A Compensation Estimation Algorithm for Multi-Sensor Attack Detection in Autonomous Driving

CHEN Xingzhou

September 22, 2025

Abstract

We investigate the impact and counter-measures of potential sensor attacks on autonomous vehicles. The attacker can inject a certain bias or increase the random noise of the positioning measurement. Since the attack signal is stealthy sometimes, it is difficult to distinguish the system error and the attack signal. To tackle this, we propose a novel Kalman-based framework to compensate for and detect attack signals on the multi-sensor system. Simulation results show that the proposed algorithm has a more flexible and convenient threshold selection than conventional detection methods. Besides, it mitigates the impact of the bad sensory data on the system before we determine the attacked sensor.

Keywords: autonomous vehicles, multi-sensor fusion, state estimate, sensor attack

Contents

1 Introduction

1.1 Background

Autonomous vehicles (AVs) have recently attracted much attention from academic and industrial domains. Today, many automobile manufacturers and IT companies, such as Google and Tesla, have implemented aggressive plans to develop AV technology. Before AV technology matures and eventually becomes available on a large scale, researchers and engineers should address its security issues well. After all, for autonomous and conventional vehicles, safety is one of the most critical issues in various vehicle applications. To benefit intelligent transportation, autonomous vehicle systems utilize multiple sensors, shown in Figure.1, such as GPS and LiDAR, to localize themselves and perceive the environment. However, this increases the risk of being attacked by malicious cyber attacks. When attacked, the vehicle may perform abnormal behavior with devastating consequences or even catastrophic accidents.

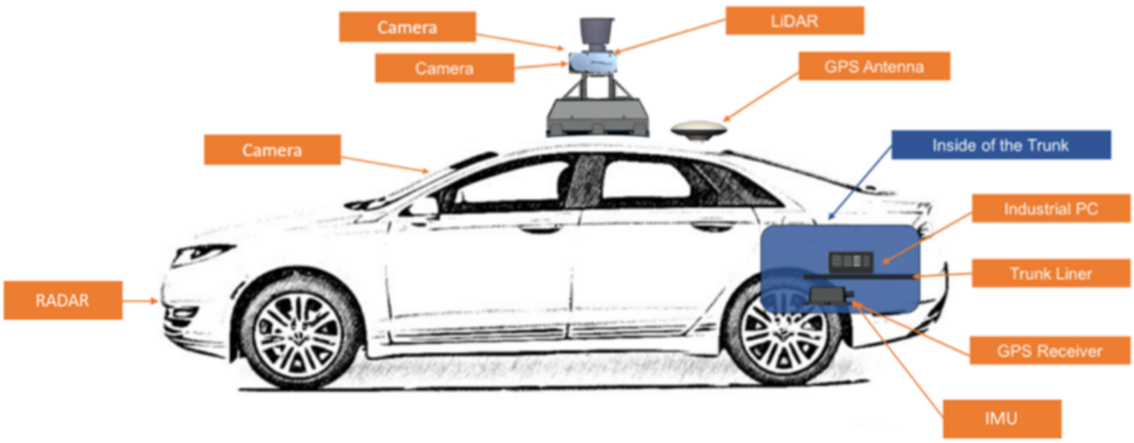

Figure 1: Important Sensors on Autonomous Vehicles [1]

1.2 Related works

Some studies have demonstrated the possibility of sensor attacks on autonomous vehicles. For example, GPS spoofing can fake GPS readings [2]. Besides, spoofing attacks on LiDAR can manipulate point clouds by erasing natural obstacles or adding false obstacles [3]. In addition, optical flow sensors, like cameras, can be tricked by spoofing attacks [4]. Moreover, the widely used robotics middleware suite, Robot Operating System (ROS), has proven vulnerable to cyber attacks such that the sensor readings can tamper with [5]. Therefore, developing relevant algorithms to protect vehicles from sensors in real time is significant.

For the cyber-security threats on autonomous vehicles, various mitigation strategies have been proposed in the past decade, especially during the past five years. Generally, approaches can be classified into two categories, information-oriented and control-oriented. Information-oriented methods apply data security techniques, such as encryption, user authentication, etc., to achieve information security [6–8]. However, these ignore the information interaction and potential cyber-attacks between the vehicle and the physical world. As a complement, control-oriented approaches focus on studying how cyber-attacks affect the control system's physical dynamics.

