WIP: Modeling and Detecting Falsified Vehicle Trajectories Under Data Spoofing Attacks

Jun Ying
Purdue University
ying29@purdue.edu

Yiheng Feng
Purdue University
feng333@purdue.edu

Qi Alfred Chen
University of California at Irvine
alfchen@uci.edu

Z. Morley Mao
University of Michigan
zmao@umich.edu

***Abstract*—Connected Vehicle (CV) and Connected and Autonomous Vehicle (CAV) technologies can greatly improve traffic efficiency and safety. Data spoofing attack is one major threat to CVs and CAVs, since abnormal data (e.g., falsified trajectories) may influence vehicle navigation and deteriorate CAV/CV-based applications. In this work, we aim to design a generic anomaly detection model which can be used to identify abnormal trajectories from both known and unknown data spoofing attacks. First, the attack behaviors of two representative known attacks are modeled. Then, Using driving features derived from transportation and vehicle domain knowledge, an anomaly detection framework is proposed. The framework combines a feature extractor and an anomaly classifier trained with known attack trajectories and can be applied to identify falsified trajectories generated by various attacks. In the numerical experiment, a highway segment with a signalized intersection is built in the V2X Application Spoofing Platform (VASP). To evaluate the generality of the proposed anomaly detection algorithm, we further tested the proposed model with several unknown attacks provided in VASP. The results indicate that the proposed model achieves high accuracy in detecting falsified attack trajectories from both known and unknown attacks.**

I. Introduction

The implementation of connected and autonomous vehicle (CAV) and connected vehicle (CV) technologies can greatly benefit the transportation system by improving safety and efficiency. CVs report their status (e.g., location and speed) to infrastructure, which utilizes the received vehicle trajectories for various safety and mobility applications. For example, CV-based adaptive signal control can effectively reduce intersection congestion [17]. CAVs also can utilize information from other vehicles or infrastructure to improve situation awareness (e.g. help detect occluded and far away objects) [21].

To guarantee the efficiency and safety of CAV and CV applications, it is crucial that the received data needs to be authentic. However, falsified data can be generated easily using various attack methods. Based on the purposes, the attack methods can be roughly categorized into two types. In the first type, the attacker generates falsified data to deteriorate CV/CAV-based applications through V2X communications such as Cooperative Adaptive Cruise Control (CACC) [13], CV/CAV-based signal control systems [4], and CV/CAV-based safety warning systems [20]. The above-mentioned methods primarily aim to disrupt V2X-based applications, resulting in reduced efficiency and safety. In the second category, the attacker spoofs the vehicle's onboard sensors, for example, GPS [9] [7] and LiDAR [3] to mislead vehicle's localization and navigation. Multi-Sensor Fusion (MSF) algorithms [15] [8], which take data from multiple sensors, are usually considered as one way to defend data spoofing attacks. However, a recent study from Shen et al. [12] found that by spoofing the GPS data only, the error in the MSF-based localization module could grow exponentially.

To identify data spoofing attacks, various detection methods have been proposed. Consistency and plausibility checks are two widely used methods. Multiple plausibility check based algorithms have been developed to check the consistency between vehicle's speed and location [11] [14]. However, the above-mentioned methods cannot identify trajectory spoofing attacks designed considering vehicle dynamics and kinematics. Wong et al. [16] proposed a hierarchical detection framework to detect falsified vehicle trajectories, considering vehicle dynamic boundaries, kinematic relationships, and the overlap among trajectories. The proposed method may fail to identify sophisticated falsified trajectories generated considering vehicle dynamics and traffic flow properties. To defend against more complicated attacks, Huang et al. [5] proposed an anomaly trajectory detection algorithm to defend against falsified trajectories generated by optimization models (i.e., consider vehicle dynamic and car-following behavior). Yang et al. [18] proposed a GPS spoofing attack detection framework that achieves high accuracy in identifying GPS spoofing attacks toward the MSF module in autonomous vehicles. However, this method (along with most existing anomaly detection methods), is mainly designed to distinguish specific attacks while the performance in detecting unknown attacks is limited.

