wevolver.com

New attack on autonomous vehicle sensors creates fake obstacles

5-6 minutos

Autonomous vehicles (AVs) rely on a number of different sensors to perceive the world around them. In most systems, cameras, RADAR, and LiDAR (or Light Detection and Ranging) work together to build a multi-faceted view of the area directly surrounding the car. The three provide different details that, combined, can safeguard against each other's limitations.

Unfortunately for the designers building these systems, this opens them up to three different sets of vulnerabilities that can be used to sabotage a vehicle mid-drive. Cameras can fall prey to simple visual confusion—even putting stickers on road signs can completely change their meaning, leading to traffic jams or accidents.

Up to this point, no attacks had been discovered targeting a car's LiDAR system—but a major new finding from researchers at the University of Michigan has demonstrated what that might look like.

A group including the labs of Profs. Z. Morley Mao and

Kevin Fu, led by CSE PhD student Yulong Cao, designed and carried out a “spoofing” attack against a LiDAR sensor, effectively tricking the system into perceiving an obstacle in its path that wasn’t really there. The attack is able to fool the car’s machine learning program meant to serve as a safeguard against misleading inputs, and can be performed without being seen by passengers on board. The project was the first security study of LiDAR-based perception in an autonomous vehicle.

The team’s proposed attack pulls together two components of a standard AV package, its LiDAR sensors and machine learning model. The means to fool the former is somewhat straightforward—LiDAR operates on light, calculating the time it takes light to travel to and from the objects around the sensor. An attacker would simply need to “spoof” an incorrect signal to the sensor by sending light signals toward it. In this case, the team used lasers.

But the machine learning package makes the attack a little more complicated.

To date there have been a number of famous examples of machine learning models being undetectably fooled while performing a variety of tasks. In a case targeting simple image classifiers, researchers found that adding unnoticeable noise to an image caused the model to identify objects as something completely different.

Not just any noise would do—randomized noise did little to fool the model, and ended up being lost in interpretation.

But with a carefully crafted pattern of noise, a hacker can actually control what the perturbed image will be detected as.

Between an AV’s perception system and the signals it gathers from the outside world, one such machine learning model interprets data from the sensors over time to both provide a detailed complete image and see past anomalies and noise. Because of this, Cao says, the team couldn’t simply spoof a random point into the LiDAR’s inputs.

“If you just randomly place spoofed points in the LiDAR sensor, it doesn’t spoof an obstacle,” says Cao. “But if we strategically find a pattern to spoof it can become more powerful and fool the model.”

To determine the correct pattern, the team formulated the attack task as an optimization problem, which had been shown to be effective in previous machine learning security studies. They worked experimentally to determine the position of data points necessary to create a harmful signal.

The researchers were able to construct two attack scenarios with this method, with a 75% success rate. In the first, an emergency brake attack, they forced a moving AV to suddenly hit the brakes. And in the second, an AV freezing attack, they forced an AV waiting for a red light to sit “frozen” in the intersection and block traffic, even after the light turns green.

This attack not only opens up a new vector for disruption in fleets of AVs, it targets the sensors so far deemed the most reliable in the whole package.

“The particular AV systems we’ve been investigating rely heavily on the LiDAR system,” says Cao. “Basically, they trust the LiDAR more.”

The paper will be presented at the ACM Conference on Computer and Communications Security (CCS) in November, 2019. Authors of the paper are CSE PhD students Yulong Cao, Chaowei Xiao, Benjamin Cyr, and Won Park, CS undergraduate Yimeng Zhou, alum Qi Alfred Chen (now Assistant Professor of CS at the University of California, Irvine), research scientist Sara Rampazzi, and Profs. Kevin Fu and Z. Morley Mao.

More information on the project can be found on the team’s website. View the paper here.