DOI: 10.18137/RNU.V9187.25.04.P.119 УДК 004.056

Сокирченко Данил Артемович
студент, Магнитогорский государственный технический университет имени Г.И. Носова, город Магнитогорск.
Электронный адрес: danilsokirchenko@gmail.com
Danil A. Sokirchenko
Student, Nosov Magnitogorsk State Technical University, Magnitogorsk.
E-mail address: danilsokirchenko@gmail.com

Лазарева Кристина Евгеньевна
студент, Магнитогорский государственный технический университет имени Г.И. Носова, город Магнитогорск.
Электронный адрес: pandetccc@gmail.com
Kristina E. Lazareva
Student, Nosov Magnitogorsk State Technical University, Magnitogorsk.
E-mail address: pandetccc@gmail.com

Кузьмина Ульяна Владимировна
кандидат технических наук, доцент, доцент кафедры информатики и информационной безопасности, Магнитогорский государственный технический университет имени Г.И. Носова, город Магнитогорск.
SPIN-код: 2335-5916, AuthorID: 659266
Электронный адрес: ylianapost@gmail.com
Ulyana V. Kuzmina
Ph.D. of Technical Sciences, Docent, Associate Professor at the Department of informatics and information security, Nosov Magnitogorsk State Technical University, Magnitogorsk. SPIN-code: 2335-5916, AuthorID: 659266
E-mail address: ylianapost@gmail.com

ЗАЩИТА ГОЛОСОВЫХ АССИСТЕНТОВ ОТ УГРОЗ ПО АКУСТИЧЕСКОМУ КАНАЛУ

Аннотация. Исследованы методы защиты голосовых помощников от несанкционированных управляющих команд, передаваемых злоумышленниками. Определены ограничения традиционных методов фильтрации и верификации, установлены уязвимости и ограничения голосовых ассистентов. Оценена эффективность средств и методов защиты, в том числе потенциальных для типовых сценариев. В ходе исследования предложен метод, использующий комбинацию программных и аппаратных мер защиты голосового ассистента, обеспечивающий защиту выполнения критически важных команд.

Ключевые слова: защита информации, акустическая безопасность, голосовые ассистенты, биометрическая верификация, пространственная локализация, машинное обучение.

Для цитирования: *Сокирченко Д.А., Лазарева К.Е., Кузьмина У.В.* Защита голосовых ассистентов от угроз по акустическому каналу // Вестник Российского нового университета. Серия: Сложные системы: модели, анализ, управление. 2025. № 4. С. 119 – 128. DOI: 10.18137/RNU.V9187.25.04.P.119

Вестник Российского нового университета
Серия «Сложные системы: модели, анализ и управление». 2025. № 4

PROTECTING VOICE ASSISTANTS FROM THREATS VIA THE ACOUSTIC CHANNEL

Abstract. The study examines methods for protecting voice assistants from unauthorized control commands issued by attackers. The limitations of traditional filtering and verification techniques are identified, along with vulnerabilities and constraints of voice assistants. The effectiveness of existing and proposed protective measures is evaluated, including their applicability in typical scenarios. As part of the research, a method was proposed that combines software- and hardware-based security mechanisms for voice assistants, ensuring the protection of critical command execution.

Keywords: information protection, acoustic security, voice assistants, biometric verification, spatial localization, machine learning.

For citation: Sokirchenko D.A., Lazareva K.E., Kuzmina U.V. (2025) Protecting voice assistants from threats via the acoustic channel. *Vestnik of Russian New University. Series: Complex Systems: Models, analysis, management*. No. 4. Pp. 119– 128. DOI: 10.18137/RNU.V9187.25.04.P.119 (In Russian).

Введение

С ростом популярности голосовых ассистентов увеличивается число атак на них. Так, установлено, что современные голосовые помощники все еще подвержены риску восприятия несанкционированных голосовых команд, поступающих не от легитимного владельца устройства и (или) аккаунта. Реализуются такие команды как по скрытым каналам передачи (акустические сигналы в ультразвуковом диапазоне, замаскированные слова), так и с помощью технологий имитации речи, а также воспроизведения заранее записанных команд. Это создает потенциальную угрозу пользовательской информации и самому пользователю из-за выполнения вредоносных действий.

