BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF INFORMATION TECHNOLOGY

Excel@FIT 2023
http://excel.fit.vutbr.cz

Resilience of Biometric Authentication of Voice Assistants against Deepfakes

Petr Kaška*

Abstract

voice assistants (Apple Siri, Amazon Alexa, Google-assistant, Cortana) supporting voice authentication offer more and more possibilities to facilitate all our daily activities. People give them access to data and information to take full advantage of all these features. Along with the rapidly developing voice deepfake technology, a great threat is emerging in the misuse of deepfakes to trick smart assistants. An attacker can record the victim's voice, synthesize the voice and create a recording of some command to trick the assistant. The aim of this paper is to investigate the current state of assistant defenses against deepfakes. The conducted experiment proves the initial hypothesis about the vulnerability of voice assistants to deepfake attacks. The results are rather alarming and require the introduction of further countermeasures to avoid risks abuse, as the number of voice assistants in use is currently comparable to the world's population.

*xkaska01@vut.cz, *Faculty of Information Technology, Brno University of Technology*

1. Introduction

Digital Assistants, also known as Virtual Assistants, Intelligent Personal Assistants, and Artificial Intelligence Assistants(VA), are becoming increasingly popular due to their growing sophistication and capabilities. These assistants are integrated into devices such as speakers, mobile phones, and web services and use advanced AI and algorithmic approaches to perform tasks for individuals, answer questions, maintain conversations with users, and retain information for issuing reminders and warnings based on environmental constraints like time and location. This makes them particularly helpful for individuals with mobility problems and the elderly.

Research by Chambers and Beaney [1] suggests that VAs can also be used to address patients' health and care needs.

With the growing popularity of virtual assistants, security, privacy and legal risks are increasing. This study aims to challenge VAs that attempt to circumvent the above limitations. As a first step, a deepfake tool was selected. The main criterion for selecting the tool was its ease of use. Subsequently, a minimal amount of voice samples were collected for the creation of a synthetic voice. In the second step, which followed the data collection, voice synthesis was performed using a third party application, Reasemble AI. Finally, the created synthesized output was replayed from another device in order to attack the VA with the appropriate commands. There are many methods to do this, e.g., the adversary plays back the so-called sound when in the vicinity of the VA, or the sound is reproduced by a smartphone or by inserting a malicious command into TV or radio adswhich triggers the VA. In [2, 3, 4], exploiting the fact that some VAs do not distinguish the source of the sound.

Around 3.25 billion VA devices were purchased globally in 2019. Projections indicate that by the end of 2024, the number of VA devices will surge to approximately 8.4 billion units, a figure equivalent to the world's population [5]. VAs have become ubiquitous in consumer electronics, from smartphones to home appliances to mobile-controlled automotive systems. As such, it is critical to comprehend the magnitude of risks associated with attacks on VAs. Typically, VAs are not restricted to isolated environments or smartphone-only applications due to their increased utility with the growing number of smart appliances. Consequently, VAs are inherently linked with the smart home, an integrated system of connected

devices and sensors that can be remotely controlled and scheduled using internet-connected devices such as smartphones, tablets, and watches. A central "gateway" connects these devices, enabling the user to manage all connected devices, including lighting, thermostat, boilers, and other functions through their personal device, even when physically remote from it. The user receives relevant notifications of any operation in the house at any given moment. However, the use of VAs in home automation poses a plethora of security risks to users. If VAs are easily tricked, it can have catastrophic effects on home automation or leakage the user's personal information. Moreover, smart home automation grants physical access, which means that attacks can extend the cyber layer to reach the physical. Figure 2 illustrates various home devices that can be manipulated by a VA, providing insight into the potential threat an attacker can pose by controlling them.

2. Experiment design

The experiment only targets user authentication in English language.All the tested assistants allow voice authentication by voiceprint, which consisted of repeating predefined phrases. That is, whether the quality of the synthetic voice created by the commercial tool is sufficient for the voice assitent to accept the authentication of the attacker using the created synthetic voice from the voice samples of the victim using the assistant. First, the basic parameters of the experiment are determined.

