UW–Madison: Researchers hack voice ID with PVC pipes

Posted on Thursday, Aug 17, 2023

MADISON — Researchers are in an arms race with hackers to prevent data theft. Their standard tools include strategies like multi-factor authentication systems, fingerprint technology and retinal scans. One type of security system that is gaining popularity is automatic speaker identification, which uses a person's voice as a passcode.

These systems, already in use for phone banking and other applications, are good at weeding out attacks that try to fake a user's voice through digital manipulation. But digital security engineers at the University of Wisconsin–Madison have found these systems are not quite as foolproof when it comes to a novel analog attack. They found that speaking through customized PVC pipes — the type found at most hardware stores — can trick machine learning algorithms that support automatic speaker identification systems.

The team, led by PhD student Shimaa Ahmed and Kassem Fawaz, a professor of electrical and computer engineering, presented their findings Aug. 9 at the Usenix Security Symposium in Anaheim, California.

The risks posed by analog security holes could be far-reaching. Ahmed points out that many commercial companies already sell the technology, with financial institutions among their early customers. The technology is also used for AI-supported personal assistants like Apple's Siri.

"The systems are advertised now as secure as a fingerprint, but that's not very accurate," says Ahmed. "All of those are susceptible to attacks on speaker identification. The attack we developed is very cheap; just get a tube from the hardware store and change your voice."

The project began when the team began probing automatic speaker identification systems for weaknesses. When they spoke clearly, the models behaved as advertised. But when they spoke through their hands or talked into a box instead of speaking clearly, the models did not behave as expected.

Ahmed investigated whether it was possible to alter the resonance, or specific frequency vibrations, of a voice to defeat the security system. Because her work began while she was stuck at home due to COVID-19, Ahmed began by speaking through paper towel tubes to test the idea. Later, after returning to the lab, the group hired Yash Wani, then an undergraduate and now a PhD student, to help modify PVC pipes at the UW Makerspace. Using various diameters of pipe purchased at a local hardware store, Ahmed, Yani and their team altered the length and diameter of the pipes until they could produce the same resonance as they voice they were attempting to imitate.

Eventually, the team developed an algorithm that can calculate the PVC pipe dimensions needed to transform the resonance of almost any voice to imitate another. In fact, the researchers successfully fooled the security systems with the PVC tube attack 60 percent of the time in a test set of 91 voices, while unaltered human impersonators were able to fool the systems only 6 percent of the time.

The spoof attack works for a couple of reasons. First, because the sound is analog, it bypasses the voice authentication system’s digital attack filters. Second, the tube does not transform one voice into an exact copy of another, but instead spoofs the resonance of the target voice, which is enough to cause the machine learning algorithm to misclassify the attacking voice.

Fawaz says part of the motivation behind the project is simply to alert the security community that voice identification is not as secure as many people think it is, though he says many researchers are already aware of the technology’s flaws.

The project has a bigger goal as well.

“We’re trying to say something more fundamental,” Fawaz says. “Generally, all machine learning applications that are analyzing speech signals make an assumption that the voice is coming from a speaker, through the air to a microphone. But you shouldn’t make assumptions that the voice is what you expect it to be. There are all sorts of potential transformations in the physical world to that speech signal. If that breaks the assumptions underlying the system, then the

system will misbehave.”