IBIMA Publishing
Journal of Information Assurance & Cyber security
https://ibimapublishing.com/articles/JIACS/2021/198312/
Vol. 2021 (2021), Article ID 198312, 11 pages, ISSN: 2165-9923
DOI: 10.5171/2021.198312

Research Article

Breaking Voice Authentication – Security Testing Approach

Olga Dziegielewska

Military University of Technology, Warsaw, Poland
olga.dziegielewska@wat.edu.pl

Received date:13 November 2020; Accepted date:14 January 2021; Published date: 18 February 2021

Academic Editor: Claudia Carstea

Abstract

A couple of years ago, along with raising popularity of smart devices that promote voice authentication mechanisms, voice recognition became a buzz topic again. Not only voice patterns recognition algorithms significantly improved their quality, but also with a rapid growth in the development of sound editing tools and an outburst of the deep-fake concepts, the attack surface of voice authenticators significantly expanded. This paper describes the approach for penetration testing of the voice recognition solutions and tackles common misconceptions of voice patterns characteristics.

Keywords: voice biometrics testing, voice biometrics threat modelling.

Introduction

In the recent years, voice authentication systems are on the raise. Along with raising popularity of smart devices voice unlock functionality, other industries started looking deeper into the possibility of voice authentication implementation for their needs as a single or a part of a multi-factor authentication, e.g., in IVR systems. As an effect, the threat landscape changed.

So far, there are no standardized documents explaining the approach to testing of the voice authenticators, therefore security researchers try to apply and adapt different techniques and processes when assessing the security of such mechanisms. This paper presents a proposition of an approach to security testing of the voice authentication systems starting with the threat modeling, proposing test scenarios and concluding with the risk evaluation for the identified issues.

Voice Authentication Systems

To better understand the threat landscape, a security researcher or assessor must first understand the basic breakdown of the voice authentication systems and their characteristics as it provides better understanding of the potential threats.

Cite this Article as: Olga Dziegielewska (2021), "Breaking Voice Authentication – Security Testing Approach", Journal of Information Assurance & Cyber security, Vol. 2021 (2021), Article ID 198312, DOI: 10.5171/2021.198312

Online systems are understood as systems which are evaluating the samples using online database that is updated in a concurrent manner, e.g. online IVR systems.

Offline systems are understood as systems which are evaluating the samples using offline database that is frequently physically stored on the input device, e.g. smart devices.

Active authentication or ***non-adaptive authentication*** is understood as a process in which a user is aware of the voice authentication process as he/she is asked to provide a voice sample during the process (Gajo, A., 2020).

Passive authentication or ***adaptive authentication*** is understood as a process in which a user may not be aware of the voice authentication process as he/she is not asked to provide a voice sample during the process, but the system analyzes the voice that is gathered through a different process, e.g. during a conversation with an online banking assistant (Gajo, A., 2020).

Playback detection or ***fingerprinting detection*** is a feature of the authentication system that allows to detect the same voice sample being replayed into the input device during two independent and separate authentication processes.

Recording detection is a feature of the authentication system that allows to detect the recording of a voice sample being provided into the input device. This type of feature frequently provides further breakdown portraying its accuracy when using ***collaborative*** and ***non-collaborative*** recordings. Collaborative recordings are understood as high quality recordings of a user saying his/her authentication quote directly to the recorder, while non-collaborative recordings are understood as low quality recordings that were grabbed without user's knowledge.

Understanding Voice Dynamics

To properly assess security of voice authentication systems, it is important to understand the physical properties of the voice dynamics, including: the linguistic properties of languages the authentication system is designed for, acoustic conditions during the sampling process and the genetic properties.

The linguistic properties that need to be taken into account as a minimum are phonemes, inflexion and intonation of the covered languages. Those three characteristics help to differentiate not only the languages, but the nationalities of the sample suppliers as non-native speakers tend to mirror the inflexion and intonation of their native languages when they speak in a different language (Crystal, D., 2008).

One characteristic that is partially a linguistic and partially genetic property is the tempo, as both of the factors influence the tempo of one's speech. Another important genetic property is the genetic similarity, which can be understood as the degree to which the voice of different people is similar. This is especially important for considering the corner cases in the threat modelling process as people who are related tend to have close proximity when it comes to the genetic voice characteristics.

