REPLAY ATTACKS AND COUNTERMEASURES AGAINST THEM

Alexandru BUJOR, Artiom BOZADJI, Gheorghe GURSCHI*

Department of Software Engineering and Automation, group FAF-231, Faculty of Computers, Informatics, and Microelectronics, Technical University of Moldova, Chisinau, Republic of Moldova

Corresponding author: Gurschi Gheorghe gheorghe.gurschi@isa.utm.md

Abstract: *Voice authentication technology uses an individual's unique voice characteristics for secure identity verification, providing a seamless method to access devices and services. However, it faces challenges in maintaining robustness against security breaches and impersonation attempts. This study examines the effectiveness of voice authentication technology, focusing on its ability to analyze and differentiate between complex voice attributes like pitch, tone, and speech patterns. The study found that advanced voice authentication systems have high accuracy in recognizing and validating users based on voice biometrics, enhancing security for various applications. The findings emphasize the importance of continuous advancements in voice authentication technology to counteract evolving security threats and ensure a safer and more reliable user experience across various sectors, including virtual assistants and customer service interfaces.*

Key words: *authentication, convolutional neural networks, replay attack, safe access, security, voice authentication.*

Introduction

Voice authentication technology provides a convenient and safe way to access devices and services by using the distinctive qualities of a person's voice to confirm their identity [1]. This biometric system is a popular option for protecting smartphones, bank accounts, and sensitive data because it analyzes a variety of hard-to-replicate aspects of a person's voice, including pitch, tone, and speech patterns. Robust voice authentication systems are increasingly important as voice commands find their way into everyday life, from virtual assistants to customer service interfaces.

Figure 1. System usage scenario

Replay attacks, however, pose a threat to the security of voice authentication systems. Replay attacks happen when an attacker records a valid transaction or data transmission—like a voice command—and then purposefully or dishonestly delays or replicates this transmission in order to gain unauthorized access to resources.

Because an attacker could record and replay the voice of an authorized user to access secured systems, this kind of attack is especially concerning for voice authentication. Voice authentication is susceptible to these kinds of attacks even with its sophisticated technology, which emphasizes the need for more security precautions. To identify such attacks, a trained convolutional neural network (CNN) can discriminate between live and recorded human speech spectrograms using picture classification. CNNs are a deep learning (DL) technique utilized in

Machine learning (ML) techniques that are commonly used for image categorization and speech recognition. They are faster than typical neural networks at classifying visual contents and require less computation.

Background and Related Work

The use of human biometric features in authentication methods has advanced significantly in the last few decades. Voice and fingerprint authentication have been widely incorporated into mobile phones and applications. The capabilities of current systems have grown to include authentication based on cardiac motion, facial recognition (as demonstrated in the iPhone X), and gait recognition using Wi-Fi signals.

The increasing integration of voice commands in everyday technology, from virtual assistants to customer service interfaces, underscores the urgent need for secure voice authentication methods. The motivation behind this study is to fortify the security measures of voice authentication systems, ensuring they remain reliable in the face of evolving cyber threats.

The primary aim of this study is to develop a method using CNNs that can accurately differentiate between live and recorded speech, thereby detecting replay attacks against voice authentication systems. Objectives include:

- Evaluating the current vulnerabilities of voice authentication to replay attacks.
- Designing and training a CNN model to recognize the unique characteristics of live versus recorded speech.
- Testing the model's effectiveness in a controlled environment and real-world scenarios [2].

Security of Voice Authentication:

Voice authentication is susceptible to replay, speech synthesis, impersonation, and voice conversion spoofing attacks. Various tactics have been used to counter this. Some approaches use the Time Difference of Arrival (TDOA) data from numerous microphones, while others analyze the variations in synthesis or conversion aspects, and still others look at the phase spectra disparities between arriving and actual speech. Techniques for countering replay assaults include examining the physical characteristics of the recording equipment, determining whether there is background or channel noise, and using extra verification methods such as video.

