Researchers probe safety of AI in driverless cars, find vulnerabilities

September 2 2024, by Laurie Kaiser

UB's autonomous Lincoln MKZ sedan is one of the vehicles that researchers have used to test vulnerabilites to attacks. Credit: University at Buffalo

Artificial intelligence is a key technology for self-driving vehicles. It is used for decision-making, sensing, predictive modeling and other tasks.

But how vulnerable are these AI systems to an attack?

Ongoing research at the University at Buffalo examines this question, with results suggesting that malicious actors could cause these systems to fail. For example, it's possible that a vehicle could be rendered invisible to AI-powered radar systems by strategically placing 3D-printed objects on that vehicle, which mask it from detection.

The work, which is performed in a controlled research setting, does not mean existing autonomous vehicles are unsafe, researchers say. Nonetheless, it could have implications for the automotive, tech, insurance and other industries, as well as government regulators and policymakers.

"While still novel today, self-driving vehicles are poised to become a dominant form of transportation in the near future," says Chunming Qiao, SUNY Distinguished Professor in the Department of Computer Science and Engineering, who is leading the work. "Accordingly, we need to ensure the technological systems powering these vehicles, especially artificial intelligence models, are safe from adversarial acts. This is something we're working on diligently at the University at Buffalo."

The research is described in a series of papers dating back to 2021 with a study published in *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security (CCS)*. More recent examples include a study from May in *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking* (more commonly known as Mobicom), and another study at this month's 33rd USENIX Security Symposium that's available on arXiv.

mmWave detection effective, but vulnerable

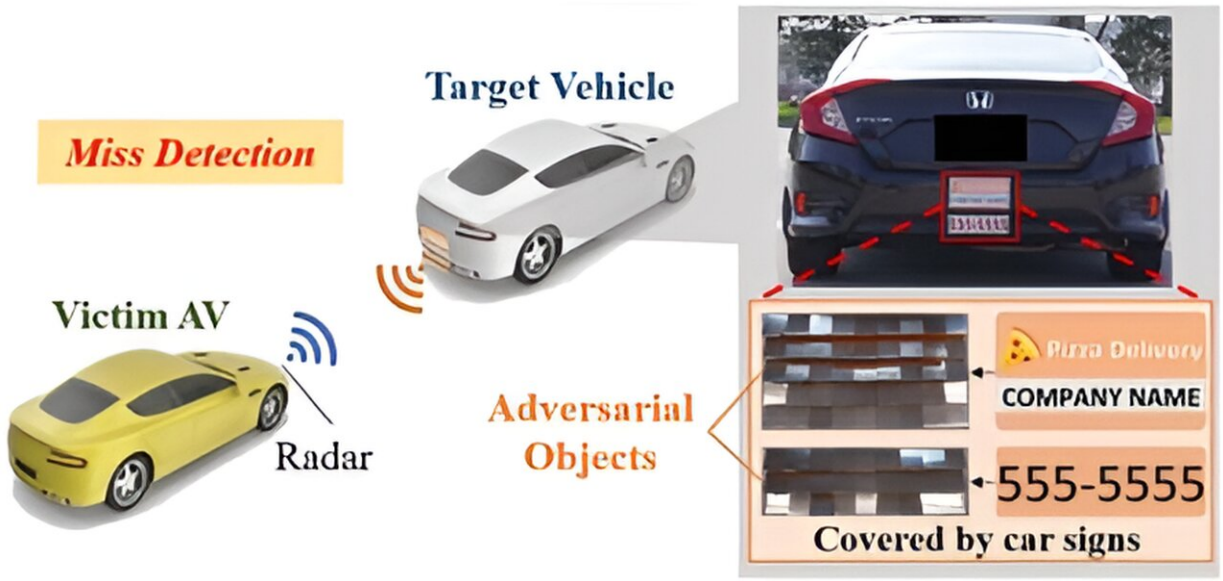
For the past three years, Yi Zhu and other members of Qiao's team have been running tests on an autonomous vehicle on UB's North Campus.

Zhu, who completed his Ph.D. from the UB Department of Computer Science and Engineering in May, recently accepted a faculty position at Wayne State University. A specialist in cybersecurity, he is a primary author of the aforementioned papers, which focus on the vulnerability of lidars, radars and cameras, as well as systems that fuse these sensors together.

"In autonomous driving, millimeter wave [mmWave] radar has become widely adopted for object detection because it's more reliable and accurate in rain, fog and poor lighting conditions than many cameras," Zhu says. "But the radar can be hacked both digitally and in person."

In one such test of this theory, researchers used 3D printers and metal foils to fabricate objects in specific geometric shapes that they called "tile masks." By placing two tile masks on a vehicle, they found they could mislead the AI models in radar detection, thus making this vehicle disappear from its radar.

The work on tile masks was published in *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security* in November 2023.



UB researchers used 3D printers and metal foils to fabricate objects in specific geometric shapes that could be strategically placed on a vehicle to make it disappear from radar detection. Credit: University at Buffalo

Attack motives may include insurance fraud, AV competition

Zhu notes that while AI can process loads of information, it also can get confused and provide incorrect information if given special instructions it wasn't trained to handle.

"Let's assume we have a picture of a cat, and AI can correctly identify this is a cat. But if we slightly change a few pixels in the image, then AI might think this is an image of a dog," Zhu says. "This is an adversarial example of AI. In recent years, researchers have found or designed many adversarial examples for different AI models. So, we asked ourselves: Is it possible to design examples for the AI models in autonomous vehicles ?"

The researchers noted that potential attackers could surreptitiously stick an adversarial object on a vehicle before the driver begins the trip, parks temporarily, or stops at a traffic light. They could even place an object in something a pedestrian is wearing, such as a backpack, effectively erasing detection of that pedestrian, Zhu says.

Possible motivations for such attacks include causing accidents for insurance fraud, competition between autonomous driving companies, or a personal desire to hurt the driver or passengers in another vehicle.

It's important to note, researchers say, that the simulated attacks assume the attacker has full knowledge of the radar object detection system of the victim's vehicle. While obtaining this information is possible, it's also not very likely among members of the public.

Security lags behind other technology

Most AV safety technology focuses on the internal part of the vehicle, while few studies look at external threats, says Zhu.

"The security has kind of lagged behind the other technology," he says.

While researchers have looked at ways to stop such attacks, they haven't found a definite solution yet.

"I think there is a long way to go in creating an infallible defense," Zhu says. "In the future, we'd like to investigate the security not only of the radars but also of other sensors like the camera and motion planning. And we also hope to develop some defense solutions to mitigate these attacks."

More information: Yang Lou et al, A First Physical-World Trajectory Prediction Attack via LiDAR-induced Deceptions in Autonomous

Driving, *arXiv* (2024). DOI: 10.48550/arxiv.2406.11707

Yi Zhu et al, Malicious Attacks against Multi-Sensor Fusion in Autonomous Driving, *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking* (2024). DOI: 10.1145/3636534.3649372

Yi Zhu et al, TileMask: A Passive-Reflection-based Attack against mmWave Radar Object Detection in Autonomous Driving, *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security* (2023). DOI: 10.1145/3576915.3616661

Provided by University at Buffalo