The control-oriented approaches achieve security by analyzing the consequences of attacks based on the vehicle's system model. We can divide existing control-oriented approaches into two classes, data-driven and model-based. The data-driven approach addresses attack detection by incorporating machine learning techniques, comparing online data with previous data records corresponding to a specific attack. Some typical machine learning frameworks, such as SVM [9], CNN [10], and reinforcement learning [11], are proposed to detect anomalies in the overall system. In contrast, model-based approaches detect the deviation of the sensor measurement. Most model-based approaches [12] [13] contain Kalman filter and residual analysis, which compare the actual measurement with the predicted state based on the system model. Conventional model-based approaches have used this detection structure, and its efficacy has been demonstrated in long-term practices.

1.3 Contributions

This report focuses on the detection and compensation of cyber attacks in positioning sensors in autonomous vehicles, particularly GPS and LiDAR, commonly used in autonomous vehicles. In this report, we propose a Kalman-based framework to detect sensor spoofing attacks and compensate for the impact of attacks on state estimation.

The main contribution of this report can be summarized as follows.

1) Based on the theory of the Kalman filter, we propose a novel multi-sensor framework that combines detection and state compensation.

2) According to the different values of χ^2, we innovatively divide the sensor status into normal, suspected, and attacked cases. Unlike traditional detection models, in which it is difficult to choose a suitable threshold that compromises the false positive rates and true negative rates, our approach provides new ideas to solve this problem.

3) The proposed algorithm is validated via simulations under different attack strategies and is shown to achieve a better detection performance compared with conventional binary-status detection methods.

The remainder of the report is organized as follows. In Section 2, we introduce the vulnerable positioning sensors. We describe the formulation of the multi-sensor system model and the attacker model in Section 3. In Section 4, we present our detection and compensation algorithms. In Section 5, we offer simulations to validate the proposed algorithm. The report ends with a conclusion in Section 6.

2 Anomaly Sensors

The primary purpose of a malicious sensor attack is to cause car crashes by spoofing the autonomous vehicle's sensor. There are two main ways to attack positioning sensors [14].

1) tamper with the sensory data via the given bias $y_t^{attack} = y_t^{measurement} + \xi_t^{bias}$

2) disrupt sensor operation by increasing random noise $y_t^{attack} = y_t^{measurement} + \xi_t^{noise}$

For case 1), the autonomous vehicle's system may mistakenly guide the vehicle to an unexpected position. For case 2), the estimation error of the multi-sensor fusion algorithm increases.

This report mainly considers two significant sensors in a multi-sensor fusion algorithm. The sensors are GPS and Lidar, as introduced as follows.

1) GPS: GPS sensors are receivers with antennas that use a satellite-based navigation system to provide position. Shown in Figure.2, GPS spoofing happens when someone uses a radio transmitter to send a counterfeit GPS signal to a receiver antenna to counter a legitimate GPS satellite signal [15].

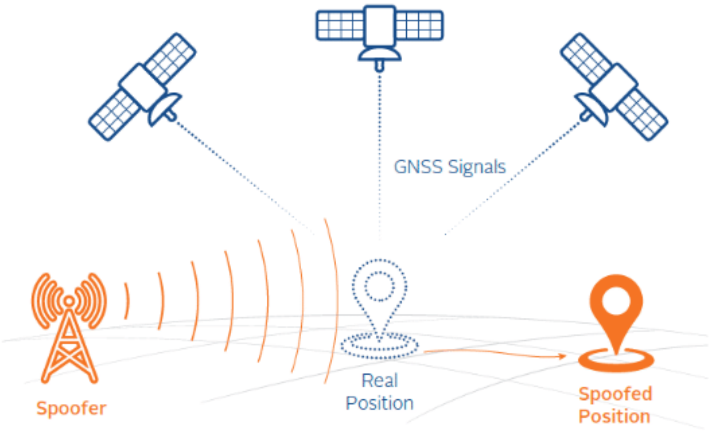

Figure 2: GPS spoofing [16]

2) Lidar: Lidar calculates the vehicle's location by obtaining points cloud map of the surrounding environment. Seen in Figure.3, Lidar spoofing occurs when the photodiode receives the laser pulses from the Lidar and activates the delay component that triggers the attacker laser to simulate real echo pulses [14].

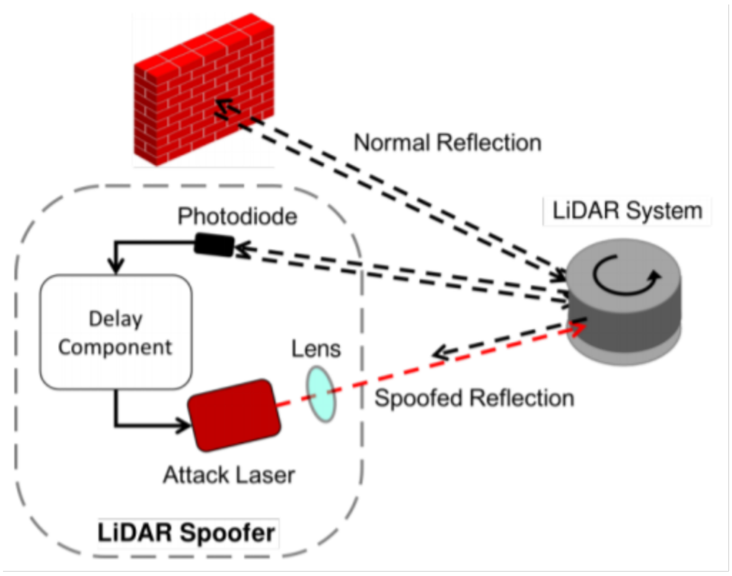

Figure 3: Lidar spoofing [14]

