IEEE ComSoc | IEEE COMPUTER SOCIETY | IEEE Signal Processing Society | VTS | IEEE Transactions on Machine Learning in Communications and Networking

Received 26 June 2023; revised 14 December 2023 and 4 April 2024; accepted 29 May 2024.
Date of publication 5 June 2024; date of current version 11 June 2024.

The associate editor coordinating the review of this article and approving it for publication was K. Zeng.

Digital Object Identifier 10.1109/TMLCN.2024.3410208

Sybil Attack Detection Based on Signal Clustering in Vehicular Networks

HALIT BUGRA TULAY (Member, IEEE), AND CAN EMRE KOKSAL (Senior Member, IEEE)

Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH 43210 USA

Corresponding author: H. B. Tulay (tulay.1@osu.edu)

ABSTRACT With the growing adoption of vehicular networks, ensuring the security of these networks is becoming increasingly crucial. However, the broadcast nature of communication in these networks creates numerous privacy and security concerns. In particular, the Sybil attack, where attackers can use multiple identities to disseminate false messages, cause service delays, or gain control of the network, poses a significant threat. To combat this attack, we propose a novel approach utilizing the channel state information (CSI) of vehicles. Our approach leverages the distinct spatio-temporal variations of CSI samples obtained in vehicular communication signals to detect these attacks. We conduct extensive real-world experiments using vehicle-to-everything (V2X) data, gathered from dedicated short-range communications (DSRC) in vehicular networks. Our results demonstrate a high detection rate of over 98% in the real-world experiments, showcasing the practicality and effectiveness of our method in realistic vehicular scenarios. Furthermore, we rigorously test our approach through advanced ray-tracing simulations in urban environments, which demonstrates high efficacy even in complex scenarios involving various vehicles. This makes our approach a valuable, hardware-independent solution for the V2X technologies at major intersections.

INDEX TERMS Cybersecurity, Sybil attack, vehicular ad hoc networks, VANET, physical layer security.

I. INTRODUCTION

IN A vehicular network, vehicles use the information from the other vehicles or infrastructure as a part of an intelligent transportation system (ITS) to increase their situational awareness and improve their automated/assisted decision making. However, these networks face numerous security challenges due to the broadcast nature of communication and the potential for malicious attacks. Ensuring the authenticity of vehicles before vehicles access any services or information is critical for many applications related to driving safety, as false information can lead to accidents [1], [2].

Among the various types of attacks in vehicular networks, the Sybil attack is considered one of the most threatening attacks. In the Sybil attack model, a single node (vehicle) declares itself with multiple identities to other nodes in the network. Therefore, the other nodes can't determine if the information comes from a single node or multiple nodes [3]. In [4], the authors show that Sybil attack can affect control algorithms in cooperative driving scenarios, and may result in accidents at high speeds. A Sybil attacker can also damage the network by consuming network resources more than benign nodes or spreading false safety messages. For example, it can send incorrect information about road conditions, and enforce other vehicles to change their routes and clear the road. Furthermore, the attacker can also take advantage of applications relying on the assumption of honest majority as in [5].

In this paper, we introduce a unique approach for detecting Sybil attacks in vehicle ad hoc networks (VANETs), a critical challenge in the field of vehicular communication security. Due to the nature of wireless propagation, the channel state information (CSI), specifically the channel frequency response (CFR), shows major spatio-temporal variations. As a result, signals collected from different vehicles tend to be clustered in isolated groups in the signal space, as shown in Fig. 1. Our proposed detection system exploits the spatial characteristics of the channel state information, and differentiates the attacker from the legitimate nodes based on the clustering of CSI samples in the signal space. A notable feature of our approach is its independence from specific wireless channel models, where the test statistic is derived according to the channel model assumption. In addition to this, we introduce certain preprocessing stages that enhance

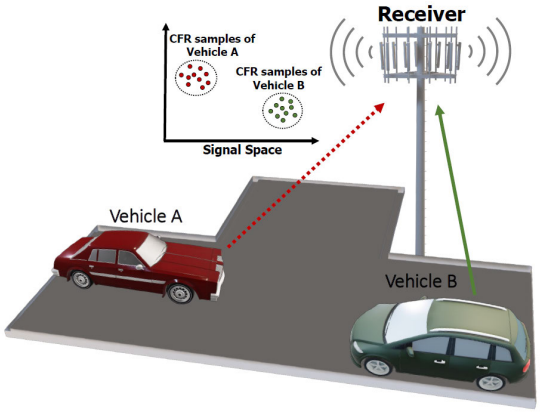


FIGURE 1. The CSI profiles of vehicles differ in the signal space.

the efficacy of our approach. Specifically, we demonstrate that preprocessed CSI samples from different vehicles are clustered around distinct points in the signal space, and the distance between the centroids of the clusters derived by the k-means algorithm is a good test statistic for effective attack detection. Our approach also allows each node to independently perform detection without the need for cooperation from neighboring nodes. This eliminates the need for establishing the credibility of these nodes, a common limitation of trust-based approaches.

We verify the efficacy of our approach, particularly for scenarios involving low-speed or stationary vehicles, by conducting dedicated short-range communications (DSRC)-based vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) experiments in a real vehicular network that includes many roadside units (RSUs) and vehicles equipped with DSRC radios as a part of Smart Columbus program [6]. Our experiments include unique configurations, such as launching a real Sybil attack using DSRC radios and a congestion generation tool in a V2V communication setting. We collect SPaT (Signal Phase and Timing) messages broadcast from real RSUs that mimics an attack in V2I communication setting. In addition to real-world experiments, we have analyzed large-scale and broader situations using the data obtained from a ray-tracing simulator in an urban environment with many vehicles. By using a ray-tracing simulator, a more sophisticated simulator tool than typical channel models in the literature, we obtain realistic signal level data under different scenarios in a complex environment. In the simulations, we evaluate the impact of the distance between an attacker and legitimate nodes, and the impact of the transmitter power on our performance. These simulations complementing our real-world experiments provide a robust evaluation of our approach across a wide range of scenarios. Our findings demonstrate over 98% detection rate in real-world experiments that emphasize the efficacy and practicality of our approach in detecting Sybil attacks in vehicular networks.

Given most major intersections will be deployed with a vehicle-to-everything (V2X) technology (either DSRC or cellular) in the near future, our approach has the potential to provide value to existing infrastructure, without a need for additional hardware to detect Sybil attacks. The main contribution of this paper can be summarized as follows:

- We introduce a unique approach for detecting misbehavior in vehicular networks. Our approach encompasses multi-layered preprocessing stages, each meticulously designed to address specific challenges encountered in vehicular communication signals, and a test statistic not derived under wireless channel model assumptions.
- We demonstrate the feasibility of our approach by collecting substantial V2I and V2V data in a real vehicular network, including a launch of a real Sybil attack with DSRC radios and Congestion Generation Tool, and data collection from real RSUs.
- In addition to real-world experiments, we evaluate the performance of our approach in complex propagation environments via extensive simulations using a ray-tracing simulator and discuss the impact of different parameters on performance.
- The proposed algorithm, which does not require additional hardware, efficiently uses wireless radio data samples, achieving an impressive 98% detection rate in real-world experiments.

The rest of this paper is organized as follows. The related work is discussed in Section II. Section III presents the details of our approach. Section IV demonstrates the experimental and simulation results. Finally, Section V summarizes our work.

II. RELATED WORK

After the introduction of Sybil Attacks by Doucer [7], researchers have comprehensively studied Sybil attack detection in the networks, and different approaches have been proposed to ensure the uniqueness of identities of the nodes in the network. We can classify different approaches for VANETs as 1) Resource testing based approaches, 2) Certification and cryptography based approaches, 3) Reputation or Trust based approaches 4) Physical-layer based approaches.

A. RESOURCE TESTING BASED APPROACHES

These approaches depend on the observation of different resources of vehicles, such as radio resources, computational, and memory resources. The main assumption of resource testing is that each node in the network has limited resources. In [8], the authors suggest several approaches including radio resource testing, identity registration, and position verification. In the context of radio resource testing, they assume that an attacker can have multiple identities but it is incapable of simultaneous transmission or reception on a single radio. Several studies [9], [10], [11] have proposed an approach where each node is required to periodically solve computational puzzles using a proof-of-work algorithm, similar to those employed in cryptocurrencies. The solutions are later verified by other nodes or RSUs before initiating communication.

The nodes failing the computational work are revoked from the network. Since an attacker needs to dedicate its limited resources to these computations, this limits the attacker's capability of launching the attack. However, the constant computation of proof-of-work for authentication of nodes restricts the network's scalability.

