Detection and Mitigation of False Data Injection in Cooperative Vehicular Scenarios

Luis Javier Puente Lam†*, Pedro M. Santos*, Luís Almeida‡*
† *CISTER Research Centre, Porto, Portugal*
‡ *Instituto de Telecomunicações, Porto, Portugal*
* *Faculdade de Engenharia, Universidade do Porto (FEUP), Porto, Portugal*
javie@isep.ipp.pt, {pmsantos, lda}@fe.up.pt

***Abstract*—In Autonomous Driving (AD), accurate positioning is essential to ensure safe, efficient, and cost-effective transportation. However, malicious cyber threats, such as False Data Injection Attacks (FDIA) through Vehicle-to-Everything (V2X) communications, pose significant risks to these objectives. FDIA can manifest as deceptive obfuscation, either translational (T-OBF) – a vehicle falsely claims to be on a parallel road –, or random (R-OBF) – the vehicle reports a random position that differs from its actual position by a few meters. Our approach tackles incorrect position data by leveraging cooperation between trusted participants. These participants perform ML-based FDIA detection by exchanging Angle-of-Arrival (AoA) measurements, and also for positioning in conjunction with an Extended Kalman Filter to reduce FDIA effects. FDIA detection achieves 89% accuracy for T-OBF and 72% for R-OBF, reducing localization errors below $10m$ in 77.74% and 81.84% of cases, respectively.**

***Index Terms*—V2X, False Data Injection, Extended Kalman Filter, Cooperative Wireless Positioning, Angle-of-Arrival**

I. Introduction

Autonomous driving (AD) requires situational awareness (learning where other road-users are and their future positions) for safe, efficient and cost-effective transportation. Relative positioning is can be achieved by on-board sensors such as cameras, radar and LiDAR, but they are limited to their detection range and resolution, thus constraining their effectiveness in some cases (e.g., around a corner). Cooperative Wireless Positioning (CWP), based on Vehicle-to-Everything (V2X) communication links such as vehicle-to-vehicle (V2V) and vehicle-to-network (V2N), can be used to share relevant information, notably position, direction and speed.

Reporting faulty or inaccurate positions over V2X links can lead to serious accidents. One source of errors is the inaccuracy or lack of availability of positioning technologies. The Global Navigation Satellite Systems (GNSS), in particular the Global Positioning System (GPS), are the primary technology for vehicle localization and navigation today, but their availability and accuracy may be limited in covered and dense urban environments. Other potential cause is intentional misreporting, i.e. the Tx vehicle deliberately inserts incorrect position data in V2X messages, also called **False Data Injection Attack (FDIA)** [1] – see Fig. 1. FDIA can take the form of *deceptive obfuscation*, e.g. **translational obfuscation (T-OBF)**, when a vehicle indicates that it is on a road parallel to its actual position (to simulate being inside or outside a certain area or to avoid a surveillance zone), or **random obfuscation (R-OBF)**, when it changes its position by a few meters (to simulate having stopped at a crosswalk or performing a maneuver).

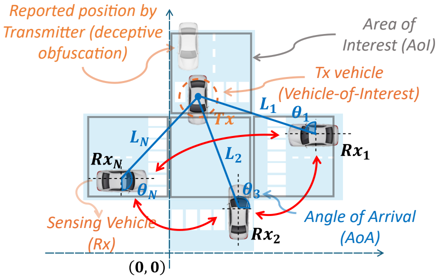

Fig. 1: Scenario with a Vehicle-of-Interest (VoI) Tx located in a certain Area-of-Interest (AoI). VoI was hacked with FDIA. A number of n sensing vehicles Rx_1,..., Rx_n, which measure the AoA of the Tx transmissions using antenna arrays. AoAs are used to detect anomalies Tx's reported positions and mitigate them by calculating intersections to estimate its location.

Detecting position misreports, caused specifically by obfuscation, from a **Vehicle-of-Interest (VoI)** can be seen as a problem of anomaly detection, for which machine learning techniques can be used. It can also be mitigated by leveraging cooperation between other road-users through Wireless Positioning (WP). Physical-layer mechanisms such as **Angle-of-Arrival (AoA)** can be used to estimate the angle of a transmitter relative to the receiver. A set of (trusted) sensing vehicles that exchange their AoA of signals transmitted by the VoI can produce an independent VoI position estimate that can be compared with the reported one. Ultimately, AoA-based CWP can mitigate the positioning errors introduced by FDIA.

In this paper, we present a CWP-based framework that offers techniques for detecting and mitigating FDIA by a set of trusted sensing vehicles. For the detection stage, we use ML methods to detect anomalies in positions reported by a VoI; the AoA observed by the trusted parties are also used as input. We compare different ML-based methods (i.e. Isolation Forest, DBSCAN, One-Class SVM and Local Outlier Factor) to detect anomalies (due to T-OBF and R-OBF) using metrics such as F1-score, recall and precision. SVM is the method that best detects anomalies in our setup and achieves an overall accuracy of 89% against T-OBF and 72% against R-OBF. When an anomaly is detection, the mitigation stage

This work was supported by national funds through FCT/MCTES (Portuguese Foundation for Science and Technology), within the CISTER Research Unit (UIDP/UIDB/04234/2020) and under the project Intelligent Systems Associate Laboratory - LASI (LA/P/0104/2020). The work of Luis Lam is supported by a Ph.D. Scholarship from FCT, grant nr. 2024.06411.BD.

is activated. We apply an Extended Kalman Filter, to which the output o mitigate the effects of FDIA. The EKF is complemented with a cooperative positioning mechanism When $N_{\mathrm{Rx}} = 5$, position estimation errors have a mean of $8.63m$ and a standard deviation of $10.39m$, with 77.74% of estimates falling within $10m$ of error against T-OBF. For R-OBF, the mean was $7.36m$, the standard deviation $7.98m$, and 81.84% of estimates were within $10m$ of error. We also discuss how the geometry of the receivers' positions relative to the VoI influences positioning accuracy. For 2, 3, and 5 receivers, 71.02%, 88.96% and 98.88% of the estimates were classified as ranging from *ideal* to *good*, respectively.

This paper is structured as follows. Sec. II reviews related literature. Sec. III introduces the scenario, attack model and solution outline. Sec. IV describes the CWP mechanism. Sec. V presents the FDIA detection and mitigation solutions. Sec. VI reports system performance, and Sec. VIII presents concluding remarks.