3 System Model

This section introduces system models and multi-sensor fusion algorithms of autonomous vehicles. Then, we present an attack model to describe cyber-attacks imposed on the sensors.

3.1 Vehicle's dynamic model

Vehicle controllers generally have a hierarchical structure in autonomous driving applications, including an upper-level planner and a low-level controller. The upper-level planner generates velocity commands, which are passed to the low-level controller for execution. We assume the upper-level planner performs a kinematic point model to simplify the system model. As this report mainly focuses on sensors related to vehicle localization, the point model is adopted to describe vehicle motion, which is given as follows.

$$x_{t+1} = A_t x_t + B_t u_t + w_t$$

where x_t and u_t represent the position and velocity of the vehicle, respectively. We assume the process noise $\{w_t\}$ has (i.i.d) distribution $\mathbb{N}(0, Q)$ with $Q \succ 0$. To illustrate the main ideas, we consider $A_t = I, B_t = \Delta t \cdot I$.

3.2 Multi-sensor fusion

In real applications, the position of the vehicle can be observed by GPS, LiDAR, IMU, camera, etc., which are corrupted with noise. The measurement equation is given as

$$\begin{bmatrix} y_t^1 \\ \vdots \\ y_t^i \\ \vdots \\ y_t^n \end{bmatrix} = Cx_t + v_t = \begin{bmatrix} C^1 \\ \vdots \\ C^i \\ \vdots \\ C^n \end{bmatrix} x_t + \begin{bmatrix} v_t^1 \\ \vdots \\ v_t^i \\ \vdots \\ v_t^n \end{bmatrix}$$

where y_t^i and v_t^i represent the pose observation and measurement noise of the i^{th} sensor, and n is the number of sensors. Assume that $\{v_t^i\}$ satisfies the i.i.d. distribution $\mathbb{N}(0, R_i)$ with $R_i \succ 0$, and is independent of others. To simplify the model, we consider $C^i = I$.

To compute a state estimate $\widehat{x}_t$, a multi-sensor fusion algorithm is implemented in vehicle based on Kalman filter. The process of state estimate update are as following

$$\hat{x}_{t+1|t} = A_t \hat{x}_{t|t} + B_t u_t \tag{1}$$

$$\hat{x}_{t+1|t+1} = \hat{x}_{t+1|t} + K_{t+1}(\begin{bmatrix} y_t^1 \\ \vdots \\ y_t^i \\ \vdots \\ y_t^n \end{bmatrix} - \begin{bmatrix} C^1 \\ \vdots \\ C^i \\ \vdots \\ C^n \end{bmatrix} \hat{x}_{t+1|t}) \tag{2}$$

with the update of the error covariance matrices

$$P_{t+1|t} = A_t P_{t|t} A_t^T + Q \tag{3}$$

$$P_{t+1|t+1} = P_{t+1|t} - P_{t+1|t} C^T (R + C P_{t+1|t} C^T)^{-1} C P_{t+1|t} \tag{4}$$

where $K_{t+1} = P_{t+1|t} C^T (R + C P_{t+1|t} C^T)^{-1}$, $R = diag(R_1, \ldots, R_i, \ldots, R_n)$.

3.3 Attack model

As GPS and LiDAR are vulnerable to various cyber-attacks, anomalous measurements may occur. Here, we assume that attacks occur over a short period and that the sensors are not attacked simultaneously. The following modified measurement equation is given to describe the i^{th} sensor observation under attack

$$\bar{y}_t^i = y_t^i + \xi_t^i = C^i x_t + v_t^i + \xi_t^i$$

where ξ_t^i represents the deterministic bias or random noise caused by attacks. Consequently, the anomalous observations of GPS and LiDAR will affect state estimate by equation (2). Therefore, the key problem is how to mitigate the impact of sensor attacks on the system while detecting them.

4 Proposed Framework

A model-based framework is proposed to address this problem mentioned in Section 3. First, we introduce why and how we classify the sensors' status into three types. We propose an algorithm that uses information from other sensors to compensate for anomalous sensory data. Then we summarize the framework for anomaly handling and detection.

