A Beamforming Signal-Based Verification Scheme for Data Sharing in 5G Vehicular Networks

VAN LINH NGUYEN1,2, (Member, IEEE), PO-CHING LIN1, (Member, IEEE), AND REN-HUNG HWANG1, (Senior Member, IEEE)

1Department of Computer Science and Information Engineering, National Chung Cheng University, Chiayi 621301, Taiwan, R.O.C.
2Department of Information Technology, TNU-University of Information and Communication Technology, Thai Nguyen 25000, Vietnam

Corresponding author: Van Linh Nguyen (nvlinh@cs.ccu.edu.tw)

This work was supported in part by the Ministry of Science and Technology of Taiwan, R.O.C., under Grant MOST 109-2811-E-194-501, Grant MOST 108-2221-E-194-022-MY3, and Grant MOST 108-2221-E-194-019-MY3; and in part by the Advanced Institute of Manufacturing with High-Tech Innovations (AIM-HI) through the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan.

ABSTRACT Vehicle-to-Everything (V2X) communications are vital for autonomous vehicles to share sensing data about the surrounding environment, particularly in non-line-of-sight (NLOS) areas where the camera and radar systems often perform poorly. However, an insider adversary such as a compromised vehicle can disseminate false sensing data that even a signature scheme cannot counter. Trusting such shared data, the surrounding vehicles may be trapped to react unexpectedly and potentially poses the risk of a fatal crash. In this work, we introduce a prospective cooperative verification scheme to support both the host vehicles and V2X edge applications in validating the truthfulness of sharing data in the fifth-generation (5G) vehicular networks. First, the detection systems at the host vehicle (local detector) and the road-side unit (RSU) (global detector) separately recreate a trajectory of a target vehicle by extracting its status and attributes from the received Cooperative Awareness Messages (CAM). Simultaneously, they also build another trajectory of the vehicle by independently fusing real-time measurement metrics from signal-based positioning. We then perform a Student’s t-test to detect any significant differences between the extracted trajectory and the corresponding measured one. Finally, the quantified evidence from the local and global detector tests will be fused through the Dempster-Shafer fusion for the final decision, i.e., whether the target vehicle is trustful. Besides the theoretical analysis of basic limits, we perform extensive evaluations of the work in cases of both sparse and heavy traffic densities. Through the simulation, this work demonstrates its significant effect in terms of detection performance and response time, particularly for detecting Sybil and false data attacks quickly.

INDEX TERMS 5G vehicular security, V2X data verification, vehicle misbehavior detection.

I. INTRODUCTION

Recently, autonomous driving has obtained many achievements and now progresses towards Level 5 (i.e., fully autonomous) of the SAE J3016 standard [1]. To enhance autonomy, particularly when a car’s radar vision is obscured, an emerging solution is to use vehicle-to-everything (V2X) communication for sharing information with nearby vehicles and road-side units (RSUs). This technology aims to expand the awareness of a vehicle about surrounding situations, particularly the zones with limited visibility, e.g., street intersections. Unfortunately, such sharing poses new risks since the receivers cannot assure whether the received data are truthful. A malicious or compromised vehicle may transmit false data and then fool the receivers, thereby potentially causing safety hazards and severe accidents. Such attacks are common and well acknowledged in the literature [2], including in V2X networks [3]. Since the host vehicle’s decision-making system may rely much on the shared data, particularly in non-line-of-sight (NLOS) areas and harsh environments, malicious vehicles must be detected and ejected from the network. Also, the high-definition (HD)
map building process in self-driving vehicles often relies on the shared information [2]; thus, shielding the receivers against an unreliable data source is worthwhile.

To tackle the security attacks, there are two primary approaches in the literature. The most common one is to use a data-centric misbehavior detection engine [4], which focuses on checking the consistency of the received data such as the relations between the received packets, the correctness of data format, or plausibility movement/average speed of the targets, to judge whether the sender node is honest. For example, the MARV-X framework [5] proposed a Kalman-filter-based method to predict the maneuver information of the vehicles and compare them with the received data from Car-to-X (C2X) communication. The authors of [6], [7] introduced a method to detect vehicle anomaly by verifying the consistency of the report data frequency, whether they are in line with protocol specifications. Similarly, the authors propose an internal check from the host vehicle’s multiple sensors [8] or machine learning-based models to identify outliers in the traffic density information [9] to verify the integrity of the received data. However, a major drawback of this approach is that poor modeling of underlying data can significantly reduce system performance. Moreover, this method is also highly susceptible to be misled by sophisticated attacks where several colluding attackers surround a victim consistently report wrong information. The second approach is node-centric [4], which typically collects knowledge from multiple participants to figure out whether the target is misbehaving. For example, the authors of [10] proposed to collect the radar checks of neighboring vehicles to identify an honest vehicle. Another common type of this approach is the trust-based detection [11]–[15], which includes reputation mechanisms to vote on the correctness of the information. Unfortunately, the challenge of this method is to detect Sybil attacks and respond quickly. In Sybil attacks, an attacker can create multiple identities (e.g., using pseudonyms) and then abuse the reputation mechanism to broadcast its bias assessment to the network. Consequently, if the receivers trust this baseless assessment, the legitimate nodes can be excluded from the network. Similarly, if a vehicle suddenly starts abnormal (e.g., due to damaged sensors), the trust-based mechanisms require substantial time to detect (e.g., up to 20-30 seconds [16]).

Requesting an independent measurement metric, particularly in the physical layer [17], is another emerging approach in detecting misbehaving vehicles and false sharing data. For example, the authors of [18] introduced an enhanced verification relying on the angle of signal arrival (AoA). However, the approach can face high inaccuracy for position offset attacks. Close to our approach, in [19], the authors presented a sufficient forged location reports verification with the help of a single physical signal-based measurement. In this work, we further consider multiple sensing sources and propose a cooperative judicious decision-level fusion mechanism. In [20], the authors proposed a plausibility check on the frequency of messages received at the host vehicle and the estimated value by Doppler radar and cameras. However, their approach did not tackle Sybil attacks. To cope with Sybil attacks only, the authors propose to use power control identification information [21]/received signal strength indication (RSSI) [22] and machine learning-based classifier (e.g., support vector machine) for detection.

Unlike previous studies, we propose a cooperative signal-based verification scheme to enhance the reliability of data sharing in vehicular networks. Determining the reliability of a target vehicle now relies not only on the assessment of the host vehicle (local detector) but also on the likelihood of the final fusion from a collation of the detectors, including the one on the nearest RSU (global detector). Notably, the detection engines are capable of signal-based verification and respond quickly to the attacks. In this core function, the values of the target vehicle’s six-dimensional state (longitude, latitude, velocity, heading, acceleration, and yaw rate) extracted from the Cooperative Awareness Messages (CAM) (defined in the ETSI standard [23]) are cross-checked by the Student’s t-test with the corresponding values estimated by an unscented Kalman filter (UKF) in the time series. The UKF-based trajectory estimation’s reliability is significantly boosted with the measurements from independent data sources, e.g., 5G signal-based positioning and Doppler shift signals. At the global detector, the detection engine performs a fusion on the verification results of the UKF-based estimation and a multi-opinion voting module. The voting works based on nearby vehicles’ signal-based detection results about the target vehicle with a majority’s consensus. The separate detection results at the local and the global detectors are then fused at the host vehicle based on the Dempster-Shafer theory to determine whether the target vehicle is trustful finally. If the fused result’s degree of belief is low, e.g., less than 0.5, the host vehicle will mark the target vehicle as an attacker. Such violation data can be reported to the misbehaving authority in the V2X identity management infrastructure, e.g., Security Credential Management System (SCMS) [24], for further punishment. The workflow of our proposal is illustrated in Fig. 1.
Our main contributions are summarized as follows:

- We propose a prospective cooperative framework for verifying V2X sharing data in 5G networks. Specifically, our method maintains a six-dimensional state of the target vehicle with the independent measurements from a joint localization and communication model, i.e., 5G signal-based positioning, to determine whether the vehicle is moving as it claims in the CAM. Notably, the signal-based sources are available, cheap, and easily collectible from 5G V2X communications. As a result, our approach can work with affordable vehicles that are seldom fully equipped with expensive facilities such as LiDAR.

- We provide theoretical analysis on basic limits and performance evaluation to verify the proposed method in various traffic scenarios, including the traffic in the real environments and under noise/fading interference conditions. The evaluation results show the significant effects of the method in terms of detection performance, particularly for detecting Sybil [16], [25] and false data attacks in promising time.

- We provide a comprehensive analysis of the system performance for massive verification in large traffic densities and the proposed design’s computational complexity. Finally, security analysis and feature comparison of the proposal compared with some benchmark literature are also revealed.

The remainder of this article is organized as follows. Section II clarifies the attack model and the system model, while our verification method in detail is presented in Section III. The experimental setup and evaluation results are presented in Section IV. Finally, we conclude this work and brief the remaining challenges in Section V. The notations used in this work are shown in Table 1.