Цель исследования заключается в разработке теоретически обоснованного подхода к повышению устойчивости голосовых ассистентов к акустическим атакам. Рассматривается реализация механизмов безопасности при передаче команд за счёт применения методов пространственной и биометрической верификации источника. Объектом исследования выступают системы голосового управления, такие как «Яндекс Алиса», Siri, Alexa, Cortana и другие, а предметом – методы и технические средства биометрической и пространственно-акустической верификации, направленные на подавление вредоносных команд со стороны злоумышленника либо предотвращение некорректной обработки запроса.

Обзор источников по теме исследования

Изучение научных статей и практических исследований на тему угроз голосовых ассистентов позволяет выявить основные слабые места и методы противодействия им. Рассмотрим актуальные угрозы и уязвимости на сегодняшний день.

Одни из наиболее опасных атак на голосовые ассистенты – атаки скрытыми командами. Эти команды не воспринимаются человеческим слухом, но распознаются микрофонами систем анализа речи (далее – ASR). Таким методом является Dolphin Attack, впервые описанный сотрудниками Чжэцзянского университета в 2017 году 1]. При этом используется амплитудная модуляция голосовых команд на ультразвуковой несущей (часто-

та > 20 Гц). Аппаратные нелинейности микрофонных цепей демодулируют сигнал, превращая его в слышимую для ASR речь, что позволяет выполнять команды на устройствах Siri, Alexa, Cortana, Echo и др.

Следующей угрозой после Dolphin Attack стала NUIT (Near Ultrasound Inaudible Trojan), продемонстрированная учёными из Университета Техаса и Колорадо Спрингс и представленная на USENIX Security 2023 [2]. В отличие от предка, NUIT может выполняться удалённо через динамик того же (NUIT-1) или второго (NUIT-2) устройства. Аудиофайл с near ultrasound-сигналом (частоты ≈16…20 кГц) включается, например, во время воспроизведения контента через приложение или видео и инициирует команды без всякого подозрения пользователя.

Методы отличаются архитектурой: Dolphin Attack требует аппаратной генерации ультразвука и физической близости, тогда как NUIT использует возможности динамика устройства и высокую чувствительность микрофонов в ближнем ультразвуковом диапазоне. При NUIT-1 злоумышленнику достаточно загрузить скрытую команду, и она выполнится на том же устройстве без звука. При NUIT-2 динамик одного гаджета активирует микрофон другого на расстоянии, на котором команды остаются неслышными.

Исследование рассматривает угрозы, возникающие при использовании модифицированных аудиозаписей для атак на ASR. Такие атаки искажают аудиосигнал для передачи команд, незаметных для человека, но распознаваемых системой, используя маскирующий шум или изменение спектральных характеристик сигнала [3]. Осуществляются атаки дистанционно, например, через потоковый контент. Атаки воздействуют на ранние этапы обработки аудио – фильтрацию шума и извлечение признаков. Для этого не требуется специальных микрофонов, поэтому атака остается применима к широкому спектру систем.

Слабые стороны атаки: стандартные голосовые помощники не распознают команды на расстоянии более 3,5 м или при соотношении сигнал/шум менее 5 дБ из-за фильтров шумоподавления. Злоумышленники могут использовать динамики или другие устройства на расстоянии до нескольких метров.

Такие системы, как Google Speech, успешно распознают обычные команды (≈85 %) и менее эффективно – скрытые (≈60 %). Улучшение отношения сигнал/шум повышает точность до более чем 80 %.

Эксперименты показали эффективность скрытых команд. Liuetal (2019) достигли 80 % распознавания модифицированных аудиофайлов, а Qinetal (ICML 2019) – 100 % транскрипции с психоакустическим маскированием.

Fuzzy Wake Words – ложная активация голосовых ассистентов из-за восприятия слов, фонетически схожих с ключевыми (например, Alexa или OK Google). Это системная особенность архитектуры детектора, некорректно интерпретирующая звуковые сигналы, что приводит к записи и обработке случайных команд. Проблема распространена: ассистенты реагируют на искажённые или случайные слова, например, «холис» вместо Alexa или фразы, вызывающие ассистента Google.

