RSSI Sequence and Vehicle Driving Matrix Based Sybil Nodes Detection in VANET

Wei Li, Dongmei Zhang
School of Computer Science
Beijing University of Posts and Telecommunications
Beijing, China
e-mail: liw.0204@foxmail.com, zhangdm@bupt.edu.cn

Abstract—**In VANET, Sybil nodes generated by attackers cause serious damages to network protocols, resource allocation mechanisms, and reputation models. Other types of attacks can also be launched on the basis of Sybil attack, which bring more threats to VANET. To solve this problem, this paper proposes a Sybil nodes detection method based on RSSI sequence and vehicle driving matrix - RSDM. RSDM evaluates the difference between the RSSI sequence and the driving matrix by dynamic distance matching to detect Sybil nodes. Moreover, RSDM does not rely on VANET infrastructure, neighbor nodes or specific hardware. The experimental results show that RSDM performs well with a higher detection rate and a lower error rate.**

Keywords-VANET; sybil nodes detection; RSSI sequence; driving matrix

I. INTRODUCTION

In recent years, Intelligent Transportation Systems (ITS) [1] has attracted more and more attention, and Vehicular Ad-hoc Network (VANET) is an important part of ITS. VANET regards vehicles as nodes, including two forms of communication, Vehicle-to-Vehicle (V2V) and Vehicle-to-Intermediate (V2I). VANET provides a lot of applications, such as collision warnings, electronic brake lights, traffic flow control and route navigation lights to enhance road safety and improve travel quality by exchanging self-perceived information between nodes.

Wireless Access in the Vehicular Environment (WAVE) is the most widely used VANET wireless communication protocol stack developed from the Dedicated Short Range Communications (DSRC) standard, including IEEE 802.11p and IEEE 1609.x. WAVE uses the 5.9 GHz band, called the DSRC dedicated band. Other mainstream protocol stacks include the ETSI ITS G5 protocol stack developed in Europe and the ARIB STD-T109 protocol stack developed in Japan. DSRC defines two communication devices, On Board Unit (OBU) and Road Side Unit (RSU). OBU is installed in the vehicle and connects the in-vehicle network with VANET. RSU is fixed to the infrastructure and connects to the core network.

The Sybil attack was originally proposed by Douceur [2] in the peer-to-peer (P2P) network. It is an identity-based attack. The attackers obtain multiple fake identities through illegal ways to participate in network communication to achieve illegal purpose. The attackers are called the malicious nodes, and the fake identities are called the Sybil nodes in a Sybil attack. Because VANET is an open system and the topology of VANET's is changing frequently, the cost of steal and fake identity is low. Sybil nodes break the trust relationship between nodes. However, trust and cooperation are the basic assumptions of many mechanisms and applications in VANET. The attackers can use Sybil nodes to destroy network protocols, resource allocation mechanisms, and reputation models. For example, Sybil nodes can make themselves cluster head by influencing the voting result to control part of the data forwarding in a routing protocol. In addition, the attackers can use Sybil nodes to launch other types of attacks such as Denial of Service (DOS) attacks, black hole attacks, and wormhole attacks [3].

Many methods of Sybil nodes detection were proposed by researchers. These methods can be divided into three categories: resource testing based detection [4], authentication based detection [5] [6], and geographical validity based detection. Resource testing based detection methods assume that each node has the same limited resources, and verify the resources owned by the suspicious node for detection. But it is very easy to get more resources than normal nodes for malicious nodes in VANET. Meanwhile, this type of methods brings heavy loads to the normal nodes and VANET. Authentication based methods defends against Sybil attack by certificating what to correspond to a specific vehicle. This type of methods relies on the process of updating or replacing certificates, which leads to a poor scalability. And these methods are unable to detect collusion attacks.

