Detection and Localization of Sybil Nodes in VANETs [*]

Bin Xiao[1] Bo Yu[1,2]
[1]Department of Computing
Hong Kong Polytechnic University, Hong Kong
{csbxiao,csbyu}@comp.polyu.edu.hk

Chuanshan Gao[2]
[2]Department of Computer Science and Engineering
Fudan University, China
{boyu,cgao}@fudan.edu.cn

ABSTRACT

Sybil attacks have been regarded as a serious security threat to ad hoc networks and sensor networks. They may also impair the potential applications of VANETs(Vehicular Ad hoc Networks) by creating an illusion of traffic congestion. In this paper, we present a lightweight security scheme for detecting and localizing Sybil nodes in VANETs, based on statistic analysis of signal strength distribution. Our scheme is a distributed and localized approach, in which each vehicle on a road can perform the detection of potential Sybil vehicles nearby by verifying their claimed positions. We first introduce a basic signal-strength-based position verification scheme. However, the basic scheme proves to be inaccurate and vulnerable to spoof attacks. In order to compensate for the weaknesses of the basic scheme, we propose a technique to prevent Sybil nodes from covering up for each other. In this technique, traffic patterns and support from roadside base stations are used to our advantage. We, then, propose two statistic algorithms to enhance the accuracy of position verification. The algorithms can detect potential Sybil attacks by observing the signal strength distribution of a suspect node over a period of time. The statistic nature of our algorithms significantly reduces the verification error rate. Finally, we conduct simulations to explore the feasibility of our scheme.

Categories and Subject Descriptors

C.2.0 [**COMPUTER-COMMUNICATION NETWORKS**]: General—*Security and protection*

General Terms

Algorithms, Reliability, Security

Keywords

Vehicular Ad hoc Networks, Sybil Attacks, Position Verification, Signal Strength Distribution

1. INTRODUCTION

Until recently, road vehicles were the realm of mechanical engineers. But with the plummeting costs of electronic components and the permanent willingness of the manufacturers to increase road safety and to differentiate themselves from their competitors, vehicles are becoming “computers on wheels”, or rather “computer networks on wheels” [4]. VANETs (Vehicular Ad hoc Networks) have the potential to not only facilitate the decision making tasks of the drivers (e.g., trip planning based on traffic congestion on the road), but also to improve highway safety (by bringing information about catastrophic events and road conditions to the driver’s attention). However, researchers [3][4] have pointed out that VANETs are facing a number of security threats, which might impair the efficiency of VANETs and even life safety. One of these threats is Sybil attacks, in which a malicious vehicle creates an illusion of traffic congestion by claiming multiple identities. Not only does this create an illusion, it has the potential to inject false information into the networks via a number of fabricated non-existing vehicles; it can even launch further DoS attacks by impairing the normal operations of data dissemination protocols such as [15][16][18]. For example, in the application of deceleration warning systems [3], if a vehicle reduces its speed significantly, it will broadcast a warning to the following vehicles. Recipients will relay the message to vehicles further behind. However, this forwarding process can be intervened by a large number of malicious Sybil vehicles. In this way, the malicious adversary can create a massive pileup on the highway, potentially causing great loss of life.

Traditionally in ad hoc networks and sensor networks, three types of defenses against Sybil attacks are introduced, including: radio resource testing, registration, and position verification [11]. Radio resource testing is based on the assumption that a radio cannot send or receive simultaneously on more than one channel. It does not apply to VANETs since a node may cheaply acquire multiple radios. Registration alone cannot prevent Sybil attacks, because a malicious node may get multiple identities by non-technical means such as stealing. Further, strict registration causes serious privacy concerns. In position verification, the network verifies the position of each node and ensure that each physical node is bound with only one identity. A number of position (or distance) verification techniques [8][9][10][13] have been proposed recently. However, they either are designed for indoor applications or rely on fixed base stations or specific hardwares. None of them would be suitable for the highly mobile context of vehicular networks. To our knowledge, there are few works addressing the security threat of Sybil attacks in VANETs.

The motivation behind this paper is that we can estimate a node’s position by analyzing its signal strength distribution and then verify whether its position claim is consistent with the estimated position.

[*]This work is partially supported by HK POLYU A-PA2F and ICRG A-PG52.

DIWANS’06, September 26, 2006, Los Angeles, California, USA.
Copyright 2006 ACM 1-59593-471-5/06/0009 ...$5.00.

In traditional sensor networks, we can not rely on signal-strength-based position verification for two reasons. First, signal-strength-based position estimation can only provide limited accuracy. We can not distinguish two nodes which are close to each other. Second, it's difficult to ensure that the position estimation process is not intervened by potential Sybil nodes. However, the unique properties of VANETs, such as traffic patterns, base station support, and high mobility, present us new opportunities to address the problem from different aspects. In addition, we would prefer a distributed and localized scheme, not relying on roadside base stations (detection should not be performed by base stations), for in the near future base stations can only be sparsely deployed and most sections of roads will be still not covered by base stations.

In this paper, we propose a lightweight security scheme for detecting and localizing Sybil nodes in VANETs. We first investigate the feasibility of using signal strength measurement to verify vehicles' positions. As we expected, the simulation illustrates that given the unstable nature of radio propagation, signal-strength-based position verification can only afford quite limited accuracy. Moreover, this verification technique is vulnerable to fabricated measurements by Sybil nodes. Thus, to adapt signal-strength-based position verification to VANETs, one essential step is to ensure that all signal strength measurements originate from honest physical nodes instead of fabricated Sybil nodes. Then, we present a technique to remove false measurements from potential Sybil nodes. In the technique, we take full advantage of the inherent properties of VANETs, such as high mobility, traffic pattern, and roadside base stations. To compensate for the limited accuracy of signal-strength-based position verification, we then propose a statistic approach, enhanced position verification algorithm, which is based on statistic analysis of signal strength distribution of a potential Sybil node over a period of time. This approach can estimate the physical position of the Sybil node and even obtain its corresponding trajectory. We also present another statistic approach, Sybil node classification algorithm, intended to find other accomplice Sybil nodes originating from the same malicious physical node after a potential Sybil node is detected. Simulation results show that our scheme can achieve a detection rate over 95%, even given a short observation period.