In this paper, we aim to design a generic anomaly detection model that can be used to identify abnormal trajectories from both known and unknown attacks. Known attacks refer to the attacks that have been previously launched and the abnormal patterns are known and used to train the anomaly detection model. In comparison, unknown attacks refer to attacks that may happen in the future and their abnormal patterns are

Symposium on Vehicles Security and Privacy (VehicleSec) 2024
26 February 2024, San Diego, CA, USA
ISBN 979-8-9894372-7-6
https://dx.doi.org/10.14722/vehiclesec.2024.23035
www.ndss-symposium.org

unknown. Two existing attack models (i.e., ETA attack and MSF attack) are selected and modeled as known attacks. We call both attacks "sophisticated" because the physical boundaries of vehicle kinematic parameters, vehicle kinematic motion consistency, and traffic flow properties are considered in the attack models, which greatly increase the difficulties in detection. The ETA attack generates falsified trajectories that do not obey car-following rules [4]. The MSF attack generates falsified trajectories that fluctuate around the lane center and gradually deviate from the road [12]. These two attacks represent longitudinal and lateral abnormal behaviors from the first and second attack categories respectively and thus are selected. To generate the ETA attack trajectory, we propose an optimization-based model, considering vehicle dynamics and neighboring vehicle position. We further propose an innovative modeling paradigm to mimic the Multi-Sensor Fusion (MSF) attack behavior and generate falsified trajectories with similar lateral deviation patterns and achieve the same attack goal. Then a candidate feature set is derived from transportation and vehicle domain knowledge and a greedy algorithm is applied to select representative features from the candidate feature set. Based on the selected features, several machine learning-based classifiers are trained to learn from the two representative attack trajectories. Given an observed trajectory, a feature extractor is applied first to obtain critical features, and the pretrained anomaly classifier is utilized to distinguish whether the trajectory is abnormal. The proposed algorithm achieves high accuracy (97.70%) in detecting MSF and ETA attacks, with low false positive and low false negative rates and 100% accuracy in detecting other unknown attacks from the VASP platform. The main contributions are listed as follows:

(1) A generic detection framework is proposed. The detection framework combines a feature extractor and a classifier. The feature extractor can be customized based on various driving scenarios such as highways, and signalized intersections. The detection framework is trained with two known attacks but can detect other unknown attacks.

(2) We propose a feature set that effectively represents normal driving behavior using transportation and vehicle domain knowledge. The proposed feature set can be combined with different machine learning-based classifiers, such as SVM, random forest, and decision tree, and achieves high accuracy in detecting both known and unknown attacks.

(3) We propose a new paradigm to model cyber attacks that can accurately represent vehicle-level attack behaviors without constructing complicated attack pipelines. Compared with the original MSF attack model in [12], our proposed method needs much less computation resources and only needs information from the GPS receivers.

The remainder of the paper is arranged as follows. Section II introduces how the ETA and MSF attacks are modeled. Section III presents the detection algorithm. Numerical experiments are introduced in section IV. Section V concludes the work and lays out future research directions.

II. Attack Modeling

In this section, two approaches are introduced to model the ETA attack and MSF attack.

A. Modeling Estimated Time of Arrival (ETA) Attack.

1) Attack Introduction: The ETA attack has been proven to be a threat in CV-based traffic control systems (CV-TSC), for instance, the I-SIG system [6]. **Figure 1** denotes the attack concept. The figure contains two types of vehicles, the vehicles under attack (red vehicle), and the normal vehicles (yellow vehicle). Both vehicles are CV and broadcast BSMs, including the vehicle's current speed and position to infrastructure. The signal timing plan is broadcast through Signal Phasing and Timing (SPaT) messages to the CVs. When the vehicle is not under attack, its longitudinal movement follows a certain car following pattern, for example, Intelligent Driver Model (IDM), and sends out BSMs the same as the ground truth trajectory. When the vehicle is under attack, its actual movement still follows the same car-following model, as shown in the blue color. However, it sends out falsified BSMs to fulfill the attack objective, for example, decelerating without apparent reason as shown in the red color. The ETA of the falsified BSM trajectory is then longer than the ground truth. The falsified ETA will mislead the infrastructure to generate a non-optimal signal timing plan, for example, unnecessary extension for the current green phase. To model the falsified trajectory under the ETA attack, an optimization model is formulated as shown in **Equation 1**. The generated vehicle trajectory is a sequence of trajectory points at consecutive time steps. At each time step, the trajectory point contains 5 elements, including x_t,y_t,v_t,a_t,ψ_t, which are variables of the optimization problem. x_t,y_t denote the vehicle's longitudinal and lateral position at time step t. v_t,a_t,ψ_t are vehicle kinematic-related variables, denoting the speed, acceleration and heading angle at time step t. The length of the trajectory is denoted as N, which is decided by the signal update time.

$$\begin{aligned} \min_{x_t,y_t,v_t,a_t,\psi_t} \quad & \theta^T f(x_t, y_t, v_t, a_t, \psi_t, \mathbf{e}_t) \\ \text{s.t.} \quad & \text{vehicle safety constraints (\textbf{Equations 8-9})} \\ & \text{vehicle dynamics constraints} \end{aligned} \tag{1}$$

In the objective function, f is the function of features extracted from the generated trajectory. $\mathbf{e}_t$ denotes the environmental parameters, including the road heading, the leading and following vehicles' current state and predicted future state. θ is the weight vector of the selected features.