4.1 Sensor status classification

In previous works, most common and effective detection method is based on analysis of Kalman innovation. The Kalman innovation, $r_t = y_t - C\hat{x}_{t+1|t}$, depicts the difference between expectation and real measurement. In normal case, Kalman innovation sequences $\{r_t\}$ are (i.i.d) with known distribution $\Sigma_{r_t} = R + CP_{t+1|t}C^T$. The probability of the y_t under nominal conditions is

$$\mathbb{P}(y_t) = \left[\frac{1}{\sqrt{2\pi}\,|\Sigma_{r_t}|}\right]^{\tau} e^{-\frac{1}{2}\chi^2}$$

where

$$\chi^2 = r_t^T \Sigma_{r_t}^{-1} r_t.$$

When this probability of y_t appearing is low, it means that the system is likely to be attacked. The value of χ^2 and the probability of y_t are negatively correlated. Thus to check the probability, we only need to compute χ^2, and then compare it to a fixed threshold.

However, this previous detection method based on Kalman innovation has one obvious drawback. Because traditional methods aim to classify the sensor status into under attack and without attack, the detection performance is very sensitive to the choice of threshold. A small threshold will increase the false positive rate of the attack, while a large threshold can easily overlook some potential attacks. It is difficult to distinguish between a significant system error and a stealthy attack signal by choosing a proper threshold.

	$\chi^2 \le threshold_1$	$threshold_1 < \chi^2 \le threshold_2$	$threshold_2 < \chi^2$
system error	✓	✓	
sensor attack		✓	✓

To overcome this drawback, we consider classifying sensor status into three categories: normal, suspected, abnormal.

- normal: When $\chi^2 \le threshold_1$, we trust this sensor and judge that the data is normal. In this case, sensor data can be used to multi-sensor fusion algorithm directly.

- suspected: When $threshold_1 < \chi^2 \le threshold_2$, we suspect this sensor may be under attack or have a large system error. In this case, before this sensor data is used for state estimate, we compensate the sensor data from other functioning sensors.

- abnormal: When $threshold_2 < \chi^2$, we mark the sensor data as abnormal. In this case, we will ignore this data and not use it for state estimation updates.

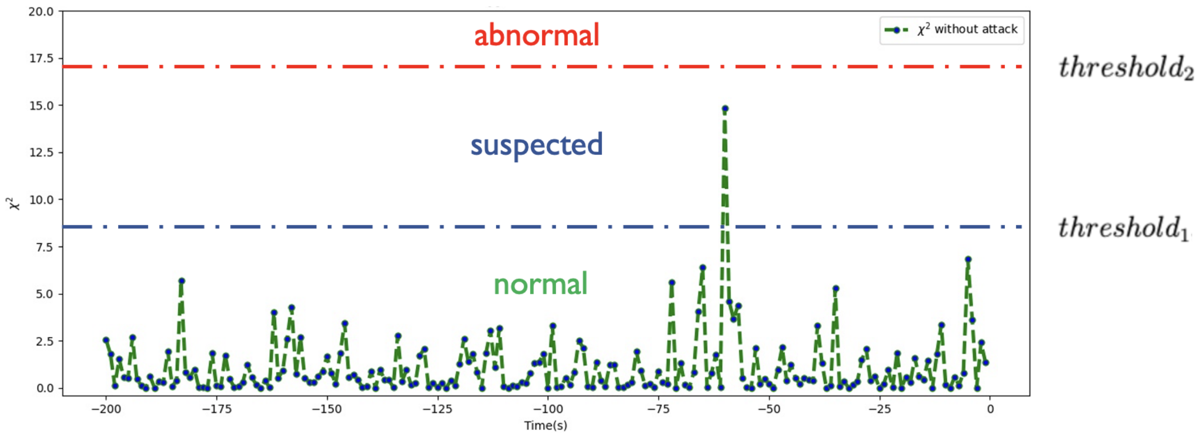

Figure 4: classification of sensor data

Our new categorization method makes the selection of the threshold less difficult. Even when we cannot distinguish between large system errors and stealthy attacks, we are able to use compensation to mitigate its impact on the system.

4.2 Compensation

Here, we assume that abnormal data has been deleted, leaving two types of data: normal and suspected. Multiple sensors often observe the same information about the system state for improved accuracy, so we can compensate for suspected sensors with normal sensors.