### TABLE 1. Notations used in this work.

| Symbol | Definition |
|--------|------------|
| $\|\cdot\|_2$ | The Euclidean norm of a vector |
| $\|\cdot\|_F$ | The Frobenius norm of a vector |
| $^T$ | The transpose of a vector or matrix |
| $\mathcal{CN}(\cdot)$ | A circularly symmetric complex Gaussian distribution |
| $I_n$ | Identity matrix of size $n$ |
| $E[\cdot]$ | Expectation operator |
| $N_{R}, N_T$ | Number of antenna elements of the host vehicle (Rx) and the target vehicle (Tx) OBU |
| $K_{R}, K_{T}$ | Number of receiving and transmitting antenna arrays |
| $\rho_{R}, \rho_{T}$ | The positions of the Rx and the Tx |
| $N_s$ | Number of sub-carriers |
| $L$ | Number of propagation paths |
| $X[\cdot]$ | The baseband channel for transmitting signals at the Tx |
| $Y[\cdot]$ | The baseband channel for received signals at the Rx |
| $H[\cdot]$ | The channel coefficients matrix |
| $s_k$ | State vector of the target vehicle at the time $k$ |
| $z_k$ | The measurement vector at the current state |
| $s_k$ | The mean of $s_k$ |
| $\delta_k(1:k)$ | The mean of $s_k$ given $s_{k-1}$ |
| $f(\cdot)$ | The non-linear function describing the process |
| $h(\cdot)$ | The non-linear function describing the measurement |
| $\eta_k$ | The process noise |
| $\sigma_k$ | The measurement noise |
| $m$ | Number of dimensions in the state vector $s$ |
| $\alpha$ | Significance level |

### II. ATTACK SCENARIO AND SYSTEM MODEL

This work aims to exploit physical signal sources to filter V2X falsification data in 5G networks and then verify whether the senders are honest. Due to the deep relevance to physical signal processing, in this section, we present the details of communication environment assumption, antenna configuration of the vehicle on-board units, communication channel, and the geometry model of the vehicles’ relative locations. In the next section, we then describe how the verification mechanism works based on the information extracted from the channel modeled in this section.

#### A. ATTACK MODEL

This work primarily targets to tackle the insider adversaries who gain full control or compromise one or more vehicles to disseminate false data. Such false data can vary from false GPS location/velocity/acceleration to inaccurate measurements from damaged high-end sensors. The attackers can modify the outgoing CAM by using the legitimate pseudonym certificates to access the vehicle’s on-board unit (OBU) and manipulate the payload. Note that the attack messages will easily pass the signature verification since they are signed with legitimate credentials (from a compromised/stolen vehicle, for example). We consider the following primary attack types:

1) **Lone wolf attack**: An attacker intentionally manipulates the positioning information in the CAM and disseminate them to the surrounding receivers. For instance, the attack vehicle may report moving at 90 km/s, but it stops at a roadside. There are many variants of this type of attack, e.g., broadcasting false position/speed offset values, eventual stop, stale messages, and arbitrary location appearance [4]. Without loss of generality, we assume the attack vehicle can send fake CAM with wrong information of any of the six-dimensional states $(x, y, v, a, h, \omega)$, where $(x, y, v, a, h, \omega)$ denote its coordinates (longitude and latitude), velocity, heading/orientation, acceleration, and turning rate.

2) **Sybil attack**: An attack vehicle uses multiple compromised identities or pseudonyms to broadcast false data. As a result, the Rx may see multiple virtual vehicles moving, but in fact, they are all from the same vehicle [16], [25]. This attack is well known in the literature, particularly in the networks that demand high privacy - commonly known as **Ghost attack** [24], [26]. In our simulation, the attack vehicle generates multiple CAM with different credentials but uses an OBU only.

3) **Collusion attack**: An attacker coaches the vehicles moving near the host vehicle to report false data together. Unlike the above two types, detecting this attack remains a challenge for most data-centric and node-centric detection approaches [4].

Fig. 2 illustrates a few attack cases and risky situations of both sparse and heavy traffic densities. In the first case, the attacker (Tx1) can quickly force the receiver (Rx1) that
has changed lane to switch back behind the truck with the sudden appearance of a fake vehicle (marker 1). This case can cause a potential rear-end collision with the vehicle behind. In the case of heavy traffic density, the attacker (Tx2) can create a tribe of vehicles ahead (marker 2) to fool the vehicles behind with an unreal traffic jam. The attacker can also fake a sudden stop (marker 3), and an accident like pileup may occur since many cars are tailgating at high speed. In the intersections or street junctions, the attacker can make the approaching vehicles hesitate to go through with a fake warning that a vehicle is moving at high speed (90km/h) and is about to merge into the lane (marker 4).

Finally, there may be many ways to attack physical signals, e.g., drone spoofing attacks/Global Positioning System (GPS) spoofing (deception), and jamming attacks (disruption). While more efforts are still required to prevent these attacks completely, the countermeasures such as GSG-series GPS/Global Navigation Satellite System (GNSS) signal generation and jamming attack resistance [27] can complement our verification system to mitigate such attacks. That means the resilience modules can be pre-installed on the host vehicle to prevent physical signal spoofing while our system will conduct an extensive check to filter false application data in the CAM.

B. ASSUMPTION AND SYSTEM MODEL
Vehicle equipment availability plays a crucial role in realizing the proposed system. A modern vehicle may have high-end facilities and built-in sensors such as high-resolution cameras, LIDAR, or even infrared systems. However, we argue that installing many such facilities and sensors into a vehicle may significantly increase its cost and probably exceed the aims of many affordable vehicles [28]. In contrast, due to low cost, we assume that future vehicles can be equipped with at least an OBU with V2X communication inside, i.e., chipsets that support Dedicated Short-Range Communication (DSRC)/Cellular-V2X communications. Notably, V2X communications are vital for helping the vehicles to share sensing data about the surrounding activities when the camera and radar systems perform poorly, e.g., due to visibility obstruction by truck, buildings, or heavy fog. By supporting the verification, even with the configuration of low-cost OBUs, we hope that our proposal’s application is affordable for most people. Also, according to the standard [27], the PKI is the fundamental infrastructure to protect the data exchange’s integrity and will be supported in wireless communications, including V2X. Unfortunately, the truthfulness of data cannot be warranted by the PKI. A compromised vehicle can use legitimate/stolen credentials to bypass the integrity check and authentication. As a result, the received data should not merely be trusted by default. The host vehicle must have the capability to verify the data’s truthfulness independently, i.e., our work.

Regarding the communication model, we assume that the receiver (Rx) is a host vehicle (or an RSU) is near the transmitter (Tx), i.e., in V2X one-hop communication range. The Tx can communicate with the Rx via vehicle-to-vehicle (V2V) communications (e.g., 5G sidelink) and with RSUs via 5G cellular networks (e.g., New Radio (NR) bands from 24–100GHz). According to the specification [29], the CAM from a Tx vehicle always includes a unique field, so-called Vehicle ID (or pseudonym ID). When the Rx receives these messages, it can extract this unique ID and associate all relevant verification of the sender with it, i.e., using it as an index key. Similarly, the RSU also has such a unique ID of the sender. Therefore, the “nearby vehicles” term in this work means the ones appearing in the one-hop communication with the Rx and having validated IDs. Following this model, the target vehicle can be any in the nearby vehicles of the Rx. In the cases of heavy traffic, we may need to perform a massive verification (evaluated in Section IV-D).

When assistive components such as the RSUs are used, we assume that they are trusted devices. In practice, these devices are administered by the authorized agencies that reduced the risk of direct inference from unauthorized users. It is also common that these RSUs are equipped with hardware security modules (HSM) by default [2], which consist of secure cryptoprocessors to prevent physical attacks. A physical inference, if any, e.g., reverse engineering, thus is difficult to be successful. Moreover, the network of the RSUs can extensively maintain trust among the nodes in various ways, e.g., using trust-based [11], [14] or blockchain-based mechanisms. Due to the scope of this work, the discussion of the mechanisms is beyond this article. Finally, the RSUs in a region can connect and outsource heavy computations to nearby edge servers [30]. The extendable high-performance edge servers can support dozens of verification instances or assist the host vehicle/RSUs for such massive verification.

