Demo: Remote Adversarial Attack on Automated Lane Centering*

Yulong Cao*, Yanan Guo†, Takami Sato‡, Qi Alfred Chen‡, Z. Morley Mao*, and Yueqiang Cheng§
*University of Michigan †University of Pittsburgh ‡University of California, Irvine §NIO

***Abstract*—Advanced driver-assistance systems (ADAS) are widely used by modern vehicle manufacturers to automate, adapt and enhance vehicle technology for safety and better driving. In this work, we design a practical attack against automated lane centering (ALC), a crucial functionality of ADAS, with remote adversarial patches. We identify that the back of a vehicle is an effective attack vector and improve the attack robustness by considering various input frames. The demo includes videos that show our attack can divert victim vehicle out of lane on a representative ADAS, Openpilot, in a simulator.**

I. Introduction

Automated Lane Centering (ALC) is a Level-2 driving automation technology that can automatically adjust the vehicle's steering based on the velocity and position of the vehicle to keep the vehicle in the lane. ALC has been available in vehicles from many manufacturers due to its convenience. However, though ALC serves as a convenient feature to have for the drivers, it is also safety-critical and could lead to severe consequences once went wrong. Since ALC relies on machine learning models to detect the lane and machine learning models are shown vulnerable to adversarial attack, the security of ALC is at stake.

In this demo, we illustrate the effects of remote adversarial patch attacks on ALC. While Sato et al. have demonstrated that ALC is vulnerable to a road patch attack [3], remote adversarial patch attack could be more easily achieved. Instead of paving a large and long patch on the road of a specific location, the proposed attack only requires the attacker to place an adversarial patch on the back of a vehicle. This demo builds upon previous works by demonstrating end-to-end attack impacts through a production-grade simulator, Carla, and on a commercial level ADAS, Openpilot [1]. By using a specialized Expectation over Transformation (EoT) [2], we improve the remote patch robustness to achieve high attack success rates on diverting the victim vehicle.

II. Threat Model and Attack Goal

Threat Model. We assume the attacker has full knowledge of the ADAS used on the victim vehicle as prior works. Using a white-box setting explores the limit of a powerful attacker and sheds light on future defenses against it. We also assume the attacker can place an adversarial patch on the back of a vehicle, and control the vehicle to drive in front of the victim.

Fig. 1. RAP-ALC attack that perturbs the victim ALC planned path.

Attack Goal. The attacker's goal is to cause the ALC functionality to fail and drive the victim vehicle out of the lane. This threatens the safety of the passengers in the victim vehicle and potentially in the vehicles on the next lane.

III. Attack Design

We propose a **R**emote **A**dversarial **P**atch attack on **ALC** (RAP-ALC) by exploiting the large receptive field of modern models. We add an adversarial patch on the back of the attacker's vehicle in front of the victim vehicle to bend the predicted planning path of the ALC module in Openpilot. By incorporating the EoT method in terms of different perspective transformations and distances, we generate robust patches that fool the ALC for continuous input frames to increase the end-to-end attack success rates.

IV. Demonstration

We plan to demonstrate the proposed RAP-ALC attack in the Carla simulator on the Openpilot victim system. We use the V0.8.6 of Openpilot as the target victim system. We assume the attacker is able to control a Coca-Cola van and print the adversarial patch on the back of the van. By driving an attacker vehicle in front of the target victim, we demonstrate that the victim vehicle will be driven out of the lane. As shown in Figure 1, Openpilot predicted the correct path to stay in lane (in red color) while perturbed to the next lane under the RAP-ALC attack. More demo videos could be found at our website [1].

Acknowledgment

This research was supported in part by the NSF under CNS-1930041, CNS-1850533, CNS-1932464, CNS-0929771, CNS-2145493, USDOT UTC Grant 69A3552047138 and the National AI Institute for Edge Computing Leveraging Next Generation Wireless Networks, Grant # 2112562.

References

[1] "OpenPilot: Open source driving agent," https://github.com/commaai/openpilot, 2018.

[2] A. Athalye, L. Engstrom, A. Ilyas, and K. Kwok, "Synthesizing robust adversarial examples," in *International conference on machine learning*. PMLR, 2018, pp. 284–293.

[3] T. Sato, J. Shen, N. Wang, Y. Jia, X. Lin, and Q. A. Chen, "Dirty Road Can Attack: Security of Deep Learning based Automated Lane Centering under Physical-World Attack," in *Proceedings of the 29th USENIX Security Symposium (USENIX Security '21)*, 2021.

*This work was conducted while the first a uthor w as d oing i nternship at NIO

[1] https://sites.google.com/view/rap-alc/home

Workshop on Automotive and Autonomous Vehicle Security (AutoSec) 2022
24 April 2022, San Diego, CA, USA
ISBN 1-891562-75-4
https://dx.doi.org/10.14722/autosec.2022.23015
www.ndss-symposium.org