Given that Sybil nodes cannot be geographically mapped to physical nodes, many researchers proposed geographical validity based detection methods. Park [7] proposed a detection method based on RSU timestamp sequence. Normal nodes should have different RSU sequences since very few nodes have the similar route with each other in a period of time. But malicious nodes can create a valid sequence by selecting a subset of the timestamps. Therefore, this method has a high miss rate. Grover [8] proposed a detection method based on the similarity of neighbor node sets. This method is RSU-free but it requires a lot number of trusted neighbor nodes, which is contradictory to the detection of Sybil nodes. Garip [9] and Shrestha [10] proposed methods by comparing the position claimed in the beacon and the estimated position of the sender, which is calculated by predefined radio propagation model and Received Signal Strength Indicator (RSSI).

978-1-7281-2184-0/19/$31.00 ©2019 IEEE

Yao [11] used the RSSI time sequences as the voiceprint and compared the similarities among all received sequences to detect Sybil nodes. But malicious nodes can adjust the RSSI transmit power and transmit interval, and even use the wireless propagation model to make themselves look more valid. Therefore, the miss rate of these detection methods is also high. Jin [12] and Xin [13] proposed a detection method based on safe distance. Firstly, the location of nodes was obtained. Then the safe distance of nodes was calculated according to the speed, environment and density of nodes. When nodes appeared in the safe distance of other nodes, they were judged as Sybil nodes. However, normal nodes may also appear in the safe distance of other nodes, so this detection method has a high error rate when the distance between vehicles is closer.

In this paper, we propose a Sybil nodes detection method based on RSSI sequence and vehicle driving matrix - RSDM. Different from most methods based on geographical validity, RSDM generates RSSI sequence and vehicle driving matrix by receiving beacon, and then it evaluates the difference between the two sequences. If the RSSI sequence of one node is quite different from its driving matrix, the node is a Sybil node. Because in order not to be found, an attacker needs to keep some distance from the Sybil node. Moreover, the main contributions of this paper are as follows:

1. RSDM does not rely on centralized deployment of infrastructure, trust relationships between neighbors, or professional hardware support.
2. Dynamic distance matching is used to evaluate the difference between RSSI sequence and vehicle driving matrix to avoid the time deviation.

The rest of the paper is organized as follows: Section II illustrates the attack model. Section III reveals the method we propose in detail. The simulation result analysis is represented in Section IV. And the conclusion of our work is represented in Section V.

II. ATTACK MODEL

We assume that a normal vehicle only has one valid identity, and a malicious vehicle can virtualize one or more Sybil identities. In this paper, vehicles can also be regarded as nodes. Sybil nodes can be generated by faking identity and colluding with other malicious nodes. A Sybil node has a latent period and represents no malicious behaviors during the latent period. We consider that the attack is successful when a Sybil node communicates with a normal node while the normal node does not recognize the Sybil node.

We define the normal vehicle or node as V, the malicious node as M, and the Sybil node as S. Assuming that the influence of environmental factors such as road conditions and weather conditions are combined as ξ, then $d_{safe}(\xi)$ is the safe distance under ξ. The mathematical expectation of the distance between two normal nodes should meet the rule of $E(d) \geqslant d_{safe}(\xi)$. There are two types of possible geographical relationships between two Sybil nodes, or between Sybil nodes and malicious nodes: $E(d) < d_{safe}(\xi)$ and $E(d) \geqslant d_{safe}(\xi)$. In the first case, $d_{safe}(\xi)$ can be calculated according to different environmental factors to detect Sybil nodes. The second case is considered in this paper.

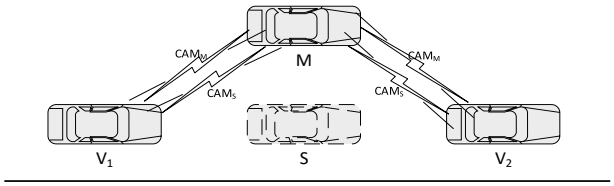

Figure 1. Sybil attack scenario in VANET.

The process of malicious nodes delivering Sybil nodes in VANET is shown in Figure1. Cooperative Awareness Message (CAM) is broadcast periodically between nodes. A CAM message includes these fields: node ID, message ID, timestamp, position, speed, acceleration, direction, and so on. Normal nodes V_1 and V_2 receive CAM messages cam_M and cam_S. cam_M represents that the motion vector of M is $\mathbf{u}_M(p_M, v_M, a_M)$. cam_S represents that the motion vector of S is $\mathbf{u}_S(p_S, v_S, a_S)$, where p, v, and a represent position, velocity, and acceleration respectively. RSSI of the two CAM messages are $rssi_M$ and $rssi_S$. M can adjust the transmission power, so that $rssi_S$ is matched better with $\mathbf{u}_S$. If there is no Sybil nodes detection method, V_1 and V_2 will trust S and add it to their neighbor nodes list.