For the vehicle antenna configuration modeling, given the fast movement of vehicles and the balance between the cost and the system complexity [31], in this work, we adopt a hybrid beamforming model [32], [33]. Note that mmWaves and hybrid beamforming promise to be widely used in the coming V2X networks, particularly in the era of 5G [28].
(even 6G [34]). Following this, suppose that the Rx and Tx vehicles are equipped with \( K_T \) and \( K_R \) arrays (radio frequency chains), and each Tx/Rx array fully connects to \( N_T/N_R \) antenna elements (as illustrated in Fig. 3). The antenna arrays are located at four corners of the vehicles, as illustrated in Fig. 4. Without loss of generality, we configure the antennas with Uniform Linear Array (ULA) type and adopt orthogonal frequency division multiple access (OFDMA) technique [35] with \( N \) sub-carriers. Also, given the challenge of guaranteeing time global synchronization in regular clocks - much cheaper than atomic clocks in high-precision GPS systems - we assume that the clocks of the Tx and the Rx are not synchronized. As a result, only the time-difference of arrival (TDoA) measurements can be estimated, while it is impossible to measure the Time of Arrival (ToA).

### FIGURE 3.
The illustration of hybrid beamforming antenna configuration in the Rx and Tx OBUs.

![Hybrid Beamforming Antenna Configuration](image)

### FIGURE 4.
The illustration of geometric model of V2X multiple antenna arrays among the transmitter vehicle (Tx) and the receiver vehicle (Rx).

![Geometric Model of V2X Multiple Antenna Arrays](image)

Regarding the geometry model, as shown in Fig. 4, we assume the reference positions of the Rx and the Tx at the time \( k \) are \( P^R(k) \) and \( P^T(k) = P^R(k) + d \), where \( d = ||P^R - P^T||_2 \) denotes a shift from the Rx to the Tx. \( \phi^R \) and \( \phi^T \) denote the orientation of the Rx and Tx vehicle. Similar to [36], we also define \( d^{R/T} \) is the distance between the \( i \)th array to the centroid of the Rx, and \( \psi^R_i \) and \( \psi^T_i \) are the angles between the horizontal axis of the vehicles and the centroid of the \( i \)th Rx array and the \( j \)th Tx array, where \( i = 1, \ldots, K_R \) and \( j = 1, \ldots, K_T \). The centroid location of the \( i \)th Rx array \( P^{R,i}_k \) thus can be expressed by \( P^{R,i}_k = P^R + d^{R,i}_k u(\psi^R_i + \phi^R) \), where \( u(x) = \begin{bmatrix} \cos(x) \\ \sin(x) \end{bmatrix} \). Since the locations of the antenna arrays are fixed on the vehicles, the position of the \( r \)th antenna element of the \( i \)th Rx array is \( P^{R,i}_r = P^{R,i}_k + d^{R,i}_k u(\psi^R_i + \phi^R) \), where \( d^{R,i}_k \) and \( \psi^R_i \) are the distance and angle from the centroid of the \( i \)th Rx array, \( r = 1, \ldots, N_R \). The coordinate of the \( j \)th array of the Tx \( P^{T,j}_k \) and the \( q \)th antenna element \( P^{T,q,j}_k \) of \( j \)th Rx array and the corresponding angles \( \psi^T_j \) and \( \psi^T_{q,j} \) can be calculated similarly, where \( q = 1, \ldots, N_T \). If the Rx is an RSU, the operation of calculating the location and orientation is the same. Moreover, given the limited OBU size (~ 1m), we assume that the aperture of the arrays is much smaller than the distance between the \( i \)th Rx array and the \( j \)th Tx array (up to tens of meters). To reduce the complexity, e.g., compensation for the difference in the geometry model's reference point, in this work, we assume the Rx's location is set at the origin. When the Rx verifies the other Tx vehicles, it will set itself as the center of the relationship. Then the propagation time from the \( q \)th antenna of the \( j \)th Tx array to the \( r \)th antenna of the \( i \)th Rx array in a LOS area is

\[
\tau_{q,j} \triangleq \tau_{i,j} - \tau_{r,i}(\phi^R_{i,j}) - \tau_{q,j}(\phi^T_{i,j}),
\]

where

\[
\begin{align*}
\tau_{i,j} &= d/c = ||P^R_i - P^T_j||_2, \\
\tau_{r,i}(\phi^R_{i,j}) &= \frac{d^{R,i}_k u(\psi^R_i) u(\phi^R)}{c}, \\
\tau_{q,j}(\phi^T_{i,j}) &= \frac{d^{T,j}_q u(\psi^T_q) u(\phi^T)}{c}, \\
\phi^R_{i,j} &= \arctan(P^R_i(y) - P^T_j(x), P^R_i(x) - P^T_j(x)), \\
\phi^T_{i,j} &= \phi^R_{i,j} + \pi,
\end{align*}
\]

and \( c \) is the speed of light.

Suppose a signed message is divided into \( n_p \) packets, and then each packet is sent through the 5G V2X physical layer in the form of \( N_s \) OFDM symbols (\( N_s \leq K_T \)). The discrete transmitted signal after precoding from the Tx over subcarrier \( p \) at the time \( k \) can be expressed as follows:

\[
X_k[p] = F_{RF} F_{BB} s_k[p],
\]

where \( p = 1, \ldots, N \), and \( F_{RF} \) and \( F_{BB} \) denote the analog precoder and the digital baseband precoder. In a fully connected hybrid precoding approach, the Tx is supposed to apply \( K_T * N_s \) digital baseband precoders for \( F_{BB} \) and \( N_T * K_T \) analog precoders for \( F_{RF} \). The total power of the Tx is constrained by normalizing \( F_{BB} \) as \( ||F_{RF} F_{BB}||^2_2 = N_s \). Also, \( s_k[p] \) denotes the \( N_s \times 1 \) symbol vector and can be normalized as in [32].

The received signal at the \( i \)th Rx array over subcarrier \( p \) at the time \( k \) is

\[
Y_{i,k}[p] = \sqrt{\rho} H_{i}[p] X_k[p] + \eta_{i,k}[p],
\]

where \( \rho \) is the average received power, \( \eta_{i,k}[p] \) denotes the vector of i.i.d \( \mathcal{CN}(0, \sigma^2 I_{N_s}) \) white Gaussian noise (processed in digital signal components). \( H_{i}[p] \in \mathbb{C}^{N_R \times N_T} \) is the channel matrix. Since there may be limited scattering in an mmWave channel, i.e., only a few scattering clusters, we assume that the channel is modeled with \( M \) scattering clusters, and the channel matrix can be expressed by
\[ H_i[p] = \gamma \sum_{m=1}^{M} \sum_{l=1}^{L} g_R(\phi_{m,l}^R) h_{m,l} a_T(\phi_{m,l}^T) e^{-j2\pi f_c (\tau_h + \Delta_1 - \eta)} / M \]

where \( L \) is the number of propagation paths, \( \gamma = \sqrt{N_r N_t / M^2} \).

\( g_l = \frac{v \tau_c}{\pi} \), \( \tau_c \) is the timing offset of the coarse synchronization from the TOA of the shortest path, \( v \) is the current velocity of the Rx, \( T_s \) is the observation time, \( h_{m,l} \) is the complex gain of the link between the Tx and the Rx array in the \( m \)-th cluster and the \( l \)-th path, \( f_c \) is the carrier frequency.

Besides, due to the assumption of imperfect synchronization at the receiver side, \( \Delta_1 \) denotes the clock synchronization error. Also, the antenna array response vectors at the transmitter \( a_T(\phi_{m,l}^T) \) and receiver \( g_R(\phi_{m,l}^R) \) evaluated at the \( l \)-th path and the \( m \)-th cluster can be represented as follows:

\[ a_T(\phi_{m,l}^T) = [e^{j2\pi f_c \tau_1 (\phi_{m,l}^T)}, \ldots, e^{j2\pi f_c \tau_{M} (\phi_{m,l}^T)}] \]

\[ g_R(\phi_{m,l}^R) = [e^{j2\pi f_c \tau_1 (\phi_{m,l}^R)}, \ldots, e^{j2\pi f_c \tau_{M} (\phi_{m,l}^R)}] \]

In ULA antennas, the responses does not depend on the elevation angle \( \alpha \).

**C. Problem Statement**

Given the input data of the received signals (Eq. 4) and the received CAM as well as the defined models and assumptions above, our goal is to figure out whether the Tx vehicle is moving consistently as it claims in CAM. To tackle the problem, we first extract data of the \( \text{six-dimensional} \) vector of records \( (x, y, v, a, h, \omega) \) from the received CAM in a time interval of \( T \). Simultaneously, we maintain a second data collection of the same dimensional vector but from our built-in UKF-based trajectory estimation process. Notably, the estimation process is periodically refined with the measurement from a signal-based localization (and Doppler radars if equipped).

With this independent measurement, the system can avoid falling into the trap of referring to the potentially manipulated shared data only. In the last step, we perform a Student’s t-test on the data from the two stages to verify the consistency of the reported CAM. To enhance the verification, we perform a high-level fusion on the results of these core verifiers (at the host vehicles and RSUs) for the final decision, i.e., whether the target vehicle is trustworthy. We eventually present the details of the system architecture and the aforementioned functions in the following section.

**III. Multi-Dimensional Signal-Based Verification**

This work aims to exploit multiple data sources along with the help of an RSU to verify whether the vehicles are moving, honestly as they claimed. In this section, we present the design of our proposed system and such verification mechanisms in detail.