Исследование FakeWake [4] под руководством профессора Ахмада-Резы Садеги показало возможность автоматизированного создания «фаззи»-слов для активации голосовых помощников без аудиоданных и внутренних моделей. Используя эволюционный алгоритм, учёные сгенерировали 965 уникальных слов, которые активировали 8 платформ, включая Amazon Echo и Google Nest. Более 40 % этих слов имели коэффициент ложного

Вестник Российского нового университета
Серия «Сложные системы: модели, анализ и управление». 2025. № 4

срабатывания выше 0,8, включая модификации слов типа Alexa (например, Alexi, Alive, Olexa).

Уязвимость связана с упрощёнными нейросетевыми моделями-детекторами в устройствах, которые ориентируются на ключевые фонемы. Даже существенные отклонения в произношении и шумовой фон не мешают активации. Это создаёт риск утечки данных или выполнения нежелательных команд.

Существующие меры защиты

Для разобранных ранее угроз разработаны соответствующие меры защиты. В настоящем исследовании проанализирована каждая из мер защиты и проведен сравнительный анализ эффективности против упомянутых угроз.

Голосовая биометрия представляет собой процедуру атрибуции личности, использующую анализ фонетических образов, что демонстрирует высокую чувствительность к физиологическим и артикуляционным особенностям человеческой речи [5]. В ходе проверки сигнал подвергается спектрально-статистическому анализу, после образуется вектор, который сверяется с хранящимися на сервере эталонами.

Традиционные системы используют вероятностные модели GMM-UBM (Gaussian Mixture Model – Universal Background Model), формируя фоновый профиль речи и адаптируя его под пользователя. Для устранения влияния акустического окружения применяется JFA (Joint Factor Analysis), разделяющий вариативные компоненты фоновых шумов. Следом были внедрены гибридные архитектуры, такие как GMM+SVM, где вероятностные параметры подаются на вход SVM-классификаторам [6]. В настоящее время используются глубинные нейросети x-vector и d-vector, обученные на больших фонетических корпусах [7].

Механизмы защиты основаны на сочетании физиологических и поведенческих маркеров: форма голосового тракта, резонанс, интонация и темп создают сложный трудновоспроизводимый профиль. Алгоритмы распознают аномалии в записанных или синтезированных голосах. Динамические протоколы «вызов – ответ» и шифрование голосовых эталонов повышают устойчивость.

Уязвимости – системы подвержены спуфингу, синтетическим голосам и deepfake. Для защиты используются:

- LCTM-анализ для выявления аномалий;
- динамический запрос;
- контекстный анализ;
- многоканальная биометрия.

В голосовых помощниках используется единый подход к ведению журналов. Звуковой поток преобразуется в текст с помощью модуля обработки естественного языка. Затем данные запроса – время, идентификатор пользователя, идентификатор устройства и намерение – шифруются и сохраняются в хранилище. Шифрование реализовано средствами AWSKMS, логи Cloud Watch и S3-объекты защищены ключами KMS, а доступ к ключам строго регулируется через IAM-политики. Временные лимиты хранятся согласно требованиям законодательства или сервисным целям, пользователь может самостоятельно запускать удаление истории, тем самым снижая потенциальную уязвимость данных.

Тем не менее реальные случаи показали наличие рисков. В июле 2019 года Google Home непреднамеренно фиксировал отрывки разговоров и передавал их подрядчикам:

около тысячи фрагментов попало к внешним транскрибаторам[1]. В июне 2020 года Check Point выявил уязвимость CORS/XSS в Alexa, позволяющую по одному клику получить доступ к личной истории и устанавливать навыки от имени пользователя[2].

Физическая защита от скрытой активации микрофона заключается в аппаратном отключении питания или сигнала микрофонного устройства, гарантированно лишающего систему возможности записывать звук. Поскольку программный контроль не влияет на разомкнутую цепь, такое отключение сохраняет абсолютную надежность даже при наличии вредоносного ПО.

Программное закрепление блокировки микрофона (через ОС или драйверы) не даёт аналогичной гарантии: злоумышленники с root-доступом способны обойти ограничения, визуальные индикаторы могут быть подменены, а микрофон активирован программно даже при отключении прав доступа в интерфейсе.

Аппаратные переключатели устанавливают чёткую грань между безопасным и уязвимым состоянием устройства. В Amazon Echo или Purism Librem, как и в MacBook с T2-чипом, включение активного LED-индикатора сигнализирует об отключении микрофона аппаратно, что резко снижает риск компрометации [8]. Дополнительно такие решения сочетаются с сокращением привилегированного кода и использованием доверенных модулей (например, Lockdown Mode в QubesOS и Purism), что исключает программную подмену функций системы.