2) Objective Function: The objective function is designed to achieve two goals: 1) generate a falsified trajectory with a specified longer ETA than ground truth, by modifying vehicle's speed and position. 2) generate a falsified trajectory close to normal driving behavior. To achieve the attack goal, we selected 5 representative driving features, representing normal car following behavior, trajectory smoothness, and desired ETA. The selected features are demonstrated as follows:

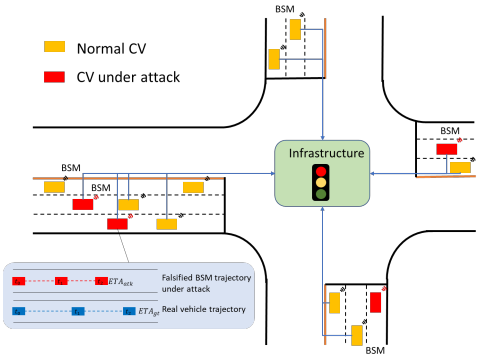

Fig. 1. ETA Threat Model

$$f_1 = \frac{1}{N}\sum_{t=1}^{N} a_t^2 \tag{2}$$

$$f_2 = \frac{1}{N-1}\sum_{t=1}^{N-1}\left(\frac{\psi_{t+1}-\psi_t}{\Delta t}\right)^2 \tag{3}$$

$$f_3 = \frac{1}{N}\sum_{t=1}^{N}(\psi_{t+1}-\psi^{\text{road}})^2 \tag{4}$$

$$f_4 = \frac{1}{N}\sum_{t=1}^{N}(d_t^{\text{des}}-d_t)^2 = \frac{1}{N}\sum_{t=1}^{N}(v_t\cdot t_{\text{headway}}+d_s-d_t)^2 \tag{5}$$

$$f_5 = \frac{d_N^{\text{stop}}}{v_N} - \text{ETA}^{\text{des}} \tag{6}$$

Equations 2-4 denote the smoothness of the generated trajectory. f_1 penalizes large acceleration and deceleration. Δt is the time interval between two consecutive time steps. f_2 calculates vehicle's heading rate. f_3 denotes the difference between the vehicle's heading and road heading. f_2 penalizes large deviations in the vehicle heading and f_3 avoids the vehicle deviating from the road. **Equation 5** calculates the difference between the desired space headway (d_t^{des}) and the actual space headway (d_t) at time step t. $t_{headway}$ denotes the desired time headway. d_s is the minimum safety gap between the two vehicles. **Equation 6** denotes the difference between the ETA of the generated trajectory and the desired ETA at the end of the planning horizon. d_N^{stop} is the distance to the stop bar at time step N. The desired ETA is calculated in **Equation 7**, which is the current ETA and a constant offset.

$$\text{ETA}^{\text{des}} = \text{ETA}^{\text{init}} + \text{ETA}^{\text{offset}} = \frac{d_0^{\text{stop}}}{v_0} + \text{ETA}^{\text{offset}} \tag{7}$$

3) Vehicle Safety and Dynamics Constraints: To generate a trajectory as realistic as possible, **Equations 8-9** are applied to guarantee the minimum space headway between the falsified trajectory and the leading/following vehicle position. Otherwise, the generated trajectory can be easily identified by cross validation using neighboring vehicle trajectories.

$$(d_i^{\text{follow\_travel}} + d_s) - \delta_i^{\text{fo}}\cdot M \le d_i^{\text{travel}}, \quad i \in (1,2,\ldots,N) \tag{8}$$

$$d_i^{\text{travel}} \le (d_i^{\text{front\_travel}} - d_s) + \delta_i^{\text{fr}}\cdot M, \quad i \in (1,2,\ldots,N) \tag{9}$$

M is a constant large number. δ_i^{fr} and δ_i^{fo} are two binary parameters indicating the existence of the leading and following vehicles. **Equations 8-9** are valid when the leading/following vehicle exists. d_i^{travel}, $d_i^{\text{front\_travel}}$, and $d_i^{\text{follow\_travel}}$ denote the traveled distance of the ego/leading/following vehicle at time step i. **Equations 8-9** guarantee the safety distance between the ego vehicle and the leading/following vehicle in the same lane and forbid the ego vehicle from exceeding the leading vehicle or overlapping with the following vehicle.