**A. The System Design and Workflow**

As the workflow illustrated in Fig. 1, our system consists of two verification engines in a client-server model: the local detector at the Rx (client) and the global detector at the RSU (server). Initially, the local detector at the Rx verifies the target vehicle through a signal-based verification scheme and reports the results to the RSU. Meanwhile, the RSU also performs its own verification with the same signal-based verification scheme. However, unlike the local verification at the Rx, the engine at the RSU can constantly collect and fuse the opinions from its associated vehicles and then reply to the assistance requests from the host vehicles with its verification results. Finally, the Rx performs a fusion on its detection result and the result in the RSU’s response to determine whether the target vehicle is truthful. The following subsections present signal-based data sources, the core signal-based verification model, and the fusion model in detail.

**B. Signal-Based Multiple Data Sources**

In this work, we use passive positioning information collected from V2X signal-based localization, acceleration, and velocity directly from Doppler radar in range. Unlike conventional localization, the signal-based localization for multi-array antennas does not require multiple anchors to locate a moving object [35], [36] and can work simultaneously with communication operations. In this method, the Tx is the target of localization and the source of signal transmission. The location of the Tx can be derived by the DCS-SOMP/SAGE algorithm [36], [37] to estimate the value of the vector \( \varphi \in \mathbb{R}^{4+2K_r K_h} \) as \( [\mu_k, \omega_k, \Delta_k, \eta_k] \) from received signal patterns (Eq. 4) and channel information (Angle-of-Arrival, Time Difference of Arrival), where \( \mu_k \) denotes the position of the Tx at the time \( k \), \( \Delta_k \) is the clock synchronization error, \( \omega_k \) the orientation, and \( \eta_k \) nuisance parameters. The nuisance parameters denote the channel gains of the link between a Tx and an Rx antenna array (in Eq. 5). We assume that our self-estimation vector is \( \hat{\varphi} \). We use the Cramer-Rao Lower Bound (CRLB) derivative to know the bounds of the estimation error. According to the CRLB theory, the estimated lower bound error of the vector \( \varphi \) can be expressed as follows:

\[ \mathbb{E}[(\hat{\varphi} - \varphi)(\hat{\varphi} - \varphi)^T] \geq J^{-1} \]

Without loss of generality, we demonstrate how to find the bound for the position by using the equivalent FIM (EFIM) to extract the information of interest. According to [36], [38], the EFIM for the position parameter can be expressed by

\[ J_{po} = \sum_{j=1}^{K_r} \sum_{i=1}^{K_h} g_{i,j} \left( \frac{2\pi f_c^2 S_i (\hat{\mu}_j^R)^2}{c^2 d^2} v_\varphi v_\varphi^T - Q_{i,j} \right) \]

\[ g_{i,j} = \frac{2N_r N_t (h_{i,j})^2 \rho}{\sigma^2}, \]

\[ v_r = [u_T^T(\phi_{i,j}^R) u_T^T(\phi_{i,j}^R) u_T^T(\phi_{i,j}^R) u_T^T(\phi_{i,j}^R)]^T, \]

\[ v_\varphi = [u_T^T(\phi_{i,j}^R) u_T^T(\phi_{i,j}^R) u_T^T(\phi_{i,j}^R) u_T^T(\phi_{i,j}^R)]^T, \]

\[ Q_{i,j} = \sum_{j=1}^{K_r} \sum_{i=1}^{K_h} g_{i,j} \beta, \]
Here \( u_L(x) = u(x - \frac{\lambda}{2}) \), \( \beta_L \) is the effective baseband bandwidth of the signal, and \( S_t(\phi_t^m) \) denotes the squared array aperture function defined in [36]. Then measuring the achievable position estimation accuracy can be expressed by the position error bound (PEB) as

\[
PEB = \sqrt{\text{lateral PEB}^2 + \text{longitudinal PEB}^2}.
\]

(11)

Antenna design and spatial correlation are the major factors to jeopardize this localization approach’s accuracy. The spatial correlation can degrade the localization’s performance since short-distance antenna arrays can increase the probability of getting similar signal components at adjacent antennas [39]. Fortunately, utilizing the antennas placement, e.g., placing antenna arrays at four corners of a vehicle, can significantly mitigate this issue because the distance between the corners will increase the antenna separation and maximize the spatial degrees of freedom in the AoA/DoA resolution. The other influential factors on the channel-based measurement, such as LOS blockage analysis, path loss, and shadow fading, can be referred to in [29], [39]–[41]. In the next subsection, we present the Tx trajectory estimation, which relies much on these signal data sources for updating its measurement process.

C. INDEPENDENT TX TRAJECTORY ESTIMATION WITH SIGNAL-BASED STATE MEASUREMENT UPDATES

This part presents the method to build an independent Tx trajectory for checking whether it is approximate to the trajectory extracted from the Tx’s CAM. To estimate the Tx’s location from independent measurements, we need to define the state model for tracking. The dynamic state of a moving object can be modeled as follows:

\[
s_{k+1} = F_k s_k + u_k,
\]

(12)

where \( F_k \) is the state transition model, and \( u_k \) denotes the noise. Note that \( u_k \) is assumed to be a zero-mean multivariate normal distribution, \( N \), with co-variances \( Q, u_k \sim N(0, Q) \).

A motion model (physical laws of motion) is required to gain an accurate estimation of the Tx’s dynamic state. Several motion models have been proposed in the literature, such as constant velocity model (CV), constant acceleration (CA), and constant turn rate and acceleration (CTRA) [42]. For example, when we use the CTRA model \((\omega_k + 1 = \omega_k; a_{k+1} = a_k)\), the final process model is [43]

\[
s_{k+1} = s_k + u_{k+1} + \int^n_{n} \dot{s}_{t}|_{t} dt
\]

(13)

where the time interval between two sequential scans, i.e., from \( t_k \) to \( t_{k+1} \), is \( T \), \( \dot{s}_{t}|_{t} \) is the derivative of the state \( s \), e.g., position, with respect to time. The non-linear state transition matrix \( F_k \) follows [43]. Applying the derivative to six parameters of the Tx state, i.e., \([x y v a h \omega] \), we can calculate the value of each parameter at the time \( k + 1 \) as follows:

- \( x_{k+1} = x_k + v_k^x \sin(\omega_k T) + v_k^y (\cos(\omega_k T) - 1) + a_k (\cos(\omega_k T) - \cos(h_k)) + a_k T \sin(h_k + \omega_k T) \)
- \( y_{k+1} = y_k + v_k^y \sin(\omega_k T) - v_k^x (\cos(\omega_k T) - 1) + a_k (\sin(h_k + \omega_k T) - \sin(h_k)) + a_k T \cos(h_k + \omega_k T) \)
- \( v_{k+1} = v_k + a_k T \)
- \( \omega_{k+1} = \omega_k + 1 \)
- \( h_{k+1} = h_k + a_k T \) (given that \( \omega = \frac{\partial h}{\partial \tau} \))

The state parameters are periodically predicted at every time interval with the correlation from the actual measurement (modeled in detail later in this section) and prior knowledge (Eq. 12). In the LOS areas, we can update the measurement, e.g., \( v_k^x, v_k^y \), by using Doppler radar [10]. In the NLOS areas, the relative velocity can be estimated through calculating \( d_k/T \), where \( d_k = \sqrt{(x_{k+1} - x_k)^2 + (y_{k+1} - y_k)^2} \) or collected from the signal-based localization. Similarly, the value of \( h_k \) and \( \omega_k \) can be estimated through channel estimation in the signal-based positioning. However, if the vehicle is moving in a straight line, the calculation of \( x_{k+1} \) and \( y_{k+1} \) will be divided by \( \omega_k = 0 \). To solve the problem, we select a fast-switch mechanism like the interacting multiple model (IMM) [44], which can deal with various road shapes and vehicle moving patterns, including straight motion [45].

The challenge of realizing the above model (Eq. 12) and get an optimal prediction is to calculate the distribution of uncertainty in time series as the state transition model is non-linear. We select the UKF to deal with this non-linear problem. Besides the advantage of excellent performance in non-linear systems, UKF also supports well with the aforementioned motion models and multiple targets tracking in real-time [46]. The UKF can be conceptualized into two phases: predict and update. The predict phase is \textit{a priori} state estimate as it does not include the measurement for refining the estimated results. In contrast, the update phase is \textit{a posterior} process, where the prediction estimate is refined with the actual measurement information. Without the updates, the Tx positions from the prediction will blindly follow a pre-defined motion model and drift away from the real location.