The rest of this paper is organized as follows. In Section 2, we define the attack model and system assumptions. Section 3 presents the basic position verification scheme based on signal strength measurement. Section 4 first introduces a technique to ensure that all measurements originate from physical nodes and then proposes two statistic approaches to detect potential Sybil nodes. Section 5 presents simulation evaluation for our scheme. Section 6 provides attack analysis and introduces several unique features of our scheme. Finally, we introduce the related work in Section 7 and conclude the paper in Section 8.

2. ATTACK MODEL AND ASSUMPTIONS

In this section, we define the attack model of Sybil attacks and then present the system assumptions which would be appropriate for future applications of VANETs.

2.1 Attack Model

The Sybil attack refers to a malicious node illegitimately taking on multiple identities [7]. In wireless networks, mobile nodes usually discover new neighbors by periodically broadcasting beacon packets, in which they claim their identities. However, given the invisible nature of wireless communication, a malicious node can easily claim multiple identities without being detected. Identity authentication doesn't help prevent Sybil attacks in VANETs (Vehicular Ad hoc Networks), since a malicious driver can still get additional identity information by non-technical means such as stealing, or simply borrowing from his friends. The goal of detecting Sybil attacks is to ensure that each physical node is bound with only one legal identity.

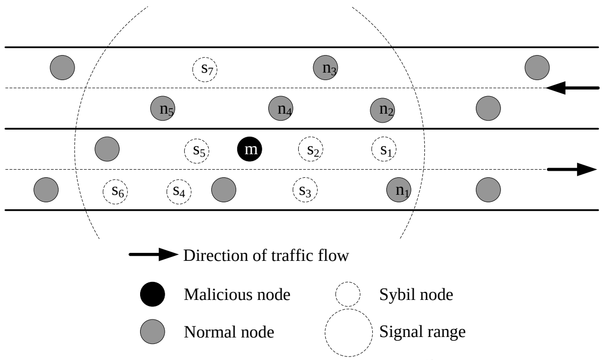


Figure 1: An example VANET under Sybil attacks.

In this paper, we refer to a vehicle as a node in the context of VANETs. We refer to a physical node claiming multiple identities as *a malicious node* and, correspondingly, the malicious node's additional identities as *Sybil nodes*.

Sybil attacks can incur great security threats to VANETs. First, Sybil nodes may cause an illusion of traffic congestion. A greedy driver may convince the neighboring vehicles that there is considerable congestion ahead, so that they will choose alternate routes and allow the greedy driver a clear path to his/her destination [4]. Second, Sybil nodes may directly or indirectly inject false data into the networks, greatly impacting on the data consistency of the system. For example, VANETs may rely on multiple vehicles voting to generate a traffic status report. However, if some of the voters are Sybil vehicles, the report may be deviated from the fact, depending on the benefits of the malicious. Finally, Sybil nodes may launch further DoS attacks such as channel jamming attacks and message suppression attacks [3]. Data dissemination protocols for VANETs such as [15] [16] [18] can be easily cracked by Sybil attacks.

2.2 Assumptions

The following assumptions would be appropriate for a future VANET application. First, we assume that there are a certain amount of vehicles travelling independently on roads and most drivers (vehicles) can be trusted. Only a few greedy drivers (vehicles) may perform Sybil attacks in order to achieve their malicious goals. Second, we assume that all vehicles, including malicious vehicles, are equipped with same radio modules, one for each. The radio module may be based on any RF(Radio Frequency) communication technique providing RSSI(Received Signal Strength Indicator), such as DSRC [1]. Third, we assume that each vehicle is equipped with GPS devices, and GPS positions are supposed to be accurate. Finally, we assume that roadside base stations are sparsely deployed along roads, and the identity authentication infrastructure such as ELP(Electronic License Plate) [5] has been implemented for the whole network. Identity authentication prevents a malicious vehicle from unlimitedly fabricating false identities. Of course, as we afore mentioned, identity authentication alone cannot prevent Sybil attacks. Please also note that the main detection mechanisms of our scheme are not implemented in roadside base stations, but we do require indirect support from base stations.

3. BASIC SIGNAL-STRENGTH-BASED POSITION VERIFICATION

The detection of Sybil attacks usually relies on three categories of approaches, namely, radio resource testing, identity registration, and position verification [7] [11]. Position verification seems to be a promising approach for VANETs, whereas radio resource testing requires special radio modules such as multi-channel radio and identity registration doesn't work very well in VANETs.

In this section, we propose a basic scheme for verifying position claims by signal strength analysis and then explore the feasibility of this scheme through simulations. For simplicity of analysis, we assume that, in this section, all nodes in a network are static.

3.1 Scheme

Our position verification scheme relies on monitoring the signal strength of periodical beacons. For clarity of description, we define three categories of nodes' roles: *claimer*, *witness*, and *verifier*. Each node would periodically play all these roles, that is, each node is a claimer, a witness as well as a verifier but at various moments and for various purposes.

1. Claimer. Each node periodically broadcasts a beacon message at *beacon intervals*, t_b, for the purpose of neighbor discovery. In the beacon message, it claims its identity and position such as GPS position. At this moment, we name the node as a claimer. The goal of our scheme is to verify its claimed position.

