УДК 004.056.523 DOI https://doi.org/10.30837/IYF.PDICIMT.2024.069

ОРГАНІЗАЦІЙНО-ТЕХНІЧНІ МЕТОДИ ПІДВИЩЕННЯ ЯКОСТІ ЗАХИСТУ В СИСТЕМАХ ГОЛОСОВОЇ АВТЕНТИФІКАЦІЇ

Квашенко В. Р., Пастушенко М.С.
Науковий керівник – к.т.н., проф. Пастушенко М. С., каф. ІКІ
Харківський національний університет радіоелектроніки
м. Харків, Україна
e-mail: vladyslav.kvashenko@nure.ua, Mykola.Pastushenko@nure.ua

With the rapid development of artificial intelligence technologies, industries must quickly adapt to new challenges. It is known that 62% of IT executives expressed concerns about the security threats posed by AI and deepfakes, as there have been cases where large banks' authentication systems were bypassed using synthesized voices. Voice authentication systems are a convenient and effective method of user identity verification. However, the development of modern voice synthesis technologies creates significant security threats. Criminals can generate synthetic voices that precisely imitate another user's voice, leading to unauthorized access. To address this issue, organizational and technical measures can be implemented. For example, the use of a changeable vocabulary, issuing a one-time password on the screen, and limiting the time for password entry are steps towards strengthening the security of voice authentication systems.

Зі стрімким розвитком технологій штучного інтелекту, все більше галузей мають так само швидко адаптуватись до нових викликів. Так, згідно з [1], 62% IT-керівників повідомили про занепокоєння щодо загроз безпеці, які несуть штучний інтелект та deepfake, через випадки, коли в великих банках обходили системи автентифікації використовуючи синтезований голос.

Системи голосової автентифікації є зручним та ефективним методом підтвердження особи користувача. Однак, розвиток сучасних технологій синтезу голосу створює значні загрози для безпеки. Зловмисники можуть генерувати синтетичні голоси, які точно імітують голос іншого користувача, що призводить до неавторизованого доступу. Для вирішення цієї проблеми, окрім технічних заходів, можна впровадити організаційно-технічні заходи.

Зазвичай процес голосової автентифікації складається з наступних кроків [2]:

1. Реєстрація голосу – користувач записує зразок голосу, який система використовує для створення унікального шаблону.
2. Виділення ознак – система виокремлює характерні риси зі зразка голосу, такі як висота, тон і швидкість мовлення.
3. Навчання моделі – виділені ознаки зберігаються в базі даних, для подальших порівнянь.

4. Автентифікація – на етапі автентифікації користувач надає новий зразок свого голосу. Система ідентифікує ознаки з цього зразка і порівнює їх зі збереженим шаблоном голосу. Якщо збіг перевищує певний поріг, користувач проходить автентифікацію.

Організаційно-технічні методи застосовуються під час етапу автентифікації. Розглянемо випадок, коли зловмисники змогли дістати достатньо вхідних даних для створення правдоподібного зразку голосу особи. Можна вирізнити декілька методів запобігання атакам синтезу голосу.

Використання змінного словника – замість запровадження сталого секретного слова, використовується динамічна система словника, в якій фраза для автентифікації завжди змінюється. Це ускладнює зловмисникам підготовку синтезованого зразка голосу заздалегідь.

Виведення одноразового пароля (OTP) на екран – поєднання секретного слова з відображенням випадково згенерованого OTP на екрані користувача на короткий час. Користувач, потім, повинен прочитати цей OTP для голосової автентифікації. Оскільки OTP є непередбачуваним та швидкозмінюваним, це унеможливлює сценарій авторизації з попередньо синтезованим секретним словом та зменшує ризик успішних атак синтезу голосу.

Якщо автентифікація відбувається в якомусь контрольованому приміщенні – є сенс встановити камери відеонагляду, для унеможливлення використання будь яких сторонніх інструментів для синтезу голосу, або, якщо автентифікація відбувається віддалено – встановити часове обмеження на введення паролю.

Обмеження часу на введення пароля – встановлення чіткого часового ліміту для введення OTP після того, як він з’явиться на екрані. Таке обмеження часу додає додатковий рівень безпеки, оскільки зменшує вікно можливостей для зловмисників використовувати синтезований зразок голосу.

Наприклад, модель штучного інтелекту VALL-E від Microsoft може клонувати голос з трисекундного аудіокліпу і генерувати новий голос відносно швидко [3]. Хоча точний час, необхідний для синтезу речення, явно не вказаний, здатність моделі клонувати голос з такого короткого кліпу дозволяє припустити, що процес синтезу може бути досить швидким, потенційно протягом декількох секунд. Тому, потрібно визначити баланс між складністю OTP паролю та часом на його проголошення.

Згідно з [4], швидкість читання для дорослих становить близько 183 слів на хвилину при читанні вголос, а фактичний час, необхідний для синтезу, також залежатиме від таких факторів, як довжина OTP-пароля, обчислювальна потужність системи та ефективність механізму Text-to-Speech, тому можна обрати довжину паролю в 2 слова та ліміт в 5 секунд для проголошення паролю.

В доповіді наводяться чисельні характеристики, які показують ефективність запропонованих організаційно-технічних заходів.

Системи голосової автентифікації, хоча й ефективні, стають уразливими перед атаками синтезу голосу. Запровадження таких організаційно-технічних заходів безпеки як, змінний словник, використання одноразових паролів та обмеження часу на введення пароля є кроками до зміцнення безпеки голосових систем автентифікації.

Подальші наукові дослідження будуть орієнтовані на розробку заходів підвищення захисту систем голосової автентифікації.

Список використаних джерел:

1. Daon Unveils xSentinel to Combat Voice Deepfakes as Part of AI.X Family. 2023. URL: https://www.daon.com/resource/daon-unveils-xsentinel-to-combat-voice-deepfakes-as-part-of-ai-x-family/.

2. Pastushenko M. KrasnozheniukY, Lemeshko O. Analysis of voice signal phase data informativity of authentication system // Proceedings of The Third International Workshop on Computer Modeling and Intelligent Systems (CMIS-2020), Zaporizhzhia, Ukraine, April 27–May 1, 2020. P. 1040–1053.

3. Microsoft's new VALL-E AI can clone your voice from a three-second audio clip. 2023. URL: https://techmonitor.ai/technology/ai-and-automation/vall-e-synthetic-voice-ai-microsoft.

4. How many words do we read per minute? A review and meta-analysis of reading rate. 2019. URL: https://www.researchgate.net/publication/335174808_How_many_words_do_we_read_per_minute_A_review_and_meta-analysis_of_reading_rate.