II. Related Work

FDIA detection has motivated various approaches. Machine Learning (ML) is widely used for detecting anomalies as it learns a pattern and identifying irregularities against it. A SVM-based classifier was implemented to recognize *malicious* or *trustworthy* data packets from other vehicles in [2]. A Digital Twin-based Asynchronous Federated Learning (AFL) with a Gated Recurrent Unit (GRU) model for identifying FDIA in Vehicular Cyber Physical Systems (VCPS) was presented in [3]. The authors in [4] propose a local-global fuzzy clustering method to detect FDIA based on Deep-Learning. In [5], seven representative classification algorithms for detecting common injection attacks on in-vehicle networks were discussed. In [6], Long Short-Term Memory (LSTM) based on a vehicle's speed and acceleration is used to detect FDIAs. In [7], authors propose the detection of *tampered* data in vehicles within a platoon using a Residual Function (RF) defined based on a Detection Unknown Input Observer (UIO), which outputs a high value if the vehicle has been hacked and close to zero otherwise.

The use of positioning techniques that complement GPS can be used to *mitigate FDIA* in the positions reported by a vehicle. Cooperative positioning from AoA estimates has been addressed in [8] using LSTM-NN in scenarios where base stations (BS) track a mobile user (unlike in our setup where vehicles (nodes) in motion estimate the VoI's position). Furthermore, the use of LSTM implies a temporal relationship in the data that we do not consider in our scenario. LSTM have also been used in [9], [10] and Transformer Models [11], but all of them focus on the prediction of trajectories, while our work focuses on the improvement of runtime accuracy. In [12], authors propose a vehicle self-localization system based on the integration of on-board sensors with a KF. In [9], cooperative positioning is explored, but AoA from Device-to-Device (D2D) channels is not considered; instead, data from cellular base stations are used. Many studies on localization algorithms and accuracy improvement techniques highlight that they become less effective or fail when the receivers' geometry is not favorable [13]. However, this issue is overlooked by many of the previous authors.

III. Scenario, Attack Model and Solution Outline

A. Scenario

We consider a vehicular scenario with a set of n sensing vehicles (Rx_1,..., Rx_n) and a Vehicle-of-Interest (VoI) located within a specific area of the scenario, referred to as the Area-of-Interest (AoI). The true position of the VoI must be determined due to potential FDIA suspicion. To this end, sensing vehicles first estimate the AoA and then analyze potential anomalies in the VoI's reported positions using ML-based methods. These methods leverage both the estimated AoA and the claimed AoI to assess whether the reported position is anomalous. If such anomalies exist, they can mitigate the FDIA effects through a localization procedure based on exchanged angle estimates. We assume all vehicles are in Line-of-Sight (LoS) conditions, and communication between sensing vehicles is instantaneous. Fig. 1 represents the scenario.

B. FDIA Model

An FDIA is a type of cyberattack in which the intruder injects false data into the real data stream of a system [14], resulting in erroneous decisions by the autonomous vehicle, such as unsafe driving behavior, misunderstanding of the environment, or inability to respond effectively. FDIA can manifest as deceptive obfuscation, where position data is manipulated to disguise the true location of a vehicle. Two types of obfuscation were considered in this work:

1) **Translation Obfuscation (T-OBF):** The attack shifts the reported position by a fixed offset so that the vehicle appears at a different location (e.g. on a parallel road).
2) **Random Obfuscation (R-OBF):** The position is unpredictably disturbed within a certain range, which reduces the consistency of localization.

We also consider z_{kin} as the position of the VoI according to its kinematic model, and also as its exact localization. The relative position that the vehicle reports at each time t is z_{rep}. In the absence of FDIA the reported position (obtained e.g. from GPS) is $\mathrm{z}_{rep} = \mathrm{z}_{GPS}$ where $\mathrm{z}_{GPS} = \mathrm{z}_{kin} + \vec{\mathrm{n}}_{GPS}$, and $\vec{\mathrm{n}}_{GPS} \sim \mathcal{N}(\mu_{GPS}, \sigma_{GPS})$. An FDI attack, launched at time t_{FDIA}, is modelled as $\mathrm{z}_{rep} = \mathrm{z}_{GPS} + \vec{\mathrm{n}}_{FDIA}$. The shift $\vec{\mathrm{n}}_{FDIA}$ can be of different nature, depending on the attacker's intention. For example, $\vec{\mathrm{n}}_{FDIA}$ would be set at a systematic offset w.r.t. to the VoI's position (e.g., 100 meters to the right or left) if the attacker wants to simulate that the VoI is on a parallel road (T-OBF). Or a random vector can be assigned to $\vec{\mathrm{n}}_{FDIA}$, e.g. $\vec{\mathrm{n}}_{FDIA} \sim \mathcal{N}(\mu_{FDIA}, \sigma_{FDIA})$, if the goal is just to obfuscate one's location (R-OBF).

C. System Architecture

The two stages of our solution rely on techniques described next, partly based on outputs of the CWP mechanism.

1) **Detection via Machine Learning.** Interested vehicles use ML-based methods to detect anomalies in the VoI's reported positiong, leveraging exchanged AoA values and the general area where VoI's reports itself.
2) **Mitigation via Kalman Filter.** Interested vehicles use an EKF and VoI position estimates from the CWP mechanism to reduce FDIA-induced position errors.

Fig. 2 shows the overall architecture. CWP is explained in Sec. IV, and detection/mitigation in Sec. V.

IV. Cooperative Wireless Positioning

A. Angle-of-Arrival Estimation at Individual Vehicle

Consider $\mathbf{n}$ vehicles: a transmitter Tx (the VoI) and $\mathbf{n}-\mathbf{1}$ receivers Rx. The transmitter emits signals isotropically. The receiving vehicles are equipped with a Uniform Circular (antenna) Array (UCA) with radius r, consisting of M antennas ($M=4$ in our scenario) which are arranged at a distance d_{ant}. The wavelength of narrowband is denoted as λ with $\lambda = \frac{c}{f}$, and $d_{\text{ant}} < \lambda$. The received signal $\vec{\mathbf{x}}$ is given by: $\vec{\mathbf{x}} = \vec{\mathbf{a}}s + \vec{\mathbf{n}}_{\text{AoA}}$ or, more precisely:

$$\vec{\mathbf{x}} = [a(\theta_1) \quad \dots \quad a(\theta_D)] \begin{bmatrix} s_1(k) \\ s_2(k) \\ \vdots \\ s_D(k) \end{bmatrix} + \vec{\mathbf{n}}_{\text{AoA}}(k) \tag{1}$$

where incident signals from the D transmitters, along with their amplitude and phase, are represented by complex quantities $s_1, s_2, ..., s_D$, from directions $\theta_1, \theta_2, ..., \theta_D$. For simplicity, we set $D=1$. Incoming signals vary over time, and therefore we take k samples over time of the incoming signal.