Vehicle dynamics constraints represent the physical limits of variables and their kinematic motion relations. Details of these constraints can be found in previous studies such as [19].

4) Example attack trajectories: **Figure 2** denotes the time-space diagram of the ego vehicle, leading vehicle and following vehicle trajectories. The x-axis denotes the time steps, and the y-axis denotes the traveled distance. The red curve denotes the BSM trajectory sent by the ego vehicle when it is under attack, the dashed red curve denotes the ego vehicle's actual trajectory. The black and blue curves denote the trajectories of the leading and following vehicles. The figure demonstrates that the attack trajectory decelerates when it is still far from the leading vehicle in order to fulfill the ETA attack goal. Besides, the figure also demonstrates that the attacked trajectory keeps safe distances from the neighboring vehicles, increasing difficulty for detection.

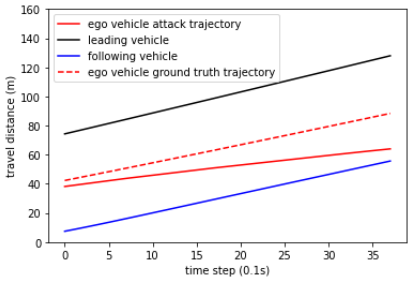

Fig. 2. Ego vehicle, leading vehicle and following vehicle trajectories

B. Modeling Multi-Sensor Fusion (MSF) Attack

MSF algorithms are widely used on AVs [15] [8]. It combines input from multiple sensors, including GPS, IMU (Inertial Measurement Unit) and LiDAR to provide precise localization results. MSF algorithms are usually considered to be one defense method for spoofing attacks because it is highly unlikely that all sensors are compromised at the same time. However, recent study proposed by Shen et al. [12] indicates that by modifying the output from GPS receiver only, the performance of the MSF algorithm is deteriorated. The MSF attack contains two stages, the vulnerability profiling stage, and the aggressive spoofing stage. In vulnerability profiling, constant deviations are added to the original GPS data to identify vulnerabilities. After the vulnerability is found, in the aggressive spoofing stage, exponentially growing deviation is added. The deviation added to the original GPS data causes

large deviations in the MSF output and causes the AV driving off the road. The proposed MSF algorithm [12] is evaluated using LGSVL simulator [10], an autonomous vehicle simulator with Apollo 5.0 as the autonomous driving platform [2]. The entire simulation process is time-consuming and complicated, making it challenging to scale up.

Since our primary objective is to model the attack's impact on safety and mobility at the transportation application level, which is mainly reflected by the falsified trajectories. As a result, we only need to model the trajectory level attack behaviors without replicating the complicated attack pipeline. For trajectory level attack behavior, the MSF attack generates falsified trajectories with lateral deviations increasing exponentially [12]. Such lateral deviation can cause vehicles to cross the lane boundaries, forcing neighboring vehicles to decelerate and potentially lead to crashes. By analyzing the falsified trajectories, we observe that the lateral deviations in the vulnerability profiling stage follow the Gaussian distribution. The lateral deviations in the aggressive spoofing stage follow a family of exponential functions, with different parameters in different driving scenarios. Similar to the original attack model, we also model the lateral deviations in two stages.

1) Vulnerability profiling stage modeling: Falsified trajectories in the vulnerability profiling stage are modeled considering the lateral deviation from the ground truth and attack duration. The lateral deviation follows the Gaussian distribution with mean equals to -0.020 and standard deviation equals to 0.048, as shown in **Figure 3a**. The majority of lateral deviation ranges between (-0.2m, 0.2m), with most of the lateral deviation being close to zero.