В результате аппаратное отключение микрофона работает как последний рубеж защиты: если питание отрезано, никакое ПО или микропрошивка не сможет его активировать, что обеспечивает физическую безопасность от скрытых аудиоатак.

В Таблице 1 представлен сравнительный анализ рассмотренных мер защиты голосовых ассистентов от угроз по акустическому каналу.

Таблица 1

Существующие меры защиты

Метод	Покрываемые угрозы	Преимущества	Недостатки
Голосовая биометрия (фильтрация команд по голосовому отпечатку)	– Несанкционированный доступ к функциям помощника; – атаки социальной инженерии; – подмена пользователя (имперсонация)	– Повышенная точность аутентификации; – пассивная и незаметная идентификация; – устойчивость к поддделкам и социальной инженерии; – интеграция в системы IVR, мобильные и голосовые интерфейсы; – выявление мошенников по голосовому «черному списку»	– Уязвимость к атакам с использованием сэмплов голоса (voice replay/spoofing); – неэффективность в шумной среде; – требуется специализированная инфраструктура; – потенциальные проблемы с приватностью (биометрические данные)

[1] *Verheyden T., Baert D., Van Hee L., Van Den Heuvel R.* Google employees are eavesdropping, even in your living room, VRT NWS has discovered // VRT NWS. 2019. 10 July. URL: https://www.vrt.be/vrtnws/en/2019/07/10/google-employees-are-eavesdropping-even-in-flemish-living-rooms/ (дата обращения: 08.10.2025).

[2] *O'Donnell L.* Amazon Alexa 'One-Click' Attack Can Divulge Personal Data // 2020. URL: https://threatpost.com/amazon-alexa-one-click-attack-can-divulge-personal-data/158297 (дата обращения: 08.10.2025).

Вестник Российского нового университета
Серия «Сложные системы: модели, анализ и управление». 2025. № 4

Окончание таблицы 1

Метод	Покрываемые угрозы	Преимущества	Недостатки
Лог-фильтры и проверка истории действий	– Несанкционированное выполнение команд; – нарушение прозрачности и контроля; – хранение чувствительной информации без ведома пользователя	– Повышение доверия и прозрачности; – контроль за сохранёнными данными; – возможность удаления истории; – настройка автоматической очистки; – широкая поддержка производителями	– Не предотвращает прослушивание/выпол-нение в реальном времени; – требует ручного взаимодействия; – может быть отключено пользователем/злоумышленником; – не идентифицирует личность исполнителя команд
Методы «чистой зоны» (физическая блокировка микрофона)	– Пассивное прослушивание; – утечка аудиоданных; – непреднамеренная активация устройства; – сбор личной информации и поведенческих данных	– Полное исключение прослушивания; – простая реализация (механическая кнопка); – независимость от ПО; – не требует дополнительных настроек	– Потеря функциональности голосового помощника; – требуется физическое взаимодействие; – пользователи могут забыть включить обратно; – не защищает при включенном микрофоне

Источник: здесь и далее таблицы составлены авторами

Материалы и методы

В целях повышения защищенности голосовых ассистентов от несанкционированного управления со стороны посторонних лиц или с использованием воспроизводимых записей в настоящей статье предлагается метод, предусматривающий выборочную дополнительную верификацию пользователя при выполнении команд с высокой значимостью. Подход основан на комплексной оценке подлинности пользователя с использованием биометрических и пространственных признаков. Рассмотрим ключевые компоненты метода (см. Рисунок 1).

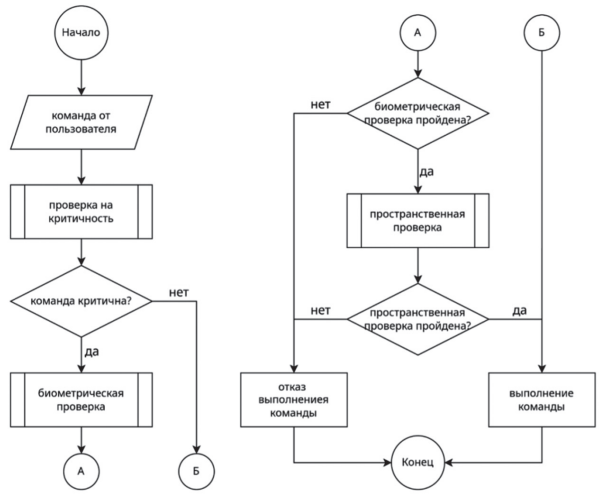

Рисунок 1. Архитектура метода
Источник: рисунок выполнен авторами.