To model a UKF system, let \( \hat{s}_{k|k} \) be the estimated \( s \) at time step \( k \) given the measurements up to \( k \), also called the posterior. Then when updating the sigma points, we will determine the \textit{prior} \( \dot{s}_{k+1|k} \), the estimated state vector for the next time step \( k + 1 \). In the predict stage, predicting \( \dot{s}_{k+1|k} \) and \( P_{k+1|k} \) by UKF is expressed as follows:

\[
\begin{align*}
\dot{s}_{k+1|k} &= F_k \dot{s}_{k|k} - 1 + u_k \\
P_{k+1|k} &= F_k P_{k|k} F^T - 1 + Q
\end{align*}
\]

(14)

\( \dot{s}_{k|k} \) and \( P_{k|k} \) can be initialized as follows:

\[
\begin{align*}
\dot{s}_{0|0} &= E[s_0] \\
&P_{0|0} = E[(s_0 - \dot{s}_0)(s_0 - \dot{s}_0)^T]
\end{align*}
\]

(15)
Given the state \( s_k \) with \( m \) dimensions (in this work, \( m = 6 \)), \( 2m + 1 \) sigma samples are required. Also, \( \hat{s}_{a|a-1} \) and the covariance matrix \( P_{k|k-1} \) are as follows:

\[
\begin{align*}
\hat{s}_{a|a} &= \hat{s}_k ; a = 0, \\
\hat{s}_{a|a-1} &= \hat{s}_{k-1} + (\gamma \sqrt{mP_{k-1}})_a ; a = 1, 2, \ldots, m, \\
\hat{s}_{a|a-1} &= \hat{s}_{k-1} - (\gamma \sqrt{mP_{k-1}})_{a-m} ; a = m + 1, \ldots, 2m, \\
P_{k|k-1} &= \sum_{a=0}^{2m} w_a (s_a - \hat{s}_a)(s_a - \hat{s}_a)^T + Q,
\end{align*}
\]

(16)

where \( \gamma \) denotes the parameter value to control the sigma points range around the mean, \( (\gamma \sqrt{mP_{k-1}}) \) means the ith column of \( \sqrt{mP_{k-1}} \), \( e^{-\frac{4}{\gamma}} \leq \gamma \leq 1 \). Also,

\[
w_a = \begin{cases}
\frac{\lambda}{m + \lambda} & a = 0 \\
\frac{m + \lambda}{2(m + \lambda)} & a = 1 \ldots 2m,
\end{cases}
\]

(17)

where \( \lambda \) is the spreading parameter.

So far, we have estimated the Tx state based on the motion model. To refine the prediction, we need an actual measurement from sensors and the associated measurement model (e.g., signal-based localization). The measurement model at time step \( k \), namely \( z_{k|k-1} \) of the true state \( s_k \), is defined as follows [44], [47]:

\[
z_{k|k-1} = H_k s_k + \vartheta_k,
\]

(18)

where \( H_k \) is the measurement model that maps the true state space into the observed space, and \( \vartheta_k \) is the measurement noise, which is assumed to be zero-mean Gaussian white noise with covariance \( R: \vartheta_k \sim \mathcal{N}(0, R) \).

The next step is to determine the mean and covariance matrix of the expected measurement from the transformation of the predicted sigma points into the measurement space. Similar to the state prediction, the mean \( \hat{z}_{k|k-1} \) and covariance \( P_{z_k|k-1} \) of the sigma points \( Z_{k|k-1,i} \) in the measurement space are calculated by

\[
\begin{align*}
\hat{z}_{k|k-1} &= \sum_{a=0}^{2m} w_a Z_{k|k-1,a} \\
P_{z_k|k-1} &= \sum_{a=0}^{2m} w_a (Z_{k|k-1,a} - \hat{z}_{k|k-1,a})(Z_{k|k-1,a} - \hat{z}_{k|k-1,a})^T + R
\end{align*}
\]

(19)

The final step in the UKF is to update the Tx state estimate vector and covariance matrix from the predicted mean and covariance along with the measurement’s mean and covariance (ground truth \( z_k \)) as follows:

\[
\begin{align*}
\hat{s}_{k+1|k} &= \hat{s}_{k|k-1} + K_k (z_k - \hat{z}_{k|k-1}) \\
P_{k+1|k} &= P_{k|k-1} - K_k^T Z_{z_{k|k-1}} K_k
\end{align*}
\]

(20)

where \( K_k = P_{k|k-1}^{-1} Z_{z_{k|k-1}} \) denotes the Kalman gain.

**D. BASIC LIMITS OF THE UKF-SIGNAL-BASED ESTIMATION**

Since the signal-based estimation is a crucial part of the solution to verify the target vehicle’s trajectory, understanding its fundamental limits is essential to optimize the overall system and tune the relevant parameters. For this purpose, we introduce the CRLB-based estimation method. According to the CRLB theory, the covariance of \( \hat{s}_{k|k} \) has a lower bound expressed by

\[
C_k \triangleq \mathbb{E}[(s_k - \hat{s}_{k|k})(s_k - \hat{s}_{k|k})^T] \geq J_k^{-1}
\]

(21)

where \( J_k^{-1} \) is the \( m \times m \) FIM with the elements

\[
J_k^{bc} \triangleq \mathbb{E}[(\nabla_{s_k} \log(p(s_k, z_k)))(\nabla_{s_k} \log(p(s_k, z_k)))^T].
\]

(22)

where \( m \) is the dimension of the target state defined above \((m = 6, b, c = 1, \ldots, m)\). \( p(s_k, z_k) \) refers to a joint probability density function (pdf) of the true state and the measurement at the time step \( k \), and can be written as follows:

\[
p(s_k, z_k) \approx p(s_0) \prod_{r=1}^{k} p(z_r|s_r) \prod_{r=1}^{k} p(s_r|s_{r-1}),
\]

(23)

where

\[
p(s_0) = \frac{1}{\sqrt{2\pi Q_0}} \exp\left[\frac{-(s_0 - \pi_0)^2}{Q_0}\right],
\]

(24)

\( Q_0, \pi_0 \) are the initial values.

Given the assumption that the measurement and the true state estimation follow the normal distribution, we can write the following conditional probabilities:

\[
\begin{align*}
p(z_k|s_k) &= \frac{1}{\sqrt{2\pi R}} \exp\left[\frac{-(z_k - s_k)^2}{R}\right] \\
p(s_k|s_{k-1}) &= \frac{1}{\sqrt{2\pi Q}} \exp\left[\frac{-(s_k - s_{k-1})^2}{Q}\right]
\end{align*}
\]

(25)

Then the CRLB of the state error estimate is given by

\[
J_{k+1} = D_{k+1}^{22} - D_{k+1}^{21}(J_k + D_{k+1}^{11})^{-1} D_{k+1}^{12},
\]

(26)

where

\[
\begin{align*}
D_{k+1}^{11} &= \mathbb{E}[(\nabla_{s_k} \log(p(s_{k+1}|s_k)))(\nabla_{s_k} \log(p(s_{k+1}|s_k)))^T] \\
D_{k+1}^{12} &= \mathbb{E}[(\nabla_{s_{k+1}} \log(p(s_{k+1}|s_k)))(\nabla_{s_k} \log(p(s_{k+1}|s_k)))^T] \\
D_{k+1}^{21} &= \mathbb{E}[(\nabla_{s_k} \log(p(s_{k+1}|s_k)))(\nabla_{s_{k+1}} \log(p(s_{k+1}|s_k)))^T] \\
D_{k+1}^{22} &= \mathbb{E}[(\nabla_{s_{k+1}} \log(p(s_{k+1}|s_k)))(\nabla_{s_{k+1}} \log(p(s_{k+1}|s_k)))^T]
\end{align*}
\]

(27)

According to [48], the initial FIM \( J_0 \) can be easily calculated from the prior pdf \( p(s_0) \) (Eq. 24) mentioned above as follows:

\[
J_0 = \mathbb{E}[(\nabla_{s_0} \log(p(s_0)))(\nabla_{s_0} \log(p(s_0)))^T]
\]

(28)
E. CONSISTENCY VERIFICATION FOR THE SIGNAL-BASED VERIFICATION

The trajectory consistency check on whether the Tx’s movement is honest, as it claimed, is the next step. Like the attack modeling in [49], we consider the claimed location of the Tx to the Rx at time slot \( t \) is \( P^T_c \), and the information should be verified (truthful or not). We assume that the value from our signal-based estimation is \( \hat{P}^T_c \). Given two chances to know the Tx’s exact location, we use the Student’s t-test to test the evidence of the consistency on the estimated data \( \hat{P}^T_c \) and the claimed data \( P^T_c \) (extracted from the CAM). Let the null hypothesis \( H \) represent that the Tx is at the position as claimed in its CAM. \( H \) is rejected if the calculated p-value is smaller than a specified significant level of \( \alpha \), e.g., 0.05, meaning an attack is likely to happen. The degree of certainty in this testing can be considered as \( 1 - \alpha \). Moreover, since the signal-based verification module is the core of both the local and the global detectors, the result in this step will be one of the inputs into the fusion stages. The detail of our fusion mechanism is presented in Section III-G.