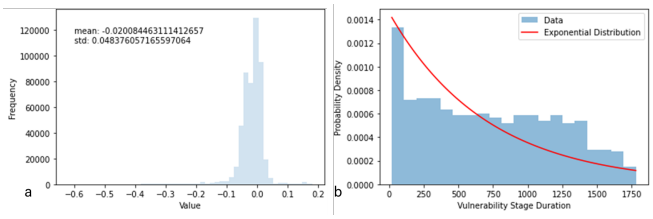

Fig. 3. Vulnerability Profiling Stage Lateral Deviation Distribution

The vulnerability profiling stage duration is modeled as an exponential distribution, as shown in **Figure 3b**. The fitted exponential function is represented in **Equation 19**, with the scale equals to 705.15. As shown in the figure, the exponential function doesn't fit the data perfectly. However, lateral deviation in the vulnerability profiling stage only deviates the vehicle from the lane center within a limited range and does not cross the lane boundary. The influence of the attack vehicle to other vehicles in the neighboring lane can be ignored. As a result, the error between the data and the fitted exponential curve is acceptable.

$$f(x) = \frac{1}{\text{scale}} \cdot e^{-\frac{x}{\text{scale}}} \tag{10}$$

2) Aggressive spoofing stage modeling: The lateral deviation of the aggressive spoofing stage trajectory follows exponential growth. **Equation 20** is applied to fit the curve. x represents the x_{th} attack point and $f(x)$ denotes the lateral deviation of the x_{th} point. a,b,c are function parameters. **Figure 4** demonstrates six types of lateral deviation profiles from the original attack study and the fitted curve in the same type. The x-axis denotes the time step. The y-axis denotes the lateral deviation. By analyzing trajectories collected from different driving scenarios, it is revealed that the lateral deviation pattern remains consistent for trajectories within the same scenario but varies among different driving scenarios. Therefore, for trajectories in each scenario, we use an exponential function to fit all the lateral deviation profiles. Trajectories in different scenarios are fitted separately, as shown in **Figure 4**.

$$f(x) = c \cdot a^x + b \tag{11}$$

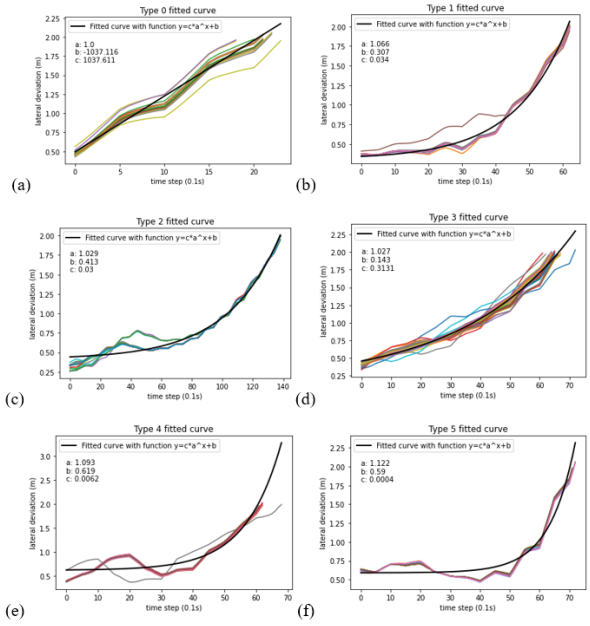

Fig. 4. Lateral Deviation Profile and Fitted Exponential Curve

III. DETECTION METHODOLOGY

A. Problem Statement

An anomaly detection framework is designed to distinguish falsified trajectories. The proposed detection framework is demonstrated in **Figure 5**. The framework combines a feature extractor and a machine learning algorithm based anomaly classifier. Two types of falsified trajectories modeled in the previous section, are used to train the anomaly classifier. Given the collected ETA attack trajectories (D_E) and MSF attack trajectories (D_M), a feature extractor is applied to select representative driving behavior related features from the trajectories. Then, the selected features are applied to train the classifier to distinguish between normal and abnormal trajectories. Given an observed trajectory, the same feature extractor is applied to identify critical features f_{D_o}, which

are fed into the anomaly classifier to distinguish whether the observed trajectory is attacked or not.

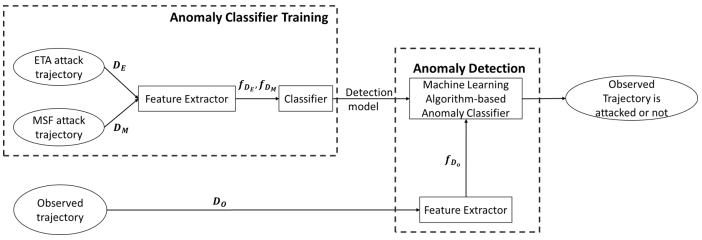

Fig. 5. Anomaly Detection Framework

B. Feature extractor

Thirteen features are designed to capture various aspects of normal driving, considering both longitudinal and lateral driving behaviors. The features are described as follows:

1) Mean acceleration (absolute value) $f_1 = \frac{1}{N}\sum_i |a_i|$. f_1 calculates the average absolute acceleration value.