B. CERTIFICATION AND CRYPTOGRAPHY BASED APPROACHES

These approaches request vehicles to provide their certificates issued by a trusted authority in order to communicate with other vehicles and RSUs. The authentication is performed with encryption/decryption of the messages between vehicles, and digital certificates [12]. There are currently two standards used at the application layer for security in ITS, namely the IEEE 1609.2 standard [13] and the European Telecommunication Standard Institute (ETSI)-ITS standard [14], [15]. Both standards adopt the public key infrastructure (PKI), e.g., the policy governing the issuance of digital certificates and asymmetric cryptography techniques. IEEE 1609.2 standard classifies entities into two categories, Certificate authority entities and End entities. It defines a secure message format and methods for them to secure application messages. ETSI TS 102 940 [14] and ETSI TS 102 941 [15] standards specify the security services and architectures along with ETSI TSI 103 097 standard defines V2X message security header and certificate formats. Since PKI-based approaches rely on an infrastructure for the administration and revocation of certificates, they incur high cost and face scalability challenges in large ad-hoc networks [16]. Furthermore, the overhead associated with the signature and certificate is typically larger than the main message content. For example, each 69-byte message should be attached by a 125-byte certificate and a 56-byte signature [17]. This creates a significant latency for transmission.

C. REPUTATION OR TRUST BASED APPROACHES

Although cryptography-based approaches have been widely proposed in the literature, they are vulnerable to insider attackers who are already authenticated in the system. Therefore, reputation or trust based approaches have been proposed to address this problem wherein vehicles assess the other vehicles based on their reputation scores. The reputation score of a vehicle is used as an indicator of its likelihood to engage in misbehavior, such as launching a Sybil attack. The vehicles can then use this information to decide whether to trust or distrust the other vehicles in the network [18]. In [19], the authors propose a machine learning and reputation-based scheme for detecting misbehavior in vehicular communication networks. The proposed scheme utilizes reputation scores computed based on the trustworthiness of vehicles and features extracted from vehicular network communication to train a machine learning model for misbehavior detection. ETSI 102 941 standard [15] also defines trust-related functions and procedures based on the security architecture defined in [14].

D. PHYSICAL-LAYER BASED APPROACHES

This class of approaches utilizes physical layer information like received signal strength indicator (RSSI), CSI, and angle of arrival (AoA). RSSI has been extensively proposed to detect Sybil Attacks. Most of the previous RSSI-based approaches compute the absolute position [20], [21] or relative distance [22], [23] from RSSI values. The position information is then used to discern and detect Sybil attacker posing with multiple identities. In [23], a receiving node records RSSI values of nearby nodes and calculates the distance between the RSSI vectors of two nodes to find a similarity between them. If the similarity exceeds a certain threshold, the corresponding nodes are considered to belong to the same vehicle (i.e., a Sybil attacker). Parallel to this approach, the authors in [24] proposed an approach that uses RSSI time series. This approach is based on the observation in real-world experiments that the RSSI time series of Sybil nodes have very similar patterns, and compares the similarity among all received series to detect Sybil attacks. The authors in [25] proposed an RSSI-based localization technique that uses mobile nodes for localizing another mobile node and adjusts itself based on the heterogeneous interference levels in the environment. In [26], the authors proposed a scheme that uses a set of sensor nodes to record RSSI values of a wireless node and the nodes that have similar RSSI distributions are declared as Sybil nodes.

Although widely used, the accuracy of RSSI-based approaches hinges on the employed propagation models employed, predominantly the free space path loss model in [22] which is most effective in environments with strong line-of-sight (LOS) conditions. The performance can be substantially degraded in non-LOS scenarios, typically in urban or obstructed areas. Furthermore, the approaches are susceptible to an attacker that purposely changes its transmission power, potentially skewing the RSSI readings and thereby evading detection, and the slight changes in propagation environment. Due to these limitations, researchers have proposed utilizing other types of information such as AoA. The authors in [27] propose to use AoA information for detecting a Sybil attacker that falsifies GPS locations of different identities. In this approach, vehicles calculate the AoA of the received signals and compare it with the angle calculated from the GPS location of the transmitter where a mismatch indicates a Sybil attack. The proposed approach in [28] utilizes both RSSI and AoA information where an RSU calculates the location of vehicles from RSSI and AoA of received signals. It deduces a Sybil attack if identical values are obtained from different identities. Similarly, the authors in [29] proposed an algorithm that combines AoA and CSI amplitude for static attacks in an indoor setting. Although AoA is proposed to be used for enhancing the reliability of Sybil attack detection, the performance of subspace-based or super-resolution estimators (e.g., MUSIC, ESPRIT) degrades in multipath environments because of strong reflectors [30]. So, the performance of the proposed approach can be significantly reduced due to the multipaths from the surrounding environment.

Given the limitations of RSSI and AoA, the channel frequency response as a type of CSI is proposed to be used for Sybil attack detection. CFR can better capture the multipath characteristics and convey richer information about the propagation environment [31]. In [32], the authors use CFRs and build a hypothesis test based on a test statistic derived according to Rayleigh fading to detect Sybil attacks. Therefore, the test statistic is chosen on the prior assumption about the attacker's power and the channel model. Given that Rayleigh fading is a reasonable assumption in a rich scattering environment without a line-of-sight path, it becomes less appropriate for LOS communication scenarios. In [33], the authors proposed an approach that utilizes the Pearson correlation coefficient between CFR of different vehicles for detecting Sybil attacks. If the correlation coefficient is above a certain threshold, it indicates that these nodes are indeed a single node, a Sybil attacker.

In this study, we diverge from conventional RSSI and AoA-based methods, introducing a novel CFR-based approach for Sybil attack detection in VANETs. This approach uniquely exploits the spatial characteristics of CFR, and offers a robust solution that is independent of traditional wireless channel models. Another significant merit of our approach is its compatibility with the existing infrastructure which eliminates the requirement for any additional equipment. To validate the feasibility and robustness of our approach, we have conducted real-world experiments, including DSRC-based V2V and V2I experiments. Complementing these experiments, we employed a ray-tracing simulator to evaluate the system's performance in large-scale and diverse urban environments, taking into account variables such as the distance between the attacker and transmitter, and the power of the transmitted signals. These evaluations allow us to verify the efficacy of our approach in a broad spectrum of scenarios that sets our study apart from previous studies that primarily relied on only simulated data based on certain channel models. Moreover, our system's capability for independent detection by individual nodes, without requiring cooperation or credibility establishment of neighboring nodes, further distinguishes our work from existing trust-based or cooperative detection methods.

III. ATTACK AND CHANNEL MODEL

The communication between vehicles and other ITS nodes is referred to as V2X communication which comprises V2V that enables vehicles to exchange data, and V2I communication that enables the exchange of information between vehicles and road infrastructure. V2X applications heavily rely on the periodic broadcast of basic safety message (BSM) specified by SAE J2945/1 standard [34], or cooperative awareness messages (CAM) which is equivalent to BSM specified by ETSI [35]. In a Sybil attack model, an attacker illegitimately uses multiple identities, and it broadcasts each message with different identities, while legitimate nodes use a single identity at a time. We refer to the vehicle claiming multiple identities as a Sybil attacker, and the corresponding

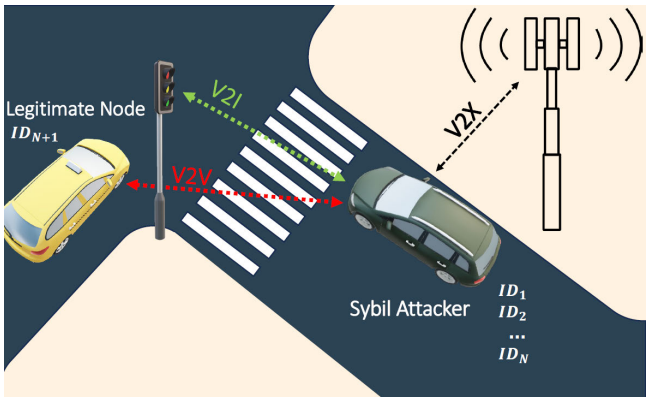


FIGURE 2. **The communication model in a VANET.**

identities as Sybil nodes. It is important to note that our model does not make specific assumptions about how these identities are obtained by the attacker - they could be either self-fabricated or stolen from legitimate vehicles. As long as the attacker has valid identities, identity-based authentication schemes will fail in this attack model. In this paper, we study an attack scenario including legitimate nodes and an attacker node in V2V communication and V2I communication, as shown in Fig. 2. For simplicity, we assume that all nodes in the network are equipped with a single antenna.

A. WIRELESS CHANNEL MODEL

The complex baseband representation of a wireless channel response can be described as:

$$h(t,\tau) = \sum_i a_i(t) e^{-j2\pi f_c \tau_i(t)} g(\tau - \tau_i(t))$$

where $a_i(t)$, $\tau_i(t)$, f_c are the path attenuation, the delay of path i at time t and the carrier frequency, respectively. Here, $g(\tau)$ is the impulse response of the transmit and receive filters. The corresponding CFR can be calculated as the Fourier transform of $h(t,\tau)$ with respect to τ:

$$\begin{aligned} H(t,f) &= \int_{-\infty}^{\infty} h(t,\tau) e^{-j2\pi f \tau} \, d\tau \\ &= G(f) \sum_i a_i(t) e^{-j2\pi (f+f_c)\tau_i(t)} \end{aligned} \tag{1}$$

where $G(f)$ is the frequency response of the transmit and receive filters and it can be assumed to be constant in the presence of guard subcarriers on both sides of the spectrum [36]. We use the orthogonal frequency division multiplexing (OFDM) in this work since New Radio (NR) C-V2X as a part of 3GPP Rel. 16 and DSRC adopt OFDM modulation [37].