F. CALIBRATION FOR THE GLOBAL DETECTOR

The nearby vehicles can significantly support the Rx verification in heavy traffic density. For example, in some cases, the Tx is not in the LOS area of either the Rx or RSU, but nearby vehicles. Because the attacker can be one of such neighbors, we propose using a multi-opinion voting mechanism to find the nearby vehicles’ consensus. A vehicle is eligible to vote if it is within a given distance around the Rx and either the RSU or the Rx can independently verify the location through the signal-based localization. The distance threshold should be flexibly adjusted to balance between the communication range (where the signal-based localization can accurately perform) and the threat level to the Rx’s safety. It is highly encouraged that the Rx may only take care of the nearby targets (e.g., < 300m). The farther distance the target is, the less threat it poses to the safety of the Rx. Suppose the number of vehicles with a consensus on the honesty of the Tx is \( O_m \), and the total number of received opinions is \( O_t \). If \( \frac{O_m}{O_t} > 0.5 \) and the major opinions are positive, we consider the Tx as an honest vehicle with a degree of belief at \( \frac{O_m}{O_t} \); otherwise, it is misbehaving with the same level of belief. The detail of the fusion and the potential case of the majority of voting vehicles compromised are presented in Section III-G.

G. FUSING THE VERIFICATION RESULTS FOR THE FINAL DECISION

We use the Dempster-Shafer (DS) fusion and its rule to collate two independent observers’ detection results. In this work, the first fusion is performed at the global detector for the signal-based estimation results and the multi-opinion voting module. The second fusion is between the local and the global detector’s detection results. We define \( A \) and \( \bar{A} \) to specify two elements to build the power set in the DS theory [50]. Let \( A \) represent the honesty of the Tx reported from a detector, \( \bar{A} \) denotes the opposite, and \( \mathcal{U} \) denotes either \( A \) or \( \bar{A} \). Given the separate detection from each detector, the fused value of two degrees of belief is calculated as follows:

\[
\begin{align*}
\text{bel}_A &= m_1(A) \oplus m_2(A), \\
\text{bel}_{\bar{A}} &= m_1(A) \oplus m_2(\bar{A}),
\end{align*}
\]

where \( \oplus \) is the orthogonal sum. \( m_1(A) \) is the belief assignment of the local detector for \( A \), \( m_1(\bar{A}) = 0 \), and \( m_1(\mathcal{U}) = 1 - \rho \). The belief assignment of the global detector by \( m_2 \) is defined similarly.

Dempster’s rule for combination [50] is expressed by

\[
m(\mathcal{A}) = \frac{m_1(\mathcal{A})m_2(\mathcal{A}) + m_1(\mathcal{A})m_2(\mathcal{U}) + m_1(\mathcal{U})m_2(\mathcal{A})}{K},
\]

where

\[
K = m_1(A)m_2(A) + m_1(A)m_2(\mathcal{U}) + m_1(\mathcal{U})m_2(A) + m_1(\mathcal{A})m_2(\mathcal{A}) + m_1(\mathcal{A})m_2(\mathcal{U}) + m_1(\mathcal{U})m_2(\mathcal{A}) + m_1(\mathcal{U})m_2(\mathcal{U}).
\]

The degree of certainty in the hypothesis testing \( (1 - \alpha) \) at the local detector and the global detector’s fusion is taken as the degree of belief. Let us consider an example in which the degrees of belief of detecting an attack vehicle are 0.95 for the local detector and 0.75 for the global detector. Both detectors claim the occurrence of an attack. In this case, \( K = 0.95 \ast 0.75 + 0.95 \ast 0.25 + 0.05 \ast 0.75 + 0.05 \ast 0.25 = 1 \). \( m(\mathcal{A}) = (m_1 \oplus m_2)(\mathcal{A}) = 0.9875 \), implying that we believe that Tx is untrustworthy at 98.75% or likely an attacker. The fusion will increase the reliability significantly if both sources agree on the same hypothesis.

If the majority of voting vehicles are compromised and intentionally issue false decisions for a given Tx vehicle, the system will likely give a false decision. However, such false decisions may not significantly impact the final result due to the RSU’s conclusion in the signal-based detection module (i.e., the second input component for DS-fused fusion \( m_2(A) \)). In the worst case, if the signal-based detection module gives a result contrary to the multi-opinion module, according to DS theory, the belief will be reduced due to two contrary inputs, but by no means entirely negative. Moreover, we also believe that compromising all the vehicles near the RSU, while possible, may not be common in practice due to the high expense of doing so.

To avoid redundant verification, if a vehicle has been marked as an illegitimate one, it will be added to a blacklist called certificate revocation list (CRL). Then the host vehicle will discard all upcoming CAM from the vehicles in this list. The RSUs can also maintain a similar list, probably larger in range. To share the knowledge about the misbehaving/attack vehicles for all the vehicles of interest, the RSUs can broadcast the list via the WAVE advertisement services, and any vehicle in range can receive and decode the records.
H. COMPLEXITY ANALYSIS

The proposed design’s complexity comes from the most complex processing of the following components: the signal-based localization, the Tx trajectory estimation, the multi-opinion voting, and the DS-fusion module. Without loss of generality, the estimation is performed for single vehicle verification. For the signal-based localization, the computational tasks include (1) the channel information generation $O(M \times L)$, i.e., channel matrix in Eq. 5; (2) beamspace generation and precoder observation (i.e., subspace decomposition and spectral search) for the received signals $O(N_R \times N_s \times N)$ (Eq. 3); (3) distance computation in each path $O(L)$. The overall complexity of the signal-based localization for a Tx vehicle is $O(N_R \times N_s \times N)$. For the Tx trajectory estimation, with $m$ states under tracking, the state/covariance estimation (Eq. 14 and Eq. 16) and the state prediction (Eq. 12) cost $O(m^3)$, while the measurement updates and Kalman gain calculation only require $mK^2 + m^2K$ (Eq. 18-20) flops, where $K$ is the number of data sources. The complexity of the UKF-based estimation can be found in [47]. For the multi-opinion voting, a linear calculation for quantifying the degree of belief from $p$ opinions costs $O(p)$ only. For the DS-based fusion, the power set computation for $x$ observers costs $O(2^x)$, according to the DS theory [50]. However, we only set two observers (i.e., $x = 2$) for the power set in this work: $A$ and $\bar{A}$. Therefore, theoretically, this task can run in linear time due to the small dimensions. In conclusion, the most complex processing is in the signal-based localization. Since the proposed method involves limited spectral search and a reduced-dimension noise subspace (e.g., exploiting the advantage of the sparse structure of beamforming uniform linear arrays); hence it is quite computationally efficient. However, massive verification for dozens of vehicles at the same time is a challenge. We address this challenge in Section IV-D.

IV. PERFORMANCE EVALUATION

Given the lack of a real and open testbed (e.g., 5G massive array antennas-supported OBUs in vehicles) and a limited deployment of 5G V2X in practice at the time of this work, we implement and evaluate the system on MATLAB with the help of the VANET/5G/Communication toolboxes. Note that MATLAB is also widely acknowledged in simulating beamforming mmWave/NR signal processing (e.g., 5G and phased array system toolbox) at the microscopic level. The advantages are essential for our physical signal-based verification. For traffic model generation, we use SUMO with a diversity of vehicle types (e.g., trucks, vans), departing routes (e.g., highway, street junctions). The simulated traffic is modeled up to 100 seconds. The vehicles randomly appear on the roads, and the attack vehicles can be any. However, to have a ground truth for assessment, we specify the attack vehicles by a pre-defined list of vehicle IDs and locate them on pre-defined road segments. For a close-to-the-real-world test, the Luxembourg SUMO traffic (LuST) scenario with almost 930km of roads of various types and 203 intersections regulated by the traffic signals [51] is also used, as illustrated in Fig. 5. The scenario records the traffic mobility in both the morning and evening rush hour peaks around (e.g., 08:15 and 18:30), and a lower peak period around lunchtime. We set $K_R = K_F = 4$ for the antenna configuration and assume that the four antennas’ centroid is at the vehicle’s center. The other parameter configuration is listed in Table 2. We follow [41] for the path-loss and shadowing configuration of V2X communications in various conditions.

In the lone wolf attack, the Tx reports various variants of the false position offsets (e.g., the offsets $\Delta x$, $\Delta y$ are randomly chosen from $[−100, 100]$). Similarly, the offset can be injected to false acceleration $a + \Delta a$, false turn rate $\omega + \Delta \omega$ ($\Delta a$, $\Delta \omega$ uniformly random from $[−20, 20]$). Also, in the eventual stop attack, the Tx transmits its position repeatedly while it stops at a roadside, e.g., from the 10th to 15th second. We set from 1 to 4 out of 10 nearby Tx vehicles periodically disseminating false CAM at specific time and positions on the pre-defined roads in the collusion attack. Finally, in the Sybil attack, we set a vehicle to send false messages on behalf of various vehicle IDs. We set the sending rate of the vehicles at 10 messages/s. All the attack traces are mixed with the traffic traces generated from SUMO. However, the ground truth is known through accessing the list of vehicle IDs used to launch the attacks on the pre-defined roads, e.g., the highway and several street junctions near the Rx in the LuST scenario. Finally, we set a default value of $T = 0.5s$, $\Delta k = 2ms$.