2. Witness. All neighboring nodes, within the signal range of the claimer, would receive the previous beacon message. They measure the signal strength and save the corresponding neighbor information in their memory. Next time they broadcast a beacon message, they will attach their neighbor list, including the signal strength measurements for each received beacon, to the beacon message. We name these nodes performing measurement and reporting measurements as witnesses.

3. Verifier. After receiving a beacon message, a node waits for a *verifying interval*, t_v, during which it collects enough signal strength measurements concerning the previous beacon message from neighboring witnesses. t_v may be a little longer than the beacon interval t_b, since after another interval of t_b, each neighboring witness should have broadcasted a beacon containing the expected measurements. With the collected measurements, the node can locally compute an estimated position of the claimer, for example, by performing MMSE(Minimum Mean-Square Error) on the collected signal strength and a pre-defined radio model. We call a node performing verification a verifier.

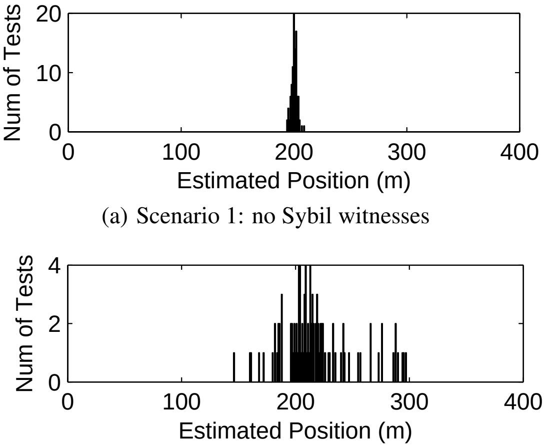
To obtain the estimated position, we first calculate the mean square error:

$$MSE(p) = \frac{\sum_{i=1}^{k}(S_r(w_i) - S_m(w_i, p))^2}{k}$$

where p is a potential position of the claimer, k is the number of witnesses, S_r is the received signal strength at witness w_i, S_m is the calculated signal strength at w_i obtained from radio propagation model. By varying p, we can minimize MSE and finally get the optimal estimated position $\hat{p}$.

If the estimated position of a claimer is far away from its claimed position, we regard this node as a suspect node. Note that due to the unstable and irregular nature of RF(Radio Frequency), we still cannot assert, based on the results of this simple computation, that a Sybil attack is happening.

We take Figure 1 as an example. Node s_1, a claimer, broadcasts a beacon, claiming its identity and position. Node n_1, a verifier, collects all signal strength measurements from neighboring witnesses which have received the beacon. Obviously, the final estimated position of s_1 would be near the position of node m, instead of the position s_1 claimed, as node s_1 and m are physically the same vehicle.



(a) Scenario 1: no Sybil witnesses

(b) Scenario 2: 30% Sybil witnesses

Figure 2: Estimated position distribution.

The beacon message can be in the following format:

$$\{NodeID, Beacon\#, Position, NebList, Signature\}$$

$$NebList : \{NodeID_i, Beacon\#_i, RSS_i\}$$

where $NodeID$ is the claimer's identity, $Beacon\#$ is a beacon sequence number, $Position$ is the sender-claimed position, $NebList$ is the sender's most recent neighbor list containing signal strength measurements, $Signature$ is the digital signature for the whole packet. In each item of $NebList$, RSS_i is the Received Signal Strenth of beacon $Beacon_i$ recently received from neighboring node $NodeID_i$.

3.2 Simulation

Radio model. In our simulation, we apply a widely-used radio propagation model, the shadowing model [14], which consists of two parts. The first one is known as path loss model, which also predicts the mean received power at distance d. The second part of the shadowing model reflects the variation of the received power at a certain distance. The overall shadowing model is represented by

$$\left[\frac{P_r(d)}{P_r(d_0)}\right]_{dB} = -10\beta log(\frac{d}{d_0}) + X_{dB}$$

where d_0 is a reference position, d is the position where the signal strength is measured, β is called the path loss exponent, and X_{dB} is a Gaussian random variable with zero mean and standard deviation σ_{dB}.

Scenario I. In this scenario, we suppose that the signal range is 200m, the physical position of a claimer is at the point of 200m, and all 10 witnesses distributed at random positions faithfully report the measured signal strength from the claimer. Since the signal range (200m) is much larger than the road width(for example, 20m), we also suppose that all vehicle are distributed on a line. Our simulation runs independently for 100 times with different random witness positions, and the distribution of estimated position is shown in Figure 2(a). From the figure, we can find that the estimated positions of most tests ($> 99\%$) are within 10m of the real position

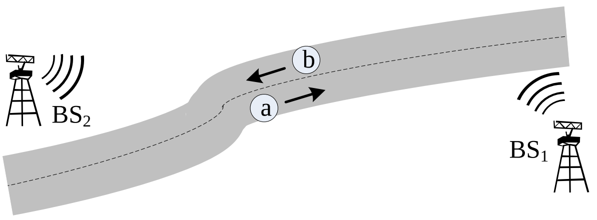


Figure 3: A scenario with roadside infrastructures.

of the claimer, thereby suggesting that the positioning accuracy is about 10m.

Scenario II. The basic configurations of Scenario II is the same as Scenario I. However, we assume that the claimer is a malicious node, which claims that its position is at the point of 300m instead of its real position at 200m. What's more, 30% of the witnesses are sybil vehicles. In order to change the computed result to match the fabricated position (at 300m), these sybil witnesses would report fictitious measurements instead of the real measured ones. The results in Figure 2(b) shows that much of the estimated positions are detracted from the real position (at 200m), moving closer to the fabricated position (at 300m).

3.3 Conclusion

We can reach two conclusion based on the above simulation results. First, the signal-strength-based verification accuracy is rather limited, with an error about 10m. That is to say, if two physical vehicles are very close to each other, less than 10m, we can not ensure whether they are two neighboring vehicles or two sybil vehicles. Second, Sybil witnesses might largely impact the results of position verification. If all witnesses tell the real measured signal strength, the estimated position is very close to the claimer's physical position.