The acoustic conditions also play a huge role in voice sampling as the background noise or the echo that is present during different phases of the voice authentication can later influence the overall security of the system.

Threat Modelling

The threat modelling process for voice authentication systems testing follows typical steps for such activities, i.e., at first the scope of the test must be determined, then the documentation analysis should be performed, and, as a last step, the test scenarios are created.

Determining the scope is a crucial part of each test, but in case of the biometrics

testing, it is especially important to establish the project limitations as there are no standardized systematic approaches to tackle each and every aspect of the biometrics mechanisms security testing, e.g., similar to the OWASP ASVS. The main question to ask is what the purpose of the test is, as different scenarios would be developed accordingly depending on the purpose. In case of the voice biometrics, the key aspect is the validation of the biometrics engine's susceptibility to attacks performed from a perspective of an external attacker. The two most important factors here are: to determine what kind of attackers pose the biggest threat to the tested environment, and to understand the characteristics of the tested solution.

What naturally follows is determining what kind of testing can be performed, i.e., white-box, grey-box or black-box. When performing testing for the biometrics product owners, the white-box approach is a feasible option, however in most cases the tester will not have access to the source code of the solution, therefore grey-box and black-box are more common.

Regardless of the selected testing approach, it must be confirmed at early stages of the threat modelling which biometrics engine is used and what are its features. Getting to know what kind of engine is in use allows to craft tool-specific scenarios.

What also helps in developing dedicated test cases is the documentation of the system, understood not only as the technical specification of the product itself, but also as the actual implementation and post-implementation documentation, the risk analyses and procedural guides. The relevant supporting documentation should be identified as one of the first steps of this part of the assignment as any gaps in the received material may impact the outcome of the test, therefore should be addressed before actual testing phase. In case of the full black-box testing approach, it is most likely that there will not be any kind of documentation provided beforehand, therefore the threat modelling and the scenarios development must be based on previous experiences of this sort as no systematic standard for testing voice-based biometric authentication is formally developed yet.

The key elements to look for in the system's documentation are:

- What kind of features provides the system?
- What is the type of accepted input and the type of the provided output?
- How is the voice pattern stored in the database? How is the integrity of the pattern over time measured?
- Where is the pattern stored (product owner's side, integrator's side or the client's side)?
- Does the tool provide:
    - playback detection, also described as fingerprinting,
    - recording detection and if so, how its effectiveness differs for collaborative and non-collaborative recording types,
    - genetic similarity detection?
- If the engine can be used in the telephony systems:
    - What are the accepted telephony protocols?
    - Does the engine provide the same level of resilience against attacks using different kinds of telephony protocols?
- How the supporting infrastructure architecture looks like and what kind of security protection mechanism is in use?

The last question is especially important, as in the voice biometrics systems, there are usually three key elements that need to be somehow connected: the voice samples recording system, the biometric engine and the database. The interesting case are the telephony systems which use the telephony devices with conjunction of the telephony protocols as their recording system, therefore introducing more threat vectors to the initial assessment.

Except from the technical layer, the procedural layer may be also assessed as a part of the assignment. Nonetheless, it must be noted that the procedural part may help in understanding the business need of biometrics use cases for the authentication system and also to properly assign the risk levels after a gap or vulnerability is identified but should not state a sole subject of the biometrics systems' testing. The key elements to look for at the procedural level are:

- What kind of business processes are protected by the biometric authentication?
- How does the decision tree for the processes look like?
- What are the stages of biometric pattern registration?
- What are the stages of the biometric authentication?
- Have any risk analyses been performed for the process and the selected tool? What are the results? What elements have been analyzed?

The important factor that needs to be mentioned here is establishing the relation between the product owner, i.e., the biometric engine legal owner, and the entity that requested the test of the engine. If the product owner requested the test, the case is pretty transparent.

If the requestor is a different legal entity, then it is worth confirming with the client that the third-party provider is notified that the test activities will take place and that the agreement between the client and the third-party allows for such activities. It is also important to confirm the legal aspects, as in case of a gap or vulnerability identified in the solution itself; the risk mitigation falls onto the product owner and the requestor loses the control over the mitigation process, what must be taken into account.

Depending on the project scope established in the previous steps and information gathered during the documentation analysis, the threat models and the test scenarios strictly related with them can be designed.