2) Mean heading rate (absolute value) $f_2 = \frac{1}{N}\sum_i |(\dot{\psi}_i)|$. f_2 measures the smoothness of the vehicle's heading angle. The change in the heading angle should be small during normal car following and lane-changing maneuvers.

3) Mean car following distance difference (absolute value) $f_3 = \frac{1}{N}\sum_i |d_{i,des} - d_{i,act}| = \frac{1}{N}\sum_i |d_s + v_i \cdot t_{\text{headway}} - d_{i,act}|$. f_3 measures the fluctuation of the difference between $d_{i,des}$ and $d_{i,act}$.

4) Mean speed $f_4 = \frac{1}{N}\sum_i v_i$. f_4 calculates the average driving speed.

5) Standard deviation of speed $f_5 = \sqrt{\frac{\sum_i (v_i - \bar{v})^2}{N-1}}$. f_5 calculates the standard deviation. f_4,f_5 denotes the distribution of vehicle speed.

6) Mean acceleration $f_6 = \frac{1}{N}\sum_i a_i$. f_6 calculates average acceleration.

7) Standard deviation of acceleration $f_7 = \sqrt{\frac{\sum_i (a_i - \bar{a})^2}{N-1}}$. f_1,f_6,f_7 are vehicle acceleration related features. f_1 denotes the magnitude of vehicle acceleration while f_6,f_7 denotes the acceleration's distribution.

8) Mean heading rate $f_8 = \frac{1}{N}\sum_i (\dot{\psi}_i)$. f_8 denotes the average heading rate.

9) Standard deviation of heading rate $f_9 = \sqrt{\frac{\sum_i \left((\dot{\psi}_i - \bar{\dot{\psi}})^2\right)}{N-1}}$. f_8,f_9 denotes the distribution of heading rate.

10) Mean car following distance difference $f_{10} = \frac{1}{N}\sum_i (d_{i,des} - d_{i,act})$. f_{10} calculates the average difference between desired car following distance and the actual car following distance.

11) Standard deviation of car following distance difference. $f_{11} = \sqrt{\frac{\sum_i \left((d_i^{\text{diff}} - \bar{d_i^{\text{diff}}})^2\right)}{N-1}}$. $d_i^{\text{diff}} = d_{i,\text{des}} - d_{i,\text{act}}$. f_{10},f_{11} denotes the car following distance distribution.

12) Maximum speed $f_{12} = max_i v_i$. f_{12} calculates the maximum speed.

13) Maximum acceleration $f_{13} = max_i a_i$. f_{13} calculates the maximum acceleration.

To avoid overlapping among designed features, an iterative greedy algorithm is applied to select the most significant features from the feature list. Details of the greedy algorithm description can be found in [20]. The proposed feature set can be integrated with multiple machine learning based algorithms, for example, SVM classifier, decision tree classifier and random forest classifier. The structure of the classifiers is different and the features need to be selected separately for each classifier using greedy algorithm. Features f_2,f_4,f_7,f_{13} are selected for SVM based anomaly classifier. Features f_4,f_9 are selected for both the random forest classifier and the decision tree classifier. The selected features for the above mentioned classifiers include both longitudinal (e.g., car-following) and lateral (e.g., lane changing) behaviors.

IV. NUMERICAL EXPERIMENTS

To validate the proposed anomaly detection framework, a highway segment with a signalized intersection is built to generate the attack trajectories in the V2X Application Spoofing Platform (VASP) [1]. VASP is an open-source framework developed by Qualcomm to simulate attacks on V2X networks and applications. VASP is selected as the test platform since it contains 68 typical V2X attacks including attacks based on kinematic values and attacks on V2X applications, which are considered as unknown attacks developed by a third-party to test the proposed anomaly detection model. ETA attack and MSF attack models in Sections II are also integrated into the VASP to generate falsified trajectories for training.