Device-free, ultrasound-based measurements:

In line with our findings, ultrasound technology has made a lot of distance measurement applications possible, especially in situations when devices are not present. Projects like as LLAP use baseband signal phase changes to track finger motions in a two-dimensional space. Using chirps, Strata and FingerIO are able to track fingers or palms in two dimensions with very few errors. UltraGesture uses ultrasonic to identify different human gestures, while PCIAS gauges the rotational speed of rotating objects. In contrast to previous methods, our work focuses on the extraction of distinct voiceprint information from ultrasound signals for authentication reasons, such as vocal cord and mouth movement information.

II. Replay attack:

Replay attacks are a kind of network attack where the attacker records an authentic network transfer and then sends it again at a later time. The primary goal is to deceive the system into believing that the data being retransmitted is authentic. Replay assaults are dangerous as well since they are hard to identify. Moreover, it can succeed even in the event that the initial transmission was encrypted [3].

Replay attacks can be launched by an attacker to obtain unauthorized access to networks or systems. Replay attacks can also interfere with a system's normal functioning by bombarding it with repetitive requests. This attack can be planned by an attacker who will use network packet interception and retransmission. Furthermore, recreating an attack can be an effective way to carry out a replay [4].

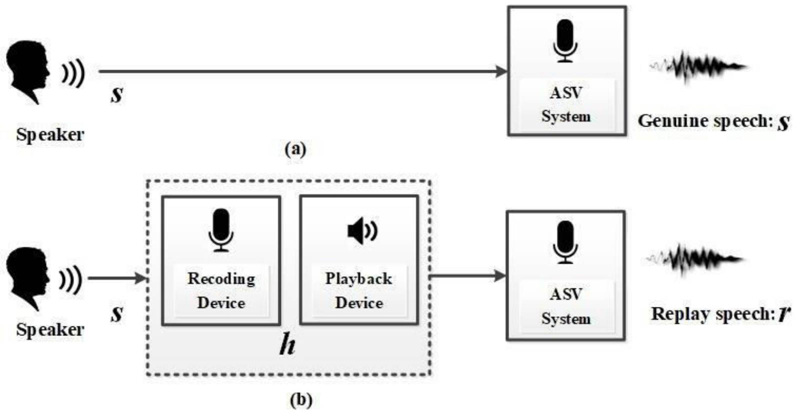

Figure 2. Replay attack in action

Since the majority of research is concentrated on deepfake speech synthesis and conversion, audio replay attacks can pose a serious threat to Automatic Speaker Verification (ASV) systems due to their low cost and effort requirements. As a result of this difficulty, voice antispoofing—including its types, origins, and preventive measures—was developed [5].

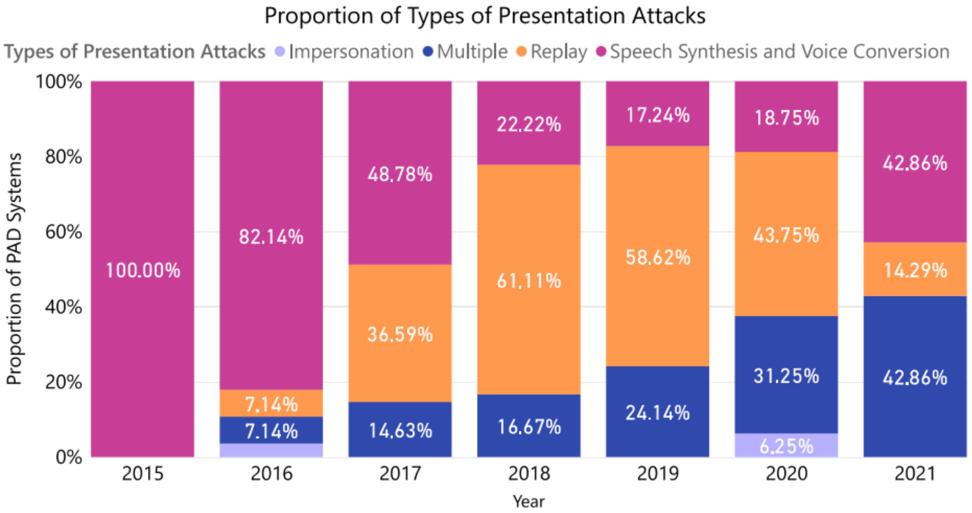

Figure 3. Antispoofing statistics in voice PAs from 2015 to 2021.