We consider the composition of Kalman innovation

$$\begin{bmatrix} r_t^1 \\ \vdots \\ r_t^i \\ \vdots \\ r_t^n \end{bmatrix} = \begin{bmatrix} y_t^1 \\ \vdots \\ y_t^i \\ \vdots \\ y_t^n \end{bmatrix} - \begin{bmatrix} C^1 \\ \vdots \\ C^i \\ \vdots \\ C^n \end{bmatrix} \hat{x}_{t+1|t} = \begin{bmatrix} C^1 \\ \vdots \\ C^i \\ \vdots \\ C^n \end{bmatrix} \Delta x_t + \begin{bmatrix} v_t^1 \\ \vdots \\ v_t^i \\ \vdots \\ v_t^n \end{bmatrix} + \begin{bmatrix} \vdots \\ \vdots \\ \xi_t^i \\ \vdots \\ \vdots \end{bmatrix} \tag{5}$$

where $\Delta x_t = x_{t+1} - \hat{x}_{t+1|t}$ represents the previous estimation error, v_t represents the measurement noise and ξ_t represents potential attack signals.

To better compensate the y_t^i under the attack, we are supposed to estimate the attack signal ξ_t^i. We formulate it to an optimization problem as follows. The complete compensation algorithm is shown in Algorithm 1.

$$\begin{aligned} \min_{\Delta x_t, \xi_t^1, \cdots, \xi_t^n} \quad & J = \alpha \left\| \Delta x_t \right\| + \sum \left\| \xi_t^i \right\| \\ \text{s.t.} \quad & \left\| r_t^i - C^i \Delta x_t - \frac{1+sgn(\left\| r_t^i \right\| - \tau^i)}{2} \xi_t^i \right\| < \tau^i, t = 1, ..., n \end{aligned} \tag{6}$$

where α is a weight parameter, and τ^i represents the threshold in the i^{th} sensor. One explanation is that when $\|r_t^i\| < \tau^i$, $\frac{1+sgn(\left\| r_t^i \right\| - \tau^i)}{2} = 0$, so this sensor will have no compensation. Else when $\|r_t^i\| > \tau^i$, $\frac{1+sgn(\left\| r_t^i \right\| - \tau^i)}{2} = 1$, so this sensor will get a compensation that pulls its χ^2 back to normal.

Algorithm 1 proposed compensation algorithm

```
Input: $x_{t+1|t}, P_{t+1|t}, \{y_t^i\}, \{C^i\}, \{Q^i\}$
Output: $x_{t+1|t+1}, P_{t+1|t+1}$
Step1:
for $i \leftarrow 1$ to $n$ do
    the residue $r_t^i = y_t^i - C^i x_{t+1|t}$
    calculate the residue's distribution $\sum_{r_t^i} = Q^i + C^i P_{t+1|t} C^{iT}$
    get the $\chi^2$ value and classify the sensor data by comparing to the threshold
end for
Step2:
        solve the optimization problem and calculate the compensation $\{\xi_t^j\}$
Step3:
        run multi-sensor fusion algorithm with sensor data $\{y_t^i\}$ and $\{y_t^j + \xi_t^j\}$
        update the state estimate and return $x_{t+1|t+1}, P_{t+1|t+1}$
```

4.3 Detection

We depict the framework in Figure. 5. From this figure, we can observe that the proposed framework comprises two modules, sensor data classification, and compensation.