A. Detection framework evaluation

To evaluate the anomaly detection framework, 1000 ground truth trajectories are collected by running the simulation without implementing attack. 276 ETA attack trajectories and 464 MSF attack trajectories are generated. Both the attack and ground truth trajectories are collected with a frequency of 10Hz. The collected trajectories are divided into training and testing sets, with proportions of 80% and 20%. The detection results are demonstrated as follows:

Anomaly detection starts when the entire trajectory is observed. The proposed detection framework achieves 97.70% accuracy for SVM based classifier. 196/203 of the ground truth trajectories and 144/145 of the attack trajectories are identified correctly. The false positive rate equals 3.45% and the false negative rate equals 0.69%. The accuracy for the decision tree classifier and random forest classifier is 100.00%. **Table I** denotes the details of the detection performance, using detection rate and false alarm rate. The results indicate that the proposed algorithm performs well in distinguishing between normal and abnormal trajectories. To further evaluate the generality of the proposed model, we selected 6 attacks provided in the VASP platform, whose falsified trajectories are not trained by the proposed model (i.e., unknown). The selected attacks are random position attack, random position offset attack, high acceleration attack, low speed attack, Braking from communication range attack and Electronic Emergency Brake Light (EEBL) attack, including attacks focus on vehicle

TABLE I
PERFORMANCE OF ANOMALY DETECTION

Classifier	FP	FN	TP	TN	Detection Rate	False Alarm Rate
SVM	7	1	144	196	144/145	7/203
Decision Tree	0	0	145	203	145/145	203/203
Random Forest	0	0	145	203	145/145	203/203

TABLE II
DETECTION ACCURACY OF ATTACKS ON VASP

Attack type	Attack Trajectory Number	SVM Accuracy	Decision Tree Accuracy	Random Forest Accuracy
Random Position	246	1.0	1.0	1.0
Random Position Offset	216	1.0	1.0	1.0
High Acceleration	225	1.0	1.0	1.0
Low Speed	225	1.0	1.0	1.0
Braking from Communication Range	990	1.0	1.0	1.0
EEBL	1055	1.0	1.0	1.0

position, speed, acceleration and V2X application. Details of the selected attacks can be found in (24).

Table II denotes the detection accuracy of attack trajectories generated by VASP. The results indicate that the proposed detection algorithm achieves high accuracy in detecting VASP attack trajectories, even though they are not used for training (i.e., unknown attacks). The reason is that the proposed model captures normal driving behaviors well. Therefore, as long as the VASP attacks change normal driving behavior, the proposed algorithm can be applied to detect the anomaly.

V. FUTURE RESEARCH

To complete this work, two types of baseline models will be added for comparison in the future, including plausibility check-based models and neural network-based models. Plausibility check has been widely used for anomaly detection and the model proposed by So et al. [14] will be used as one of the baselines. In this work, the authors selected six features to check the consistency between vehicle acceleration, speed, and position. It is interesting to see whether the plausibility check based models are able to detect sophisticated attacks (e.g., ETA and MSF attacks).

To show the importance of integrating domain knowledge with the feature selection, we also plan to compare the proposed framework with a few neural network-based anomaly detection models, where features are selected by the neural networks and may not have any physical meanings. In addition, whether neural network-based models can detect unknown attacks remains uncertain, since the features extracted from the training data (i.e., known attacks) may not represent the abnormal patterns in unknown attacks.

ACKNOWLEDGMENT

This research is supported in part by the U.S. National Science Foundation through Grant CNS-1930041, CNS-1929771, CNS-2145493 and USDOT UTC Grant 69A3552348327. The authors would like to thank Dr. Jonathan Petit and Mr. Raashid Ansari from Qualcomm for their technical support in the VASP platform. The views presented in this paper are those of the authors alone.

REFERENCES

[1] M. R. Ansari, J. Petit, J.-P. Monteuuis, and C. Chen, "Vasp: V2x application spoofing platform," in *Proceedings Inaugural International Symposium on Vehicle Security Privacy, ndss-symposium*, 2023. [Online]. Available: https://doi.org/10.14722/vehiclesec.2023.23071

[2] Apollo Auto, "Apollo: an open autonomous driving platform," https://github.com/ApolloAuto/apollo, [Internet]. Apollo Auto; 2023 [cited 2023 Jul 24].

[3] Y. Cao, C. Xiao, B. Cyr, Y. Zhou, W. Park, S. Rampazzi, Q. A. Chen, K. Fu, and Z. M. Mao, "Adversarial sensor attack on lidar-based perception in autonomous driving," in *Proceedings of the 2019 ACM SIGSAC conference on computer and communications security*, 2019, pp. 2267–2281.

