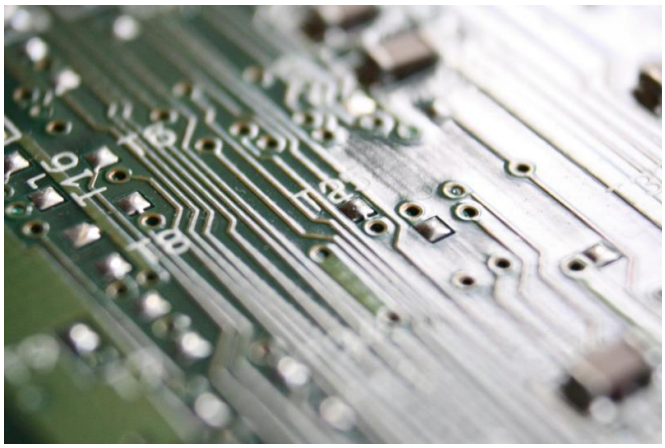
Research finds automated voice imitation can fool humans and machines

28 September 2015, by Katherine Shonesy



Credit: Public Domain

University of Alabama at Birmingham researchers have found that automated and human verification for voice-based user authentication systems are vulnerable to voice impersonation attacks. This new research is being presented at the European Symposium on Research in Computer Security, or ESORICS, today in Vienna, Austria.

Using an off-the-shelf voice-morphing tool, the researchers developed a voice impersonation attack to attempt to penetrate automated and human verification systems.

A person's voice is an integral party of daily life. It enables people to communicate in physical proximity, as well as in remote locations using phones or radios, or over the Internet using digital media.

"Because people rely on the use of their voices all the time, it becomes a comfortable practice," said Nitesh Saxena, Ph.D., the director of the Security and Privacy In Emerging computing and networking Systems (SPIES) lab and associate professor of computer and information sciences at UAB. "What they may not realize is that level of comfort lends itself to making the voice a vulnerable commodity. People often leave traces of their voices in many different scenarios. They may talk out loud while socializing in restaurants, giving public presentations or making phone calls, or leave voice samples online."

A person with potentially malicious intentions can record a person's voice by being in physical proximity of the speaker, by making a spam call, by searching and mining for audiovisual clips online or even by compromising servers in the cloud that store audio information.

This study from researchers within the UAB College of Arts and Sciences Department of Computer and Information Sciences and Center for Information Assurance and Joint Forensics Research explores how an attacker in possession of audio samples from a victim's voice could compromise the victim's security, safety and privacy.

Advances in technology, specifically those that automate speech synthesis such as voice morphing, allow an attacker to build a very close model of a victim's voice from a limited number of samples. Voice morphing can be used to transform the attacker's voice to speak any arbitrary message in the victim's voice.

"As a result, just a few minutes' worth of audio in a victim's voice would lead to the cloning of the victim's voice itself," Saxena said. "The consequences of such a clone can be grave. Because voice is a characteristic unique to each person, it forms the basis of the authentication of the person, giving the attacker the keys to that person's privacy."

As a case study for this paper, the researchers investigated the aftermaths of stealing voices in two important applications and contexts that rely upon voices as the basis for authentication.

The first application is a voice-biometrics, or speaker-verification, system that uses the potentially unique features of an individual's voice to authenticate that individual.

"Voice biometrics is the new buzzword among banks and credit card companies," Saxena said. "Many banks and credit card companies are striving for giving their users a hassle-free experience in using their services in terms of accessing their accounts using voice biometrics."

The technology has now also been deployed on smartphones as a replacement to traditional PIN locks, and is being used in many government organizations for building access control.

Voice biometrics is based on the assumption that each person has a unique voice that depends not only on his or her physiological features of vocal cords but also on his or her entire body shape, and on the way sound is formed and articulated.

Once the attacker defeats voice biometrics using fake voices, he could gain unfettered access to the system, which may be a device or a service, employing the authentication functionality.

Secondly, the research team looked at the implications stealing voices had on human communications as its other application for the paper's case study. The voice-morphing tool imitated two famous celebrities, Oprah Winfrey and Morgan Freeman, in a controlled study environment.

If an attacker can imitate a victim's voice, the security of remote conversations could be compromised. The attacker could make the morphing system speak literally anything that the attacker wants to, in the victim's tone and style of speaking, and can launch an attack that can harm a victim's reputation, his or her security, and the safety of people around the victim.

"For instance, the attacker could post the morphed voice samples on the Internet, leave fake voice messages to the victim's contacts, potentially create fake audio evidence in the court and even impersonate the victim in real-time phone conversations with someone the victim knows," Saxena said. "The possibilities are endless."

The results show that the state-of-the-art automated verification algorithms were largely ineffective to the attacks developed by the research team. The average rate for rejecting fake voices was less than 10 to 20 percent for most victims. Even human verification was vulnerable to the attacks. According to two online studies with about 100 users, researchers found that study participants rejected the morphed voice samples of celebrities as well as somewhat familiar users about half the time.

"Our research showed that voice conversion poses a serious threat, and our attacks can be successful for a majority of cases," Saxena said. "Worryingly, the attacks against human-based speaker verification may become more effective in the future because voice conversion/synthesis quality will continue to improve, while it can be safely said that human ability will likely not."

While the results of this study show how vulnerable a person can be to voice attacks, there are ways to prevent one's voice from being stolen. Saxena suggests people increase their awareness of the possibility of these attacks, and also that they be wary of posting audio clips of their voices online.

"Ultimately, the best defense of all would be the development of speaker verification systems that can completely resist voice imitation attacks by testing the live presence of a speaker," Saxena said. "Our future research will examine this and other defense strategies."

More information: esorics2015.sba-research.org/

Provided by University of Alabama at Birmingham