In an OFDM system, the channel frequency response at the k^{th} subcarrier of the n^{th} OFDM frame will be

$$H_{n,k} = H(nT, k\Delta_f)$$

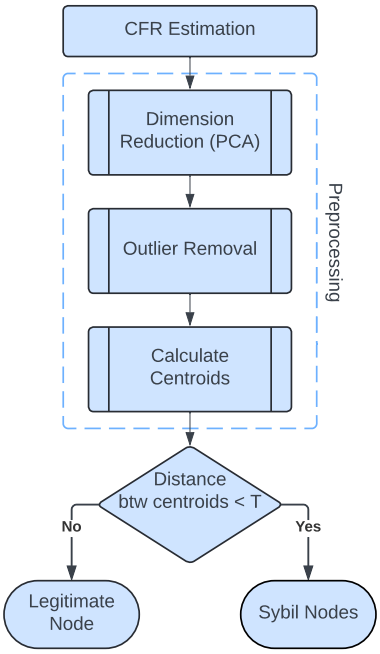


FIGURE 3. Flowchart of the proposed algorithm.

where T denotes the frame duration and Δ_f is subcarrier spacing. We can form the CFR vector using K subcarriers as:

$$\mathbf{H_n} = [H_{n,1},\ H_{n,2},\ \ldots,\ H_{n,K}]^T$$

The CFR is estimated using the preamble symbols at the beginning of each OFDM frame, and it can be assumed constant over the frame. Note that $H_{n,k}$ is a complex number represented by its magnitude $|H_{n,k}|$ and its phase $\angle H_{n,k}$ as $H_{n,k} = |H_{n,k}|e^{j\angle H_{n,k}}$. Our authentication algorithm depends on the magnitude of CFR values, $|H_{n,k}|$, instead of complex $H_{n,k}$ values. This makes it robust against the phase estimation errors and the phase noise, also eases the implementation of our approach.

IV. APPROACH

Our proposed algorithm is summarized in the flowchart given in Fig. 3. Given the prevalent high vehicle density and frequent traffic congestion in urban environments, our approach primarily targets scenarios involving low-speed or stationary vehicles where the clustering effect remains relatively stable and more accurate CFR estimation is ensured. The nodes in a vehicular network are broadcasting 10 messages per second in accordance with the standard protocol. The algorithm initiates when the receiver obtains 10 CFRs from different nodes.

The receiver determines if the frames are received from a stationary vehicle. This determination can be made based on the speed information contained within the frame's content, as specified in the basic safety messages. Alternatively, if the packet content is modified by an attacker, the receiver can determine it using the CFRs, following the approach proposed in [38]. Once the frames are received from stationary vehicles at the receiver, the algorithm proceeds with CFR estimation and preprocessing steps.

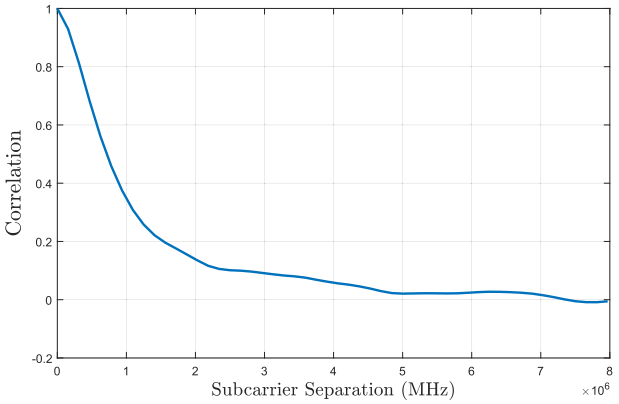


FIGURE 4. Correlation of the subcarriers with different frequency separations.

A. DIMENSION REDUCTION

In OFDM systems, the channel gains of subcarriers can be substantially correlated due to the channel coherence bandwidth. In particular, the adjacent subcarriers can be highly correlated. We can define the Pearson correlation coefficient between the subcarrier k and m using N received frames as:

$$\rho_{k,m} = \frac{1}{N}\sum_{n=1}^{N}\left(\frac{|H_{n,k}| - \mu_k}{\sigma_k}\right)\left(\frac{|H_{n,m}| - \mu_m}{\sigma_m}\right)$$

where μ_k and σ_k are the mean and standard deviation of the k^{th} subcarrier, respectively. Given a 10 MHz vehicular communication channel, the lowest and highest subcarriers are separated by about 8 MHz. We calculate the correlation between the subcarriers over 5000 frames using the real-world V2I experiment data in Sec. V-A. Fig. 4 shows the Pearson correlation between the subcarriers with different frequency separations. We observe that the correlation is significant up to 1 MHz of frequency separation that corresponds to the bandwidth of 7 subcarriers.

Principal component analysis (PCA) is a mathematical algorithm that reduces the dimensionality of the data while retaining most of the variation in the data set [39]. We utilize principal component analysis to reduce the dimensionality of the CFR vector by transforming the correlated CFRs into a new, uncorrelated set of variables called principal components (PCs). The PCs were derived from the eigenvectors of the sample covariance matrix ($\mathbf{\Sigma}$) computed from the CFRs. Specifically, it turns out that for $k = 1, 2, \ldots, p$, the k^{th} PC is given by $c_k = \boldsymbol{\alpha}_k' x$ where $'$ denotes transpose, and $\boldsymbol{\alpha}_k$ is an eigenvector of $\mathbf{\Sigma}$ corresponding to its k^{th} largest eigenvalue, λ_k. Furthermore, if $\boldsymbol{\alpha}_k$ is chosen to have unit length ($\boldsymbol{\alpha}_k^T\boldsymbol{\alpha}_k = 1$), the variance of c_k is equal to λ_k [39]. Then, the PCs were ordered according to the percentage of variance explained by each component. We can calculate the explained variance

with the ratio of eigenvalue of related principal component λ_k to the sum of all eigenvalues ($\lambda_1+\lambda_1+\cdots+\lambda_K$). We observe that the first few PCs retain most of the variation present in the CFR vectors. Therefore, we use the first three principal components whose total explained variance reaches above 90% in our experiments. Thus, we reduce the K-dimensional CFR vectors to three-dimensional vectors while retaining most of the information. This helps us to reduce the running time of the clustering algorithm, considering that the running time increases linearly with the size of the dimension [40]. Subsequently, a clustering algorithm is utilized to cluster the CFR vectors of different nodes in the three-dimensional space.

B. OUTLIER REMOVAL

Internal state transitions (e.g. transmission strength changes, rate changes) in communication systems, constructive and destructive interference can cause outliers in CFR values. The k-means clustering algorithm is sensitive to such outliers [41]. Fig. 5a shows the clustering results on the first two principal components of 10 OFDM frames collected from two Sybil nodes. We observe that a single outlier by itself can be identified as a cluster, and this degrades the performance of our approach. Because of this, the outliers should be removed before clustering. For this purpose, we detect and remove outliers by applying a Hampel filter to the principal components. In detail, given a sequence $x_1, x_2, x_3, \ldots, x_n$ and a sliding window of length w, we can define the local median and the median absolute deviation (MAD) as:

- $m_i = \text{median}(x_{i-w}, x_{i-w+1}, \ldots, x_{i+w+1}, x_{i+w})$
- $MAD_i = \text{median}(|x_{i-w} - m_i|, \ldots, |x_{i+w} - m_i|)$

where m_i and σ_i are local median and the median absolute deviation. If a sample x_i is such that

$$| x_i - m_i | > 3MAD_i$$

The Hampel filter declares x_i an outlier and replaces it with m_i. Fig. 5b shows the clustering results after applying the Hampel filter with the effective window size of 5 ($w = 2$). The window size is chosen heuristically to identify and remove outliers in each second of data (10 samples) from each node while preserving the true signal characteristics. We observe that the clustering result is significantly improved after the outlier removal.

C. CLUSTERING

In this work, we use the k-means++ algorithm [42] that differs from the k-means algorithm by choosing random starting centers with very specific probabilities. In the k-means algorithm, given a number of clusters k and a set of data points, the goal is to choose k centers (centroids) to minimize the sum of the squared distances between each point and its closest centroid. The algorithm details are as follows:

1. Arbitrary choose k initial center $\mathcal{C} = \{c_1, \ldots, c_k\}$.
2. For each $i \in \{1, \ldots, k\}$, set cluster C_i to be the set of points in $\mathcal{X}$ that are closer to c_i than they are to c_j for all $j \neq i$.