Архитектура метода предполагает предварительную классификацию команды с целью определить, является ли она критичной, и следует ли, основываясь на этой классификации, проводить дополнительную проверку. Под критичными командами понимаются те команды, выполнение которых потенциально может служить получению конфиденциальной информации сторонним лицом или нарушить комфортные и безопасные условия жизни пользователя за счет систем управления умным домом. В случае если команда не является критичной, то она сразу приходит в исполнение, в противном случае голосовой запрос сначала проходит усиленную верификацию пользователя.

Усиленная верификация пользователя состоит из двух поэтапно выполняемых компонентов: биометрического и пространственного. Если биометрический компонент подтвердил, что команда была дана голосом владельца ассистента, то проверка переходит к пространственному компоненту. Пространственный компонент анализирует пространственные характеристики источника звука, определяя, что голос пользователя является настоящим, а не его записью, не искусственно создан при помощи синтеза голоса или нейросетевой моделью. Верификация считается успешной, если оба компонента независимо подтверждают пользователя легитимным владельцем. В случае успешной верификации выполнение команды разрешается.

Ключевым этапом предлагаемого метода является определение степени критичности голосовой команды. Классификация команд осуществляется с целью минимизации избыточных проверок, повышая скорость работы и комфорт пользователя от использования ассистента и понижая требуемую вычислительную нагрузку.

Возможны несколько подходов к решению задачи классификации команд – от простых списков до машинного обучения. Содержание статических списков основывается на оценке разработчика и/или пользователя. Предпочтительный выбор подхода зависит от набора возможностей и архитектуры построения модулей голосового ассистента. Списки могут вестись одним из следующих методов: разрешить выполнение без проверки всех команд, кроме указанных, или запретить выполнение без проверки всех команд, кроме указанных. Такой подход требует ручной настройки и актуализации по мере добавления нового функционала.

Применение машинного обучения может избавить разработчиков от актуализации списков и защитить от непредвиденных сценариев, однако потребует больших усилий для обучения и увеличивает вычислительные затраты. Такой подход способен самостоятельно определить критичность команды на основе оценки семантического содержания команды, контекст предыдущих запросов, а также частоту использования команд с целью выявления аномалий в поведении.

Основная задача биометрического компонента заключается в определении соответствия между голосом произносящего команду и сохранёнными голосовыми слепками пользователей голосового ассистента. При первом использования голосового ассистента для работы компонента требуется предварительное формирование эталонной модели голоса каждого пользователя в семье путем анализа контрольных фраз. При поступлении новой команды на проведение верификации система извлекает признаки из команды и сравнивает их с эталонной моделью, вычисляя степень сходства.

Проверки в биометрическом компоненте могут осуществляться как полностью на стороне клиента, так и в гибридной модели с привлечением облачной инфраструктуры.

Вестник Российского нового университета
Серия «Сложные системы: модели, анализ и управление». 2025. № 4

Выполнение проверок только на клиентской стороне исключает требования постоянного сетевого соединения и обеспечивает лучшую реакцию ассистента. Однако с целью снижения требований к аппаратным характеристикам устройства могут использоваться вычислительные мощности облачного сервиса разработчика. В этом случае разработчику потребуется позаботиться о создании безопасного метода передачи информации по каналу связи и хранении голосовых моделей пользователей на своей стороне.

Задача пространственного компонента – убедиться, что подтвержденный голос пользователя, отдавшего команду биометрическим компонентом, является естественной речью. Техническая реализация пространственного анализа основывается на применении микрофонного массива, состоящего из нескольких микрофонов, от количества которых будет зависеть точность определения. При помощи алгоритмов пространственной локализации, таких как GCC-PHAT (Generalized Cross-Correlation with Phase Transform), сравнивается полученное направление источника звука с допустимыми зонами. Так как только оценка локализации звука не гарантирует достоверную точность, предполагается усиление метода использованием многопараметрического анализа, включающего оценку спектральных характеристик сигнала и контроль амплитудно-частотных искажений. Дополнительным критерием служит анализ динамических характеристик речевого сигнала, позволяющий выявлять артефакты цифровой обработки.