Test Scenarios

In case of biometrics mechanisms testing, potential attack scenarios prepared and agreed with the test requestor help in understanding the real threat related with used solution. In contrast with standard approach of vulnerabilities testing, the biometric mechanisms testing is based rather on story-telling than automatic-scans-like vulnerabilities detection.

As previously mentioned, the fundamental divisions of the scope areas in case of voice biometrics testing are:

- Supporting infrastructure tests,
- Biometrics engine tests.

Additionally, the perspective of an attacker must be considered, e.g.:

- External attacker with limited resources,
- External attacker with unlimited resources,
- Internal attacker with limited privileges,
- Internal attacker with unlimited privileges.

The following pages present the general outline of the test, focused on the specific areas and more detailed baseline for the biometrics engine test.

From the requestor's risk perspective, the key part of the engagement is usually the biometric engine test scenarios covering the perspective of an external attacker.

A. Threat agent: internal or external with internal access

The table below presents base scenarios for the threat agent described as an internal attacker or an external attacker who gained access to the internal system.

Olga Dziegielewska (2021), Journal of Information Assurance & Cybersecurity,
DOI: 10.5171/2021.198312

Scenario name	Description
Algorithm-level abuse	Targeted area: biometric engine Action: Analysing used algorithms in terms of effectiveness against the voice recognition attacks.
Source-code level errors	Targeted area: biometric engine Action: Analysing the source code of the biometric engine, including its configurations, to verify the effectiveness and security of the implementation of the algorithms and the system as a whole.
Voice pattern integrity abuse	Targeted area: infrastructure Action: Verification of the database server and the database security by analysing its configuration and performing exploitation of the attacks identified in the infrastructure that could affect the database integrity. Attempts to replace the patterns from the perspective of different roles in the system (e.g., administrator, regular internal user, external user).
Voice pattern capture	Targeted area: infrastructure Action: Traffic analysis between the system's elements. Measuring the probability of capturing the voice pattern by tapping the user's phone.
Voice pattern leakage	Targeted area: infrastructure Action: Traffic and configuration analysis covering the system's elements.
Man-in-the-Middle for registration or validation	Targeted area: infrastructure Action: Attempts to proxy the voice traffic between the user and the analysed system.

*B. **Threat agent: external attacker***

The table below presents base scenarios for the threat agent described as an external attacker who has only access to the input interface of the system. All of the scenarios mentioned below target the biometric engine.

Scenario name	Description
Voice pattern registration process abuse	Attempts to register a biometric pattern under various acoustic conditions (e.g., different characteristics of background noise) and technical condition (e.g., registration of samples from recordings) to verify how such cases affect the authentication mechanism.
Voice pattern reregistration	Attempts to use or bypass alternative authentication methods available in the system to modify (or overwrite) the already registered biometric pattern. Attempts to invoke parallel pattern registration for the same user ID and verifying how such action affects the authentication mechanism.
Brute-force attack	Verification of the engine tolerance when modulating / fuzzing the created pattern.
Replay attack	Performing replay attacks by recreating a previously captured pattern, including attempting to adapt the system tolerance to the patterns used (i.e., obtaining the effect of increased tolerance).

Deep-fake voice simulation	Attempts to create an AI voice avatar from voice recordings and using simulated voice in the authentication process.

The test scenarios mentioned above can be treated as a base for the voice biometrics system testing with the focus put on the biometric engine. It must be noted that there are other aspects not covered explicitly here, e.g., hardware security, network security or database security, but the scenarios targeting the infrastructure elements are partially tackling those areas. The weight was put on the biometrics aspects testing, as those are characteristic only for such solutions and other aspects can be treated as somehow supporting elements, whose security must be assessed and comply with the standards but are not the subject of this paper.

Biometric engine test scenarios implementation

Each of the scenarios targeting the biometric engine depending on the test environment can be completed using different approaches and tools. In the sections below, a workplan's steps and tooling are proposed along with appropriate commentary regarding the implementation of the scenarios in the test environment.

The first two scenarios assume that an attacker is a person with an internal access to the environment, i.e., indicating white-box or grey-box testing (depending on the scenario). The other five assume that an attacker has only access to the input interface and has no other knowledge regarding the system.

