes entered the network. This model is utilized in the edge node layer. To this end, the cuckoo filter is used because the big data is generated in the VANET. In addition, a plausibility model is proposed to detect accurate data using location verification. The proposed plausibility model is used in vehicle node layer. It is composed of three modules including LvoD, LvoT and decision-making module. Upon receiving an event message from surrounding vehicles, the authentication model firstly checks verification of sender. If the sender is authorized, the integrity and accuracy of data are evaluated by cross-checking

the plausibility of event message. The obtained results show that AFPM is highly resilient to malicious nodes. Comparison of AFPM scheme and Vouch shows that AFPM has better performance than Vouch in case of LOS and NLOS. Additionally, AFPM is more precise and accurate than Vouch on different node density, different velocity and different percentage of malicious nodes.

Acknowledgements This work was supported by Ministry of Education, Malaysia, in collaboration with the Research Management Center, Universiti Teknologi Malaysia (Grant No. Q.J130000.2451.04G80), Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia (Grant No. GGPM-2020-029), and partially supported by King Saud University (Grant No. RSP-2019/12), Riyadh, Saudi Arabia.

References

1 Anjum S S, Noor R M, Anisi M H. Review on MANET based communication for search and rescue operations. Wirel Pers Commun, 2017, 94: 31–52
2 Soleymani S A, Abdullah A H, Hassan W H, et al. Trust management in vehicular ad hoc network: a systematic review. EURASIP J Wirel Commun Netw, 2015, 2015: 146
3 Al-Sultan S, Al-Doori M M, Al-Bayatti A H, et al. A comprehensive survey on vehicular ad hoc network. J Netw Comput Appl, 2014, 37: 380–392
4 Hua L C, Anisi M H, Yee P L, et al. Social networking-based cooperation mechanisms in vehicular ad-hoc network-a survey. Vehicular Commun, 2017, 10: 57–73
5 Sedjelmaci H, Senouci S M, Abu-Rgheff M A. An efficient and lightweight intrusion detection mechanism for service-oriented vehicular networks. IEEE Int Things J, 2014, 6: 570–577
6 Bismeyer N, Mauthofer S, Bayarou K M, et al. Assessment of node trustworthiness in vanets using data plausibility checks with particle filters. In: Proceedings of 2012 IEEE Vehicular Networking Conference (VNC), 2012. 78–85
7 Manvi S S, Tangade S. A survey on authentication schemes in VANETs for secured communication. Vehicular Commun, 2017, 9: 19–30
8 Garg S, Singh A, Kaur K, et al. Edge computing-based security framework for big data analytics in VANETs. IEEE Netw, 2019, 33: 72–81
9 Engoulou R G, Bellache M, Pierre S, et al. VANET security surveys. Comput Commun, 2014, 44: 1–13
10 Soleymani S A, Abdullah A H, Zareei M, et al. A secure trust model based on fuzzy logic in vehicular ad hoc networks with fog computing. IEEE Access, 2017, 5: 15619–15629
11 Pournaghi S M, Zahednejad B, Bayat M, et al. NECPPA: a novel and efficient conditional privacy-preserving authentication scheme for VANET. Comput Netw, 2018, 134: 78–92
12 Lu R X, Lin X D. ECPP: efficient conditional privacy preservation protocol. In: Proceedings of the 27th Conference on Computer Communications, 2015. 51–70
13 Huang D, Misra S, Verma M, et al. PACP: an efficient pseudonymous authentication-based conditional privacy protocol for VANETs. IEEE Trans Intell Transp Syst, 2011, 12: 736–746
14 Tangade S, Manvi S S, Lorenz P. Decentralized and scalable privacy-preserving authentication scheme in VANETs. IEEE Trans Vehicular Tech, 2018, 67: 8647–8655
15 Chen Y M, Wei Y C. A beacon-based trust management system for enhancing user centric location privacy in VANETs. J Commun Netw, 2013, 15: 153–163
16 Lo N-W, Tsai H-C. Illusion attack on vanet applications-a message plausibility problem. In: Proceedings of 2007 IEEE Globecom Workshops, 2007. 1–8
17 Boeira F, Asplund M, Barcellos M P. Vouch: a secure proof-of-location scheme for vanets. In: Proceedings of the 21st ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems, 2018. 241–248
18 Goudarzi S, Abdullah A H, Mandala S, et al. A systematic review of security in vehicular ad hoc network. In: Proceedings of the 2nd Symposium on Work Sheet Control Number, 2013. 1–10
19 Singh A, Garg S, Kaur R, et al. Probabilistic data structures for big data analytics: a comprehensive review. Knowledge-Based Syst, 2019, 188: 104987
20 Bender M A, Farach-Colton M, Johnson R, et al. Don't thrash: how to cache your hash on ash. Proc VLDB Endow, 2012, 5: 1627–1637
21 Fan B, Andersen D G, Kaminsky M, et al. Cuckoo filter: practically better than bloom. In: Proceedings of the 10th ACM International on Conference on Emerging Networking Experiments and Technologies, 2014. 75–88
22 Pagh R, Rodler F F. Cuckoo hashing. J Algorithms, 2004, 51: 122–144
23 Soleymani S A, Abdullah A H, Anisi M H, et al. BRAIN-F: beacon rate adaption based on fuzzy logic in vehicular ad hoc network. Int J Fuzzy Syst, 2017, 19: 301–315
24 Limouchi E, Mahgoub I. BEFLAB: bandwidth efficient fuzzy logic-assisted broadcast for VANET. In: Proceedings of IEEE Symposium on Computational Intelligence, 2016. 1–8

25 Khan S, Mauri J L. Security for Multihop Wireless Networks. Boca Raton: CRC Press, 2014
26 Shaikh R A, Alzahrani A S. Intrusion-aware trust model for vehicular ad hoc networks. Secur Commun Netw, 2014, 7: 1652–1669
27 Huang Z. On reputation and data-centric misbehavior detection mechanisms for VANET. Dissertation for Ph.D. Degree. Ottawa: University of Ottawa, 2011
28 Abumansoor O, Boukerche A. A secure cooperative approach for nonline-of-sight location verification in VANET. IEEE Trans Vehicular Tech, 2011, 61: 275–285
29 Shah S, Shah B, Amin A, et al. Compromised user credentials detection in a digital enterprise using behavioral analytics. Future Gener Comput Syst, 2019, 93: 407–417
30 Davis J, Goadrich M. The relationship between precision-recall and ROC curves. In: Proceedings of the 23rd International Conference on Machine Learning, 2006. 233–240
31 Villalba L J G, Orozco A L S, Cabrera A T, et al. Routing protocols in wireless sensor networks. Sensors, 2009, 9: 8399–8421
32 Kumar N, Singh Y. Routing protocols in wireless sensor networks. In: Handbook of Research on Advanced Wireless Sensor Network Applications, Protocols, and Architectures. Hershey: IGI Global, 2017. 86–128