2.1 Attacker model

An attacker is someone who is able to create a voice deepfake by synthesizing the victim's voice and penetrate the ASV (automatic speaker verification) of selected voice assistants (VAs) in order to command the voice assistant to do malicious things or to obtain the victim's personal information, for example, the voice assistant that is part of a Smart-Home. Thus, the attacker possesses samples of the victim's voice, knows the necessary information and procedures for creating voice deepfakes, and knows the details of the ASV system and its functionality. The attacker must be able to collect the necessary number of samples of the victim's voice to be able to create his own synthesized model of the victim's voice in order to control the VA. [6]

The targets that an attacker can target with his attack can be as follows

- Security - The attacker wants to manipulate devices that support Smart-Home and cause damage. For example. manipulate door locks, window locks.
- Privacy - The attacker wants to obtain information about the victim in order to exploit it, blackmailing the victim. E.g., Obtain information about the victim's daily routine through a calendar shared among VAs, have VAs call a toll-free line, insert some malicious event into the victim's calendar, or send messages with the victim's name. (emails,... linking to the card)
- Other - attacker's ways to harm the victim are to order the assistant to visit some malicious site, buy things on the internet from the victim's account.

The only problem for the potential attacker is to obtain samples of the victim's voice, but these can be obtained by manipulating the victim into saying the phrases mentioned.

2.2 Input metrics

The Reasemble AI tool in the non-paid version was chosen for the experiment because it is a very simple tool to use. It can be said that even a person without much knowledge of information technology could create their own synthetic voice model. Sample collection consisted of repeating the selected invocation phrase 50 times. The phrases used in the experiment were selected based on the invocation phrases of each assistant. "Hey Siri" for apple, "Alexa" for Amazon Alexa, "Hey Google" for Google Assistant and "Cortana" for Windows Cortana. Subsequently, I decided to create 3 models of my (male) voice and thus performed the experiment 3 times. With each model the experiment was conducted separately to check the robustness of the experiment. Figure 2 shows the different parts of the experiment.

3. Conclusions

An experiment was designed that clearly demonstrates that the current VA protection against synthetic speech is critical. Figure 3 shows that an attacker is able to use Deepfakes to authenticate to individual systems very easily to Alexa with 93.3% success rate, to Siri with 77.3% success rate, to Cortana with 95.3% success rate, and to Google assistant with 94% success rate. We believe that for systems that are so widely used and integrated into millions of devices and connected to many more, these numbers are very serious. We want to conduct further tests of more sophisticated attack methods and their impact on authentication success. Create and test a

female voice model and compare the results with an experiment with a male voice model. And create our own voice assistant protection system.

Acknowledgements

I would like to thank my supervisor Mgr. Kamil Malinka Ph.D. and my consultant Ing. Anton Firc for their invaluable guidance and support.

References

[1] Richard Chambers and Paul Beaney. The potential of placing a digital assistant in patients' homes. *British Journal of General Practice*, 70:8–9, 2020.

[2] Jane Wakefield. Burger King advert sabotaged on Wikipedia, 2017.

[3] Efthimios Alepis and Constantinos Patsakis. Monkey says, monkey does: Security and privacy on voice assistants. *IEEE Access*, 5:17841–17851, 2017.

[4] Rui Zhang, Xiaokuan Chen, Jun Lu, Sheng Wen, Surya Nepal, and Yang Xiang. Using ai to hack ia: A new stealthy spyware against voice assistance functions in smart phones. *arXiv preprint arXiv:1805.06187*, 2018.

[5] Daniel Ruby. 65 Voice Search Statistics For 2023 (Updated Data), 2023.

[6] Chen Yan, Xiaoyu Ji, Kai Wang, Qinhong Jiang, Zizhi Jin, and Wenyuan Xu. A survey on voice assistant security: Attacks and countermeasures. *ACM Computing Surveys*, 55(4):84:7, May 2022. Publication date: 21 November 2022.