Для оценки перспективности использования пространственного анализа программно была создана модель системы локализации звука на основе двух микрофонов, расположенных на расстоянии 5 см, и использовании алгоритма GCC-PHAT для локализации звука. Модель была написана на языке программирования Python с использованием библиотек NumPy, Scipy, Pandas, Matplotlib.

В процессе выполнения программы моделировались сигналы от источников, расположенных под различными углами относительно устройства. Легитимная зона определялась как сектор 45 градусов от фронтальной стороны моделируемого устройства. Если звук источник звука попадал под эту зону, то источник считался легитимным. Для каждого тестового сценария взаимодействия вычислялись временные задержки между микрофонами под разными углами (реальные углы – под которыми располагался источник звука, оценочные углы – рассчитанные алгоритмом GCC-PHAT). В результате эксперимент показал высокую точность определения доверия для фронтального расположения источника звука; значительные погрешности при оценке углов источника сбоку и сзади (см. Таблицу 2).

Таблица 2

Результаты эксперимента

Сценарий	Реальный угол	Оценочный угол	Доверие, %	Легитимность
Легитимный пользователь (фронтально)	0	0	100	Легитимный
Легитимный пользователь (сбоку)	30	–25,4	100	Легитимный
Атака сбоку	90	–59	95	Атака
Атака сзади	180	0	100	Легитимный

Способность моделируемой системы обнаруживать потенциальные атаки составила 83,3 % при 100-процентном пропуске легитимных команд. Это показывает обусловлен-

ность добавления пространственного компонента верификации с дальнейшим его усовершенствованием.

Использование методов проверки, опирающихся на биометрические и пространственные данные, требует применения определённых вычислительных и технических средств. Идентификация по голосу, осуществимая только путём сравнения голосового шаблона пользователя, может быть реализована на устройстве. Для более детального анализа при условиях ограниченных ресурсов может потребоваться применение облачной инфраструктуры, что связано с рисками утечки биометрических данных и задержками при выполнении команд. Требование пространственного компонента к количеству дополнительных микрофонов увеличивает стоимость устройства и усложняет его конструкцию и заводскую настройку. Кроме того, каждый компонент может быть чувствительным не только к изменению окружающей среды, в которой находится устройство, но и к изменению голоса владельца вследствие его состояния здоровья или эмоционального состояния.

Заключение

В статье предложен метод, защищающий голосовые ассистенты от несанкционированного выполнения критичных команд на основе многоуровневой верификации пользователя, произносящего команду. Применение этого метода повышает устойчивость системы к современным типам атак, включая Dolphin Attack, NUIT, Hidden Voice и Fuzzy Wake Words.

Дальнейшие исследования будут направлены на разработку алгоритмов классификации команд на основе машинного обучения, определение оптимального решения реализации биометрического и пространственного компонента в связке друг с другом и поиск путей для внедрения метода в существующие голосовые ассистенты.

Литература

1. *Zhang G., Yan C., Ji X., Zhang T., Zhang T., Xu W.* DolphinAtack: Inaudible Voice Commands // Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS '17). Association for Computing Machinery, New York, NY, 2017. P. 103–117. DOI: https://doi.org/10.1145/3133956.3134052

2. *Xia Q., Chen Q., Xu S.* Near-ultrasound inaudible trojan (NUIT): Exploiting your speaker to attack your microphone // 32nd USENIX Security Symposium (USENIX Security 23) 2023. URL: https://www.usenix.org/system/files/sec23fall-prepub-261-xia-qi.pdf (дата обращения: 08.10.2025).

3. *Wu X., Rajan A.* Catch Me If You Can: Blackbox Adversarial Attacks on Automatic Speech Recognition using Frequency Masking // 2022 29th Asia-Pacific Software Engineering Conference (APSEC). Institute of Electrical and Electronics Engineers, 2023. P. 169–178. DOI: https://doi.org/10.48550/arXiv.2112.01821

4. *Chen Y., Bai Y., Mitev R., Wang K., Sadeghi A.-R., Xu W.* FakeWake: Understanding and Mitigating Fake Wake-up Words of Voice Assistants // arXiv. 2021. DOI: https://doi.org/10.48550/arXiv.2109.09958

5. *Михайлова У.В., Лукьянов Г.И., Дончан Д.М.* Анализ биометрической аутентификации на устойчивость при воздействии внешних факторов // Актуальные проблемы современной науки, техники и образования : Тезисы докладов 76-й международной научно-технической конференции, Магнитогорск, 16–20 апреля 2018 г. Т. 1. Магнитогорск : Магнитогорский государственный технический университет им. Г.И. Носова, 2018. С. 295. EDN XMEXAL.