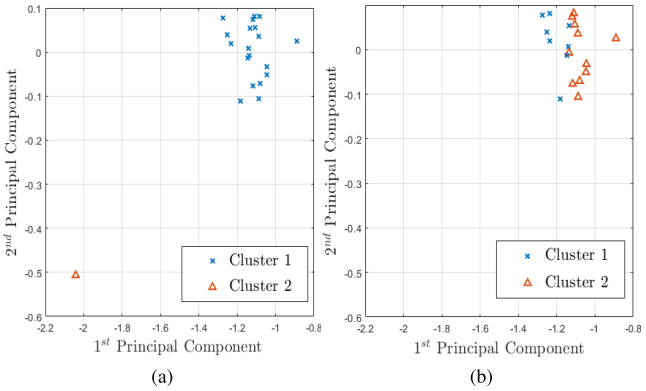


FIGURE 5. (a) Clustering results before outlier removal. (b) Clustering results after outlier removal.

3. For each $i \in \{1, \ldots, k\}$, set c_i to be center of mass of all points in C_i, $c_i = \frac{1}{|C_i|}\sum_{x \in C_i} x$.
4. Repeat Step 2 and 3 until $\mathcal{C}$ no longer changes.

The authors in [42] propose a specific way of choosing these centers in Step 1 and they show that their algorithm (k-means++) improves both the speed and the accuracy of the k-means. In this work, we use the k-means++ algorithm and calculate the distance between the centroids of each pair of nodes with K = 2. Later, the distance information is used as a test statistic for the attack detection.

D. HYPOTHESIS TESTING

We formulate the Sybil attack detection problem as a hypothesis testing problem. To this end, we build a hypothesis test for each node where the node is a legitimate node in the null hypothesis, $\mathcal{H}_0$, and the alternative hypothesis $\mathcal{H}_1$ represents the presence of Sybil attacks. So, we have

$$\mathcal{H}_0 : \text{The node } p \text{ is legitimate node.}$$
$$\mathcal{H}_1 : \text{The node } p \text{ is Sybil node.}$$

After estimating the centroid distances for each pair of nodes, where c_p and c_l represent the centroids of signals from these nodes, we calculate the standard score (a.k.a z-score) of the centroid distances as a measure of relative standing and perform thresholding on the standard scores. The standard score describes the location of an observation relative to the mean in units of the standard deviation and it can be calculated as:

$$z_{pl} = \frac{d_{pl} - \mu}{\sigma} \tag{2}$$

where $d_{pl} \triangleq ||c_p - c_l||$ is the centroid distance between node p and l, μ is the sample mean of the centroid distances and σ is the sample standard deviation of them. Our decision rule can be written as:

$$\delta_p = \begin{cases} 1, & p \neq l \; s.t \; z_{pl} \leq \tau \\ 0, & \text{otherwise} \end{cases}$$

We here claim that the distance between centroids of different nodes should be larger than a threshold distance if they are not

at the same location. If the distance is below a certain threshold, τ, we declare the presence of a Sybil Attack and deduce that these nodes are indeed a single node, a Sybil attacker. The detection rate, i.e., the percentage of detected attacks under the alternative hypothesis, and the false alarm rate, i.e., the percentage of incorrectly identified nodes under the null hypothesis, vary under different thresholds. The threshold choice is critical to decrease the false alarm rate while obtaining a decent detection performance. A negative standard score means that the relative distance between a pair of nodes is closer than the average pair-wise centroid distance. Thus, it is a possible indication of the Sybil nodes. We choose a negative threshold from our experiments empirically.

V. PERFORMANCE EVALUATION

There are currently two V2X radio access technologies; DSRC [43] and Cellular-V2X (C-V2X) which is defined by the 3GPP as part of its LTE and ongoing 5G standards families. These communications can either depend on 802.11p-based technologies or C-V2V Sidelink interface [44]. We first conduct DSRC-based real-world experiments to evaluate the performance of our approach. Our experiments include two types of scenarios:

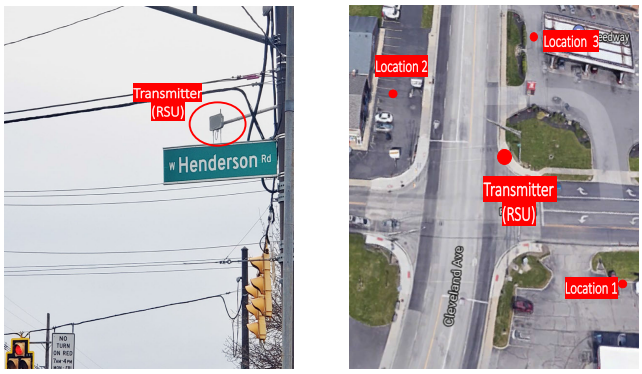
1) Vehicle-to-infrastructure experiments.
2) Vehicle-to-vehicle experiments.

Next, we evaluate the performance of the approach for large-scale situations by using a ray-tracing simulator in an urban environment.

A. VEHICLE-TO-INFRASTRUCTURE EXPERIMENTS

As a part of the Smart Columbus program, approximately 86 intersections have been equipped with RSUs as shown in Fig. 6a, and over 1500 participating private, transit, and emergency vehicles have been equipped with an onboard unit (OBU) in Ohio [6]. The RSUs broadcast signal phase and timing messages 10 times per second on DSRC channel 180 at 5.90 GHz [43], and the SPaT messages are used to provide the current signal/phase timing data for one or more signalized intersections, as well as other time of day status details.

We have collected SPaT messages broadcast from the RSU at three different locations shown in Fig. 6b. In this experiment, we aim to mimic an attack scenario where an RSU collects messages from two legitimate nodes and an attacker node. Since we don't have a chance to obtain signal level data from the RSU, we assume the reciprocity of the wireless channel that holds under sub-6 GHz [45]. We have used X300 of Ettus research with a CBX daughterboard that supports a 10 Mhz bandwidth of DSRC channel 180. We changed the IEEE 802.11p-compliant OFDM receiver implementation of [46] to obtain CFR values of subcarriers and the correct time of each frame reception. Table 1 shows the physical layer parameters of the OFDM receiver. The least-squares channel estimation algorithm is used to estimate the CFR vector using the long training sequence of the received frames.



(a) (b)

FIGURE 6. (a) Intersection with a roadside unit. (b) Messages were collected from three different locations as indicated.

TABLE 1. Physical layer parameters of OFDM receiver.

Parameter	
Bandwidth	10 MHz
OFDM subcarrier	64
Subcarrier spacing	156 kHz
OFDM symbol time	8 μs

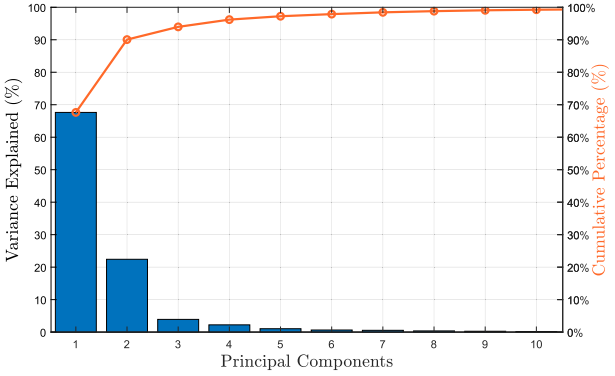
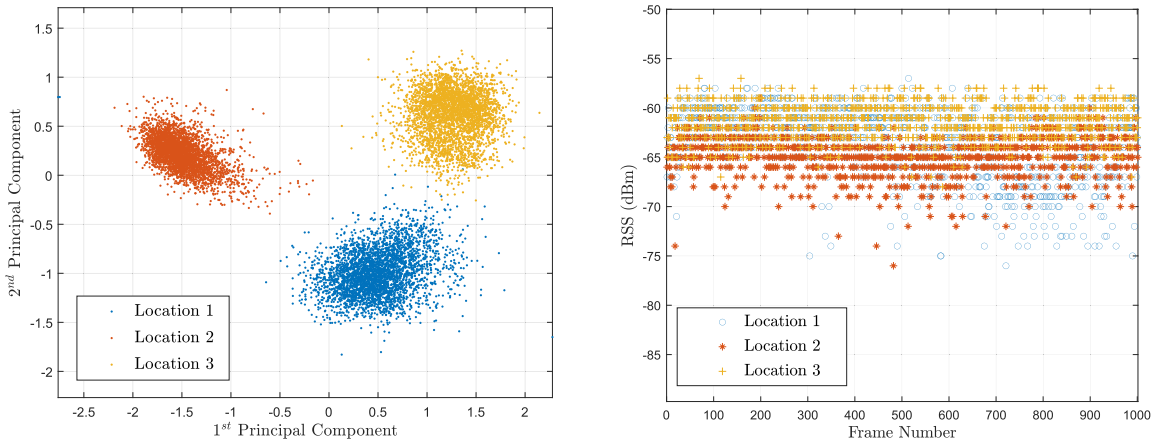


FIGURE 7. Variance explained by each principal component.

The locations are chosen 40 meters away from the RSU. Fig. 8a shows the first and second principal components for three different locations after PCA is performed on the estimated CFRs. We observe the scores are clustered around a point in the PCA space for different locations. This enables us to differentiate different vehicles at different locations and detect Sybil attacks. Fig. 7 shows the variance explained by each principal component and we observe that the first 3 PCs explain 94% of variance as our algorithm uses the first 3 PCs for clustering.