A. ***Algorithm level abuse***

Algorithm level abuse intends to find security gaps at the algorithmic level of the voice biometrics engine. This scenario covers analyzing used mechanisms against the current state of the voice recognition algorithmic designs and verifying the concept and the design of the algorithms, including allowed setup and use-cases.

There are no particular ready-to-use tools for this activity, as the review is performed based on the research and understanding of the current trends. The review checklists can be used to make the assessment more systematic, but the detailed steps will depend on the latest developments in the area, therefore rather cannot be standardized.

B. ***Source-code level errors***

Source-code level errors detection aims at the identification of the gaps in the implementation of the algorithms that were analysed as a part of the *algorithm level abuse* activity. It is important to note that even if an algorithm is proven to be secure, the improper implementation can introduce security vulnerabilities.

This activity focuses on the code review of the biometric algorithms, however, the code should be analysed as a whole, as the overall integration of the key algorithms with the system is also an important element where the security vulnerabilities may be introduced.

Another element of this activity is the biometrics engine configuration analysis. It must be understood that the implementation of the engine and its supporting system is a separate element to the configuration of the solution. In most cases, the configuration of the biometric authentication system lies in the competencies of the product user, i.e., the systems can be configured differently depending on the product user's needs. A product user is understood as a party who implements the solution into its infrastructure.

The configuration review must be performed bearing in mind that the system must not only be secure but also usable, therefore setting all the parameters to provide the smallest level of the verification errors may not be a feasible option. A proper trade-off between the usability and the security must be determined based on the characteristic of the reviewed system and

most likely this activity will almost always have to be done manually.

The source code review on the other hand provides a variety of tooling for static, dynamic and interactive analyses. The vast majority of them are automatic or semi-automatic tools that may miss the mark on the key aspect which is the implementation of the biometric algorithms. The algorithms implementation review should be done manually with support of the code review tools. Just as in the previous scenario, the review checklists can be used to make the assessment more systematic, but the business logic issues are impossible to be standardized in such form.

C. *Voice pattern registration process abuse*

Voice pattern registration process abuse targets the initiation of the biometric authentication mechanism for a particular end-user. This scenario covers:

- Registration of the voice patterns in different acoustic conditions, e.g.:
  - artificially generated or reduced background noise with varying amplitude and volume,
  - artificially generated or reduced echo,
- Registration of the voice patterns in various technical conditions:
  - pattern registration from recordings,
  - for the telephony systems: pattern registration using different parameters of telephone connection (e.g., 3G, LTE, VOIP).

There should be no direct relationship between various acoustic and technical conditions during the registration process and security of the authentication process identified, i.e., the identified authentication issues occur regardless of how the pattern was registered to pass this scenario. However, some functional errors may be identified that do not have a direct impact on system security.

To effectively execute this scenario, the tester would need access to audio devices, serving as an input (recorders) and output (speakers) and an audio editor. In case of the telephony systems testing, additionally a telephony devices and virtual audio devices allowing for coupling the audio inputs and outputs should be also used.

D. *Voice pattern reregistration*

Voice pattern reregistration attempts to find gaps in the processes responsible for changing the registered biometric patterns.

This scenario can cover attempts to invoke registration mechanism for a user with already registered biometric pattern and parallel voice pattern registration for the same end-user abusing the race condition on different devices.

Such behavior can be a result of the misconfiguration of the biometric engine or the introduction of vulnerabilities at the source code level, therefore when testing this scenario from a perspective of an external attacker some of the conditions in which the existing vulnerability would be exploitable may remain stealth.

In general, the system should not allow for reregistration or parallel voice pattern registration. In case the parallel voice registration is not blocked, then the authentication should be possible using only one pattern – the one that has been registered as the first pattern. In such case, it's also worth testing how technical errors, such as closing the connection before the pattern has been successfully registered, affect the authentication process.

E. *Brute-force attack*

Brute-force attack verifies the engine tolerance when modulating or fuzzing the created authentication pattern. The aim of the attack is to verify how the sound distortions can affect the authentication process. There should be no direct relationship between various acoustic distortions and security of the authentication process identified, i.e., the identified authentication issues occur

regardless of the distortions present to pass this scenario.

The scenario-related activities would include:

- Adding noise and sound distortion with different characteristics to the attempts to authenticate with natural voice and recordings.
- Attempts to authenticate by other people for the registered tester pattern.
- Attempts to authenticate by other people, genetically similar, e.g., twin, siblings, family members, for the registered tester pattern.