The incident signals are represented by the steering vector $\vec{\mathbf{a}} = a(\theta_i)$, where the first column, $a(\theta_1)$, corresponds to the steering vector for the first user, and so on.

$$a(\theta_1) = \begin{bmatrix} G_0 e^{j\beta p_0} \\ G_1 e^{j\beta p_1} \\ \vdots \\ G_{M-1} e^{j\beta p_{M-1}} \end{bmatrix}$$

with $\beta = \frac{2\pi}{\lambda}$, and $p_m = r \cdot cos\left(\frac{2\pi m}{M} - \theta_1\right)$, $m = 0,...,M-1$. G_m represents the antenna gain in the direction θ. As we assume isotropic antennas, $G_m = 1$.

We use the MUSIC (MUltiple SIgnal Classification) algorithm [15]to estimate the direction of the incoming signal. It achieves this by decomposing the signal correlation matrix into two orthogonal components: the signal subspace and the noise subspace. The direction can be determined from either subspace, assuming that noise in each channel is uncorrelated.

Noise Characterization in the Received Signal: The noise in the received signal $\vec{\mathbf{x}}$ is defined as in Eq. 2:

$$\vec{\mathbf{n}}_{\text{AoA}} = \sigma^2_{SNR} \cdot (\mathbf{n}_r + j \cdot \mathbf{n}_i) \tag{2}$$

where $\mathbf{n}_r$ and $\mathbf{n}_i$ are noise vectors in the real and imaginary domain (for amplitude and phase noise, respectively) generated from standard normal distributions, and $\sigma^2_{SNR} = \frac{1}{\text{SNR}_{\text{linear}}}$ is the noise variance calculated from the signal-to-noise ratio (SNR) on a linear scale. SNR in **dB** is calculated as in Eq. 3:

$$SNR_{\text{dB}} = P_{\text{Rx}} - n_{\text{Power}} \tag{3}$$

with n_{Power} as the typical noise power level (usually -90db) and $P_{\text{Rx}} = P_{\text{Tx}} - L_{\text{PL}}$. L_{PL} is the path-loss in **dB** using the two-ray ground-reflection loss model [16] (Eq. 4):

$$L_{\text{PL}} = 40 \cdot \log_{10}(d_{\text{Tx-Rx}}) - 10 \cdot \log_{10}(G \cdot (h^2_{\text{Tx}}) \cdot (h^2_{\text{Rx}})) \tag{4}$$

where $d_{\text{Tx-Rx}}$ represents the distance between the transmitter and receiver, G is the antenna gain, and h_{Tx} and h_{Rx} denote the heights of the Tx and Rx antennas, respectively. The trusted sensing vehicles share their positions and measured AoAs to the VoI θ_i to each other.

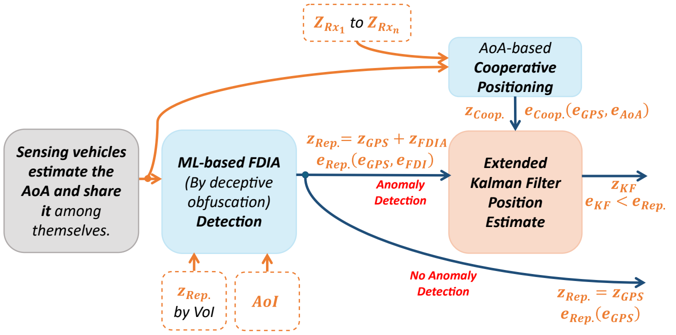
Z_{Rx_1} to Z_{Rx_n}
AoA-based Cooperative Positioning
$z_{Coop.}$
$e_{Coop.}(e_{GPS}, e_{AoA})$
Sensing vehicles estimate the AoA and share it among themselves.
ML-based FDIA (By deceptive obfuscation) Detection
$z_{Rep.} = z_{GPS} + z_{FDIA}$
$e_{Rep.}(e_{GPS}, e_{FDI})$
Anomaly Detection
Extended Kalman Filter Position Estimate
z_{KF}
$e_{KF} < e_{Rep.}$
No Anomaly Detection
$z_{Rep.} = z_{GPS}$
$e_{Rep.}(e_{GPS})$
$z_{Rep.}$ by VoI
AoI

Fig. 2: Architecture of FDIA detection and mitigation system

B. Cooperative Position Determination using SVD

Based on AoA information receceived from the other trusted sensing vehicles, each vehicle can internally represent signal trajectories as lines with respect to a common referential.

In [17], it is proposed that the relationship between each receiving vehicle Rx_i located at position $\mathrm{z}_{\mathrm{Rx}_i} = [x_i, y_i]^{\mathrm{T}}$ (in the global referential), and the transmitting vehicle Tx located at $\mathrm{z}_{\mathrm{Tx}} = [x,y]^{\mathrm{T}}$ as:

$$\mathrm{z}_{\mathrm{Tx}} = \mathrm{z}_{\mathrm{Rx}i} + d_{\mathrm{Tx-Rx}_i} \cdot \mathrm{v}_{\mathrm{Rx}i} \tag{5}$$

where $\mathrm{v}_{\mathrm{Rx}i} = [\cos\theta_i, \sin\theta_i]^{\mathrm{T}}$ is the unit vector in the L_i direction, and $d_{\mathrm{Tx-Rx}_i}$ is the distance between each receiving vehicle and the transmitter (which we assume is unknown). Eq. 5 can be rewritten as Eq. 6:

$$-x_i \sin\theta_i + y_i \cos\theta_i = -x \sin\theta_i + y \cos\theta_i \tag{6}$$

which, after some manipulation, produces the following matrix equation system: $b(\theta) \approx A(\theta) \cdot \mathrm{z}_{\mathrm{TX}}$ (we refer to [17] for details). $A(\theta)$ can be determined via Singular Value Decomposition (SVD): $A(\theta) = U\Sigma V^T$, where U is an unitary $n \times n$ (orthogonal) matrix (then $U^T = U^{-1}$), Σ is a diagonal matrix with non-negative singular values, and V^T an unitary 2×2 (orthogonal) matrix. Substituting the SVD into the system of equations:

$$U^T b(\theta) = \Sigma V^T \mathrm{z}_{\mathrm{TX}} \tag{7}$$

Let $\mathbf{z} = V^T \mathrm{z}_{\mathrm{TX}}$ and $\mathbf{z} = \Sigma^{-1} U^T b(\theta)$. The Tx's cooperative position $\tilde{z}_{\text{Coop}}$ is calculated by $\mathrm{z}_{\mathrm{TX}} = V\mathbf{z}$.

