BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF INFORMATION TECHNOLOGY

Excel@FIT 2024
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, Samsung Bixby) supporting voice control offer more and more possibilities to make all our daily activities easier. People give them access to data and information to take full advantage of all these features. Along with the rapidly developing voice deepfake technology, there is a big threat in the area of misusing deepfakes to trick smart voice assistants. An attacker can record the victim's voice, synthesize the voice and create a recording of some command to trick the assistant in order to harm the victim. The aim of this work is to design an experiment that will simulate attacks, performed by synthetic voice, on voice assistants and then evaluate their defensiveness. The conducted experiment confirms the initial hypothesis of the vulnerability of voice assistants to deepfake attacks and the results are very alarming with an overall success rate of 90% indicating insufficient defense of voice assistants and require the implementation of additional countermeasures to prevent the risk of misuse as the number of voice assistants in active use is rapidly increasing.

*xkaska01@stud.fit.vutbr.cz, *Faculty of Information Technology, Brno University of Technology*

1. Introduction

In recent years, advances in Deepfake technology have been accelerating more and more. So nowadays a user with minimal IT knowledge can create their own synthetic media within a few hours. With this medium, he can become a potential attacker and attack some systems to cause financial damage to the victim or create misinformation to influence masses of people. Thus, our main goal was to experimentally demonstrate the vulnerability of voice assistants to trendy voice synthesizers.

In already published works, usually only one voice assistant is tested, on few testing samples, and the deepfake recordings used in the experiments are made on some unknown synthesizers.

It was therefore necessary to design a large-scale experiment whose results would be replicable. We set four voice assistants and four synthesizers to test. We conducted one more small pre-experiment to verify a key feature of voice assistants, which is that the assistant applies automatic speaker verification only to the invocation phrase and not to the entire communication with the user.

The results of our experiment confirmed the vulnerability of voice assistants, since we were able to achieve an extreme success rate of simulated voice assistant attacks, compared to other works in which the highest success rate was around 30% [1, 2]. The information about the inability of voice assistants to defend could be used to motivate developers to create defense mechanisms against these types of attacks.

2. Experiment Design

For the experiment we used 4 synthesis tools. Two of them were free (XTTS, TortoiS) and two of them were paid (Resemble AI, CoquiTTS). As test assistants we chose Bixby, Alexa, Siri and Google because they are the most used voice assistants nowadays. We chose to record the respondents through the microphone of the laptop. Most of the recording of the respondents took place in a quiet room of the student halls. The other part took place on the premises of SCHOTT, which we asked to cooperate so that we would not have only stundents in the experiment.

It was also complicated to devise a testing process, given that each of the tools required differently formatted input data, which would have extended the

time the respondent needed with the experiment to over an hour, which would have been inconvenient for the respondent. We therefore devised a compromise whereby we recorded into two instruments at once. Next, the respondent was no longer needed at the experiment and the synthesis phase came in, which lasted from 10 to 80 minutes. In this phase, we created the required voice models and used the TTS method to make deepfake recordings of the invocation phrases for each assistant.

3. Experimental Process

Our experiment lasted a total of **2 months**. We managed to collect data from **72 respondents**. The gender distribution of the respondents is shown in the graph and the age distribution of the respondents is shown in the graph. Our experiment was conducted in English language and no respondent had English as their mother tongue.

4. Experiment Results

The results of our experiment confirmed the trend of vulnerability of voice assistants on a scale that has never been done before. We calculated the success rate of the attack based on a metric we created, and it was nearly 90%. The results are recorded in the graph. They show that the most vulnerable voice assistants are Alexa, Google and Siri. Ironically, these assistants are the most used.

Among the synthesizers, the commercial tool Resemble AI had the highest success rate, but we got slightly worse results with the free tool TortoiS. We further divided the results by the success rates of male VS female deepfakes. Since the representation of men and women was uneven, these results are not conclusive, but there are no big differences. Except for Bixby, where female deepfakes had many times higher success rates, which may be due to training the bixby model on male voices.

5. Conclusions

We conducted an extensive experiment as part of this work, which clearly confirms the lack of robustness of current voice assistants to replay and deepfake attacks. The findings show that there is a real danger that a potential attacker can relatively easily record the victim's voice, synthesize his speech, and then use this fake speech to manipulate the victim's voice assistant. In this way, the attacker could reveal sensitive personal information or cause financial harm.

In this context, the choice of appropriate use cases for voice assistants is a key element of decision-making. It is necessary to consider in detail when it is appropriate to use voice authentication and when it is necessary to implement more robust security measures.

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] Domna Bilika, Nikoletta Michopoulou, Efthimios Alepis, and Constantinos Patsakis. Hello me, meet the real me: Voice synthesis attacks on voice assistants. *Computers and Security*, 137:103617, 2024.

[2] Justin Ubert. *Fake It: Attacking Privacy Through Exploiting Digital Assistants Using Voice Deepfakes*. PhD thesis, 2023. Copyright - Database copyright ProQuest LLC; ProQuest does not claim copyright in the individual underlying works; Last updated - 2023-05-18.