Архитектура системы защиты от атак с использованием голосового фишинга

М. Г. Синкевич

Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В. И. Ульянова (Ленина)

mariasinkevich888@gmail.com

Аннотация. **Исследование содержит анализ методов атак с использованием голосового фишинга, существующих методов защиты, а также архитектуру и описание работы предложенной системы защиты, позволяющей анализировать звонок в реальном времени и классифицировать его как фишинговый или легитимный. Особое внимание уделено выбору технологий, которые будут использованы в дальнейшем при реализации системы.**

Ключевые слова: информационная безопасность; голосовой фишинг; вишинг; телефонное мошенничество; атаки на телефонные звонки; атаки на звонки в мессенджерах

I. Введение

Цифровые технологии интегрированы в каждую сферу нашей жизни, но особую роль играют в сфере коммуникаций: сегодня огромное количество звонков как в корпоративной сфере, так и в личной, происходят через цифровые каналы. Всесторонняя цифровизация с одной стороны помогает ускорить выполнение каждодневных задач и упростить общение между людьми из разных частей мира, но с другой стороны открывает очень много возможностей для злоумышленников, ведь новый канал для общения – новый канал для мошенничества.

Фишинг – метод мошенничества, при котором злоумышленники похищают конфиденциальные данные пользователей: логины и пароли от учётных записей, данные паспортов и банковских карт или иную информацию, открывающую доступ к финансам или другим важным ресурсам. Согласно последним статистическим данным, с каждым годом фишинговые атаки становятся всё более сложными и изощрёнными и, несмотря на существование множества средств защиты, от таких атак каждый год страдают тысячи людей. Важно заметить, что успешно проведённые фишинговые атаки несут за собой не только крупные финансовые потери частных лиц, но и репутационные риски для компаний, обслуживающих их.

В этой работе проведён анализ текущей ситуации в России в сфере противодействия голосовому фишингу путём исследования статистики, используемых методов атак и методов защиты. На основе проведённого анализа предложена архитектура системы, анализирующей звонок в реальном времени и классифицирующей его как фишинговый или легитимный.

II. Методы атак с использованием голосового фишинга

В голосовом фишинге доверие жертвы – один из важнейших факторов, определяющих успех атаки, его необходимо завоевать задолго до начала разговора. На начальном этапе атаки мошенникам необходимо, чтобы человек посчитал звонок с неизвестного номера легитимным и ответил на него. Чтобы номер не выглядел подозрительно, злоумышленники используют подмену номера телефона – вместо личного номера, с которого звонит фишер, жертва увидит, например, городской – к таким номерам доверие выше.

После того, как жертва взяла трубку, в атаку вступает социальная инженерия, именно она является основным звеном атаки и определяет её исход. Здесь мошеннику важно хорошо понимать психологию человека и правильно воздействовать на эмоции жертвы. Стоит отметить, что социальная инженерия и человеческий фактор – одни из самых уязвимых точек ИБ, поскольку на этом этапе не так просто внедрить технические средства защиты.

Чтобы атака прошла успешно, злоумышленники строят свои скрипты так, чтобы они вызывали определённые эмоции: страх, любопытство, жадность, жалость в сочетании со срочностью или авторитетом вводят жертву в состояние, в котором у неё отключается объективное восприятие реальности [1]. Во время звонка у жертвы почти нет времени обдумать и критически оценить, что ей говорят по телефону, а вызываемые эмоции и обстоятельства только усугубляют ситуацию.

Основными тенденциями в атаках с помощью голосового фишинга является использование ИИ как для генерации скриптов, так и для создания дипфейков, а также переход в мессенджеры, поскольку там более низкий уровень защиты.

III. Организационные и программно-технические методы защиты

Первым и самым главным методом защиты является работа с людьми и повышение их осведомлённости в сфере телефонного мошенничества, так как в основе фишинга лежит социальная инженерия, а на эффективность атак сильно влияет человеческий фактор. Организационные меры защиты включают:

- повышение осведомлённости граждан и распространение информации о типичных схемах телефонного мошенничества и мерах противодействия;
- разработка документа с алгоритмом действий на случай, если человек стал жертвой мошенников, и обеспечение быстрого доступа к нему для оперативного реагирования и минимизации ущерба;
- информирование граждан о новых схемах и сценариях мошенничества с помощью СМИ и социальных сетей.

К сожалению, для защиты от массовых атак на частных лиц не получится обойтись только организационными мерами, поэтому программно-технические меры являются неотъемлемой частью систем защиты. Самым распространённым и эффективным методом защиты от телефонного мошенничества на данный момент является блокировка вызовов с подменных номеров. Этот метод позволяет пресекать атаку на самых ранних этапах и блокировать такие звонки до того, как у абонента зазвонит телефон.