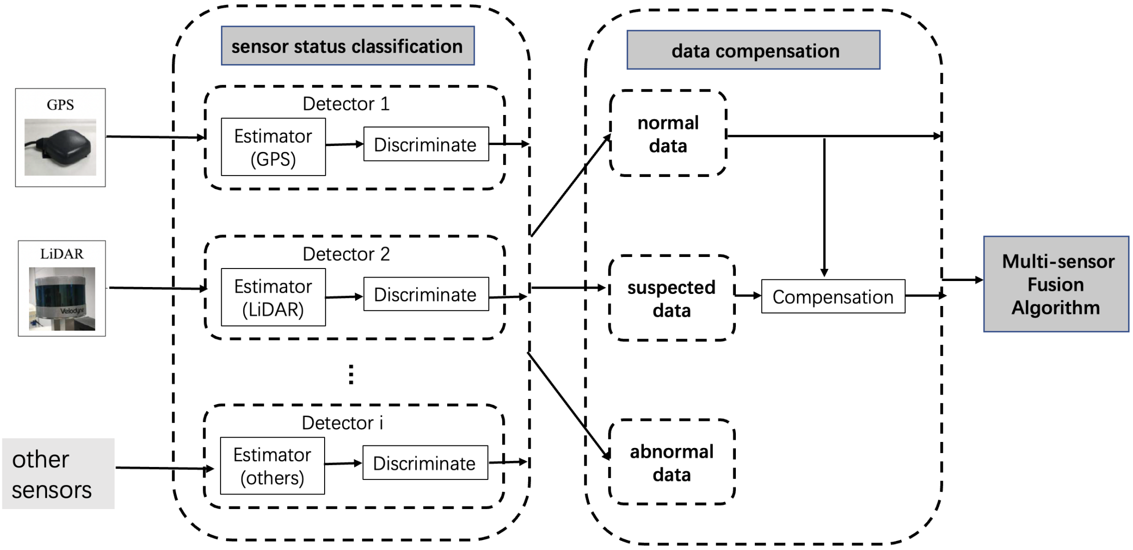

Figure 5: The proposed framework

Based on the above discussions, to facilitate further implementations, our proposed model-based framework can be summarized in Algorithm 2. For those sensors that often require significant compensation, we can consider them under attack. The detection system will send the alarm and mask the data.

Algorithm 2 Detection framework

```
Input: $x_{init}, P_{init}$
Output: state estimate and alarm
while sensors are working do
    receive sensory data $\{y_t^i\}$
    classify sensory data $\{y_t^i\}$ into three types: normal, suspected, abnormal
    compensate for suspected data $\{y_t^j = y_t^j + \xi_t^j\}$
    update state estimate with normal and compensated data $\{y_t^i\}, \{y_t^j + \xi_t^j\}$
    if a particular sensor requires persistent large compensation then
        send an alarm!
    end if
end while
```

5 Experimental Result

In this section, we illustrate the performance of our compensation and detection framework. In the experiment, we consider a scenario in which autonomous vehicles follow the lane line, maintaining uniform motion in the horizontal direction and close to zero in the vertical direction. We assume that the vehicle is equipped with GPS and LiDAR to provide redundant pose measurements, which cannot be attacked simultaneously.

5.1 Experiment 1: without attack

Experiment 1 simulates the performance of our compensation algorithm on state estimate without attack. Figure. 7 shows that our algorithm will adjust the measurements with χ^2 large values when there is no attacker. As mentioned, χ^2 in a normal case satisfies zero-mean Gaussian distribution, so the probability of a large χ^2 value is small. Figure. 6 and Figure. 8 show that few numbers of compensation have little effect on the state estimate of the normal system.

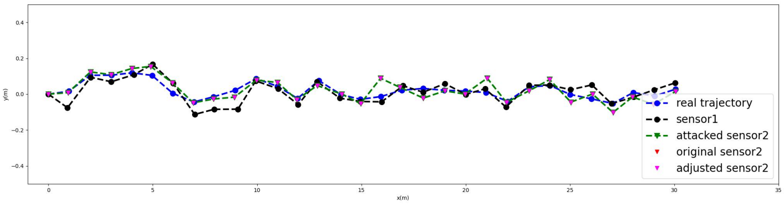

Figure 6: sensor data with attack and compensation

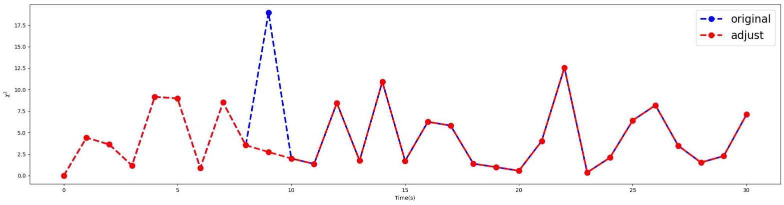

Figure 7: adjustment of χ^2 values

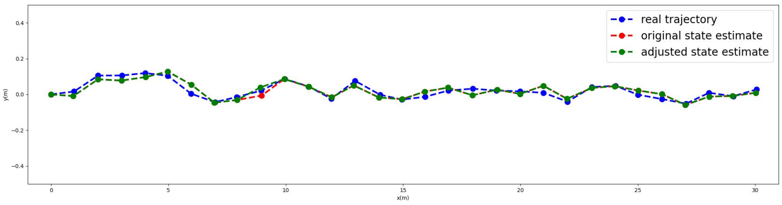

Figure 8: state estimation before and after compensation

5.2 Experiment 2: under certain bias attack

Experiment 2 simulates the attacker that injects a certain bias into the sensor measurement. Figure. 9 and Figure. 10 illustrate that our algorithm gives anomalous observations proper compensation, which helps anomaly sensor data get closer to the prediction with a lower χ^2 value. As seen in Figure. 11, our compensation avoids large drifts in the system's state estimate under attack.