It is important to emphasize that in cases of collinearity (when two receivers have the same AoA) or near-collinearity (when the AoAs are very similar, e.g. differing by only 1 degree), the previously described method cannot estimate the position of the VoI, i.e., there is no service availability. We introduced a Kalman Filter to correct position errors from FDIA and AoA while ensuring service availability.

V. FDIA Detection and Mitigation

A. FDIA Detection via Machine Learning

We created four ML-based methods for detecting FDIA with the following inputs: i) position reported by the VoI z_{Rep}; ii) the Area-of-Interest (AoI) to which its position belongs A_{VoI}; and iii) the angles $\theta_{\text{AoA}\,i}$ estimated cooperatively by the sensing vehicles, output by the first stage of the CWP mechanism (Sec. IV-A). The Areas-of-Interest (AoI) are blocks into which the map is divided; e.g., our Manhattan grid scenario was divided into blocks of $20 \times 20m$ (see Fig. 1)

AoIs provide context to the models, enabling it to learn the AoAs corresponding to each AoI, which helps detect anomalies when this correspondence fails. The models are trained on correct information, i.e., no mis-reports; during operational/inference stage, there can be mis-reports, that will be flagged as anomalies. The models are:

- **Isolation Forest:** isolates anomalies by recursively partitioning the feature space through random splits [18]. Anomalous points require fewer splits to be isolated. The anomaly score is defined by $S_{\text{IF}}(x) = 2^{-\frac{E_{\text{path}}(x)}{c(d)}}$, where $E_{\text{path}}(x)$ is the expected path length of x in the isolation trees and $c(d)$ is a normalization factor.
- **DBSCAN:** clusters data based on local density. Points with insufficient neighbors for a given range ϵ are marked as outliers [19].
- **One-Class SVM:** a support vector machine is trained on normal data to learn a decision boundary, separating normal from anomalous instances [20]. It solves $\min_{\mathbf{w},\mathbf{b},\nu}\frac{1}{2}\|\mathbf{w}\|^2$ subject to $y_i(\mathbf{w}^T\mathbf{x}_i+b)\geq 1-\xi_i, \xi_i\geq 0$, where ν controls the fraction of outliers, i.e., the proportion of points that are allowed to be misclassified as anomalies.
- **Local Outlier Factor:** LOF measures the relative density of a point compared to its neighbors [21]. The anomaly score is based on the local reachability density (LRD) $\text{LOF}(x) = \frac{\sum_{y\in N_k(x)}\frac{\text{LRD}(y)}{\text{LRD}(x)}}{|N_k(x)|}$, where $N_k(x)$ represents the k-nearest neighbors of x.

B. Extended Kalman Filter for FDIA Mitigation

1) Overview: The Kalman Filter (KF) is an iterative mathematical process that uses successive input data with a series of equations [22]. The stochastic model defines the noise characteristics, while the dynamic model shows the behavior of the state variables. These are the most important factors that influence the performance of KF. For nonlinear systems (like ours), the Extended Kalman Filter (EKF) applies a linear approximation by Taylor expansion [23]. The EKF consists of two steps as follows:

- **Predict Step:** The state is estimated using the nonlinear model $f(\mathrm{x}_{k-1})$ of the system. To approximate this nonlinear behavior, the state transition matrix $\mathbf{F}_k$ is obtained as the Jacobian of the state transition function, evaluated at $\mathrm{x}_{k-1|k-1}$, i.e., $\mathrm{F}_k = J(f(\mathrm{x}_{k-1|k-1}))$. Matrix $\mathbf{Q}$ captures prediction uncertainty, and $\mathbf{P}$ the state estimation error.

$$\mathrm{x}_{k|k-1} = f(\mathrm{x}_{k-1|k-1}) \tag{8a}$$
$$\mathrm{P}_{k|k-1} = \mathrm{F}_k\mathrm{P}_{k-1|k-1}\mathrm{F}_k^T + \mathrm{Q} \tag{8b}$$

- **Update Step:** It is calculated the residual $\mathbf{y}$, which captures the difference between observed z_k and predicted values $h(\mathrm{x}_{k|k-1})$. The observation matrix $\mathbf{H}_k = J(h(\mathrm{x}_{k|k-1}))$ is used to linearize the measurement function. $\mathbf{S}$ is residual covariance matrix and $\mathbf{R}$ is the measurement noise covariance matrix. Adjustments in the state estimate are determined by the Kalman gain matrix $\mathbf{K}$, which compensates for the effects of prediction errors and observation uncertainties. The state error covariance matrix $\mathbf{P}$ is then updated to improve the accuracy of the state estimation process. $\mathbf{I}$ is the identity matrix.

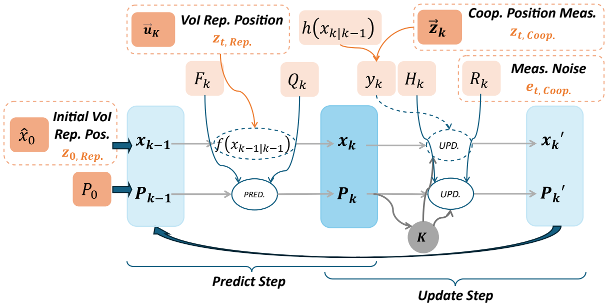

Fig. 3: Extended Kalman Filter Architecture

$$\mathrm{y}_k = \mathrm{z}_k - h(\mathrm{x}_{k|k-1}) \tag{9a}$$
$$\mathrm{S}_k = \mathrm{H}\mathrm{P}_{k|k-1}\mathrm{H}^T + \mathrm{R} \tag{9b}$$
$$\mathrm{K}_k = \mathrm{P}_{k|k-1}\mathrm{H}^T\mathrm{S}_k^{-1} \tag{9c}$$
$$\mathrm{x}_{k|k} = \mathrm{x}_{k|k-1} + \mathrm{K}_k\mathrm{y}_k \tag{9d}$$
$$\mathrm{P}_{k|k} = (\mathrm{I} - \mathrm{K}_k\mathrm{H})\mathrm{P}_{k|k-1} \tag{9e}$$

2) Application: To reduce the position error caused by FDIA, we consider an EKF that uses the positions reported by the VoI $z_{\text{Rep,t}}$ and the positions estimated by the cooperative positioning based on AoA measurements $\tilde{z}_{\text{Coop,t}}$ to improve the position estimate of the Tx at time $t+1$ (Fig. 3).

