Poster: Bypassing Physical Invariants-Based Defenses in Autonomous Vehicles

Yinan Zhao
Waseda University
yinanzhao@nsl.cs.waseda.ac.jp

Tatsuya Mori
Waseda University
mori@nsl.cs.waseda.ac.jp

I. Introduction

Autonomous vehicles (AVs), which are critical to advancing road safety, rely heavily on sensor data for navigation, raising concerns about their vulnerability to data manipulation and the need for robust data validation. To improve the security of AVs, Quinonez et al. [1] developed SAVIOR, a framework designed to defend against such threats. SAVIOR exploits nonlinear physical invariants associated with unmanned and AVs by learning system model parameters through offline System Identification (SI) and using an online Extended Kalman Filter (EKF) to predict sensor values while statistically detecting anomalies with the CUSUM algorithm.

Recognizing SAVIOR's reliance on physical models for defense, our research challenges these defenses by developing a method for attackers to perform SI using Model Predictive Control (MPC). The proposed attack strategy is to apply perturbations that are undetectable by physical invariants-based Intrusion Detection System (PI-IDS), striking a balance where the induced deviations do not exceed the anomaly detection threshold. This technique aims to perform stealthy attacks that are carefully planned to evade advanced security systems such as SAVIOR. At the core of our approach is the use of MPC, SI, and Dynamic Time Warping (DTW) to allow attackers to accurately simulate system behavior-allowing them to perform undetectable attacks on PI-IDS-secured AV systems.

In this preliminary study, we present our progress in successfully simulating MPC and SI using MATLAB Simulink.

II. Attack Model

The attack model assumes that the attacker is able to:

- Interact with the CAN network to read and write messages through the OBD-II port or other hardware attachments.
- Understand the inputs and outputs of the steering control system and create a representative state-space model (SSM) with MPC.
- Use DTW to determine fault thresholds to make manipulations remain within acceptable limits.
- Inject tailored messages into the vehicle's network to stealthily control the vehicle.

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.2024.25017
www.ndss-symposium.org

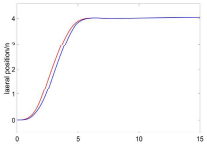

Fig. 1: Lateral Positions: Predicted (Red) vs. Actual (Blue).

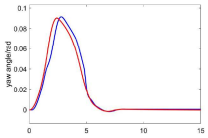

Fig. 2: Yaw Angles: Predicted (Red) vs. Actual (Blue).

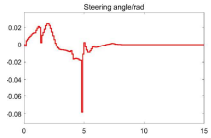

Fig. 3: Predicted Steering Angle

III. Preliminary Results

Our strategy for stealthy attacks on steering systems of AVs ensures that deviations remain below a certain error threshold to avoid detection. Using historical steering data, attackers predict optimal steering angles with MPC and determine system behavior using SSM. Finally, they use the DTW technique to set an error threshold that deviations caused by malicious inputs cannot exceed. This allows for subtle, undetectable tampering.

The accuracy of our MPC simulation has been validated in a preliminary evaluation using MATLAB Simulink, as shown by the alignment of predicted and actual values in Fig. 1 and Fig. 2, with the MPC-derived steering angles shown in Fig. 3. While we have successfully demonstrated accurate system modeling by an attacker, the evaluation of actual stealthy attacks remains a work in progress.

IV. Future Work

Our future research includes the implementation and comprehensive evaluation of the stealthy attack. We also aim to integrate the attack framework into a realistic autonomous driving simulator, such as Carla or Autoware, to evaluate the overall attack framework. We will also work on developing countermeasures to mitigate these threats.

Acknowledgement A part of this work was supported by JSPS KAKENHI 22S0604 and JST CREST JPMJCR23M4.

References

[1] R. Quinonez, J. Giraldo, L. Salazar, E. Bauman, A. Cardenas, and Z. Lin, "{SAVIOR}: Securing autonomous vehicles with robust physical invariants," in *29th USENIX Security Symposium (USENIX Security 20)*, 2020, pp. 895–912.