To mimic a Sybil attack, we choose a location as the location of a Sybil attacker, and the Sybil attacker broadcasts its messages with multiple identities from this location. We split the estimated CFR from this location between the multiple identities. The other two locations are treated as legitimate nodes' locations. We run our detection algorithm through all three possible attack locations. Table 2 shows the average



(a) (b)

FIGURE 8. (a) First two principal components of the CFRs estimated at different locations. (b) RSS values of the frames received from different locations.

TABLE 2. Detection and false positive rates under different thresholds for V2I experiments.

Threshold	Detection Rate	False Positive Rate
τ = -0.9	94.44%	0 %
τ = -0.8	96.95%	0.10%
τ = -0.7	98.20%	0.21%
τ = -0.6	98.75%	0.55%
τ = -0.5	99.08%	1.10%

detection rate and false positive rate of our approach with different thresholds. We achieve a detection rate of 99.08%, and a false positive rate of 1.10% with the threshold of -0.5.

1) COMPARISON WITH RSSI

Besides CFRs, we also log the received signal strength (RSS) values of the received frames. The received signal strength is the strength of a received signal measured at the receiver's antenna and quantized to form the received signal strength indicator (RSSI). We can write the received signal as:

$$y(t) = \mathbb{R}\{\Big[\int_{-\infty}^{\infty} h(\tau,t)x(t-\tau)d\tau\Big]e^{j2\pi f_c t}\}$$
$$= \mathbb{R}\{\Big[\int_{-\infty}^{\infty} \sum_i a_i^b(t)g(\tau-\tau_i(t))x(t-\tau)d\tau\Big]e^{j2\pi f_c t}\}$$

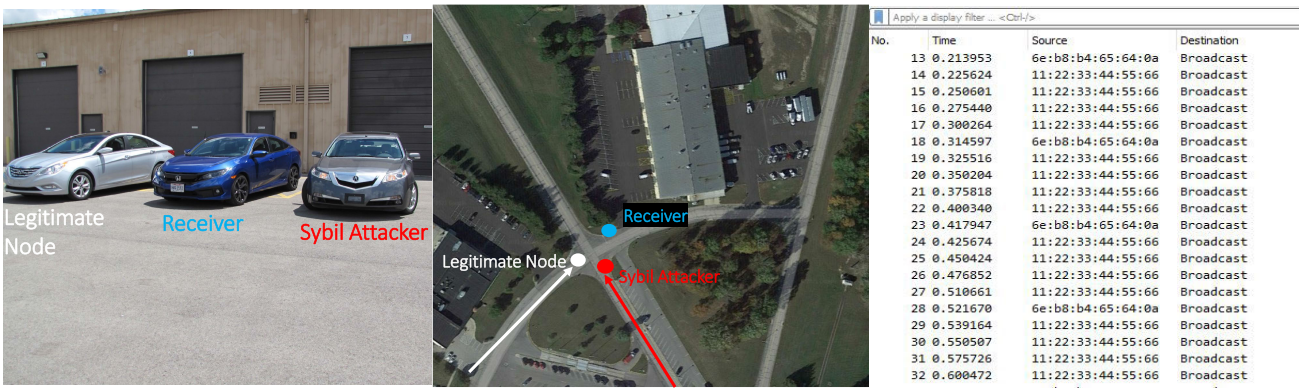
where $x(t)$ is the transmitted signal, and $\mathbb{R}$ denotes the real part. Therefore, RSS at a location is the superposition of multiple path signals through different paths. As seen from the equation above, the fundamental drawback of RSSI is that it cannot capture the multipath effects. Thus, the probability of RSSI coincidences at different locations is relatively large, which makes it difficult for RSSI to distinguish different vehicles. Whereas, in the frequency domain, CFR contains the CSI of multiple subcarriers that are shaped by the constructive and destructive effects of different paths. Therefore, it provides diversity compared to RSSI. Fig.8b shows the RSSI values of the first 1000 frames collected from three different locations. It is observed that the RSSI values of the vehicles are close to each other and RSSI itself is not a good metric to distinguish different vehicles compared to the principal components of the CFRs.

B. VEHICLE-TO-VEHICLE EXPERIMENTS

The V2V experiments have been conducted in the Transportation Research Center (TRC), the largest multi-user automotive test ground in the US. In the experiment, we have equipped a vehicle with our software-defined radio and collected BSMs broadcasted from two vehicles, the legitimate node and the Sybil attacker as shown in Fig. 9a. The legitimate node is equipped with a Denso 5900 series DSRC transceiver. We evaluate a scenario in that three vehicles approach an intersection. The Sybil attacker and the legitimate node approaching from two different directions wait at the locations indicated by the circles in Fig. 9b.

1) SYBIL ATTACKER WITH CONGESTION GENERATION TOOL

The basic safety message conveys critical vehicle state information for supporting V2V safety applications. In a benign network, vehicles are supposed to send BSMs at the rate of 10 messages per second to their neighbors [43]. In our experiment, the Sybil attacker is equipped with the Denso Congestion Generation Tool (CGT) to launch a Sybil attack by sending multiple BSMs in one message period with different identities. The CGT comprises multiple GNSS/DSRC radios that can transmit messages synchronously on channel 172, and it emulates multiple vehicles. It is essentially used to verify the operation of the congestion control mechanisms of radios under repeatable congestion scenarios. The configuration file of CGT set parameters like the number of vehicles to emulate, target channel busy percentage, etc. We only use one of the available DSRC radios to have the vehicle transmit its messages from a single radio in our experiment.



(a) (b) (c)

FIGURE 9. **(a) Vehicles used in the experiments. (b) Locations of the vehicles. (c) The attacker sends 4 BSMs in one BSM period while the legitimate node sends one BSM.**

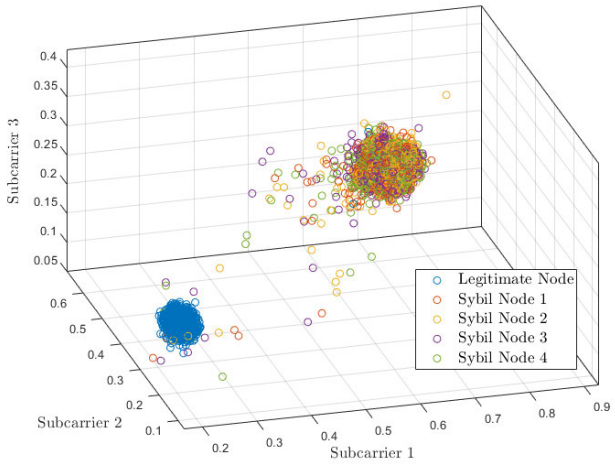


FIGURE 10. **Magnitude response of three subcarriers for unique node identities.**

The receivers collect the BSMs during the wait time. Each BSM includes a unique identity that helps us identify the source of the messages. To launch a Sybil attack, we set the number of vehicles as 4 in CGT, thus the attacker sends 4 BSMs with different identities in one BSM period (100 msec) while the legitimate node sends one BSM in a period as shown in Fig. 9c. We also record and analyze the MAC ID of each successfully decoded frame using Wireshark software [47]. During the wait time, we have collected 5090 BSMs from the Sybil attacker and 1300 BSMs from the legitimate node. Later, the unique identities of the nodes are parsed using a parser tool. Fig. 10 shows the magnitude response of 3 subcarriers out of 52 subcarriers for each unique identity.

We observe that the CFRs of Sybil nodes are clustered around the same point in the signal space although the CFRs of the legitimate node are separated from it. Our detection algorithm is run using the estimated CFRs, and we achieve a detection rate of 98.43%, and a false positive rate of 0.80% with the threshold of -0.5.

C. URBAN SIMULATIONS

In wireless communication, the wireless signals propagate from transmitter to receiver via multiple paths. The whole details of the wireless propagation can be obtained by solving Maxwell's equations with boundary conditions. However, we need to have the physical characteristics of the surfaces to solve these equations. Since it is not possible in most situations, probabilistic channel models are widely used to characterize propagation environments for mathematical convenience while designing the communication systems. However, channel modeling in a complex propagation environment is one of the main challenges for communication system design as highlighted in [48]. The well-known channel models struggle in complex scenarios with many imperfections, although they may capture some features in conventional channels. In this paper, we utilize the ray-tracing approach rather than stochastic channel models. The ray-tracing techniques represent the electromagnetic waves sent from a transmitter as a simple particle that obeys the laws of reflection and refraction, and estimates the multipaths between a receiver and the transmitter. The ray-tracing provides more accurate and spatially consistent results compared to the stochastic models. More details about the principles of ray-tracing can be found in [49]. We use Remcom's Wireless Insite [50] as a ray-tracing simulator to simulate the wireless propagation. With the help of simulations, we evaluate our approach in a complex propagation environment, and also assess the impact of the transmitter power and the distance between attacker and legitimate nodes on our performance.