With regards to the genetic similarity recognition, it must be noted that the legal factor must be taken into account when deciding to perform such activity. It should be agreed with the test requestor whether this kind of scenario should be pushed during the testing, due to the liability reasons, in case of using voice samples that are originating from the tester's relatives which are not a party in the agreement between the test's requestor and the tester.

To effectively execute this scenario, the tester would need access to audio devices, serving as an input (recorders) and output (speakers) and an audio editor. In case of the telephony systems testing, additionally a telephony devices and virtual audio devices allowing for coupling the audio inputs and outputs should be also used.

F. ***Replay attack***

Replay attacks are performed by recreating previously captured voice pattern of a user, including attempting to adapt the system tolerance to the patterns used (i.e., obtaining the effect of increased tolerance).

To understand the importance of this scenario, it must be noted that available voice recognition tools classify their resistance against such attack in different manners. Some of them would consider the "replay" attack as reusing the identical perfect recording of a voice sample, in some cases this kind of behavior is referred as sample fingerprinting or playback resilience. Other tools would define the replay attacks as the recording detection. It is important to note the difference at the definition level, as this changes the understanding of the risk factors characteristic for the tested solution.

Another important factor to consider is the adaptivity of the authentication process. Some of the solutions provide the adaptive mechanism that dynamically analyze the end-user's voice and other ones provide the only static phrases verification and they can be differentiated as follows:

- When there is no adaptive mechanism – static constant phrases are the key;
- When the mechanism is designed as passive adaptive mechanisms, the dynamically changing phrases are used as keys;
- When the voice recognition happens as the background task, i.e., during regular activities unrelated with the authentication process – an active adaptive mechanism is used.

The scenarios for the replay attack are based on the prerecorded phrases. Depending on the adaptivity of the system, the testing scenarios vary. For the non-adaptive system, the risks related with reusing prerecorded phrase are the highest among the mentioned ones, as capturing a valid pattern sample is higher.

Before performing the scenarios, sample recordings must be crafted. The voice recognition systems usually divide the recording to the collaborative and non-collaborative ones. The first type is the kind of recording when an attacker records the samples with close proximity to the victim in perfect acoustic conditions. The second type is the kind of recording captured in bad acoustic conditions, without the close access to the victim.

To test the system's actual resilience in both cases, continuous and composite recordings can be used. Continuous

recordings are understood as recordings of the complete phrase at the appropriate tempo of spoken words (approx. 2.5 words per second) and intonation for declarative, interrogative or imperative sentences of the used language. Composite recordings are understood as phrases composed of different statements of the person being recorded, from recordings of different sound quality, which did not meet the standards of the tempo of speech and intonation for declarative, interrogative or imperative sentences of the used language. This kind of differentiation is needed to fully test the effectiveness of the voice authentication mechanisms in the tested solution.

As it was mentioned above, depending on the adaptivity of the system, the test scenarios vary. The table below shows a summary of the test cases for specific systems.

Test case	**System characteristic**
Attempts to authenticate with other phrases, similar in the phonetic context and using a different set of phonemes, to verify whether the biometric engine based on the user's voice characteristics will correctly recognize it	non-adaptive
Attempts to authenticate with prerecorded continuous phrases.	Non-adaptive, passive adaptive
Attempts to authenticate with prerecorded composite phrases.	Non-adaptive, passive adaptive
Attempts to use voice simulator	Active adaptive

To effectively execute this scenario, the tester would need access to audio devices, serving as an input (recorders) and output (speakers) and an audio editor. In case of the telephony systems testing, additionally a telephony devices and virtual audio devices allowing for coupling the audio inputs and outputs should be also used.

It is also recommended to modify the recordings so that the volume of the spoken phrase exceeded the volume of background noise, what can be achieved naturally in the case of recordings from close range (collaborative recordings), or by appropriate editing (non-collaborative recordings).

In some cases, recordings are detected based on the differences of the background noise before, during and after replaying the phase, therefore a special attention must be put into those aspects in the testing phase, as the proper timing of the replay may play a key role as well.

G. ***Deep-fake voice simulation***

Deep-fake voice simulation is based on creating an AI voice avatar which simulates the natural voice of the tester. It must be noted that dependent on the languages supported by the voice recognition system, appropriate generator must be used to properly reflect the targeted phonetics. Currently available non-commercial solutions support only selected languages, therefore the risk for those languages related with attempting this attack scenario is higher.