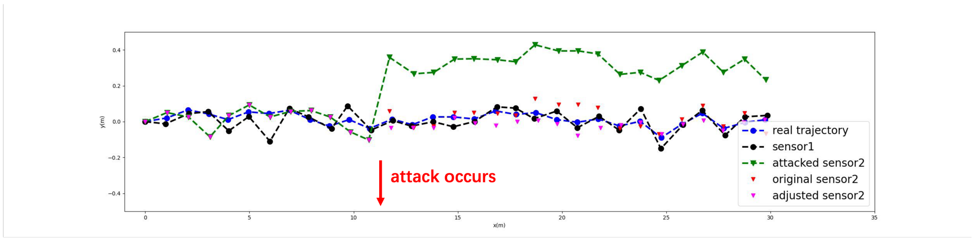

Figure 9: sensor data with attack and compensation

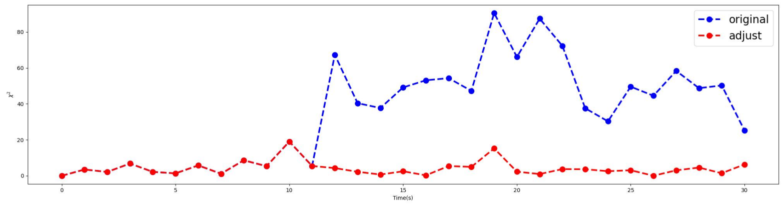

Figure 10: adjustment of χ^2 values

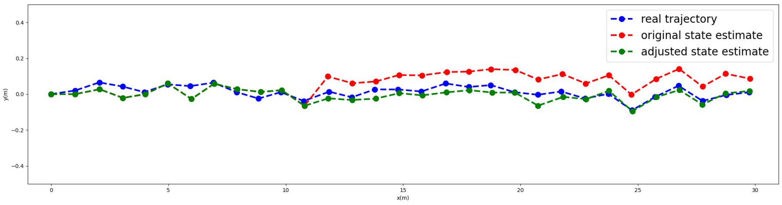

Figure 11: state estimation before and after compensation

5.3 Experiment 3: under random noise attack

Experiment 3 gives an example to illustrate how our algorithm detects and defends against the attack with random noise. In Experiment 3, we consider the attack to last longer and be more random. Figure. 12 and Figure. 13 depict the random attack signal injecting random noise to sensor 2, while our compensation algorithm corrects it and detects the attack. Figure. 14 demonstrates our defense strategy improves the accuracy of the state estimate under attack.

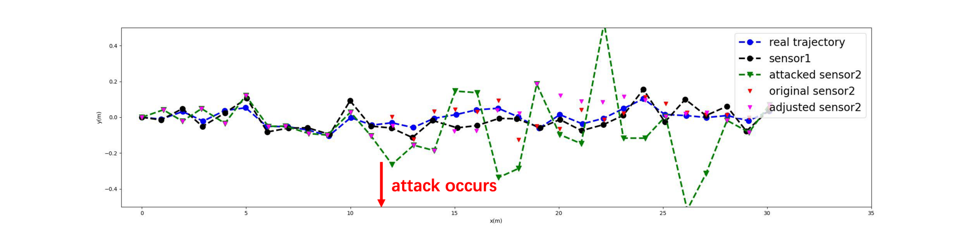

Figure 12: sensor data with attack and compensation

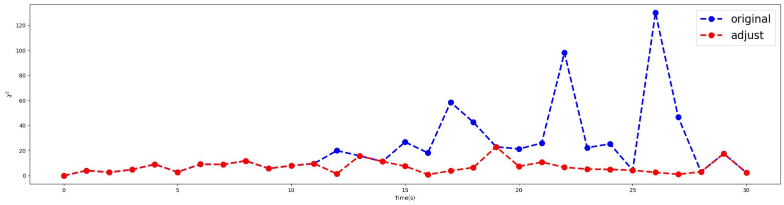

Figure 13: adjustment of χ^2 values

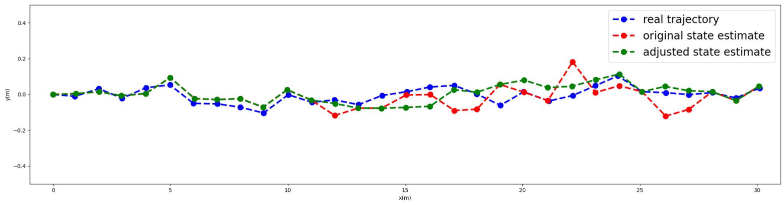

Figure 14: state estimation before and after compensation

6 Conclusion

This report proposes a Kalman-based compensation algorithm that can efficiently adjust the anomaly measurement and detect the attack. The novelty is to combine compensation with detection that improves the accuracy of state estimates even when attackers are not distinguished. Simulations show that the proposed algorithm can defend against various attacks and have a better estimation performance than previous methods. However, there are a few limitations: 1) we only validate our algorithm on linear kinetic models; 2) we propose a framework without specifically analyzing the defense strategy for stealthy attacks.