In the next section, we will show how we take full advantage of the unique features of VANETs, such as traffic patterns and high mobility, to apply signal-strength-based position verification into the detection of Sybil attacks.

4. DETECTING SYBIL NODES IN VANETS

In this section, we propose a detection scheme for ensuring that each physical vehicle is bound with only one identity. If multiple identities, claiming to be at various positions, prove to be at one physical position through position verification, then sybil attacks are likely in progress. In this section, we propose several techniques to compensate for the two weaknesses of the basic signal-strength-based position verification, namely, limited accuracy and Sybil witnesses.

4.1 A Technique to Remove Sybil Witnesses

Before a verifier performs the position-estimation computation, it should remove the signal-strength measurements originating from potential Sybil witnesses as much as possible. We present a technique for ensuring that each witness is a physical vehicle but not a Sybil vehicle. In this technique, we take full advantage of the traffic patterns and base station support in the vehicular environment.

As we assumed in Section 2.2, roadside base stations, manipulated by governments, are sparsely distributed along roads. Then, based on this assumption, we establish the following two rules:

Rule 1. *A roadside base station would issue a position certification for each vehicle passing by. The position certification contains a time stamp and location information of the base station and therefore can prove the presence of the vehicle near the base station at a certain time.*

Rule 2. *All witnesses for a claimer should consists of vehicles in the opposite traffic flow of the claimer.*

With Rule 1, we can ensure where a certain vehicle comes from. We take Figure 3 as an example. Node a can get a position certificate from base station BS_2, when passing by BS_2, and node b also get one from BS_1. When a and b meet each other, it's easy for them to prove that they come from the opposite directions by exchanging certificates.

With Rule 2, we can ensure that each witness in the opposite traffic flow is a physical vehicle instead of a Sybil one. The example in Figure 1 can illustrate how this rule works. Malicious node m fabricates 7 Sybile nodes, in which, s_7 is traveling in the opposite direction and the rest the same. When trying to verify the positions of s_1,...,s_6, we only choose witnesses in the opposite (right-to-left) traffic flow such as node n_2,...,n_5. However, with Rule 1, we would ignore node s_7, because it cannot prove that it comes from the upstream of the road, and actually it does not. In this way, we can ensure that each witness is a physical vehicle coming from the opposite direction.

With Rule 1 and Rule 2 together, we achieve the goal that the membership of witnesses consist of only physical vehicles, excluding any Sybil vehicle. A verifier wouldn't select witness nodes from the same traffic flow, because it's difficult to decide which witnesses in the same traffic flow are potential Sybil witnesses. With the help of roadside infrastructure, the impact of dishonest Sybil nodes on position verification can be effectively removed.

Please also note that our scheme only requires that base stations are sparsely deployed along the roadside. In other words, most sections of the road are not covered within the signal range of base stations. Base stations only serve as the authority to prove where a given vehicle comes from. Research on how densely base stations can be deployed for better performance will be part of our future work.

4.2 Detection Model

We propose the following model for detecting Sybil nodes in VANETs. Let $\mathcal{P}$ be a Euclidean space and let $\|P_1 - P_2\|$ denote the Euclidian distance from point P_1 to P_2.

A **node** is a triplet (N, f, f'), where:

- $N \in \mathbb{N}$ is an integer that denotes the identity of the node
- f, the claimed location function of the node, is a continuous function f: $T \to \mathcal{P}$ that indicate the position that the node claims over the lifetime $T \subseteq \mathbb{R}$.
- f', the estimated location function, is also a continuous function f': $T \to \mathcal{P}$ that indicate the estimated position, which is computed based on the signal strength measurements, over the lifetime $T \subseteq \mathbb{R}$.

Detection Model. The detection model specifies which nodes are potential Sybil nodes. Formally, let ν be the set of all vehicles and let $\mathcal{S}$ be the set of sets of Sybil nodes. The model is a function $\mathcal{D}$: $\nu \to \mathcal{S}$. We classify two Sybil nodes into one set, if they originate from one physical vehicle. The function, $\mathcal{D}$, can be implemented by any cluster algorithm. However, the challenge is how to detect a Sybil node and how to decide the correlation between two potential Sybil nodes.

4.3 Enhanced Position Verification

We present a statistic algorithm to detect whether a node honestly claims its position. Each vehicle on a road can perform this algorithm when enough signal strength measurements from nearby witnesses are collected. The basic idea of this algorithm is that although individual estimated positions for a certain normal node may be inaccurate, the estimated positions for the normal node would be very close to its claimed positions over a period of time.

Our algorithm is based on hypothesis tests. We divide the observation period, Δt_o, into discrete time intervals, t_1,...,t_n. Hence the claimed positions of a vehicle, v, is a sequence: $f(t_1)$,...,$f(t_n)$, and the estimated positions: $f'(t_1)$,...,$f'(t_n)$. We suppose that v is a normal node. The estimated position, $f'(t_i)$, usually regarded as random errors, should follow the normal distribution with the mean $\mu = f(t_i)$. Further, the difference, $d_i = f'(t_i) - f(t_i)$, is supposed to follow the standard normal distribution with mean $\mu_d = 0$ and variance σ_0^2. Thus the problem is to test the following hypotheses for the samples, d_i, $1 \leq i \leq n$:

$$H_0 : \mu_d = 0 \quad H_1 : \mu_d \neq 0$$

$$H_0' : \sigma^2 \leq \sigma_0^2 \quad H_1' : \sigma^2 > \sigma_0^2$$

When both H_0 and H_0' are true, the algorithm returns that the node honestly claims its position. Since the mean and variance are supposed to be 0 and σ_0^2 respectively, the test statistic is

$$|z| = \left| \frac{\overline{d} - 0}{\sigma_0/\sqrt{n}} \right|$$

where $\overline{d}$ is the mean of samples, d_i. If $|z| \geq z_{\alpha/2}$, H_0 is rejected; otherwise, H_0 is accepted. $z_{\alpha/2}$ is the critical value of normal distribution $N(0,1)$, given significance α. α is the significance level, a predefined parameter, denoting:

$$P\{\text{Reject } H_0 | H_0 \text{ is true}\} \leq \alpha.$$

Then we use χ-test to test the variance σ^2. The test statistic is

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

where s^2 is the variance of samples, d_i. If $\chi^2 \leq \chi_\alpha^2(n-1)$, H_0' is accepted; otherwise, H_0' is rejected. $\chi_\alpha^2(n-1)$ is the critical value of χ^2 distribution, given significance α and freedom $(n-1)$.