1) SIMULATION SETUP AND PARAMETER SELECTION

We simulate a V2I scenario in an urban environment with an RSU (receiver) at an intersection communicating with multiple vehicles (transmitters) as shown in Fig. 11. Each red square represents a possible vehicle location equipped with a quarter-wave monopole antenna and the green square represents the RSU equipped with an omnidirectional antenna.

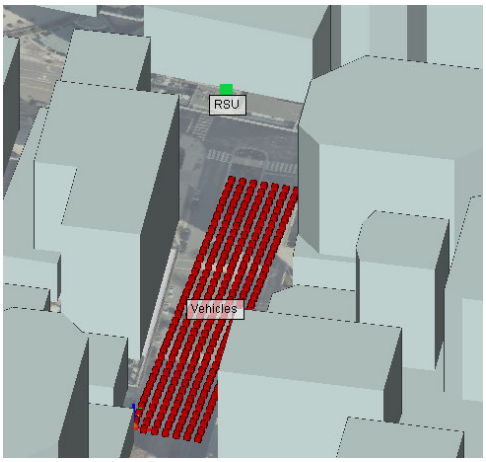


FIGURE 11. Simulation setup of an intersection in an urban environment.

TABLE 3. The ray-tracing simulator parameters.

Wireless Insite parameters	
Propagation Model	X3D
Total Number of Rays	25
Building Material	ITU Layered drywall 5GHz
Terrain Material	Asphalt
Antenna	Omnidirectional
Transmit Power	-6, -3, 0, 3 dBm
Receiver Antenna Height	20 meters
Carrier Frequency	5.89 GHz
Bandwidth	10 MHz

In the simulations, we employ the XY Grid feature of Remcom's Wireless Insite [51] to guarantee consistent spacing between transmitters (2.5 meters) on a predefined grid. These transmitters served as vehicles in the simulations. This functionality was particularly instrumental in assessing the impact of the distance between the attacker and legitimate nodes. The selection of materials for buildings, landscapes is crucial for creating realistic simulations. Wireless Insite offers a database of materials that includes typical structures, landscapes, and some basic material types. The materials are selected based on the guidelines provided by the International Telecommunication Union (ITU) [52]. Table 3 shows the parameters of the ray-tracing simulator. The ray-tracing simulator provides the rays, each corresponding to a propagation path, between a receiver and a transmitter. The simulator also provides the power, delay, and phase information of each ray. The received frequency domain signal of the n^{th} OFDM frame, after removing the cyclic-prefix and applying the inverse discrete Fourier transform, can be written in the vector form as:

$$\mathbf{Y_n} = \mathbf{H_n}\mathbf{X_n} + \mathbf{W_n} \tag{3}$$

where $\mathbf{Y_n} = [Y_{n,1}, Y_{n,2}, \ldots, Y_{n,K}]^T$ and $\mathbf{W_n} = [W_{n,1}, W_{n,2}, \ldots, W_{n,K}]^T$ represent the received signal vector and the noise vector for K number of subcarriers of n^{th} frame, respectively. $\mathbf{X_n}$ is a $K \times K$ diagonal matrix with $\langle k, k \rangle$th element given as $X_{n,k}$ where $X_{n,k}$ is a frequency domain transmitted signal. $W_{n,k}$ is assumed to be an additive Gaussian noise per subcarrier, with zero mean and variance σ_W^2. In the case of least-square (LS) estimation, the channel estimate can be calculated as:

$$\begin{aligned}\hat{\mathbf{H}}_n &= \underset{\mathbf{H_n}}{\operatorname{argmin}} ||\mathbf{Y_n} - \mathbf{X_n}\mathbf{H_n}||^2 \\ &= (\mathbf{X_n}^H\mathbf{X_n})^{-1}\mathbf{X_n}^H\mathbf{Y_n} = \mathbf{X_n}^{-1}\mathbf{Y_n} \\ &= \left[\frac{Y_{n,0}}{X_{n,0}}, \frac{Y_{n,1}}{X_{n,1}}, \ldots, \frac{Y_{n,K}}{X_{n,K}}\right]\end{aligned} \tag{4}$$

Substituting Eq. 3 into Eq. 4, the LS channel estimate can be expressed as:

$$\hat{\mathbf{H}}_n = \mathbf{H_n} + \underbrace{\mathbf{X_n}^{-1}\mathbf{W_n}}_{\mathcal{E}\text{=Estimation Error}}$$

$\mathbf{X_n}^{-1}\mathbf{W_n}$ denotes the estimation error. Since $E[\hat{\mathbf{H}}_n] = E[\mathbf{H_n}] + \mathbf{X_n}^{-1}E[\mathbf{W_n}] = E[\mathbf{H_n}]$ forms an unbiased estimator of $\mathbf{H_n}$, the covariance matrix of the estimation error can be calculated as:

$$\begin{aligned}E[\mathcal{E}\mathcal{E}^H] &= E[\mathbf{X_n}^{-1}\mathbf{W_n}\mathbf{W_n}^H(\mathbf{X_n}^{-1})^H] \\ &= \mathbf{X_n}^{-1}E[\mathbf{W_n}\mathbf{W_n}^H](\mathbf{X_n}^{-1})^H] \\ &= \sigma_W^2(\mathbf{X_n}^H\mathbf{X_n})^{-1} \\ &= \frac{\sigma_W^2}{P_T}\mathbf{I}_N = \sigma_{\mathcal{E}}^2\mathbf{I}_N\end{aligned} \tag{5}$$

where P_T is equal to the transmitter power per subcarrier (i.e., $(\mathbf{X_n}^H\mathbf{X_n}) = P_T\mathbf{I_K}$, where $\mathbf{I}$ is an identity matrix of size K) and $\sigma_{\mathcal{E}}^2$ is the error variance per subcarrier. Note that the error variance per subcarrier depends on the noise variance, σ_W^2, and the transmitter power per subcarrier. It can be calculated as:

$$\sigma_{\mathcal{E}}^2 = \frac{\sigma_W^2}{P_T} = \frac{kTN_FB}{P_T} \tag{6}$$

where $k = 1.38 \times 10^{-23}J/K$ is the Boltzmann's constant, T is the temperature in Kelvin (K), N_F is the receiver noise figure, and B is the bandwidth of a subcarrier.

In the simulations, each location is separated by 3 meters on a grid of 105m x 18m and there is a total of 252 vehicle locations. From the ray-tracing simulator, we obtain a total of 25 propagation paths between the RSU and each vehicle location. Using the power, delay, and phase information of each path provided by the ray-tracing simulator, we calculate the CFR values using Eq. 1 for 64 subcarriers, and we model the estimation error per subcarrier distributed as $\mathcal{CN}(0, \sigma_{\mathcal{E}}^2)$ where $\sigma_{\mathcal{E}}^2$ is calculated according to Eq. 6 with $T = 330$, $N_F = 2$, $B = 156.25$ kHz and $P_T = 0$ dBm. We generate 1000 CFR values for each vehicle location.

We evaluate the performance of our approach by treating one location as a source of the Sybil attack and another location as a legitimate node. To this end, we choose a vehicle location as the location of the Sybil attacker, and divide its frames between multiple identities (nodes) while the other

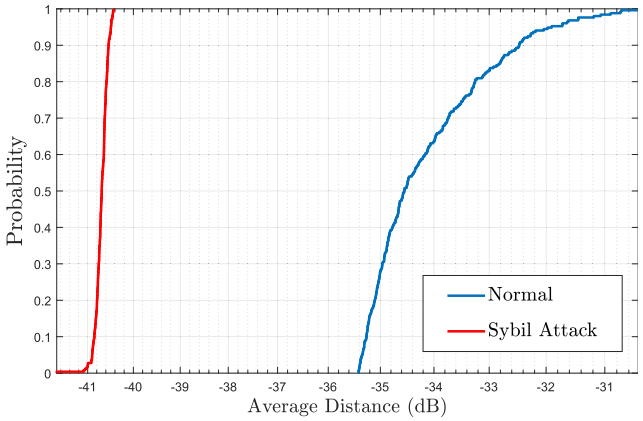


FIGURE 12. Cumulative distribution function of average cluster distances.

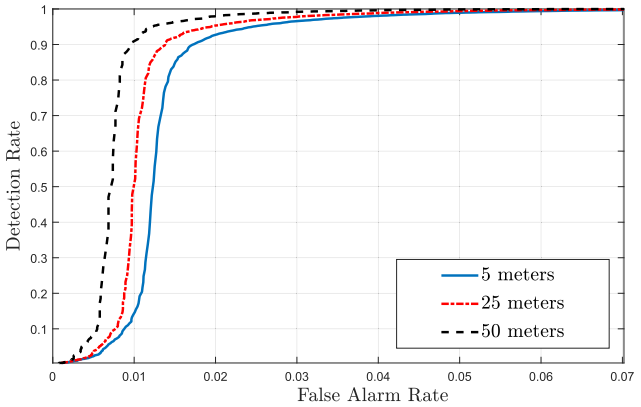


FIGURE 13. Receiver operating characteristic curves for different minimum distances.

location is considered as a legitimate node with a single identity. We run our algorithm through all the possible combinations of location pairs in the network while multiple nodes are presented at the chosen location of the Sybil attacker. The procedure is repeated by treating each location as the location of the Sybil attacker. We calculate the average cluster distance between the Sybil nodes and the legitimate nodes by averaging over location pairs. Fig. 12 shows the cumulative distribution function of the average cluster distances for the Sybil attack scenario where the frames come from the same location, and the normal scenario where the frames come from two different locations. We observe that the curves are well separated from each other, and the average distance in the Sybil attack scenario is much lower than in the normal scenario. Thus, the cluster distance of the nodes is an effective metric for detecting Sybil attacks.