[4] Q. A. Chen, Y. Yin, Y. Feng, Z. M. Mao, and H. X. Liu, "Exposing congestion attack on emerging connected vehicle based traffic signal control." in *NDSS*, 2018.

[5] S. E. Huang, Y. Feng, and H. X. Liu, "A data-driven method for falsified vehicle trajectory identification by anomaly detection," *Transportation research part C: emerging technologies*, vol. 128, p. 103196, 2021.

[6] S. E. Huang, W. Wong, Y. Feng, Q. A. Chen, Z. M. Mao, and H. X. Liu, "Impact evaluation of falsified data attacks on connected vehicle based traffic signal control," *arXiv preprint arXiv:2010.04753*, 2020.

[7] M. Kamal, A. Barua, C. Vitale, C. Laoudias, and G. Ellinas, "Gps location spoofing attack detection for enhancing the security of autonomous vehicles," in *2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall)*. IEEE, 2021, pp. 1–7.

[8] Q. Li, J. P. Queralta, T. N. Gia, Z. Zou, and T. Westerlund, "Multi-sensor fusion for navigation and mapping in autonomous vehicles: Accurate localization in urban environments," *Unmanned Systems*, vol. 8, no. 03, pp. 229–237, 2020.

[9] S. Narain, A. Ranganathan, and G. Noubir, "Security of gps/ins based on-road location tracking systems," in *2019 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2019, pp. 587–601.

[10] G. Rong, B. H. Shin, H. Tabatabaee, Q. Lu, S. Lemke, M. Možeiko, E. Boise, G. Uhm, M. Gerow, S. Mehta *et al.*, "Lgsvl simulator: A high fidelity simulator for autonomous driving," *arXiv preprint arXiv:2005.03778*, 2020.

[11] S. Ruj, M. A. Cavenaghi, Z. Huang, A. Nayak, and I. Stojmenovic, "On data-centric misbehavior detection in vanets," in *2011 IEEE Vehicular Technology Conference (VTC Fall)*. Ieee, 2011, pp. 1–5.

[12] J. Shen, J. Y. Won, Z. Chen, and Q. A. Chen, "Drift with devil: Security of {Multi-Sensor} fusion based localization in {High-Level} autonomous driving under {GPS} spoofing," in *29th USENIX Security Symposium (USENIX Security 20)*, 2020, pp. 931–948.

[13] P. K. Singh, G. S. Tabjul, M. Imran, S. K. Nandi, and S. Nandi, "Impact of security attacks on cooperative driving use case: Cacc platooning," in *TENCON 2018-2018 IEEE Region 10 Conference*. IEEE, 2018, pp. 0138–0143.

[14] S. So, P. Sharma, and J. Petit, "Integrating plausibility checks and machine learning for misbehavior detection in vanet," in *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2018, pp. 564–571.

[15] G. Wan, X. Yang, R. Cai, H. Li, Y. Zhou, H. Wang, and S. Song, "Robust and precise vehicle localization based on multi-sensor fusion in diverse city scenes," in *2018 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2018, pp. 4670–4677.

[16] W. Wong, S. Huang, Y. Feng, Q. A. Chen, Z. M. Mao, and H. X. Liu, "Trajectory-based hierarchical defense model to detect cyber-attacks on transportation infrastructure," Tech. Rep., 2019.

[17] Z. Yang, Y. Feng, and H. X. Liu, "A cooperative driving framework for urban arterials in mixed traffic conditions," *Transportation research part C: emerging technologies*, vol. 124, p. 102918, 2021.

[18] Z. Yang, J. Ying, J. Shen, Y. Feng, Q. A. Chen, Z. M. Mao, and H. X. Liu, "Anomaly detection against gps spoofing attacks on connected and autonomous vehicles using learning from demonstration," *IEEE Transactions on Intelligent Transportation Systems*, 2023.

[19] J. Ying and Y. Feng, "Full vehicle trajectory planning model for urban traffic control based on imitation learning," *Transportation research record*, vol. 2676, no. 7, pp. 186–198, 2022.
[20] J. Ying, Y. Feng, Q. A. Chen, and Z. M. Mao, "Gps spoofing attack detection on intersection movement assist using one-class classification," in *ISOC Symposium on Vehicle Security and Privacy*, 2023.
[21] X. Zhang, A. Zhang, J. Sun, X. Zhu, Y. E. Guo, F. Qian, and Z. M. Mao, "Emp: Edge-assisted multi-vehicle perception," in *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*, 2021, pp. 545–558.