III. SYBIL NODES DETECTION USING RSSI SEQUENCE AND DRIVING MATRIX

Although the malicious node can make the Sybil node look more real by changing $\mathbf{u}_S$ and $rssi_S$, there are some differences between $\mathbf{u}_S$ and $rssi_S$ when the malicious nodes keep moving. This is because RSSI is not linearly related to the distance in wireless propagation model. Therefore, the difference between driving matrix $U = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_t]^T$ and RSSI sequence $RSSI = [rssi_1, rssi_2, \ldots, rssi_t]$ is calculated to detect whether a node is a Sybil node, where $t \in [1, N]$. The difference between U and $RSSI$ is denoted by f. If the following relation is satisfied:

$$f(U, RSSI) > \theta \tag{1}$$

Then the detected node is recognized as a Sybil node.

A. RSSI Preprocessing

RSSI cannot be directly used since it is very unstable. Consequently, we first process $RSSI$ to improve its smoothness and accuracy. Kalman filter is an optimized autoregressive data processing algorithm. This algorithm is optimal, efficient, and even the most useful when solving many problems. We use the kalman filter to adjust the received RSSI sequence, which can be represented as:

$$rssi_k = Arssi_{k-1} + \omega_{k-1} \tag{2}$$

$$y_k = rssi_k + v_k \tag{3}$$

$$\hat{rssi}_k = A\hat{rssi}_{k-1} + k(y_{k-1} - A\hat{rssi}_{k-1}) \tag{4}$$

where $\hat{rssi}_k$ is the predicted value, $rssi_k$ is the theoretical value, and y_k is the observation value.

B. Matching Driving Matrix and RSSI Sequence

There are many ways to measure f, such as Euclidean distance, etc. We have noticed that U and $RSSI$ do not correspond to each other completely in a normal node. There are some time deviations due to the high speed movement of vehicles, the measurement delay of instruments, and the unavoidable measurement errors. Therefore, we introduce the Dynamic Time Warping (DTW), which calculates $f(U, RSSI)$ using the best match sequence between U and $RSSI$. To get the best match sequence, a cost matrix C is established as follows:

$$C_{i,j} = c(\mathbf{u}_i, rssi_j) \tag{5}$$

where $c(\mathbf{u}_i, rssi_j)$ is the cost between $\mathbf{u}_i$ and $rssi_j$. We define the best match sequence as $W = \{w_k \mid k \in [1, K]\}$, where $w_k = C_{i,j}$ represents $\mathbf{u}_i$ is the best match of $rssi_j$. The minimum accumulated cost of matching $\mathbf{u}_i$ and $rssi_j$ is calculated as:

$$d_{DTW}(\mathbf{u}_i, rssi_j) = \min \begin{cases} D(i-1, j) + C_{i,j} \\ D(i, j-1) + C_{i,j} \\ D(i-1, j-1) + C_{i,j} \end{cases} \tag{6}$$

The DTW algorithm starts at $w_1 = C_{1,1}$ and ends at $w_k = C_{N,N}$. Hence, the DTW distance d_{DTW} of U and $RSSI$ can be calculated as:

$$d_{DTW}(U, RSSI) = \sum_{k=1}^{K} w_k \tag{7}$$

Then the difference between U and $RSSI$ can be calculated as:

$$f(U, RSSI) = \frac{d_{DTW}(U, RSSI)}{K} \tag{8}$$

C. Calculation of Cost

Cost is used to represent the degree of behavior similarity between $\mathbf{u}_i$ and $rssi_j$. Since $\mathbf{u}_i$ and $rssi_j$ cannot be compared directly, we first discuss the relationship between $\mathbf{u}_i(p_i, v_i, a_i)$ and $rssi_j$.