In the above test, we are interested in whether there is a significant difference between claimed positions and the corresponding estimated positions for a certain node over a period of time. Each node may perform the above test for each neighboring node after an observation period of Δt_o. If a Syil node is detected, then the Sybil node classification algorithm introduced in the next subsection will be performed to find other potential Sybil nodes originating from the same physical vehicles.

How much time it takes a vehicle to collect enough data to detect potential Sybil attacks is determined by the observation period, Δt_o. Due to the statistic nature of our approach, the longer the observation time, the more accurate the verification can be.

4.4 Sybil Node Classification

Next we present a statistic algorithm intended to find other potential Sybil nodes originating from the same physical node after a Sybil node is detected by the algorithm introduced in the previous subsection. It is evident that the estimated positions of two potential Sybil nodes, belonging to one physical node, would be very close to each other over a period of time.

This algorithm can also be implemented based on hypothesis tests. v_1 and v_2 are supposed to be two potential Sybil vehicles, originating from a physical vehicle, v_0, and $f_1'(t_1)$,...,$f_1'(t_n)$ and $f_2'(t_1)$,..., $f_2'(t_n)$ are sequences of their estimated position respectively over the observation period, Δt_o. Since both $f_1'(t_i)$ and $f_2'(t_i)$ are actually estimated positions for v_0's physical position $f_0(t_i)$, we obtain the difference

$$D_i = [f_1'(t_i) - f_0'(t_i)] - [f_2'(t_i) - f_0'(t_i)] = f_1'(t_i) - f_2'(t_i)$$

which is also supposed to follow the normal distribution with mean $\mu = 0$ and variance $2 \cdot \sigma_0^2$. Thus the problem is to test the following hypotheses for the samples, d_i, $1 \leq i \leq n$:

$$H_0 : \mu_d = 0 \quad H_1 : \mu_d \neq 0$$

$$H_0' : \sigma^2 \leq 2 \cdot \sigma_0^2 \quad H_1' : \sigma^2 > 2 \cdot \sigma_0^2$$

When both H_0 and H_0' are true, the algorithm returns that these two nodes are Sybil nodes originating from the same malicious physical node; otherwise, the algorithm returns that these two nodes have no relationships with each other. The hypothesis test would be similar to the test introduced in Section 4.3.

With this Sybil node classification algorithm, we can not only find all potential Sybil node originating from the same physical node, but also localize the malicious physical node.

4.5 Overall Detection Process

In this subsection, we summarize the overall detection process. Our scheme is based on a distributed and localized approach. The overall detection process, performed by each node, includes three phases:

- **Phase 1.** Node v periodically broadcasts beacon messages and receives beacon messages from neighboring nodes. The corresponding signal strength measurements for each received beacon message is saved in its memory.

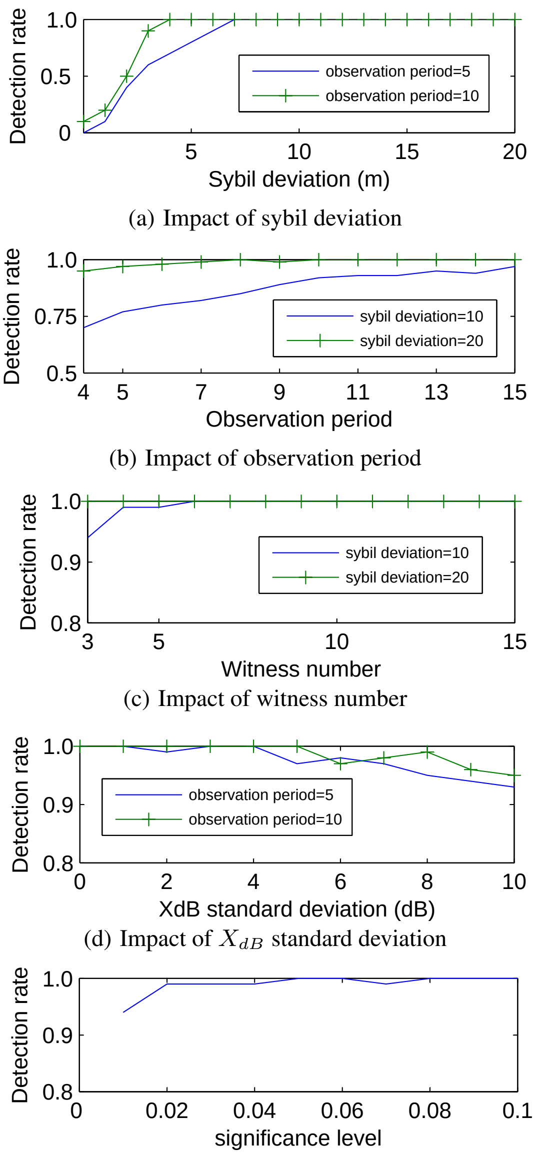
- **Phase 2.** When node v collects enough signal strength measurements for a neighboring node, s, node v performs the enhanced position verification algorithm on s.