FIGURE 5. Illustration of Luxembourg SUMO Traffic (LuST) scenario. The highways (in blue) are mostly around the city. The residential roads are in black and the rest is arterial roads (in cyan).

| Parameters          | DSRC | 5G NR |
|---------------------|------|-------|
| Center frequency $f_c$ | 5 GHz | 400 MHz |
| Bandwidth           | 10 MHz | 20 MHz |
| Antenna             | Multi-array ULA | Multi-array ULA |
| Array spacing       | 20 mm | 20 mm |
| $N_R/N_T$           | 16 | 16 |
| Sub-carrier spacing $\Delta f$ | 156.25 kHz | 120 kHz |
| $N$ (sub-carriers)   | 2048 | 2048 |
| CAM size            | 200 bytes | 200 bytes |
| CAM interval        | 100 ms | 100 ms |
| Transmit power      | +20 dBm | +31 dBm |
| Vehicle speed       | 30–144 km/h | 30–144 km/h |
We define that accuracy (Acc) and false alarm rate (FPR) for the performance metrics as follows:

$$\text{Acc} = \frac{TP + TN}{TP + TN + FP + FN}$$ (32)

$$\text{FPR} = \frac{FP}{FP + TN}$$, (33)

where true positive (TP) denotes the number of attack messages detected accurately by our verification, and false positive (FP) means the system falsely considers benign messages as from the attacker. In V2X, false negative (FN) is important since it represents the number of attack messages that bypass our verification. Finally, true negative (TN) denotes the number of honest messages that our system recognizes as they are. We use the receiver operating characteristic (ROC) curve to present the overall performance balance between the detection accuracy and false-positive rate (FPR).

A. THE VERIFICATION PERFORMANCE ON THE ATTACK CASES

Fig. 6 shows our detection performance in three attack types of LuST dataset. Through multiple testing iterations, our verification works best with accuracy over 0.98 to figure out the false positioning attacks and Sybil attacks but suffers a slight degradation (0.75 on average) in the presence of the collusion attack. In the last attack, even when the number of colluding vehicles reaches nearly half of the nearby vehicles giving opinions about the Tx, the accuracy to find the attack vehicles is not degraded significantly. The degradation results primarily from the colluding vehicles’ conflicting opinions in the DS-based opinion fusion at the global detector. If the multi-opinion voting is activated, we will have few chances to detect such collusion attacks if the vehicles giving opinions are out of the range of signal-based verification. In contrast, although the Tx tries to disseminate CAM with various pseudonyms in a Sybil attack, it is only a signal source. In this case, our detection can easily figure out the abnormal appearance of many overlapped locations reported from different pseudonyms. Compared to the data plausibility and trust-based approach [5], [10], [19], our approach has clear advantages to detect these attacks within a very short time, e.g., milliseconds (refer Section IV-D).

The other interesting result is that the larger the Rx antenna array is equipped, the better accuracy the system can gain, particularly to detect the Sybil and false data attacks. As illustrated in Fig. 7, the Rx equipped with the large scale antennas ($N_R = 100$) can get better performance than the configuration with few ones. This is because a large-scale antenna can receive signals and resolve the channel parameters such as AoA in localization better [35]. However, due to spatial correlation, the performance improvement becomes marginal with $N_R > 100$. Moreover, a large multi-array antenna system will also increase the overall cost. Finally, with single antenna configuration ($K_R = 1$) and single-beam beamforming, determining the orientation/turning rate via vehicle positioning at the Rx may not be possible, and thus the measurement for refining the Tx’s state and the signal-based verification is less accurate.

Through evaluation, we also see that operating high carrier frequency multi-array antennas, e.g., 5G NR 40GHz, can improve the accuracy of our verification in both the traffic sparsity and crowded streets. As presented in Fig. 8, the system with the antennas handling at 40GHz works 5–10% better than that at 5.9GHz within the first 300m. However, following the results, we see that the antenna configuration with such high carrier frequency may not help the system perform well if the Tx is far away from the Rx. In practice, high noise and fading are the major factors that suppress the accuracy of the signal-based localization [19]. Specifically, such factors result a highly incorrect measurement in the update stage of the UKF-based filtering and thus influence the detection accuracy as well. As shown in Fig. 8, the system performance with the local detector activated only is significantly degraded after 300m (Rx-Tx distance). These
inference factors influence more frequently if many obstacles, e.g., trucks or buildings, appear between the Tx and the Rx (the heavy traffic density case is illustrated in Fig. 2). Unfortunately, having an obstacle between the Tx and the Rx is quite likely (i.e., a probability > 0.6) within the first 200m on the highway or less likely in urban areas [39]. However, those negative influences can be mitigated significantly if more RSUs are in the streets. If Tx-Rx is far away, an accident between them is less likely, i.e., the accuracy is not so critical.

**B. DETECTION PERFORMANCE IN AN NLOS AREA**

Detecting the misbehaving vehicles in an NLOS area faces tremendous challenges [4]. Large obstacles may void the host vehicle’s radar and camera-based systems since they can significantly increase the noise/fading inference. Consequently, like many prior approaches, with the local verification activated only, the system does not perform very well to detect the attacks (the ROC of the Rx detector without the support of the RSU’s detection is shown in Fig. 8). In this work, involving the assistive participants like RSUs helps much. As shown in Fig. 8, the Rx detector with the support of the RSU’s detection can gain much higher accuracy (∼80%) and less FPR than the mode with a local detector only. This is particularly vital when the target vehicle is the host vehicle’s NLOS areas or many obstacles between them, e.g., buildings at street junctions. Fortunately, as specified in the standards [39], RSU often appears every 500m-1km road segment; thus, the system configuration with RSU’s help is likely in action. This is the case illustrated in Fig. 2, in which our system finds the attack vehicle (Tx2, Tx3) is moving near the RSUs (RSU2, RSU4). Finally, since the RSUs are connected with each other, if the global detector at an RSU cannot perform the verification, it can still ask for another nearby RSU’s contributions.

**C. THE INFLUENCE OF MOTION MODELS AND ENVIRONMENT CONDITIONS**

Since the vehicle movement relies much on the motion model, a correct motion model selection is essential to improve the system performance. If the prediction is performed on the windy roads (e.g., the residential roads in the LuST scenario) with the CV motion model only, the margin of distance error between the estimated values from the measurement and the reported values in CAM is much wider than that with an adaptive model like IMM (Fig. 9). To date, the system with the interacting multiple models activated can gain much more accurate in estimating the Tx trajectory than that with the other motion models, e.g., constant acceleration or constant velocity.

**FIGURE 8.** Accuracy performance of our verifier in NLOS with two configurations: (1) The Rx detector with the support of the RSU’s detection; (2) The Rx detector without the support of the RSU’s detection. The test is performed with the false data and Sybil attacks activated by default.

**FIGURE 9.** The illustration of the important role to select a correct motion model. The margin of distance error between the estimated values (measured by the combination of the UKF-based prediction and CV-based motion model) and the reported values in CAM is double wider than using the optimal configuration (measured by the combination of UKF and IMM).

Besides the influence of the noise/fading inference and the motion model selection to the system performance, spatial correlation phenomenon in a dense antenna system can also pose a challenge to estimate the channel information (e.g., TDoA, AoA) accurately. Consequently, the accuracy of signal-based positioning measurement for the trajectory tracking and the associated verification also suffers negative impacts. Fortunately, the motion model in our prediction and assistance of the nearby RSU can coordinate the state estimate, thus mitigating this impact remarkably. Without the help of the global detector, that positive result cannot hold (as shown in Fig. 10) if many obstacles obscure the LOS area, e.g., many trucks are ahead or a far Tx-Rx distance (> 300m). Generally, the RSUs and the cooperative verification in our approach play a crucial role to overcome the negative influences of the inference factors.

To enhance our system performance in the noisy/fading cases, there are various ways. Enhancing the signal beamforming, utilizing the antenna placement in a vehicle, and increasing the Rx array size/the number of RSUs on the street are the top selections. As illustrated in Fig. 11, with
large-scale NR-based hybrid beamforming antenna arrays inside, the Rx is capable of estimating the location and speed of two very-close vehicles accurately. However, implementing large-scale antenna arrays may come up at a high cost.