The vector $\mathrm{x}_k = [x, y, v_x, v_y, a_x, a_y]^{\mathrm{T}}$ represents the state of the system, which includes the position, velocity and acceleration in two dimensions. We assume that the movement of the vehicle is uniformly accelerated, then the state transition model $\mathbf{F}$ must reflect that the acceleration does not change over time:

$$\mathbf{F} = \begin{bmatrix} 1 & 0 & \Delta t & 0 & \frac{1}{2}\Delta t^2 & 0 \\ 0 & 1 & 0 & \Delta t & 0 & \frac{1}{2}\Delta t^2 \\ 0 & 0 & 1 & 0 & \Delta t & 0 \\ 0 & 0 & 0 & 1 & 0 & \Delta t \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

where $\Delta t = 1s$. We initialize the covariance matrix $\mathbf{P}_{6\times 6} = 5 \cdot 10^2 \cdot \mathbf{I}_{4\times 4}$ with a high value (i.e. high uncertainty in the initial system estimate). The covariance matrix of the process noise was set as $\mathbf{Q}_{6\times 6} = 10^{-3} \cdot \mathbf{I}_{4\times 4}$, assuming that changes in velocity and acceleration have a small uncertainty. The observation matrix $\mathbf{H}$ means that only the positions x and y are directly observed. In turn, the covariance matrix of the measurement noise $\mathbf{R}$ represents the uncertainty of the observations.

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \quad ; \quad \mathbf{R}_{2\times 2} = \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix}$$

The observation vector $\mathrm{z} = [\mathrm{z}_x, \mathrm{z}_y]^{\mathrm{T}}$ represents the measured positions. Our EKF is initialized with the state vector $\mathrm{x} = [x_{\text{z,Rep}}, y_{\text{z,Rep}}, v_{x,\text{Rep}}, v_{y,\text{Rep}}, a_{x,\text{Rep}}, a_{y,\text{Rep}}]^{\mathrm{T}}$ and the measurement vector $\tilde{z}_{\text{Coop}} = [\tilde{x}_{\text{Coop}}, \tilde{y}_{\text{Coop}}]^{\mathrm{T}}$. EKF can estimate the position without cooperative positions by using only the prediction step and skipping the update step.

VI. Experimental Evaluation

A. World Model & Simulation Setup

We use a Manhattan grid of 1×2 streets (one horizontal street and two vertical streets). The distances between intersections are fixed at $100m$. All roads are considered two-way, with each lane being $10m$ wide. The scenario

TABLE I: IDM parameters

Variable	*Description*	*Value*
T	Driver reaction time	1.5 s
a	Maximum acceleration	2.6 m/s^2
b	Comfortable deceleration	-4.5 m/s^2
δ	Acceleration exponent	4
s_0	Safe distance	2 m

TABLE II: AoA parameters

Par.	*Description*	*Value*
k	Number of time snapshots	500
c	Speed of light in vacuum	3×10^8 m/s
f	Signal frequency	2.442 GHz
M	Number antennas	4
D	Number incoming signals	1
d_{ant}	Spacing between antennas	10^{-1} m

TABLE III: Parameters Anomaly Detection

Method	*Parameter*	*Description*	*Value*
IF	n_estimators	Trees in the forest	200
	contamination	Expected outliers proportion	0.2
	max_samples	Sample fraction per tree	0.8
	bootstrap	Bootstrap sampling	True
	random_state	Seed for reproducibility	75
DBSCAN	eps	Max neighborhood distance	0.8
	min_samples	Min points per cluster	3
OC-SVM	kernel	Decision function type	rbf
	nu	Upper bound on outliers	0.2
	gamma	Kernel coefficient	0.4
LOF	n_neighbors	Nearest neighbors	12
	novelty	Prediction mode	True
	contamination	Expected outliers proportion	0.2

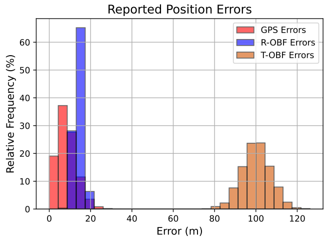

Fig. 4: Error distribution reported pos.

TABLE IV: Accuracy reported pos.

Reported	*Mean (m)*	*Std. dev. (m)*	*% < 10m*
GPS	8.61	4.61	65.30
T-OBF	14.33	2.01	1.40
R-OBF	100.17	7.08	0.00

was divided into 32 areas of interest ($20\times20m$) where the VoI could be located. Our scenario covers a total area of $340\times220m$. We consider $N_{\text{TX}}=1$ (VoI) and $N_{\text{RX}}=\{2,3,5\}$.

Both the start and end positions of each vehicle are drawn arbitrarily from the list of grid edge points, following an uniform distribution (u.d.). The vehicles move according to the Intelligent Driver Model on random paths (u.d.) selected from the list of possible paths to reach the destination point. The maximum speed is $50km/h$ (maximum speed allowed in urban areas). The length of traversed trajectories vary across runs due to the stochastic nature of the generation process.

The simulation time step is 1s, and vehicles produce one AoA estimate per (simulated) second. We conducted 180 experimental runs with a total of 5000 samples (average of 28.81 samples per run and a standard deviation of 1.38 samples). All samples from uniform distributions were produced using Python's *random.choice()* function.