- **Phase 3.** If s proves to be a Sybil node in Phase 2, node v performs the Sybil node classification algorithm on s and other neighboring nodes, attempting to find all potential Sybil nodes originating from the same malicious physical node.

During Phase 3, we can find the estimated physical position of the malicious node and even its movement trajectory, which would be very helpful for further intrusion response decisions.

5. SIMULATION EVALUATION

In this section, simulation tests are conducted to evaluate the enhanced position verification algorithm and the Sybil node classification algorithm introduced above. We are interested in how our algorithms perform with various values of system metrics, including *Sybil deviation*, *observation period*, *witness number*, X_{db} *standard deviation*, *significance level*, and *relative speed*. Sybil deviation is the difference in distance between a claimed position and the corresponding estimated position. Observation period, Δt_o, as introduced in Section 4.3, denotes the time period between the moment the first beacon from a claimer arrives and the moment the position verification algorithm is performed. For example, if observation period is equal to n units, it means that there are n estimated positions calculated for the claimer during the observation period. Witness number is defined as the average number of witnesses when



(a) Impact of sybil deviation

(b) Impact of observation period

(c) Impact of witness number

(d) Impact of X_{dB} standard deviation

(e) Impact of significance level

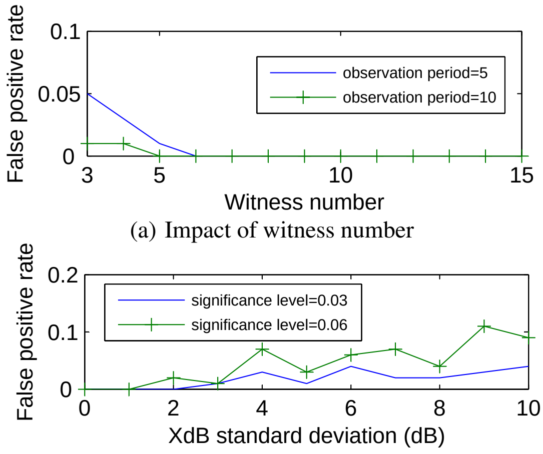
Figure 4: Evaluation of enhanced position verification algorithm

a claimer sends out a beacon. X_{dB} standard deviation denotes the standard deviation of X_{dB} in our radio model, suggesting the unstable level of radio channels. Significance level is the system parameter, α, in our hypothesis test. Relative speed is referred to as the relative speed between two vehicles heading for one direction.

5.1 Enhanced Position Verification

In the first test, we evaluate the detection rate and the false positive rate for our enhanced position verification algorithm. The detection rate indicates the probability that a Sybil vehicle can be detected, and the false positive rate denotes the percentage of normal vehicles which are mistakenly regarded as Sybil vehicles by our algorithm.

Firuge 4 shows the evaluation of detection rate for our enhanced position verification algorithm. Figure 4(a)(b) shows that the detection rate increases as Sybil deviation or observation period increases. When Sybil deviation is larger than or equal to 10m and observation time larger than or equal to 10 units, the detection rate is larger than 90%. It's easy to understand that a claimed position far from the physical position makes it more evident to detect, and the statistic approach effectively compensates for the inaccuracy of individual estimations at the cost of increased observation time (detection time). As shown in Figure 4(c), by increasing witness number, the detection rate will increased too. A large number of witness vehicles can increase the accuracy of estimated positions, thereby increasing the detection rate. Usually, a witness number, 5, would be enough to achieve a high detection rate (≥99%). Figure 4(d) shows that the more stable the radio channel, the higher the detection rate would be. As we expected, if the channel is stable, the estimated position for a potential Sybil node would be more accurate, thereby increasing the detection rate. Figure 4(e) indicates that the detection rate can also be increased by increasing significance level. An increase in significance level can increase the critical values of $z_{\alpha/2}$ and $\chi^2_{\alpha}(n-1)$ and accordingly increase the probability that H_0 and H'_0 are rejected.



(a) Impact of witness number

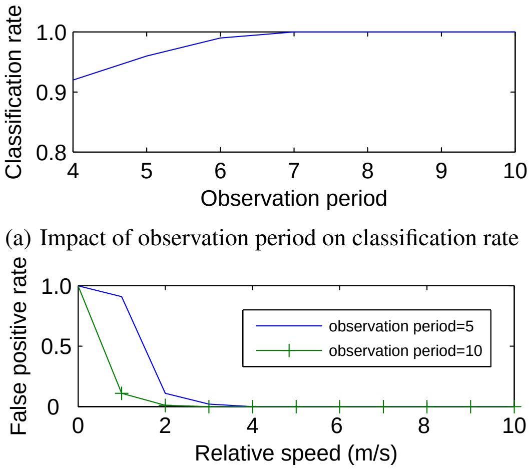
(b) Impact of X_{dB} standard deviation

Figure 5: False positive rate of enhanced position verification algorithm

Figure 5 shows the evaluation of false positive rate for our enhanced position verification algorithm. Figure 5(a) illustrates that the false positive rate mainly depends on observation period and witness number. A larger witness number can increase the accuracy of individual estimations, while a longer observation period would improve the final accuracy by using statistic approaches. Figure 5(b) indicates the impact of X_{dB} standard deviation on the false positive. The more unstable the radio channel, the more inaccurate the estimated positions, and therefore the more chances that a normal node is mistakenly regarded as a Sybil node. The figures also indicates that false positive rate would also increase as significance level increases. Therefore, although an increase in significance level can improve detection rate, we have to make a tradeoff for significance level between false positive rate and detection rate according to specific application requirements.