The relationship between RSSI and distance can be determined by radio attenuation model. On account that RSSI is easily affected by environmental factors, different radio attenuation models should be used in different scenarios. Many researchers have carried out experiments and statistics on the complex and changeable environment in VANET. According to the results of their experiments, the model often used in VANET is:

$$PL(d) = \begin{cases} PL(d_0) + 10n_1 \log_{10}(\frac{d}{d_0}), d_0 \le d < d_b \\ PL(d_0) + 10n_1 \log_{10}(\frac{d}{d_0}) + \\ \quad a10n_2 \log_{10}(\frac{d_b}{d_0}) \end{cases} , d \ge d_b \tag{9}$$

where $PL(d_0)$ is the received signal strength at distance d. d_0 is the reference distance and d_0 is usually 1 meter. d_b is the critical distance. The value of RSSI is small when d is large because RSSI produce more obvious deviation at long distance. So we only consider the case of $d_0 \le d < d_b$. When $d \ge d_b$, the cost is $c = c_{MAX}$.

Assuming that the distance between the detection node and the detected node is d_i. We denote the RSSI at distance d_i by $rssi_i^d$ according to the attenuation model. There are three main factors influencing the matching degree between $rssi_i^d$ and $rssi_j$: distance, slope and convexity-concavity. Distance ensures that the matched points are within a certain range, which can be calculated by using Euclidean distance. Slope ensures that the changing trend of the matched points is consistent. And slope can be expressed by the first order differential.

$$\frac{d(rssi_i^d)}{dt} = \frac{10n_1 v_i}{d_i \ln 10}, d_0 \le d < d_b \tag{10}$$

$$\frac{d(rssi_j)}{dt} = \frac{rssi_j - rssi_{j-1}}{t_j - t_{j-1}} \tag{11}$$

The convexity-concavity ensures that the changing trend of slope is consistent. The results of calculating the second order differential of t:

$$\frac{d^2(rssi_i^d)}{dt^2} = \frac{10n_1(a_i d_i \ln 10 - v_i^2 \ln 10)}{(d_i \ln 10)^2}, d_0 \le d < d_b \tag{12}$$

$$\frac{d^2(rssi_j^d)}{dt^2} = \frac{\frac{d(rssi_j^d)}{dt} - \frac{d(rssi_{j-1}^d)}{dt}}{t_j - t_{j-1}} \tag{13}$$

The calculation process of $c(\mathbf{u}_i, rssi_j)$ is as follows:

$$c(\mathbf{u}_i, rssi_j) = \begin{bmatrix} \omega_0 & \omega_1 & \omega_2 \end{bmatrix} \begin{bmatrix} \left| rssi_i^d - rssi_j \right| \\ \left| \frac{d(rssi_i^d)}{dt} - \frac{d(rssi_j)}{dt} \right| \\ \left| \frac{d^2(rssi_i^d)}{dt^2} - \frac{d^2(rssi_j)}{dt^2} \right| \end{bmatrix} \quad (14)$$

Where ω_0, ω_1 and ω_2 represent the weight of distance, slope and convexity-concavity respectively. According to the fluctuation of *RSSI*, the weight should also be changed accordingly. Weights should be determined by experiments under different conditions. It is assumed that U is coherent and reasonable in kinematics. If U is unstable like *RSSI*, it should be preprocessed before use.

D. Sybil Nodes Recognition

Under different speeds and densities, the probability distribution of f is different. For example, the correlation between U and *RSSI* decreases in high speed scenarios. Therefore, the threshold also needs to be changed according to the scenarios. Thresholds can be trained in machine learning models using the results of multiple experiments. In this paper, the probability distribution of the threshold is tested in Section IV. We define the threshold as a function of the speed expectation and density expectation:

$$\theta = g(E_{vel}, E_{den}) \quad (15)$$

Algorithm 1 is the description of RSDM:

Algorithm 1 RSDM

Input: $RSSI, U, \omega_0, \omega_1, \omega_2, E_{vel}, E_{den}$
Output: $isSybil$

for $i = 1 \rightarrow N$ **do**
  $\hat{rssi}_i \leftarrow \text{kalman\_filter}(rssi_i, \hat{rssi}_{i-1})$
end for
$d_{DTW}(1,1) = C_{1,1} = c(\boldsymbol{u}_1, rssi_1)$
for $i = 1 \rightarrow N$ **do**
  for $j = 1 \rightarrow N$ **do**
    if $i > 1$ **then**
      $D1 = D(i-1, j) + c(\boldsymbol{u}_i, rssi_j)$
    end if
    if $j > 1$ **then**
      $D2 = D(i, j-1) + c(\boldsymbol{u}_i, rssi_j)$
    end if
    if $i > 1 \&\& j > 1$ **then**
      $D3 = D(i-1, j-1) + c(\boldsymbol{u}_i, rssi_j)$
    end if
    $d_{DTW}(i,j) = \min(D1, D2, D3)$
  end for
end for
$\theta = g(E_{vel}, E_{den})$
if $D(N,N) > \theta$ **then**
  return $true$
else
  return $false$
end if

IV. Simulation Evalustion

A. Simulation Setup

We used SUMO 0.32, OMNet++ 5.4 and Veins 4.7 for simulation experiments. SUMO is a simulation software for vehicle motion. OMNet++ is a network simulation software. Veins is a VANET communication simulation framework based on OMNet++. By modifying the Veins source code, the received beacons and corresponding RSSI of each vehicle were derived. Our algorithm was implemented in Matlab. We experimented on custom roads and set different velocities and densities. By setting the probability, Sybil nodes are generated randomly when the nodes join the network. Driving matrix in Sybil beacon is inconsistent with that in malicious beacon, and accords with basic kinematics equation. The transmission power of the malicious node was dynamically adjusted according to the position and velocity in the beacon. The parameters used in the experiment and the parameters of the RSSI decay model can be seen in Table I.

TABLE I. PARAMETERS USED IN THE EXPERIMENT

paramters	*value*
Road length	5Km
Number of lanes	4
ω_0	0.6
ω_1	0.2
ω_2	0.2
$PL(d_0)$	-76.1
n_1	-1.66
n_2	-3.18
ω_2	6.12
Vehicle Communication Radius	500m
beacon frequency	1Hz~10Hz

B. Results and Analysis

We first carried out experiments at different velocities and densities, and calculated the probability density of f under different conditions. The experimental results are demonstrated in the Figure 2. The distribution of probability density is divided into two parts: the probability distribution of normal vehicle on the left and the probability distribution of Sybil nodes on the right. These two parts are separated by an interval, which is the value range of θ. With the increase of velocity and density, the value space of θ decreases gradually.

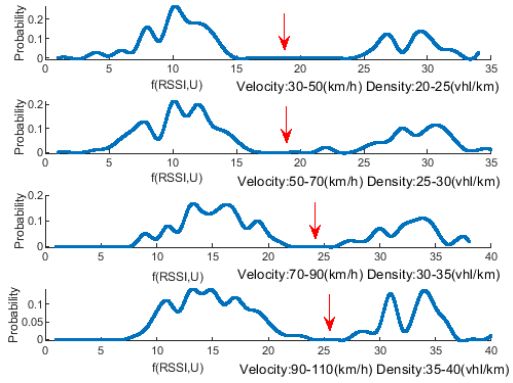

Figure 2. Probability density of f.

We compared the method presented in this paper with that proposed in [14] Vehicle Driving Pattern Based Sybil Attack Detection (DPSA). The experimental results counted two evaluation indicators: true positive rate (TPR) and false positive rate (FPR). TPR represented the proportion of Sybil nodes that were correctly detected by the method. FPR indicated the proportion of normal nodes recognized as Sybil nodes. The experimental results are shown in the Figure 2.

It can be seen in figure 3 that the method in this paper is superior to DPSA in both TPR and FPR indicators. This is because that DPSA only used the driving pattern of node broadcasting to classify, while the message broadcasted by Sybil nodes is not credible. The method proposed in this paper achieved a higher TPR and a lower FPR by combining the driving matrix and the RSSI sequence. However, as the speed increases, the boundary between the normal node and the Sybil node becomes more blurred. Therefore, the test results have dropped slightly.