B. Vehicle Kinematic Model & GPS Positioning Error

We consider the widely used Intelligent Driver Model (IDM) [24] as the kinematic model of the vehicle. IDM is able to simulate realistic vehicle movements under different traffic conditions (e.g. traffic jam and free-flowing traffic) by two ordinary differential equations:

$$\dot{x}=\frac{\mathrm{d}x}{\mathrm{d}t}=v\ \ ;\ \ \dot{v}=\frac{\mathrm{d}v}{\mathrm{d}t}=a\left(1-\left(\frac{v}{v_0}\right)^{\delta}-\left(\frac{s^*(v,\Delta v)}{s}\right)^2\right)$$

where $s^*(v,\Delta v)=s_0+vT+\frac{v\Delta v}{2\sqrt{ab}}$ is the safe distance, $\dot{x}$ represents the vehicle speed, $\dot{v}$ denotes the acceleration, v is the current speed of the vehicle, v_0 indicates the desired speed, a is the maximum allowed acceleration, δ refers to the acceleration exponent (usually $\delta=4$), s is the distance to the leading vehicle, $\Delta v=v-v_{\text{lead}}$ describes the speed difference between the vehicle and the leading vehicle, s_0 is the minimum safety distance, T denotes the driver reaction time, and b represents the comfortable deceleration. Table I shows the parameters used for IDM simulation.

The speed at the next time step $t+\Delta t$ is obtained by integrating the acceleration: $v(t+\Delta t)=v(t)+a(t)\Delta t$ Similarly, the position of the vehicle is updated by integrating the speed: $x(t+\Delta t)=x(t)+v(t)\Delta t+\frac{1}{2}a(t)(\Delta t)^2$. The GPS position, i.e. the position known and shared by the vehicles, is calculated by adding a Gaussian error $\vec{\mathrm{n}}_{GPS}\sim\mathcal{N}(0,7)$ to the kinematic positions.

C. FDIA Modelling

FDIA were modeled as described in Section III-B. The term $\vec{\mathrm{n}}_{FDIA}=100m$ was added to the x_{GPS} or y_{GPS} position for the T-OBF depending on the direction of movement. In the case of the R-OBF, both components were affected by $\vec{\mathrm{n}}_{FDIA}\sim\mathcal{N}(10,2)$. Table IV shows the characterization of the errors in the reported positions (GPS, T-OBF and R-OBF). The error distributions of GPS and R-OBF positions are similar, as shown in Fig. 4.

D. Cooperative Wireless Positioning & HDOP

For the purpose of AoA estimation, the signal reception at multiple antennas is modelled according to Eq. 1. As explained in Sec. IV-A, in our work, we model noise $\vec{\mathbf{n}}_{\text{AoA}}$ using Eq. 2, where $\mathbf{n}_r$ and $\mathbf{n}_i\sim\mathcal{N}(0,1)$. SNR values were calculated using Eq. 3 and 4. Given that $2\leq d_{\text{Tx-Rx}}\leq342.15$ m, the maximum and minimum SNR values obtained are 105dB and 15.67dB, respectively. The remaining AoA simulation parameters are shown in Table II. We found that MUSIC had difficulty estimating the AoA when the value was very close to 0° or 360° (edges of the MUSIC spectrum). To avoid this, we considered setting the estimated AoA to 0° for angles less than 1° and greater than 359°. All the above details were implemented in a simulator for mobility and wireless channel, and paired with a code version of the MUSIC algorithm. The implementation uses Python v3.12.0.

We assess the impact of the geometry of the receiver vehicle positions on position estimation using the Horizontal Dilution of Precision (HDoP). HDOP is a measure of the influence of receiver geometry on the accuracy of horizontal position estimates [25]. It is calculated using the geometry matrix $\mathbf{G}$, which encodes the spatial configuration of the receivers relative to the estimated Tx position. $\mathbf{G}$ is defined as:

$$\mathbf{G}=\begin{bmatrix}\frac{x_{\text{Rx}1}-\tilde{x}}{d_1} & \frac{y_{\text{Rx}1}-\tilde{y}}{d_1} & 1\\ \frac{x_{\text{Rx}2}-\tilde{x}}{d_2} & \frac{y_{\text{Rx}2}-\tilde{y}}{d_2} & 1\\ \vdots & \vdots & \vdots\\ \frac{x_{\text{Rxn}}-\tilde{x}}{d_n} & \frac{y_{\text{Rxn}}-\tilde{y}}{d_n} & 1\end{bmatrix} \qquad (10)$$

where $(\tilde{x},\tilde{y})$ denotes the estimated transmitter (Tx) position via EKF, $(x_{\text{Rx}_i},y_{\text{Rx}_i})$ are the positions of the receivers, and d_i is the Euclidean distance between each receiver Rx_i and Tx.

To calculate the HDoP, the submatrix $\mathbf{G}_{\text{hdop}}$ is extracted, which consists of the first two columns of the matrix $\mathbf{G}$

TABLE V: Performance of anomaly detection

	Method	***Precision***		***Recall***		***F1***		***Accur.***
		-1	**1**	**-1**	**1**	**-1**	**1**	
T-OBF	IF	0.46	0.87	0.51	0.85	0.48	0.86	0.78
	DBSCAN	0.23	0.81	0.32	0.73	0.27	0.77	0.65
	OC-SVM	0.65	1.00	1.00	0.86	0.79	0.93	0.89
	LOF	0.64	0.93	0.74	0.90	0.69	0.91	0.86
R-OBF	I. Forest	0.27	0.81	0.23	0.85	0.25	0.83	0.72
	DBSCAN	0.23	0.81	0.32	0.73	0.27	0.77	0.65
	OC-SVM	0.39	0.92	0.75	0.71	0.52	0.80	0.72
	LOF	0.20	0.80	0.11	0.89	0.14	0.84	0.73

corresponding to the horizontal components (*x* and *y*). HDoP is then computed as: $\text{HDoP} = \sqrt{\text{tr}(G_{\text{hdop}}^{-1} G_{\text{hdop}}^{-1}{}^{T})}$.

E. Training Procedure for FDIA Detection

For FDIA detection, we divided the set of 5000 samples into 4500 for training and 500 for testing. Of the 500 test samples, 100 were affected by FDIA. For both training and testing, the Python library *scikit-learn* was used. For the training process, the following features were used: the position reported by VoI (obtained from GPS), the AoI based on its location, and the AoAs measured by each of the receiver vehicles. All the data were normalized using the *StandardScaler* function. Table III shows all the parameters used for training in each method.

VII. Results

We report performance results of our technique in a simulated scenario. The positioning error is the Euclidean distance between estimate and actual location, i.e., $|z - \tilde{z}|$.