To effectively execute this scenario, the tester would need access to the voice simulator of the appropriate phonetics, audio devices, serving as an input (recorders) and output (speakers) and an audio editor. In case of the telephony systems testing, additionally a telephony devices and virtual audio devices allowing for coupling the audio inputs and outputs should be also used.

Biometric engine test scenarios implementation

The test environment setup is a crucial part of performing tests effectively and most likely will be common for the internal threat scenarios (i.e., scenarios A-B) and the external threat scenarios (i.e., scenarios C-G).

For the tests focused on the algorithmic and the source code level, the key element is the conceptual work, therefore, no particular configuration will be required. In detecting the source-code level error, automatic tools can be used to facilitate the identification of the threats that can affect the overall security of the system and the environment should be appropriately prepared to serve the needs of used tools. Except from the infrastructure part, it also must be noted that different source-code review tools require different formats of the input data, e.g., compilable source code, compiled version of the solution, or only packed sources.

For the tests targeting the external threat the setup is more complex. First of all, a set of audio recording and playing tools must be determined. Ideally, more than one device should be used to mitigate the risk of false negative cases for the device-sensitive scenarios, e.g., replay attacks. A useful tool especially for the telephony systems verification are virtual audio devices that allow to redirect the audio device's output into the audio device's input (e.g., VB-cable virtual audio device). This kind of behavior allows to obtain the same method of compression for the recordings used in the replay attacks, what proves to be a more effective solution than using external devices. When testing telephony systems, the tester must also have access to tools or devices that are able to connect to the system using different protocols (e.g., 3G, LTE, VOIP).

It is also important to select an appropriate audio editor that provides the mechanisms to introduce sound distortions, modulate the background noise levels, combine multiple recordings into one and modify their acoustic characteristics, so that they are nearly identical. Additionally, for the deep-fakes voice simulation, a voice avatar generator is required. One of the most popular ones currently available for the wide public is Lyrebird, however at the current stage of the development, it supports only English phonetics.

Apart for the tester's environment, it must be noted that the test requestor must set up his environment for the testing activities accordingly as well. The details should be addressed at the planning phase of the engagement, however, the key part is that the environment available for testing should reflect the production setup that is under the test.

Risk Evaluation and Mitigation Steps

As it was mentioned before, the risk evaluation for the biometric authentication mechanisms strongly depends on the procedural part. Even if a seemingly high-risk vulnerability is detected when testing the biometric engine itself, the actual risk for the organization must be evaluated taking into account the business risks related with this issue. The methodology that is usually the most useful for that kind of activities is OWASP RRM which introduces the adaptable tradeoff between the technical and the business impacts.

Proposing the mitigation steps for the identified gaps and vulnerabilities in case of the biometric engine testing can be affected by the fact that the test requestor may not be the tested product's owner. It must be established what is the relation between the test requestor, the system's configuration owner, the system's owner and the product owner. Without this, the proposed mitigation techniques cannot be feasible to introduce by the test's requestor.

Conclusions

This paper proposes an approach that can be followed during the security

assessments of the voice authentication systems. It should be treated as a guide during the process of determining the scope of the assessment, deploying test scenarios and estimating the risks of the identified issues. However, it must be stressed that due to the fast-evolving area which is biometric authentication, the proposed approach should be reviewed and reevaluated on a regular basis to ensure the quality of the security assessment.

References

- Crystal, D. (2008). A Dictionary of Linguistics and Phonetics. Malden, MA: Blackwell.
- Gajo, A., (2020), "Active Versus Passive Verification" [Online], [Retreived December 6, 2020], https://aurayasystems.com/2020/02/13/active-versus-passive-authentication/
- Lyrebird tool, [Retreived December 6, 2020], https://www.descript.com/lyrebird
- OWASP Application Security Verification Standard [Online], [Retreived December 6, 2020], https://owasp.org/www-project-application-security-verification-standard/
- OWASP Risk Rating Methodology [Online], [Retreived December 6, 2020], https://owasp.org/www-community/OWASP_Risk_Rating_Methodology
- VB Cable tool, [Retreived December 6, 2020], https://vb-audio.com/Cable/