2) IMPACT OF THE DISTANCE BETWEEN THE ATTACKER AND LEGITIMATE NODES

We investigate the impact of the physical distance between the two vehicles on the detection and the false positive rate. We observe that the CFR structure of a certain location is different from the CFRs of the other locations with a high probability. However, the CFR structures become similar to each other when the two vehicles physically get closer to each other. Therefore, the resulting cluster distance between two vehicles gets smaller. When it is less than the threshold value, we declare a Sybil attack even though two vehicles are legitimate nodes. Thus, it increases the false positive rate.

We measure the performance of our approach for a certain minimum distance between the vehicles. We calculate the detection rate and false positive rates for the locations that are separated by more than the defined minimum distance in the physical space. We evaluate with the minimum distances of 5 meters, 25 meters, and 50 meters. Fig. 13 shows the receiver operating characteristic curves for three distances by plotting the detection rate against the false positive rate under various thresholds. We observe that the curve shifts to the right when decreasing the minimum distance. This confirms our observation that the spatial similarity of the CFRs is increasing as the distance between the attacker and legitimate nodes is decreasing, and therefore, the false positive rate increases when the two vehicles get closer to each other. Specifically, the false positive rate almost doubled, from 2.79% to 5.11%, with the detection rate of 99% from the minimum distance of 5 meters to 50 meters.

3) IMPACT OF THE TRANSMITTER POWER

We investigate how the transmitter power per subcarrier affects the CFR values, and therefore the performance of our system. The transmitter power affects the resulting CFR in two ways. First, an attacker can try to deceive our attack detection system by changing the power of subcarriers instead of keeping the power level fixed for adjacent packet transmissions. This will result in different CFRs of Sybil nodes. To address this problem, we normalize the CFRs with their power, and the normalized CFRs are invariant to any power level changes. Second, the channel estimation error per subcarrier is inversely proportional to the transmitter power as shown in Eq. 6. Hence, the channel estimation improves with an increase in transmitter power. We analyze the effect of this by varying the transmitter power per subcarrier between -6 dBm to 3 dBm. Again, we calculate the detection rate and false positive rate for different threshold values. Fig. 14 shows the receiver operating characteristic curves for 4 different power levels (-6 dBm, -3 dBm, 0 dBm, 3 dBm). We observe that we achieve an almost perfect detection with 3 dBm transmitter power and the performance degrades when we decrease the transmitter power. Table 4 shows the detection rate and the false positive rate with a threshold of -0.5.

We observe that there is a significant increase in the false alarm rate while the transmit power decreases. Most notably, the false alarm rate jumps to 10.96% with -6 dBm transmit power. It is expected because the increase in the estimation error variance leads to a bigger variation around the centroids of the clusters in the signal space. Therefore, the separation between the centroids of legitimate nodes decreases and this results in a higher false alarm rate. Also, we observe that the

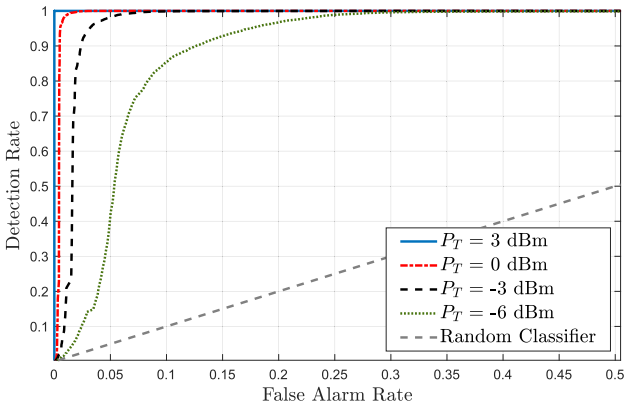


FIGURE 14. Receiver operating characteristic curves for different transmit power levels.

TABLE 4. Detection and false positive rates with a threshold of −0.5 for different transmit power levels.

Transmit Power	Detection Rate	False Positive Rate
P_T = 3 dBm	99.86%	0.01%
P_T = 0 dBm	98.36%	0.91%
P_T = -3 dBm	95.50%	3.28%
P_T = -6 dBm	87.54%	10.96%

detection rate decreases with the transmit power since the higher variation increases the probability of having two Sybil nodes separated from each other in the signal space.

VI. CONCLUSION

This paper presents a novel approach for detecting Sybil attacks in vehicular ad hoc networks, relying on the CFR of vehicles obtained at a receiver. Through extensive DSRC-based experiments and simulations, our approach achieves a remarkable detection rate of 99.08% (with a false positive rate of 1.1%) in the V2I setting, and 98.43% (with a false positive rate of 0.8%) in the V2V setting.

Moreover, we conducted in-depth investigations into the impact of various factors, such as the distance between the attacker and legitimate nodes, as well as the transmitter power, within a complex urban environment using ray-tracing simulations. Our comprehensive analysis shows that our approach exhibits superior performance with increasing node distance and higher transmitter power.

Additionally, its seamless integration into existing network architectures, without requiring additional hardware, makes it a highly cost-effective solution for enhancing the security of vehicular networks against Sybil attacks. As a valuable area for further investigation, future research could explore the impact of even more dynamic scenarios, such as high-speed mobility patterns or rapidly changing network topologies.

ACKNOWLEDGMENT

The authors express their gratitude to Dick Hoover and Guogang Xu for their valuable assistance in conducting the experiments at the Transportation Research Center (TRC) facility. Furthermore, they extend their appreciation to Frank Barickman (NHTSA), John Martin (NHTSA), Sughosh Rao (TRC), Gavin Howe (TRC), and Keith A. Redmill (OSU) for their insightful discussions during their meetings, which greatly contributed to the development of their research.

REFERENCES

[1] R. Lu, L. Zhang, J. Ni, and Y. Fang, "5G vehicle-to-everything services: Gearing up for security and privacy," *Proc. IEEE*, vol. 108, no. 2, pp. 373–389, Feb. 2020.

[2] A. K. Malhi, S. Batra, and H. S. Pannu, "Security of vehicular ad-hoc networks: A comprehensive survey," *Comput. Secur.*, vol. 89, Sep. 2020, Art. no. 101664.

[3] F. Sakiz and S. Sen, "A survey of attacks and detection mechanisms on intelligent transportation systems: VANETs and IoV," *Ad Hoc Netw.*, vol. 61, pp. 33–50, Jun. 2017.

[4] F. Boeira, M. P. Barcellos, E. P. de Freitas, A. Vinel, and M. Asplund, "On the impact of Sybil attacks in cooperative driving scenarios," in *Proc. IFIP Netw. Conf. Workshops*, Jun. 2017, pp. 1–2.

[5] M. Asghar, L. Pan, and R. Doss, "An efficient voting based decentralized revocation protocol for vehicular ad hoc networks," *Digit. Commun. Netw.*, vol. 6, no. 4, pp. 422–432, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352864819300161

[6] *Final Report for the Smart Columbus Demonstration Program*. Accessed: Apr. 3, 2024. [Online]. Available: https://smart.columbus.gov/programs/smart-city-demonstration

[7] J. R. Douceur, "The Sybil attack," in *Proc. Int. Workshop Peer-Peer Syst.* Cham, Switzerland: Springer, 2002, pp. 251–260.

[8] J. Newsome, E. Shi, D. Song, and A. Perrig, "The Sybil attack in sensor networks: Analysis & defenses," in *Proc. 3rd Int. Symp. Inf. Process. Sensor Netw.*, Apr. 2004, pp. 259–268.

[9] F. Li, P. Mittal, M. Caesar, and N. Borisov, "SybilControl: Practical Sybil defense with computational puzzles," in *Proc. 7th ACM Workshop Scalable Trusted Comput.*, 2012, pp. 67–78.

[10] S. Kim, J. Kim, J. H. Cheon, and S.-H. Ju, "Threshold signature schemes for ElGamal variants," *Comput. Standards Interfaces*, vol. 33, no. 4, pp. 432–437, 2011.

[11] G. Raj, "Detection of Sybil attack in VANET," *Karpagam JCS*, vol. 14, no. 2, pp. 397–408, 2020.

[12] S. Hamdan, A. Hudaib, and A. Awajan, "Detecting Sybil attacks in vehicular ad hoc networks," *Int. J. Parallel, Emergent Distrib. Syst.*, vol. 36, no. 2, pp. 69–79, Mar. 2021.