1) Антифрод-системы. Они анализируют звонок до того, как он дойдёт до абонента, и на основе собранных данных принимают решение – пропустить звонок, пометить его как подозрительный или заблокировать. В зависимости от структуры и масштаба такие системы анализируют разные данные и характеристики: какие-то ограничиваются информацией о трафике и активности абонентов, какие-то объединяются с банковскими антифрод-системами и формируют единую систему для обмена данными о номерах и транзакциях. Благодаря такому взаимодействию появляется возможность предупреждать пользователей о мошеннических транзакциях и своевременно блокировать их.

2) Антиспам-системы и выявление массовых обзвонов. Несмотря на то, что мошеннические и спам-звонки – разные понятия, они неразрывно связаны, так как часто мошенники также осуществляют массовые обзвоны с одного и того же номера, чтобы в итоге найти жертву. Такие системы анализируют номер звонящего и проверяют его по базам данных номеров, в которых содержится полная информация о них.

3) Автоответчик/секретарь. Технология, получившая распространение в последние годы – виртуальный ассистент сам принимает спам-звонки и разговаривает с операторами колл-центров, а потом присылает расшифровку абоненту.

4) Защита во время разговора. Эта технология является одной из самых прорывных и продвинутых за последнее время, она работает на основе нейросетей и алгоритмов обработки естественного языка и осуществляет анализ звонка в реальном времени. Система получает аудиосигнал речи звонящего, преобразовывает его из голоса в текст и сравнивает его по смыслу с типичными мошенническими скриптами. Если система фиксирует подозрительные слова, словосочетания, смысловые конструкции, она уведомляет об этом абонента с помощью звукового сигнала. Стоит заметить, что комплексное применение этой технологии и вышеописанных методов защиты помогает практически стопроцентно определять мошеннические звонки. Такие системы очень актуальны в борьбе с голосовым мошенничеством, поскольку во время таких атак жертвы почти всегда находятся под психологическим давлением, поэтому им сложно критически оценивать информацию, которую они слышат по телефону. Звуковой сигнал отрезвит человека, успокоит его и защитит от потери средств.

IV. Предлагаемая система защиты от атак с использованием голосового фишинга

Большинство существующих методов защиты позволяют предотвратить звонок до того, как он дойдёт до жертвы. Если мошенники смогли обойти эту защиту или выбрали другой канал атаки, например, мессенджеры, жертва оказывается практически беззащитна. Предлагаемая система позволит защитить человека во время звонка. Она будет анализировать звонок в реальном времени и классифицировать его как фишинговый или легитимный. Система состоит из двух модулей: модуля преобразования аудиопотока звонка в текст и модуля обработки и классификации текста (рис. 1).

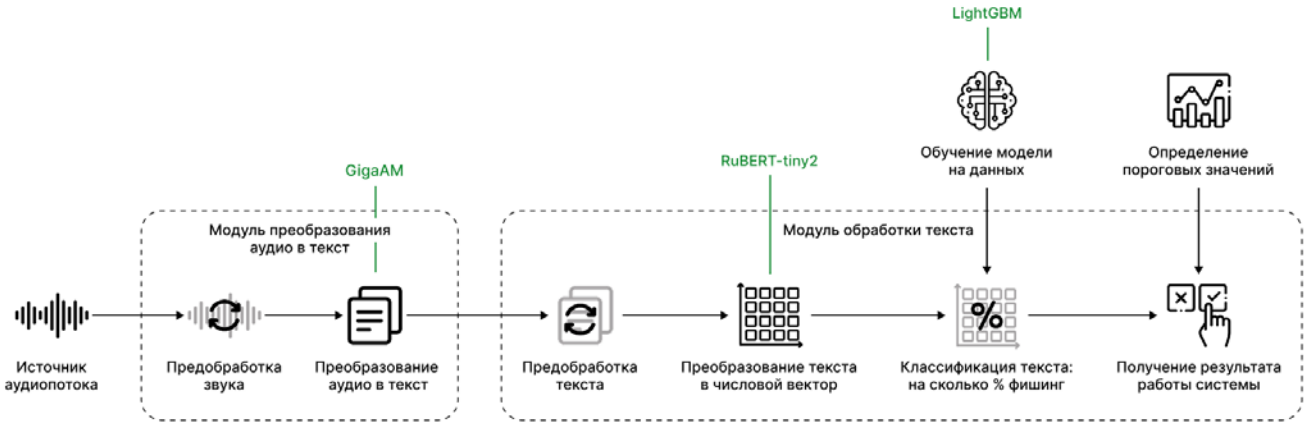

Рис. 1. Архитектура системы

Модуль преобразования аудио в текст получает на вход аудиопоток звонка, выполняет лёгкую предобработку звука и с помощью акустической модели преобразовывает звук в текст.