A. Detection: Performance of ML Methods

We first assess the performance of each ML-based method using the values from Table III, considering both T-OBF and R-OBF position anomalies. Results are shown in Table V. In this baseline scenario, we observe that the One-Class SVM method provides the most effective discrimination between normal and anomalous position data, regardless of whether T-OBF or R-OBF anomalies are present. However, anomaly detection (using any of the methods) is less effective in the presence of R-OBF, as the error distributions in normal GPS positions and those induced by R-OBF anomalies exhibit significant similarity (Fig. 4), making it challenging to differentiate between normal and anomalous data.

B. Cooperative Wireless Positioning

1) Impact of SNR on AoA Estimation: We characterize the AoA estimation error as a function of the Signal-to-Noise Ratio. We separate into ranges: $\text{SNR} < 20$ dB, $20 \leq \text{SNR} \leq 50$, and $\text{SNR} > 50$. Fig. 5 shows shows the empirical cumulative distribution function (ECDF) of AoA errors. The error was characterized using 25000 samples (5000 per receiver vehicle). The maximum AoA error obtained was 2.68°. The statistical characterization is shown in Table VI. As expected, higher SNR values indicate lower noise presence, and therefore a more accurate AoA estimation.

TABLE VI: Impact of SNR on AoA Estimation Error

SNR (dB)	***Samples***	***Mean (degrees)***	***Std. dev. (degrees)***
<20	570	0.57	0.47
$20 \leq \text{SNR} \leq 50$	19236	0.27	0.21
>50	5194	0.12	0.16

TABLE VII: Accuracy of Cooperative Position Estimation

Rx Vehicles	***Unavailable Service***	***Position Error (m)***		***% < 10m***
		Mean	***Std. dev.***	
2	470	30.01	48.29	36.60
3	93	17.27	30.07	55.37
5	0	8.84	11.99	75.48

TABLE VIII: HDoP Value distribution vs. Number of Rx

HDoP Value	Rating	***% Samples***		
		2 Rx	***3 Rx***	***5 Rx***
HDoP < 1	Ideal	0.00	0.00	38.22
1≤ HDoP < 2	Excellent	35.56	68.94	52.80
2≤ HDoP < 5	Good	32.20	18.92	7.86
5≤ HDoP < 10	Moderate	8.46	5.74	1.00
10≤ HDoP < 20	Fair	6.70	3.30	0.10
HDoP > 20	Poor	17.08	3.10	0.02

2) Impact of Number of Rx Vehicles: Cooperative positions were calculated with the model described in Sec. IV-B. We analyze the influence of the number of receiver vehicles on the VoI position estimation. For this purpose, we characterized the position error when estimating with AoA data from 2, 3 and 5 receiver vehicles (Fig. 6). The results are shown in Table VII. When $N_{\text{Rx}} = 2$ there are many cases of service unavailability (due to collinearity or near-collinearity), while with $N_{\text{Rx}} = 5$ the service is always available. As expected, positioning accuracy increases with the number of Rx.

3) Impact of Receiver Geometry: We analyze how the geometric arrangement of the receiving vehicles can affect positioning performance. Table VIII shows the distribution of HDoP values for the cooperative position estimates when varying the number of receiving vehicles (i.e., $N_{\text{Rx}} = 2{,}3{,}5$). In the case of $N_{\text{Rx}} = 2$, 17.08% of the estimates have an HDoP above 20, which includes the 470 cases of collinearity. The rest correspond to other unfavorable geometric situations (e.g., small intersection angles or large distances between the receivers). With $N_{\text{Rx}} = 5$ vehicles, 98.88% of the estimates are rated between *good* and *ideal*, since 5 receiver vehicles are able to overcome the geometry problems that could occur between 2 or 3 of them.

C. Mitigation: Performance of EKF

We also evaluate the positioning accuracy based on EKF when 2, 3, or 5 receiver vehicles are involved, in the presence of T-OBF or R-OBF. Since the EKF operates even when cooperative positions are unavailable, it ensures 100% service availability. Although R-OBF positions may resemble GPS positions due to overlapping error distributions (see Fig. 4), the EKF performs better against this FDIA because the error covariance between reported and cooperative positions is much smaller. Fig. 7 shows the ECDF of EKF position errors against R-OBF. For deeper insight, a specific run (i.e. one of the 180 experimental runs, with $N_{\text{Rx}} = 5$) is discussed. In Fig.8, we see that the AoA-based cooperative positioning (orange color) provides a good estimate during R-OBF, but with some errors (due to the modeling of propagation

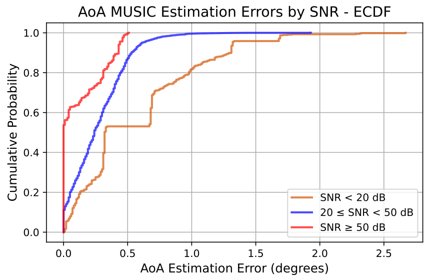

Fig. 5: AoA estimation errors vs SNR

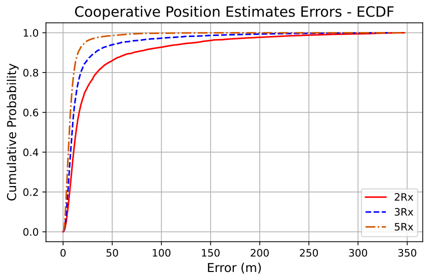

Fig. 6: Coop. position estimate vs. # rx

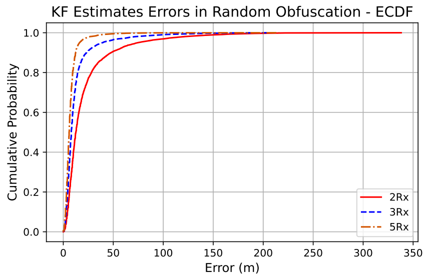

Fig. 7: EKF pos. estimate errors vs. # rx

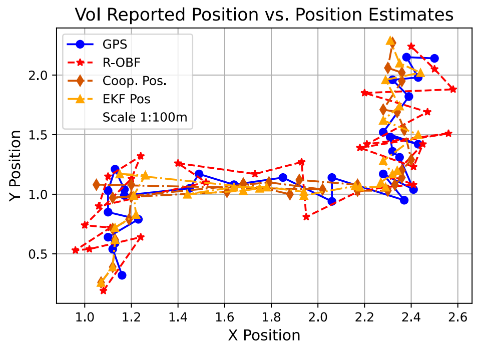

Fig. 8: Position estimates against R-OBF (example scenario)

TABLE IX: Accuracy EKF position estimates vs. T/R-OBF

	T-OBF			***R-OBF***		
Rx	***Position Error (m)***		***%<10m***	***Position Error (m)***		***%<10m***
Vehicles	***Mean***	***Std. dev.***		***Mean***	***Std. dev.***	
2	22.71	28.47	36.95	21.82	28.34	38.35
3	13.86	18.74	59.56	12.68	17.59	62.31
5	8.63	10.39	77.74	7.36	7.98	81.84

phenomena). EKF position estimates (in yellow) improve cooperative positioning in 64.92%.

