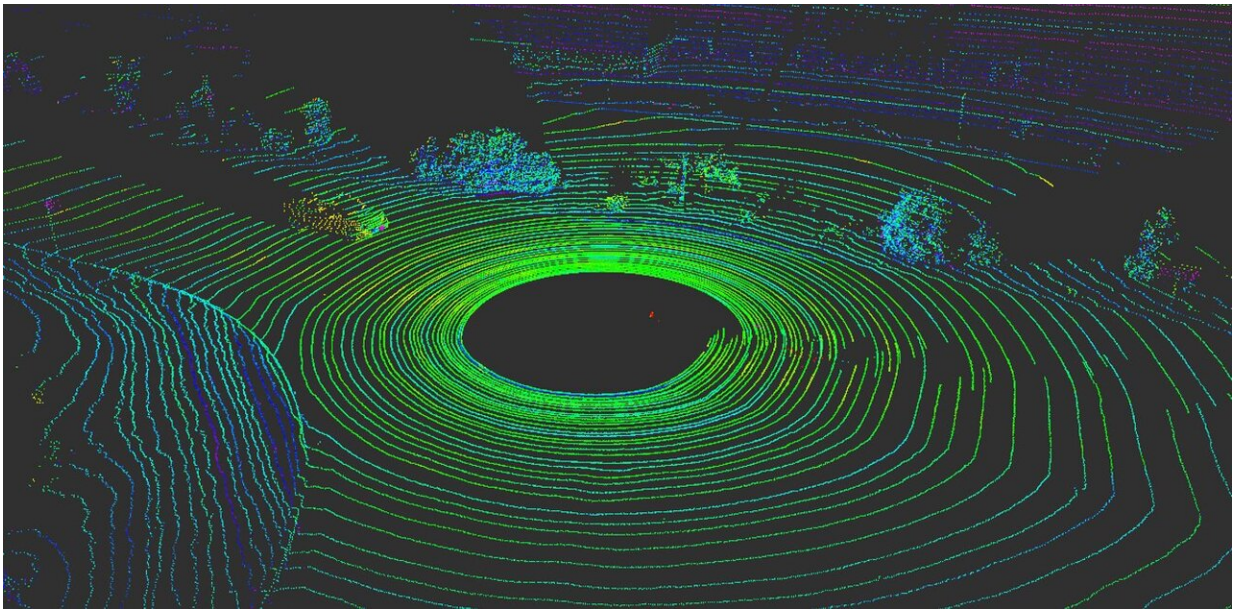
Autonomous vehicles can be fooled to 'see' nonexistent obstacles

March 6 2020, by Yulong Cao and Z. Morley Mao



LiDAR helps an autonomous vehicle ‘visualize’ what’s around it. Credit: Yulong Can with data from Baidu Apollo, CC BY-ND

Nothing is more important to an autonomous vehicle than sensing what's happening around it. Like human drivers, autonomous vehicles need the ability to make instantaneous decisions.

Today, most autonomous vehicles rely on multiple sensors to perceive the world. Most systems use a combination of cameras, radar sensors and

LiDAR (light detection and ranging) sensors. On board, computers fuse this data to create a comprehensive view of what's happening around the car. Without this data, autonomous vehicles would have no hope of safely navigating the world. Cars that use multiple sensor systems both work better and are safer—each system can serve as a check on the others—but no system is immune from attack.

Unfortunately, these systems are not foolproof. Camera-based perception systems can be tricked simply by putting stickers on traffic signs to completely change their meaning.

Our work, from the RobustNet Research Group at the University of Michigan, has shown that the LiDAR-based perception system can be comprised, too. By strategically spoofing the LiDAR sensor signals, the attack is able to fool the vehicle's LiDAR-based perception system into "seeing" a nonexistent obstacle. If this happens, a vehicle could cause a crash by blocking traffic or braking abruptly.

Spoofing LiDAR signals

LiDAR-based perception systems have two components: the sensor and the machine learning model that processes the sensor's data. A LiDAR sensor calculates the distance between itself and its surroundings by emitting a light signal and measuring how long it takes for that signal to bounce off an object and return to the sensor. This duration of this back-and-forth is also known as the "time of flight."

A LiDAR unit sends out tens of thousands of light signals per second. Then its machine learning model uses the returned pulses to paint a picture of the world around the vehicle. It is similar to how a bat uses echolocation to know where obstacles are at night.

The problem is these pulses can be spoofed. To fool the sensor, an

attacker can shine his or her own light signal at the sensor. That's all you need to get the sensor mixed up.

However, it's more difficult to spoof the LiDAR sensor to "see" a "vehicle" that isn't there. To succeed, the attacker needs to precisely time the signals shot at the victim LiDAR. This has to happen at the nanosecond level, since the signals travel at the speed of light. Small differences will stand out when the LiDAR is calculating the distance using the measured time-of-flight.

If an attacker successfully fools the LiDAR sensor, it then also has to trick the machine learning model. Work done at the OpenAI research lab shows that machine learning models are vulnerable to specially crafted signals or inputs—what are known as adversarial examples. For example, specially generated stickers on traffic signs can fool camera-based perception.

We found that an attacker could use a similar technique to craft perturbations that work against LiDAR. They would not be a visible sticker, but spoofed signals specially created to fool the machine learning model into thinking there are obstacles present when in fact there are none. The LiDAR sensor will feed the hacker's fake signals to the machine learning model, which will recognize them as an obstacle.

The adversarial example—the fake object—could be crafted to meet the expectations of the machine learning model. For example, the attacker might create the signal of a truck that is not moving. Then, to conduct the attack, they might set it up at an intersection or place it on a vehicle that is driven in front of an autonomous vehicle.

Two possible attacks

To demonstrate the designed attack, we chose an autonomous driving

system used by many car makers: Baidu Apollo. This product has over 100 partners and has reached a mass production agreement with multiple manufacturers including Volvo and Ford.

By using real world sensor data collected by the Baidu Apollo team, we demonstrated two different attacks. In the first, an "emergency brake attack," we showed how an attacker can suddenly halt a moving vehicle by tricking it into thinking an obstacle appeared in its path. In the second, an "AV freezing attack," we used a spoofed obstacle to fool a vehicle that had been stopped at a red light to remain stopped after the light turns green.

By exploiting the vulnerabilities of autonomous driving perception systems, we hope to trigger an alarm for teams building autonomous technologies. Research into new types of security problems in the autonomous driving systems is just beginning, and we hope to uncover more possible problems before they can be exploited out on the road by bad actors.

Provided by The Conversation