Вестник Российского нового университета
Серия «Сложные системы: модели, анализ и управление». 2025. № 4

6. Kenny P., Boulianne G., Ouellet P., Dumouchel P. Joint Factor Analysis Versus Eigenchannels in Speaker Recognition // *IEEE Transactions on Audio, Speech, and Language Processing*. 2007. Vol. 15. No. 4. Pp. 1435–1447. DOI: https://doi.org/10.1109/TASL.2006.881693
7. Li L., Tang Z., Shi Y., Wang D. Gaussian-Constrained training for speaker verification // arXiv. 2019. DOI: https://doi.org/10.48550/arXiv.1811.03258
8. *Pathak S., Sheikh Ariful Islam, Jiang H., Xu L., Tomai E.* A survey on security analysis of Amazon echo devices // High-Confidence Computing. 2022. Vol. 2. No. 4. Article no 100087. DOI: https://doi.org/10.1016/j.hcc.2022.100087

References

1. Zhang G., Yan C., Ji X., Zhang T., Zhang T., Xu W. (2017) DolphinAtack: Inaudible Voice Commands // Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS '17). Association for Computing Machinery, New York, NY. Pp. 103–117. DOI: https://doi.org/10.1145/3133956.3134052
2. Xia Q., Chen Q., Xu S. (2023) Near-ultrasound inaudible trojan (NUIT): Exploiting your speaker to attack your microphone. In: *32nd USENIX Security Symposium (USENIX Security 23)*. URL: https://www.usenix.org/system/files/sec23fall-prepub-261-xia-qi.pdf (accessed 08.10.2025).
3. Wu X., Rajan A. (2023) Catch Me If You Can: Blackbox Adversarial Attacks on Automatic Speech Recognition using Frequency Masking. In: *2022 29th Asia-Pacific Software Engineering Conference (APSEC)*. Institute of Electrical and Electronics Engineers, 2023. Pp. 169–178. DOI: https://doi.org/10.1109/APSEC57359.2022.00029
4. Chen Y., Bai Y., Mitev R., Wang K., Sadeghi A.-R., Xu W. (2021) FakeWake: Understanding and Mitigating Fake Wake-up Words of Voice Assistants. *arXiv.* DOI: https://doi.org/10.48550/arXiv.2109.09958
5. Mikhailova U.V., Lukyanov G.I., Donchan D.M. (2018) Analysis of Biometric Authentication Robustness under the Influence of External Factors. In: Golubeva A.O., Osintsev N.A., Chernyshova E.P., et al. (Eds) *Aktual'nye problemy sovremennoi nauki, tekhniki i obrazovaniya* [Current problems of modern science, technology and education] : Conference papers for the 76th International scientific and technical conference, Magnitogorsk, 16–20 April 2018. Vol. 1. Magnitogorsk : Nosov Magnitogorsk State Technical University Publ. Pp. 295. (In Russian).
6. Kenny P., Boulianne G., Ouellet P., Dumouchel P. (2007) Joint Factor Analysis Versus Eigenchannels in Speaker Recognition. In: *IEEE Transactions on Audio, Speech, and Language Processing*. Vol. 15. No. 4. Pp. 1435–1447. DOI: https://doi.org/10.1109/TASL.2006.881693
7. Li L., Tang Z., Shi Y., Wang D. (2019) Gaussian-Constrained training for speaker verification. *arXiv.* DOI: https://doi.org/10.48550/arXiv.1811.03258
8. Pathak S., Sheikh Ariful Islam, Jiang H., Xu L., Tomai E. (2022) A survey on security analysis of Amazon echo devices. *High-Confidence Computing*. Vol. 2. No. 4. Article no 100087. DOI: https://doi.org/10.1016/j.hcc.2022.100087

Поступила в редакцию: 19.09.2025	Received: 19.09.2025
Поступила после рецензирования: 29.10.2025	Revised: 29.10.2025
Принята к публикации: 10.11.2025	Accepted: 10.11.2025