VIII. Conclusions and Future Work

We propose a mitigation scheme for FDIA in a vehicular scenario, integrating an EKF with cooperative wireless positioning using AoA to reduce localization errors caused by FDIA. The approach is effective when CWP involves 3 or more vehicles, as two vehicles are insufficient due to the collinearity issue. Future works will include the simulation of shadowing effects due to urban obstacles (e.g. buildings) and the modeling of multipath reception over multiple antennas. We will develop ML-based algorithm to learn geo-referenced PDP and identify LoS/NLoS and multipath conditions.

References

[1] D. Adhikari, W. Jiang, J. Zhan, D. B. Rawat, and A. Bhattarai, "Recent advances in anomaly detection in Internet of Things: Status, challenges, and perspectives," *Computer Science Review*, vol. 54, p. 100665, 2024.

[2] J. D. Cabelin, P. V. Alpano, and J. R. Pedrasa, "SVM-based Detection of False Data Injection in Intelligent Transportation System," in *ICOIN*, 2021, pp. 279–284.

[3] S. Safavat and D. B. Rawat, "Digital Twin Based Asynchronous Federated Learning Enabled IDS for False Data Injection Attacks in Vehicular CPS," in *8th ICCSM*, 2024, pp. 19–23.

[4] S. A. Almalki and F. T. Sheldon, "Deep learning to improve false data injection attack detection in cooperative intelligent transportation systems," in *IEEE 12th IEMCON*, 2021, pp. 1016–1021.

[5] H. Ji, L. Wang, H. Qin *et al.*, "In-vehicle network injection attacks detection based on feature selection and classification," *Automotive Innovation*, vol. 7, pp. 138–149, 2024.

[6] B. Ko and S. H. Son, "An Approach to Detecting Malicious Information Attacks for Platoon Safety," *IEEE Access*, vol. 9, pp. 101 289–101 299, 2021.

[7] N. Ahmed, A. Ameli, and H. Naser, "Detection, Identification, and Mitigation of False Data Injection Attacks in Vehicle Platooning," *IEEE TVT*, vol. 74, no. 1, pp. 1296–1309, 2025.

[8] P. M. Santos and L. J. P. Lam, "Cooperative AoA Wireless Positioning using LSTM Neural Network: Preliminary Results," in *IEEE 100th VTC2024-Fall*, 2024, pp. 1–5.

[9] A. Ip, L. Irio, and R. Oliveira, "Vehicle Trajectory Prediction based on LSTM Recurrent Neural Networks," in *IEEE 93rd VTC2021-Spring*, 2021, pp. 1–5.

[10] L. Rossi, A. Ajmar, M. Paolanti, and R. Pierdicca, "Vehicle trajectory prediction and generation using LSTM models and GANs," *Plos one*, vol. 16, no. 7, 2021.

[11] F. Jin, K. Liu, C. Liu, T. Cheng, H. Zhang, and V. C. Lee, "A Cooperative Vehicle Localization and Trajectory Prediction Framework Based on Belief Propagation and Transformer Model," *IEEE TCE*, vol. 70, no. 1, pp. 2746–2758, 2024.

[12] J. Liu and G. Guo, "Vehicle Localization During GPS Outages with Extended Kalman Filter and Deep Learning," *IEEE TIM*, vol. 70, pp. 1–10, 2021.

[13] Y. Ji, C. Yu, and B. D. Anderson, "Geometric Dilution of Localization and Bias-Correction Methods," in *IEEE ICCA*, 2010, pp. 578–583.

[14] F. J. Niroumand, P. Ansari Bonab, and A. Sargolzaei, "Security of Connected and Autonomous Vehicles: A Review of Attacks and Mitigation Strategies," in *IEEE SoutheastCon*, 2024, pp. 1197–1204.

[15] R. Schmidt, "Multiple Emitter Location and Signal Parameter Estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[16] T. S. Rappaport, *Wireless Communications: Principles and Practice*. Cambridge University Press, 2024.

[17] A. Pages-Zamora, J. Vidal, and D. H. Brooks, "Closed-Form Solution for Positioning based on Angle of Arrival Measurements," in *The 13th IEEE PIMRC*, vol. 4, 2002, pp. 1522–1526.

[18] J. Lesouple, C. Baudoin, M. Spigai, and J.-Y. Tourneret, "Generalized isolation forest for anomaly detection," *Pattern Recognition Letters*, vol. 149, pp. 109–119, 2021.

[19] H. V. Singh, A. Girdhar, and S. Dahiya, "A Literature survey based on DBSCAN algorithms," in *6th IEEE ICICCS*, 2022, pp. 751–758.

[20] H. J. Shin, D.-H. Eom, and S.-S. Kim, "One-class support vector machines—an application in machine fault detection and classification," *Computers & Industrial Engineering*, vol. 48, no. 2, pp. 395–408, 2005.

[21] O. Alghushairy, R. Alsini, T. Soule, and X. Ma, "A Review of Local Outlier Factor Algorithms for Outlier Detection in Big Data Streams," *Big Data and Cognitive Computing*, vol. 5, no. 1, p. 1, 2020.

[22] G. Welch, G. Bishop *et al.*, "An Introduction to the Kalman Filter," 1995.

[23] M. I. Ribeiro, "Kalman and Extended Kalman Filters: Concept, Derivation and Properties," *Institute for Systems and Robotics*, vol. 43, no. 46, pp. 3736–3741, 2004.

[24] M. Treiber, A. Hennecke, and D. Helbing, "Congested Traffic States in Empirical Observations and Microscopic Simulations," *Physical Review E*, vol. 62, no. 2, p. 1805, 2000.

[25] R. B. Langley *et al.*, "Dilution of Precision," *GPS world*, vol. 10, no. 5, pp. 52–59, 1999.