Demo: Attacking Tesla Model X's Autopilot Using Compromised Advertisement

Ben Nassi[1], Yisroel Mirsky[1,2], Dudi Nassi[1], Raz Ben-Netanel[1], Oleg Drokin[3], Yuval Elovici[1]
[1] Ben-Gurion University of the Negev, [2] Georgia Tech, [3] Independent Researcher
{nassib, yisroel, razx, nassid, elovici}@post.bgu.ac.il, green@linuxhacker.ru

Video Demonstration: https://youtu.be/-E0t_s6bT_4

***Abstract*—In this demo, we demonstrate how attackers can remotely apply split-second phantom attacks by embedding phantom road signs into an advertisement presented on an Internet connected digital billboard which causes Tesla's autopilot to suddenly stop the car in the middle of a road.**

IN this demo, we identify a limitation of AI models that can be exploited by attackers to perform a new type of attack called a "split-second phantom attack." This attack exploits a weakness of advanced driving assistance systems (ADASs) where (1) digitally displayed imagery is perceived as a real object (i.e., phantom imagery), and (2) the imagery only has to appear briefly (for a few milliseconds) in order to be detected by the ADAS. One might argue that the authenticity of phantoms can be determined by commercial ADASs that use sensor fusion by cross-correlating the camera sensor with data obtained from depth sensors (e.g., radar, LiDAR, and ultrasonic sensors), however we show that the most advanced semi-autonomous vehicle, the Tesla Model X, resolves the disagreement between the camera and the integrated depth sensors regarding a phantom by treating the phantom as real, despite the fact that it does not have any depth.

We demonstrate the attack via a digital billboard by embedding a phantom into an existing advertisement and displaying it for a few milliseconds in a way that the imagery itself is hard to perceive. We picked a random McDonald's advertisement from YouTube. We embedded a stop sign into nine consecutive frames (500 ms) of the advertisement in order to attack the Tesla Model X.

The experiment was conducted on a road on our university campus after we received the proper approvals from the security department. The attacker used a 42 inch TV screen (which was used to simulate a digital billboard and was plugged into another car for electricity) that was placed in the middle of the road. We used this setup to demonstrate the attack, since we had no intention of hacking a real digital billboard in this study. The experiment was conducted as follows: We engaged Tesla's autopilot at the beginning of the road. The Tesla approached the middle of the road where the TV screen presented the compromised McDonald's advertisement (with the embedded 500 ms stop sign). Fig. 1 presents a snapshot from the 500 ms that the phantom was presented on the upper left corner on the screen. The attack was successful, since Tesla's autopilot identified the phantom stop sign and immediately triggered the car to stop in the middle of the road. The reader can view a video of the experiment and the compromised advertisements in link above.

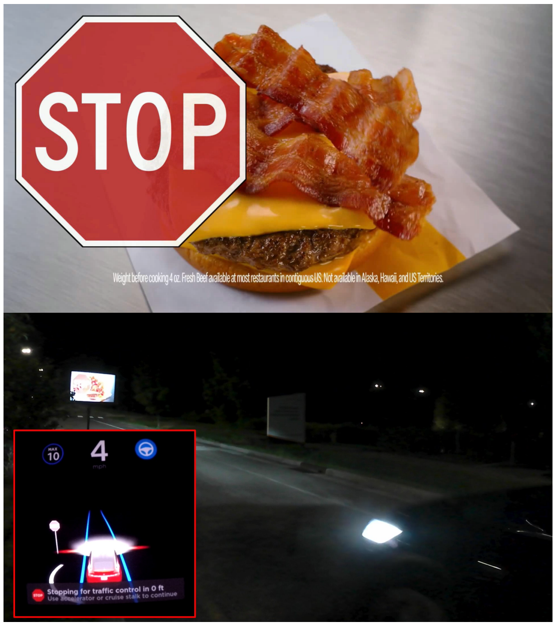
STOP
Weight before cooking 4 oz. Fresh Beef available at most restaurants in contiguous US. Not available in Alaska, Hawaii, and US Territories.
MAX
10
4
mph
STOP
Stopping for traffic control in 0 ft
Use accelerator or cruise stalk to continue

Fig. 1. Top: The compromised frame in the advertisement with the embedded phantom road sign. Bottom: Tesla's autopilot triggers the car to stop suddenly after a phantom stop sign was detected in the upper left corner of the digital billboard.

The extended version of this paper, that includes the algorithm that was used to detect the most suitable frame to embed the phantom, can be found in [1], [2].

References

[1] B. Nassi, Y. Mirsky, D. Nassi, R. Ben-Netanel, O. Drokin, and Y. Elovici, "Phantom of the adas: Securing advanced driver-assistance systems from split-second phantom attacks," in *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 293–308. [Online]. Available: https://doi.org/10.1145/3372297.3423359

[2] B. Nassi, D. Nassi, R. Ben-Netanel, Y. Mirsky, O. Drokin, and Y. Elovici, "Phantom of the adas: Phantom attacks on driver-assistance systems."

Workshop on Automotive and Autonomous Vehicle Security (AutoSec) 2021
25 February 2021, Virtual
ISBN 1-891562-68-1
https://dx.doi.org/10.14722/autosec.2021.23004
www.ndss-symposium.org