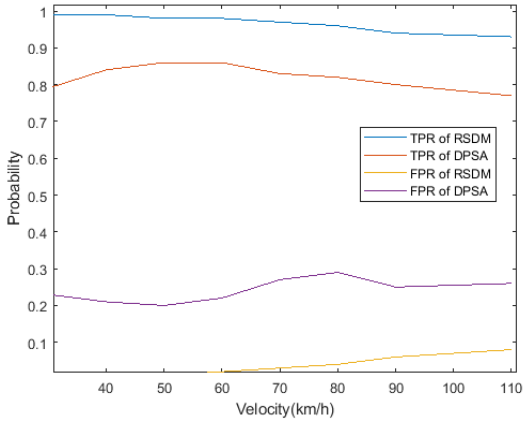

Figure 3. Detection result of RSDM and DPSA.

V. CONCLUSION

This paper proposed a Sybil nodes detection method based on RSSI sequence and driving matrix. This method detects node anomalies by evaluating the difference between the RSSI sequence and the driving matrix of the vehicle. The results of the simulation experiments demonstrate that our detection method can detect more than 90% of the Sybil nodes, and the error rate is below 10%. In future work, we hope to be able to use more features to describe vehicles, and to use training classification algorithm to distinguish between normal nodes and Sybil nodes.

REFERENCES

[1] Alam, Muhammad, Joaquim Ferreira, and José Fonseca. "Introduction to intelligent transportation systems." Intelligent Transportation Systems. Springer, Cham, 2016.

[2] Douceur, John R. "The sybil attack." International workshop on peer-to-peer systems. Springer, Berlin, Heidelberg, 2002.

[3] Alsharif, Nizar, Albert Wasef, and Xuemin Shen. "Mitigating the effects of position-based routing attacks in vehicular ad hoc networks." Communications (ICC), 2011 IEEE International Conference on. IEEE, 2011.

[4] Newsome, James, et al. "The sybil attack in sensor networks: analysis & defenses." Proceedings of the 3rd international symposium on Information processing in sensor networks. ACM, 2004.

[5] Rahbari, Mina, and Mohammad Ali Jabreil Jamali. "Efficient detection of sybil attack based on cryptography in VANET." arXiv preprint arXiv:1112.2257 (2011).

[6] Zhou, Tong, et al. "P2DAP—Sybil attacks detection in vehicular ad hoc networks." IEEE journal on selected areas in communications 29.3 (2011): 582-594.

[7] Park, Soyoung, et al. "Defense against Sybil attack in the initial deployment stage of vehicular ad hoc network based on roadside unit support." Security and Communication Networks 6.4 (2013): 523-538.

[8] Grover, Jyoti, et al. "A sybil attack detection approach using neighboring vehicles in VANET." Proceedings of the 4th international conference on Security of information and networks. ACM, 2011.

[9] Garip, Mevlut Turker, et al. "INTERLOC: An interference-aware RSSI-based localization and Sybil attack detection mechanism for vehicular ad hoc networks." Consumer Communications & Networking Conference (CCNC), 2017 14th IEEE Annual. IEEE, 2017.

[10] Shrestha, Rakesh, Sirojiddin Djuraev, and Seung Yeob Nam. "Sybil attack detection in vehicular network based on received signal strength." Connected Vehicles and Expo (ICCVE), 2014 International Conference on. IEEE, 2014.

[11] Yao, Yuan, et al. "Multi-channel based Sybil Attack Detection in Vehicular Ad Hoc Networks using RSSI." IEEE Transactions on Mobile Computing (2018).

[12] Jin, Dongxu, and JooSeok Song. "A traffic flow theory aided physical measurement-based sybil nodes detection mechanism in vehicular ad-hoc networks." 2014 IEEE/ACIS 13th International Conference on Computer and Information Science (ICIS). IEEE, 2014.

[13] Xin, Yan, and Li Tingting. "Position related lightweight Sybil detection approach in VANET" JoumalonCommunications. 2017.

[14] Gu, Pengwenlong, et al. "Vehicle driving pattern based sybil attack detection." High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS), 2016 IEEE 18th International Conference on. IEEE, 2016.