The proportion of each type of attack detected in that year is indicated by the percentages on the bars. For example, all attacks in 2015 were impersonation attacks. The range of attack types has grown over time, and by 2021, speech synthesis/voice conversion, replay, and impersonation attacks are all represented, with impersonation remaining the most common but less so than in prior years [6].

By giving every encrypted element a distinct session ID and component number, replay attempts can be avoided. Due to its dual-layered operation, which is independent of one another, vulnerabilities are successfully mitigated. The possibility of duplicating a prior run of the program is greatly decreased by creating a random session ID for every program running. Consequently, because the session ID varies with every run, an attacker would have difficulty carrying out a replay assault.

Replay attacks can be avoided by using session IDs, also known as session tokens. The following actions are usually included in the process of creating a session ID:

- John sends Alice a one-time token, which she uses to complete the password translation process and then sends John the outcome.
- John uses the session token to carry out the identical calculation on his end.
- Login success is contingent upon the matching of John’s and Jenifer's computed values.

In the event that an attacker like Eve captures this value and attempts to use it in another session, John would assign a different session token. Consequently, when Eve tries to reuse her captured value, it will not match Bob's computation, thereby indicating to John that it's not Janifer attempting to authenticate [7].

It's crucial that session tokens are generated through a random process, typically using pseudorandom methods. This prevents Eve from posing as Bob by predicting future tokens and convincing Alice to incorporate them into her transformation. By replaying her response at a later time using the predicted token, Eve could trick Bob into accepting the authentication.

Conclusions

In conclusion, this study underscores the critical significance of addressing the vulnerability of voice authentication systems to replay attacks. By leveraging convolutional neural networks (CNNs), we have proposed a sophisticated methodology aimed at discerning between live and recorded human speech, thereby enhancing the resilience of authentication mechanisms against malicious exploitation. As voice commands become increasingly integrated into various facets of daily life, from virtual assistants to customer service interfaces, the imperative to safeguard against such attacks grows ever more pressing. Our research not only contributes to the ongoing efforts to bolster the security of voice authentication technology but also emphasizes the indispensable role of continuous innovation in confronting emerging cybersecurity challenges. Moving forward, sustained advancements in this field are paramount to ensuring the integrity and reliability of voice authentication systems, thus fostering a safer and more trustworthy digital environment for users worldwide.

References:

[1] Z. Meng, M. U. B. Altaf, Biing-Hwang Juang“Active voice authentication” [Online]. Available: https://www.hypr.com/security-encyclopedia/voice-authentication

[2] “What are convolutional neural networks?” [Online]. Available: https://www.ibm.com/topics/convolutional-neural-networks

[3] “What Is a Replay Attack?” [Online]. Available: https://www.kaspersky.com/resource-center/definitions/replay-attack

[4] B. Zhang, B. Tondi , M. Barni *Adversarial examples for replay attacks against CNN-based face recognition with anti-spoofing capability* 9 May 2020

[5] H. Dai, W. Wang, Alex X. Liu, K. Ling, J.S Sun *Speech Based Human Authentication on Smartphones.* 16th *Annual IEEE International Conference on Sensing, Communication, and Networking* 2019.

[6] C. B. Tan, M. H. A. Hijazi, N. Khamis, P. N. E. binti Nohuddin, Z. Zainol, F. Coenen, A. Gani *A survey on presentation attack detection for automatic speaker verification systems: State-of-the-art, taxonomy, issues and future direction.* 4 August 2021. Available: https://link.springer.com/article/10.1007/s11042-021-11235-x

[7] “What are Session Replay Attacks ?” 08 Aug, 2022. [Online]. Available: https://www.geeksforgeeks.org/what-are-session-replay-attacks/.