[13] *IEEE Standard for Wireless Access in Vehicular Environments—Security Services for Applications and Management Messages*, IEEE Standard 1609.2-2016, 2016, pp. 1–240.

[14] *Intelligent Transport Systems (ITS); Security; ITS Communications Security Architecture and Security Management*, Standard ETSI TS 102 940, 2012.

[15] *Intelligent Transport Systems (ITS); Security; Trust and Privacy Management*, Standard ETSI TS 102 941, 2018.

[16] P. Cincilla, O. Hicham, and B. Charles, "Vehicular PKI scalability-consistency trade-offs in large scale distributed scenarios," in *Proc. IEEE Veh. Netw. Conf. (VNC)*, Dec. 2016, pp. 1–8.

[17] S. Ibrahim and M. Hamdy, "A comparison on VANET authentication schemes: Public key vs. symmetric key," in *Proc. 10th Int. Conf. Comput. Eng. Syst. (ICCES)*, Dec. 2015, pp. 341–345.

[18] R. Hussain, J. Lee, and S. Zeadally, "Trust in VANET: A survey of current solutions and future research opportunities," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 5, pp. 2553–2571, May 2021.

[19] S. Gyawali, Y. Qian, and R. Q. Hu, "Machine learning and reputation based misbehavior detection in vehicular communication networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 8, pp. 8871–8885, Aug. 2020.

[20] B. Xiao, B. Yu, and C. Gao, "Detection and localization of Sybil nodes in VANETs," in *Proc. Workshop Dependability Issues Wireless Ad Hoc Netw. Sensor Netw.*, Sep. 2006, pp. 1–8.

[21] M. Kabbur and V. A. Kumar, "MAR_Sybil: Cooperative RSU based detection and prevention of Sybil attacks in routing process of VANET," *J. Phys., Conf. Ser.*, vol. 1427, no. 1, Jan. 2020, Art. no. 012009.

[22] M. S. Bouassida, G. Guette, M. Shawky, and B. Ducourthial, "Sybil nodes detection based on received signal strength variations within VANET," *IJ Netw. Secur.*, vol. 9, no. 1, pp. 22–33, 2009.

[23] R. Shrestha, S. Djuraev, and S. Y. Nam, "Sybil attack detection in vehicular network based on received signal strength," in *Proc. Int. Conf. Connected Veh. Expo (ICCVE)*, 2014, pp. 745–746.
[24] Y. Yao et al., "Multi-channel based Sybil attack detection in vehicular ad hoc networks using RSSI," *IEEE Trans. Mobile Comput.*, vol. 18, no. 2, pp. 362–375, Feb. 2019.
[25] M. T. Garip, P. H. Kim, P. Reiher, and M. Gerla, "INTERLOC: An interference-aware RSSI-based localization and Sybil attack detection mechanism for vehicular ad hoc networks," in *Proc. 14th IEEE Annu. Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2017, pp. 1–6.
[26] Y. Chen, J. Yang, W. Trappe, and R. P. Martin, "Detecting and localizing identity-based attacks in wireless and sensor networks," *IEEE Trans. Veh. Technol.*, vol. 59, no. 5, pp. 2418–2434, Jun. 2010.
[27] A. Abdelaziz, R. Burton, F. Barickman, J. Martin, J. Weston, and C. E. Koksal, "Enhanced authentication based on angle of signal arrivals," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 4602–4614, May 2019.
[28] J. Grover, M. S. Gaur, and V. Laxmi, "Multivariate verification for Sybil attack detection in VANET," *Open Comput. Sci.*, vol. 5, no. 1, pp. 60–78, Dec. 2015.
[29] C. Wang et al., "Accurate Sybil attack detection based on fine-grained physical channel information," *Sensors*, vol. 18, no. 3, p. 878, Mar. 2018.
[30] S. Wielandt, J.-P. Goemaere, L. De Strycker, and B. Nauwelaers, "Performance simulations of a 2.4 GHz indoor angle of arrival system for multipath components," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Oct. 2015, pp. 1–8.
[31] Z. Yang, Z. Zhou, and Y. Liu, "From RSSI to CSI: Indoor localization via channel response," *ACM Comput. Surv.*, vol. 46, no. 2, pp. 1–32, Dec. 2013.
[32] L. Xiao, L. J. Greenstein, N. B. Mandayam, and W. Trappe, "Channel-based detection of Sybil attacks in wireless networks," *IEEE Trans. Inf. Forensics Security*, vol. 4, pp. 492–503, 2009.
[33] H. B. Tulay and C. E. Koksal, "Robust Sybil attack detection in vehicular networks," in *Proc. IEEE 94th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2021, pp. 1–7.
[34] *On-board System Requirements for V2V Safety Communications*, Standard J2945/1, 2020. [Online]. Available: https://saemobilus.sae.org/content/j2945/1_201603
[35] *Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Part 2: Specification of Cooperative Awareness Basic Service*, Standard 302 637-2 v1.3.1, ETSI, 2014.
[36] Y. Liu, Z. Tan, H. Hu, L. J. Cimini, and G. Y. Li, "Channel estimation for OFDM," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 4, pp. 1891–1908, 4th Quart., 2014.
[37] K. Ansari, "Joint use of DSRC and C-V2X for V2X communications in the 5.9 GHz ITS band," *IET Intell. Transp. Syst.*, vol. 15, no. 2, pp. 213–224, Feb. 2021.
[38] S. Zhou, X. Zhang, Y. Zhao, T. Zhao, and T. O. Korhonen, "Fast and accurate velocity estimation for OFDM systems based on channel frequency response," in *Proc. IEEE 73rd Veh. Technol. Conf. (VTC Spring)*, May 2011, pp. 1–5.
[39] I. T. Jolliffe, *Principal Component Analysis* (Springer Series in Statistics), 2nd ed. New York, NY, USA: Springer, 2002, pp. 10–27.
[40] O. Bachem, M. Lucic, S. H. Hassani, and A. Krause, "Approximate k-means++ in sublinear time," in *Proc. 30th AAAI Conf. Artif. Intell.*, 2016, pp. 1459–1467.
[41] S. Chawla and A. Gionis, "K-means–: A unified approach to clustering and outlier detection," in *Proc. SIAM Int. Conf. Data Mining*, May 2013, pp. 189–197.
[42] D. Arthur and S. Vassilvitskii, "k-means++: The advantages of careful seeding," in *Proc. 18th Annu. ACM-SIAM Symp. Discrete Algorithms*, 2007, pp. 1027–1035.
[43] J. B. Kenney, "Dedicated short-range communications (DSRC) standards in the United States," *Proc. IEEE*, vol. 99, no. 7, pp. 1162–1182, Jul. 2011.
[44] Z. MacHardy, A. Khan, K. Obana, and S. Iwashina, "V2X access technologies: Regulation, research, and remaining challenges," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 1858–1877, 3rd Quart., 2018.
[45] S. Haile, "Investigation of channel reciprocity for OFDM TDD systems," M.S. thesis, Dept. Elect. Comput. Eng., Univ. Waterloo, Waterloo, ON, Canada, 2009.
[46] B. Bloessl, M. Segata, C. Sommer, and F. Dressler, "An IEEE 802.11 A/G/P OFDM receiver for GNU radio," in *Proc. 2nd Workshop Softw. Radio Implement. Forum*, 2013, pp. 9–16.
[47] *Wireshark Website*. Accessed: Apr. 3, 2024. [Online]. Available: https://www.wireshark.org/
[48] T. Wang, C.-K. Wen, H. Wang, F. Gao, T. Jiang, and S. Jin, "Deep learning for wireless physical layer: Opportunities and challenges," *China Commun.*, vol. 14, no. 11, pp. 92–111, Nov. 2017.
[49] Z. Yun and M. F. Iskander, "Ray tracing for radio propagation modeling: Principles and applications," *IEEE Access*, vol. 3, pp. 1089–1100, 2015.
[50] Remcom. *Wireless Insite 3D Wireless Prediction Software*. Accessed: Apr. 3, 2024. [Online]. Available: https://www.remcom.com/wireless-insite-em-propagation-software
[51] *Wireless Insite 3D Wireless Propagation and Simulation Software: Version 3.3.5*, Remcom, State College, PA, USA, 2020.
[52] H. B. Tulay and C. E. Koksal, "Road state inference via channel state information," *IEEE Trans. Veh. Technol.*, vol. 72, no. 7, pp. 8329–8341, Jun. 2023.

HALIT BUGRA TULAY (Member, IEEE) received the B.S. degree in electrical and electronics engineering from Hacettepe University, Turkey, in 2016. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, The Ohio State University. His research interests include wireless communication, cybersecurity, and machine learning.

CAN EMRE KOKSAL (Senior Member, IEEE) received the M.S. and Ph.D. degrees in electrical engineering and computer science from MIT in 1998 and 2002, respectively. Since 2006, he has been a Professor with the Department of Electrical and Computer Engineering, The Ohio State University. His research interests include wireless communication, cybersecurity, information theory, and stochastic processes. He served as an Associate Editor for IEEE TRANSACTIONS ON INFORMATION THEORY, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and *Computer Networks*.