Future work includes: 1) extending our algorithm to non-linear kinetic models and testing the proposed algorithm on actual vehicles; 2) providing theoretical guarantees on the detection rate and speed; 3) proposing follow-up countermeasures to defend against stealthy attacks.

References

[1] Baidu. Baidu apollo survery car. `https://developer.apollo.auto/developer.html`.

[2] Andrew J Kerns, Daniel P Shepard, Jahshan A Bhatti, and Todd E Humphreys. Unmanned aircraft capture and control via gps spoofing. *Journal of Field Robotics*, 31(4):617–636, 2014.

[3] Yulong Cao, Chaowei Xiao, Benjamin Cyr, Yimeng Zhou, Won Park, Sara Rampazzi, Qi Alfred Chen, Kevin Fu, and Z Morley Mao. Adversarial sensor attack on lidar-based perception in autonomous driving. In *Proceedings of the ACM SIGSAC conference on computer and communications security*, pages 2267–2281, 2019.

[4] Drew Davidson, Hao Wu, Robert Jellinek, Vikas Singh, and Thomas Ristenpart. Controlling uavs with sensor input spoofing attacks. In *Proceedings of the 25th USENIX Conference on Security Symposium*, 2016.

[5] Nicholas DeMarinis, Stefanie Tellex, Vasileios P Kemerlis, George Konidaris, and Rodrigo Fonseca. Scanning the internet for ros: A view of security in robotics research. In *Proceedings of the International Conference on Robotics and Automation*, pages 8514–8521. IEEE, 2019.

[6] Anatolij Bezemskij, George Loukas, Richard J Anthony, and Diane Gan. Behaviour-based anomaly detection of cyber-physical attacks on a robotic vehicle. In *Proceedings of the 15th International Conference on Ubiquitous Computing and Communications and International Symposium on Cyberspace and Security*, pages 61–68. IEEE, 2016.

[7] Anatolij Bezemskij, George Loukas, Diane Gan, and Richard J Anthony. Detecting cyber-physical threats in an autonomous robotic vehicle using bayesian networks. In *Proceedings of the International Conference on Internet of Things (iThings) and Green Computing and Communications (GreenCom) and Cyber, Physical and Social Computing (CPSCom) and Smart Data (SmartData)*, pages 98–103. IEEE, 2017.

[8] Matteo Olivato, Omar Cotugno, Lorenzo Brigato, Domenico Bloisi, Alessandro Farinelli, and Luca Iocchi. A comparative analysis on the use of autoencoders for robot security anomaly detection. In *Proceedings of the International Conference on Intelligent Robots and Systems*, pages 984–989. IEEE, 2019.

[9] Yiyang Wang, Neda Masoud, and Anahita Khojandi. Real-time sensor anomaly detection and recovery in connected automated vehicle sensors. *Transactions on Intelligent Transportation Systems*, 22(3):1411–1421, 2020.

[10] Franco Van Wyk, Yiyang Wang, Anahita Khojandi, and Neda Masoud. Real-time sensor anomaly detection and identification in automated vehicles. *Transactions on Intelligent Transportation Systems*, 21(3):1264–1276, 2019.

[11] Aidin Ferdowsi, Ursula Challita, Walid Saad, and Narayan B Mandayam. Robust deep reinforcement learning for security and safety in autonomous vehicle systems. In *Proceedings of the 21st International Conference on Intelligent Transportation Systems*, pages 307–312. IEEE, 2018.

[12] Giedre Sabaliauskaite, Geok See Ng, Justin Ruths, and Aditya Mathur. A comprehensive approach, and a case study, for conducting attack detection experiments in cyber–physical systems. *Robotics and Autonomous Systems*, 98:174–191, 2017.

[13] Azarakhsh Keipour, Mohammadreza Mousaei, and Sebastian Scherer. Automatic real-time anomaly detection for autonomous aerial vehicles. In *Proceedings of the International Conference on Robotics and Automation*, pages 5679–5685. IEEE, 2019.

[14] Yuanzhe Wang, Qipeng Liu, Ehsan Mihankhah, Chen Lv, and Danwei Wang. Detection and isolation of sensor attacks for autonomous vehicles: Framework, algorithms, and validation. *Transactions on Intelligent Transportation Systems*, 23(7):8247–8259, 2021.

[15] Nils Ole Tippenhauer, Christina Pöpper, Kasper Bonne Rasmussen, and Srdjan Capkun. On the requirements for successful gps spoofing attacks. In *Proceedings of the 18th ACM Conference on Computer and communications security*, pages 75–86, 2011.

[16] Tyler Hohman. The inside scoop on gps spoofing. `https://www.orolia.com/the-inside-scoop-on-gps-spoofing/`.