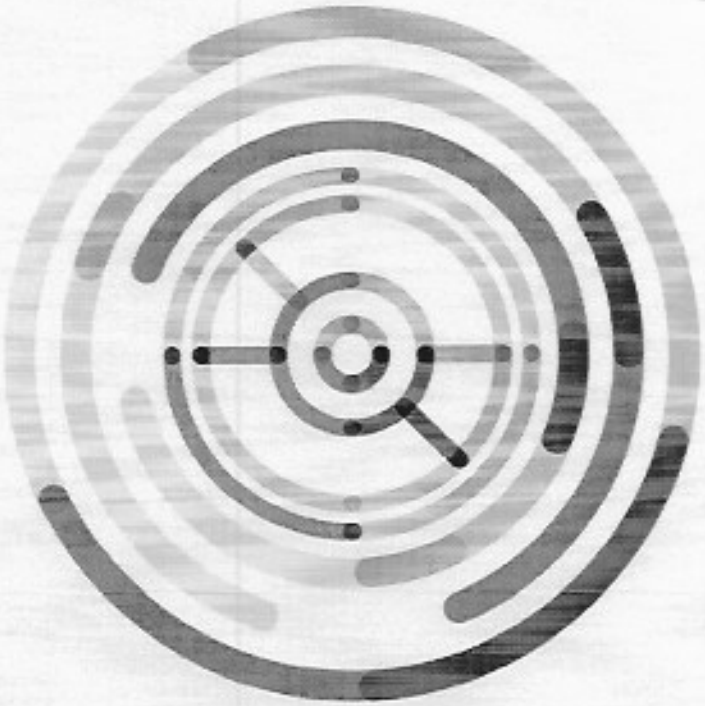
Author: Andy Greenberg. Andy Greenberg Security

Date of Publication: 05.25.16.

**A Car's Computer Can 'Fingerprint' You in Minutes Based on How You Drive**



The way you drive is surprisingly unique. And in an era when automobiles have become data-harvesting, multi-ton mobile computers, the data collected by your car—or one you rent or borrow—can probably identify you based on that driving style after as little as a few minutes behind the wheel.

In a study they plan to present at the Privacy Enhancing Technology Symposium in Germany this July, a group of researchers from the University of Washington and the University of California at San Diego found that they could "fingerprint" drivers based only on data they collected from internal computer network of the vehicle their test subjects were driving, what's known as a car's CAN bus. In fact, they found that the data collected from a car's brake pedal alone could let them correctly distinguish the correct driver out of 15 individuals about nine times out of ten, after just 15 minutes of driving. With 90 minutes driving data or monitoring more car components, they could pick out the correct driver fully 100 percent of the time.

"With very limited amounts of driving data we can enable very powerful and accurate inferences about the driver's identity," says Miro Enev, a former University of Washington researcher who worked on the study before taking a job as a machine-learning engineer at Belkin. And the researchers argue that ability to pinpoint could have unexpected privacy implications: Everything from letting insurance companies punish drivers who loan their cars to their teenage kids, to confirming the identity of a driver who violated traffic laws or caused a collision.

With very limited amounts of driving data we can enable very powerful and accurate inferences about the driver's identity. Miro Enev

The ability to identify a driver based on a car's data may not seem like the creepiest privacy invasion. But the fingerprinting study, Enev argues, should serve as a more general warning to car owners about the sensitivity of the data that travels across their vehicles' internal networks. The same data that tells their insurance company when they've let their 16-year-old kid take their car to prom might just as easily be used to identify drunk driving or a medical condition that's altered someone's driving ability, tests Enev claims would actually be simpler than trying to distinguish a driver's identity.

In fact, drivers are increasingly sending that sensitive data to the cloud with gadgets like Hum, Vinli, Automatic and Zubee, designed to be plugged into their cars' CAN networks via a port under the vehicle's dashboard. Other OBD2 devices are offered by insurance companies, like Progressive and Metromile, in exchange for lower rates, giving those firms access to a car's wealth of digital output. And as cars become increasingly connected to the internet, driving data may also be uploaded directly by cars themselves, as Tesla already does. "To me the whole concern is more about the risk surface that's exposed by these continual sensors, and the fact that not many people are thinking about this," says Enev. "Instead they're just giving this data from their car to third parties."

Here's how the study worked: Researchers asked 15 individual test subjects to drive around a parking lot on the University of Washington campus in Seattle, to the Space Needle around five miles away, and finally to another destination 50 miles further, all while a laptop was plugged into the car's dashboard to collect its CAN network data. Then the researchers tried using a machine learning algorithm to analyze each portion of those drivers' routes for every driver. In each case, the researchers' algorithm would use 90 percent of the driving data as material to "learn" from, and then try to determine based on the remaining data which driver that 10 percent matched with.

In the end, the researchers found that they didn't even need the longest portion of the driving test to reliably identify each of the 15 drivers. Using the full collection of the car's sensors—including how the driver braked, accelerated and angled the steering wheel—the researchers found that their algorithm could distinguish each of the drivers, with 100 percent accuracy, based on only 15 minutes of the driving data. Even with data from the brake pedal alone, they found that they could guess at the correct driver with 87 percent accuracy.

That driver detection could actually have positive applications, like detecting theft. If the car itself were able to identify an unknown driver, it could potentially alert the car's owner. But in their paper, the researchers propose other situations in which it might represent a privacy violation. A red light camera could combine its images with the car's sensor data to identify a driver who ran a red light even his or her face was obscured. Or a car rental company could detect that someone who wasn't authorized to drive in the rental agreement is behind the wheel, and charge the renter a fee.

The driver detection research is only the latest study to point to the danger of internet-connected cars, and particularly internet-connected devices plugged into cars' CAN networks. Last summer, a group at the University of California San Diego that included one of the same researchers from this driver detection study showed that they could hack into one of those dashboard dongles over the internet to disable the brakes of a Corvette the dongle was plugged into—a far scarier prospect.

But in both cases, Enev argues, the studies point to a more fundamental problem with automotive security. Instead of making all of a car's data and sensitive systems available to any device connected to their CAN bus, vehicles should have permission systems, just as operating systems like iOS or

Android do. A gadget meant to track your fuel efficiency, for instance, shouldn't be able to track every exact push of your brake pedal or turn of the wheel, he says. "There should be a permission structure built around every sensor stream," Enev says. "You should approach every new application that you expose your data to on a need-to-know basis."