**D. MASSIVE VERIFICATION**

Our system can serve up to 100 vehicles in parallel and suffer a peak delay of 1.54s, as presented in Fig. 12. The delay denotes the elapsed time to get the verification results for the input CAM. Given the challenge of quantifying the details of the response time through propagation delay and transmission delays calculation accurately on different computation systems, the elapsed time is equivalent to the total amount of time from when receiving the CAM to the time of getting the verification results for that message. Apparently, besides including the delay values mentioned above, this elapsed time also consists of potential delays from data and verification processing at the Rx and RSUs (processing time). In the worst case, e.g., up to 100 vehicles under monitoring and the attacker is verified last, the elapsed time can be up to seconds.

However, we believe that the case of such large traffic densities may seldom occur, particularly in highway scenarios where the Rx and the Tx may pass through each other in a very short time. The possible case is in a traffic jam where many Tx vehicles are near the Rx. Fortunately, the risk of potential accidents, in this case, is dramatically reduced due to the slow movement of the vehicles. Moreover, if we limit the number of vehicles for verification in the near distance (e.g., 100m ahead) or the system checks only specific vehicles in range (where the camera cannot track), the delay is much shorter. Although the delay in this work is a little higher than the in-vehicle approaches [19], it still satisfies well the requirement of major V2X road-safety applications [27]. Admittedly, there is a trade-off between massive verification and the response time to protect the Rx. A longer delay in verifying massive information makes the Rx more vulnerable to the attacks. However, given the assumption that modern vehicles and RSUs are equipped with high-performance computing facilities, we believe that the challenge of massive verification can be partially solved. Moreover, the host vehicle or RSUs can outsource the high computation tasks to nearby edge servers, which are often equipped with large-scale high-performance facilities. Finally, implementing multi-threading and parallel processing models can significantly accelerate massive verification.

**E. THE INFLUENCE OF OTHER FACTORS**

Bandwidth and memory usage are also important metrics for measuring system efficiency. Our approach does not generate and pour much traffic to the network since the verification results can be embedded in CAM as well. For storage, given the default configuration of the sending rate (10 messages/s per vehicle) and a massive number of vehicles (e.g., 100) to be verified in parallel, the memory size for storing the six-dimensional state vector and testing variables per vehicle gets a peak at 416KB/s or 24MB/min. However, if the memory storage on the vehicles is limited, the unnecessary records (e.g., those of the vehicles that no longer exchange data with the Rx) can be removed every minute. After this period, the system must re-establish the whole verification stages to assess the Tx vehicle’s trustworthiness if the Rx meets again. This is to avoid falling into the attacker’s trap
when the Tx suddenly becomes misbehaving or launches an attack. However, if a vehicle is classified as an illegitimate one, its ID can be stored longer in the CRL. We believe that this is reasonable because the number of attack vehicles in practice may not be so large, and storing just the IDs does not consume much memory space.

Also, network congestion or packet loss, while rarely occurs in 5G V2X networks due to the QoS requirement of V2X and 5G Ultra-Reliable Low-Latency Communication, does not influence the performance of our system. The reason is that the filtering and refinement for trajectory correction are extracted from wireless signals rather than from application messages. Unlike ours, packet loss phenomena can significantly degrade the performance of the data plausibility approaches, e.g., [5], [10]. In the worst case, if there is a sudden interruption in signal communications or no available RSU at all, the self-verification at the local detector of the host vehicle in our model can still provide a minimum verification service. However, we believe that gNBs can be the perfect alternatives if RSU deployment is limited. Notably, the density of gNBs in 5G (particularly to support NR bands) is quite high.

F. SECURITY ANALYSIS

Our system could be the target of an attack such as denial of service to the verification. As stated in Section II-A, our system could be the target of an attack such as denial of service. However, we believe that gNBs can be the perfect alternatives if RSU deployment is limited. Notably, the density of gNBs in 5G (particularly to support NR bands) is quite high.

REFERENCES

[1] SAE, Standard SAE J3016. Accessed: Sep. 28, 2020. [Online]. Available: https://www.sae.org/news/2019/01/sae-updates-j3016-automated-driving-graphic
[2] K. Ren, Q. Wang, C. Wang, Z. Qin, and X. Lin, “The security of autonomous driving: Threats, defenses, and future directions,” Proc. IEEE, vol. 108, no. 2, pp. 357–372, Feb. 2020.
[3] 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.
[4] R. W. van der Heijden, S. Dietzel, T. Leinmuller, and F. Kargl, “Survey on misbehavior detection in cooperative intelligent transportation systems,” IEEE Commun. Surveys Tuts., vol. 21, no. 1, pp. 779–811, 1st Quart., 2019.
[5] J. Firl, H. Stubing, S. A. Huss, and C. Stiller, “MARV-X: Applying maneuver assessment for reliable verification of Car-to-X mobility data,” IEEE Trans. Intell. Transp. Syst., vol. 14, no. 3, pp. 1301–1312, Sep. 2013.
[6] F. Guo, Z. Wang, S. Du, H. Li, H. Zha, Q. Pei, Z. Cao, and J. Zhao, “Detecting vehicle anomaly in the edge via sensor consistency and frequency characteristic,” IEEE Trans. Veh. Technol., vol. 68, no. 6, pp. 5618–5628, Jun. 2019.
[7] S. So, J. Petit, and D. Starobinski, “Physical layer plausibility checks for misbehavior detection in V2X networks,” in Proc. 12th Conf. Secur. Privacy Wireless Mobile Netw., May 2019, pp. 84–93.
[8] R. Changalvala and H. Malik, “LiDAR data integrity verification for autonomous vehicle,” IEEE Access, vol. 7, pp. 138018–138031, 2019.
[9] M. A. Javed, M. Z. Khan, U. Zafar, M. F. Siddiqui, R. Badar, B. M. Lee, and F. Ahmad, “ODPV: An efficient protocol to mitigate data integrity attacks in intelligent transport systems,” IEEE Access, vol. 8, pp. 114733–114740, 2020.
[10] M. Sun, M. Li, and R. Gerdes, “A data trust framework for VANETs enabling false data detection and secure vehicle tracking,” in Proc. IEEE Conf. Commun. Netw. Secure (CNS), Oct. 2017, pp. 1–9.
[11] S. Huang, J. Gui, T. Wang, and X. Li, “Joint mobile vehicle–UAV scheme for secure data collection in a smart city,” Ann. Telecommun., to be published.
[12] J. Kamel, M. R. Ansari, J. Petit, A. Kaiser, I. B. Jemaa, and P. Urien, “Simulation framework for misbehavior detection in vehicular networks,” IEEE Trans. Veh. Technol., vol. 69, no. 6, pp. 6631–6643, Jun. 2020.
[13] C. Allig, T. Leinmuller, P. Mittal, and G. Wanielik, “Trustworthiness estimation of entities: A collective perception,” in Proc. IEEE Veh. Netw. Conf. (VNC), Dec. 2019. [14] S. Huang, A. Liu, S. Zhang, T. Wang, and N. Xiong, “BD-VTE: A novel baseline data based verifiable trust evaluation scheme for smart network systems,” IEEE Trans. Netw. Sci. Eng., early access, Aug. 7, 2020, doi: 10.1109/TNSE.2020.3014455.
[15] W. Mo, T. Wang, S. Zhang, and J. Zhang, “An active and verifiable trust evaluation approach for edge computing,” J. Cloud Comput., vol. 9, no. 1, pp. 1–19, Dec. 2020.
[16] T. Zhou, R. R. Choudhury, P. Ning, and K. Chakrabarty, “P2DAP—Sybil attacks detection in vehicular ad hoc networks,” IEEE J. Sel. Areas Commun., vol. 29, no. 3, pp. 582–594, Mar. 2011.
[17] B. M. ElHalawany, A. A. El-Banna, and K. Wu, “Physical-layer security and privacy for Vehicle-to-Everything,” IEEE Commun. Mag., vol. 57, no. 10, pp. 84–90, Oct. 2019.
[18] 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.
[19] V.-L. Nguyen, P.-C. Lin, and R.-H. Hwang, “Enhancing misbehavior detection in 5G vehicle-to-vehicle communications,” IEEE Trans. Veh. Technol., vol. 69, no. 9, pp. 9417–9430, Sep. 2020.
[20] M. Nasrani and A. Lewis, “Detecting misbehavior in vehicle-to-vehicle (V2V) communications,” U.S. Patent 9,665 168, May 15, 2018.
[21] Y. Yao, B. Xiao, G. Yang, Y. Hu, L. Wang, and X. Zhou, “Power control identification: A novel sybil attack detection scheme in VANETs using RSSI,” IEEE J. Sel. Areas Commun., vol. 37, no. 11, pp. 2588–2602, Nov. 2019.
[22] J. Wang, Y. Shao, Y. Wang, Y. Ge, and R. Yu, “Physical layer authentication based on nonlinear Kalman filter for V2X communication,” IEEE Access, vol. 8, pp. 163746–163757, 2020.
[23] ETSI, Specification of Cooperative Awareness Basic Service, ETSI EN 302 637-2 V1.4.0, 2018.