5.2 Sybil Node Classification

In the second test, we investigate the classification rate and the false positive rate for our Sybil node classification algorithm. The classification rate denotes the probability that two Sybil nodes, one of which might be a malicious physical node, are correctly classified into one group. The false positive rate is the probability that



(a) Impact of observation period on classification rate

(b) Impact of relative speed on false positive rate

Figure 6: Evaluation of sybil node classification algorithm

one Sybil node and one innocent node are incorrectly classified into one group.

Figure 6(a) shows that the classification rate can be increased as the observation period increases. As shown in the figure, even when observation period is equal to 5, the classification rate is larger than 95%. It's easy to understand that if the estimated positions of a Sybil node keep very close to another node as time elapses, it is very likely that they are either two Sybil nodes or one Sybil node and one malicious physical node. Thus, the classification rate mainly depends on observation period. Figure 6(b) illustrates the variance of false positive rate over relative speed. Here, relative speed is referred to as the relative speed between a Sybil node and an innocent normal node nearby. If a normal vehicle keeps very close to a malicious node for a long time, actually, it is difficult to distinguish them by position estimation, due to the limited accuracy. However, that would be a small probability event based on the assumption that each vehicle moves independently. As shown in the figure, given observation time=10, we only required the relative speed to be larger than 1m/s to keep the false positive rate less than 10%.

6. DISCUSSION

6.1 Attack Analysis

In this subsection, we are interested in the potential strategies the adversary may apply to crack our detection scheme. In order to make the final estimated position of a Sybil vehicle closer to its claimed position, a malicious node may apply two potential strategies: spoof transmitting power and witness penetration. The former attempts to impact the final estimated position from the signal source aspect, while the latter from the witness aspect.

In spoof transmitting power, malicious physical nodes may deliberately decrease or increase the RF(Radio Frequency) power for broadcasting a beacon message in order to impact the signal strength measurements. However, this attempt is destined to fail, because our scheme is based on analysis of signal strength distribution but not direct distance measurement. We take Figure 7 as an example. The malicious node, m, broadcasts a beacon claiming to be node s_1 at increased power. Thus, node n_1,n_2,n_3, and n_4 get the enlarged measurements for the beacon. By performing MMSE on both the measurements of n_1,n_2,n_3,n_4 and m's signal strength distribution pattern, we can still compute the actual position of s_1 or m.

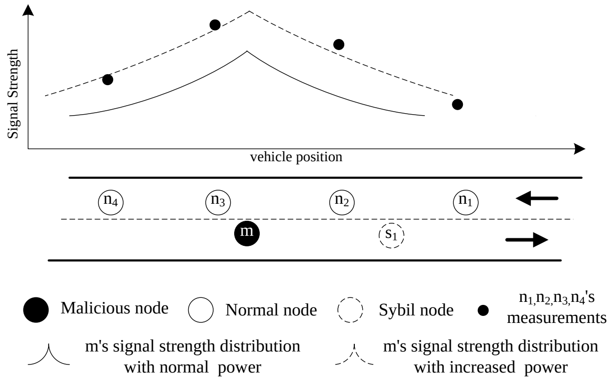

Signal Strength
vehicle position
n4
n3
n2
n1
m
s1
Malicious node
Normal node
Sybil node
n1,n2,n3,n4's measurements
m's signal strength distribution with normal power
m's signal strength distribution with increased power

Figure 7: Spoof signal strength.

In witness penetration, Sybil nodes play the role of witnesses and refuse reporting the real signal strength measurements so that they can cover up for each other. However, we may disable this attempt by using the technique proposed in Section 4.1. In this technique, all witnesses are selected from the opposite traffic flow. Our motivation behind this technique is based on the fact that physical witnesses in the opposite traffic flow can prove that they have come from the upstream of the road, whereas Sybil witnesses cannot.

6.2 Features of Our Scheme

Our detection scheme presents several unique features. First, with our detection scheme, we can have the potential to suppress Sybil attacks in VANETs, preventing a malicious vehicle from fabricating tens or hundreds of Sybil vehicles. Because each vehicle is supposed to occupy a considerable amount of space, a malicious vehicle can fabricate only a few Sybil vehicles around itself and therefore the security threats to the whole network is significantly reduced. Due to the limited accuracy of the signal-strength-based approach, we still can not detect subtle small-scale attacks. Second, based on analysis of signal strength distribution, our detection scheme can be resistant to spoof attacks. It is difficult for a malicious physical vehicle to change its signal strength distribution. Unlike traditional radio-resource-testing-based approaches[7], we don't reply on assumptions on specific RF(Radio Frequency) devices. Even when a malicious node is equipped with multiple radio modules, we still can detect potential attacks by analyzing its signal strength distribution. Finally, our scheme has the potential to detect all Sybil nodes originating from the same physical node and localize the malicious physical node, thereby facilitating further intrusion response mechanisms.

7. RELATED WORK

Considerable attention from the research community has been attracted by emerging vehicular networks. There have been several proposals pointing out the importance of security in vehicular networks[2][3][4][5]. In [3][4], a common security threat is introduced. That is Sybil attacks, in which a malicious node creates an illusion of traffic congestion by claiming multiple identities. Newsome et al.[7] introduces several techniques to detect Sybil attacks in ad hoc networks, including radio resource testing, registration, and position verification. Whereas radio resource testing replies on specific assumptions on radio modules and registration alone is not effective enough, position verification comes to be a more promising approach for vehicular networks. The use of received radio signal strength for positioning is proposed in [9]. It is designed for indoor applications, relying on establishing a signal-strength-distribution map in advance. Brands et al.[8] propose a distance bounding protocol that can be used to verify the proximity of two devices connected by a wired link. Sastry et al.[10] propose a new distance bounding protocol, based on ultrasound and radio wireless communication. The protocol can only make a rough decision about whether or not a claimer is within a certain region. Golle [11] presents a mathematical model to detect malicious data and potential fabricated vehicles. However, it doesn't concern detailed detection mechanisms. Capkun et al.[13] present a secure positioning scheme, which suppose that nodes are static and relies on multiple base stations as reference points. This scheme will not fit the highly mobile context of VANETs.

Research on other protocol layers of VANETs is also in progress. Korkmaz et al.[18] introduce a multi-hop broadcast protocol, designed to address the broadcast storm, hidden node, and reliability problems of multi-hop broadcast in urban areas. Yadumurthy et al.[15] present a reliable MAC broadcast protocol for VANETs, in which, for each broadcast packet, acknowledges from all receivers are returned to the sender. Moreno et al.[17] present a bandwidth sharing protocol for VANETs. Wu et al.[16] propose a data dissemination protocol for VANETs. This protocol relies on nodes' cooperation to forward packets toward the destinations. However, a malicious node can easily crack these protocols by using its large number of Sybil nodes and then launches further DoS attacks.

8. CONCLUSION

In this paper, we propose a lightweight security scheme for detecting and localizing Sybil nodes in VANETs. Unlike traditional approaches in ad hoc networks or in sensor networks, our scheme intend to suppress Sybil attacks in VANETs rather than eliminate individual attacks, for small-scale Sybil attacks can only make limited threats to VANETs. Our scheme is based on statistic analysis of signal strength distribution, not relying on specific hardware. We make use of the unique properties of VANETs to help us address new challenges in system design. Simulation results show that when the witness number is larger than 5 and the observation period is 10 units, we can achieve a high detection rate ($\geq$95%) as well as a low false positive rate ($\leq$5%).

Extensive work is still required in the future. First, we plan to implement a prototype system for our scheme and then investigate its feasibility in real vehicular settings. Second, a more realistic radio propagation model, suitable for the highly mobile context of VANETs, expects to be defined. Finally, we are also interested in how roadside infrastructures can provide more security supports for VANETs.

9. REFERENCES

[1] 5.9GHz DSRC. http://grouper.ieee.org/groups/scc32/dsrc/index.html.

[2] J.J. Blum, A. Eskandarian, and L.J. Hoffman.Challenges of intervehicle ad hoc networks. IEEE Transactions on Intelligent Transportation Systems, Volume 5, Issue 4, Dec. 2004, pp. 347 - 351.

[3] B. Parno and A. Perrig. Challenges in Securing Vehicular Networks. In Proc. of the Fourth Workshop on Hot Topics in Networks (HotNets-IV), 2005.

[4] M. Raya and J.-P. Hubaux. The Security of Vehicular Networks. In Proc. of the 3rd ACM workshop on Security of ad hoc and sensor networks (SASN 2005), pp. 11-21, 2005.

[5] J.P. Hubaux, S. Capkun, and J. Luo. The Security and Privacy of Smart Vehicles. IEEE Security and Privacy Magazine, 2(3):49-55, May-June 2004.

[6] Y. Zhang and W. Lee. Intrusion Detection in Wireless Ad-Hoc Networks. In Proc. ACM MobiCom, 2000.

[7] J. Newsome, E. Shi, D. Song, and A. Perrig. The Sybil Attack in Sensor Networks: Analysis & Defenses. In Proc. of the Third International Symposium on Information Processing in Sensor Networks (IPSN 2004), pp. 259-268, 2004.

[8] S. Brands and D. Chaum. Distance-Bounding Protocols. In Proc. of Workshop on the theory and application of cryptographic techniques on Advances in cryptology. Springer-Verlag, Inc., pp. 344-359, 1994.

[9] P. Bahl and V.N. Padmanabhan. RADAR: An In building RF-based User Location and Tracking System. In Proc. of IEEE INFOCOM 2000, pp. 775-784, 2000.

[10] N. Sastry, U. Shankar, and D. Wagner. Secure Verification of Location Claims. In Proc. of the 2003 ACM workshop on Wireless Security (WiSe 2003), pp. 1-10, 2003.

[11] P. Golle, D. Greene, and J. Staddon. Detecting and Correcting Malicious Data in VANETs. In Proc. of the 1st ACM international workshop on Vehicular ad hoc networks (VANET 2004), pp. 29-37, 2004.

[12] G. Zhou, T. He, S. Krishnamurthy, J.A. Stankovic. Impact of Radio Irregularity on Wireless Sensor Networks. In Proc. of the 2nd international conference on Mobile systems, applications, and services (MobiSys 2004), pp. 125-138, 2004.

[13] S. Capkun and J.-P. Hubaux. Secure Positioning of Wireless Devices with Application to Sensor Networks. In Proc. of Infocom 2005, pp. 1917-1928, 2005.

[14] T. S. Rappaport. Wireless communications, principles and practice. Prentice Hall, 1996.

[15] R.M. Yadumurthy, A. Chimalakonda, M. Sadashivaiah, and R. Makanaboyina. Reliable MAC Broadcast Protocol in Directional and Omni-directional Transmissions for Vehicular Ad hoc Networks. In Proc. of the 2nd ACM international workshop on Vehicular ad hoc networks (VANET 2005), pp. 10-19 , 2005.

[16] H. Wu, R. Fujimoto, R. Guensler, and M. Hunter. MDDV: A Mobility-Centric Data Dissemination Algorithm for Vehicular Networks. In Proc. of the 1st ACM international workshop on Vehicular ad hoc networks (VANET 2004), pp. 47-56, 2004.

[17] M. Torrent-Moreno, H. Hartenstein, P. Santi. Fair Sharing of Bandwidth in VANETs. In Proc. of ACM Workshop on Vehicular Ad Hoc Networks (VANET 2005), pp. 49-58, 2005.

[18] G. Korkmaz and E. Ekici. Urban Multi-hop Broadcast Protocol for Inter-Vehicle Communication Systems. In Proc. of the 1st ACM international workshop on Vehicular ad hoc networks (VANET 2004), pp. 76-85, 2004.