Для выбора модели распознавания речи, которая будет использована в системе, можно обратиться к исследованию «Сравнительный анализ моделей распознавания русскоязычной речи на примере телефонных звонков» [2], его авторы сравнили самые популярные модели, обрабатывающие речь на русском языке. Особенно ценно, что авторы анализировали распознавание на датасете, состоящем исключительно из телефонных звонков, а также оценивали устойчивость моделей к шуму. По итогам исследования лучшей моделью для распознавания звонков можно признать GigaAM-RNN от СБЕРа, особенно значимыми для предлагаемой системы оказались её следующие характеристики:

- высокая скорость – модель быстро работает с аудио длиной до 30 секунд [3],
- качественная работа с русским языком – модель обучена на 50 000 часов русскоязычных данных [3],

- модель показывает высокую устойчивость к шуму [2], что особенно важно при анализе звонков,
- модель поддерживает обработку данных в реальном времени, что идеально подходит к концепции работы системы.

Модуль обработки текста и классификации звонка является ключевым для системы, так как именно он обеспечивает анализ смысловой части звонка и выдаёт результат о том, является ли звонок мошенническим. Этот модуль состоит из трёх составных частей:

- функция, предобрабатывающая текст для большей эффективности его дальнейшего семантического анализа,
- функция, выполняющая семантический анализ,
- функция, выполняющая классификацию.

Далее будут более подробно рассмотрены технологии, выбранные для реализации составляющих, описанных выше.

1) Предобработка текста. Этот этап необходим, поскольку модуль распознавания голоса выдаёт сплошной текст, который может содержать информацию, мешающую анализу. Чтобы привести текст к стандартизированному виду, необходимо выполнить лёгкую предобработку: удалить лишние пробелы, междометия, служебные символы. При этом стандартная токенизация на этом этапе не нужна, потому что она выполняется с помощью специального токенизатора, соответствующего модели для семантического анализа. Приводить слова к базовой форме с помощью, например, лемматизации тоже не надо, поскольку сохранение исходных форм позволит контекстной модели лучше распознавать смысл. Исследование про сравнение моделей BERT [4] также подтверждает, что легкая обработка без лемматизации и стемминга даёт лучшие результаты.

2) Формирование эмбеддингов и семантика. После проведённой ранее лёгкой предобработки текста необходимо провести токенизацию, чтобы разделить его на токены (слова) и субтокены (части слов), а затем преобразовать в числовые идентификаторы для дальнейшей передачи в модель. После передачи данных на вход в модель, она формирует эмбеддинги – преобразовывает множество идентификаторов слов в единый числовой вектор фиксированной длины. По мере обработки модели данные проходят через слои внимания, и в итоге каждый элемент получает своё контекстное числовое представление, учитывающее предыдущие и следующие элементы, а также смысл всего предложения. После этого из набора векторов всех элементов формируется один общий вектор – он компактный, но при этом информативный, в нём заложены ключевые смысловые связи текста. Полученный вектор далее подаётся на вход классификатора. Для того, чтобы выбрать оптимальную модель-энкодер, которая будет обрабатывать полученный набор токенов и формировать итоговый вектор, необходимо учесть несколько факторов: модель должна хорошо работать с русским языком и обязательно анализировать контекст. Для таких целей отлично подходят контекстные модели типа BERT – они создают эмбеддинги, в которых каждое слово получает уникальный вектор в зависимости от его окружения в предложении. Для выбора конкретной модели можно обратиться к исследованию «Сравнение моделей-трансформеров BERT при выявлении деструктивного контента в социальных медиа» [4], его авторы рассмотрели несколько моделей из семейства BERT с целью найти самую эффективную для выявления деструктивного контента. В исследовании были рассмотрены популярные модели обработки естественного языка. Лучшей признана RuBERT-tiny2, она показала лучшие результаты на этапе преобразования текста в вектор.

3) Классификация. Этот этап является ключевым, так как именно качественная настройка классификатора влияет на то, какой будет результат работы у всей системы. Для классификации выбрана модель LightGBM. Если остальные модели, описанные в этом разделе, не требуют дополнительного обучения, поскольку работают с аудио и текстом на основе правил, связанных с языком, то LightGBM, в отличие от них, требует предварительного этапа обучения. Для этого необходимо собрать набор текстовых данных, состоящий из фишинговых скриптов, а также из текстов обычных легитимных разговоров. Обучение модели происходит так: из каждого текста из датасета строится эмбеддинг, множество полученных векторов подаётся на вход модели, она анализирует эти данные и находит закономерности. После обучения LightGBM сможет находить отличия между обычными и мошенническими текстами и выдавать вероятностный ответ, показывающий, на сколько процентов текст похож на фишинговый. Здесь появляется новая задача – определить порог фишинга и понять, в каком диапазоне вероятностей текст можно считать легитимным, в каком – подозрительным, а в каком – точно фишинговым. К сожалению, этот порог невозможно точно вычислить заранее, он будет зависеть в первую очередь от датасета, на котором обучена модель. Однако, как только модель будет обучена, можно будет собрать статистические данные на разных порогах и выбрать порог в зависимости от конкретных потребностей.

V. Модель LightGBM и градиентный бустинг

Модель LightGBM (Light Gradient-Boosting Machine) использует метод градиентного бустинга, он комбинирует несколько базовых моделей, чтобы получить более точную, при этом осуществляя последовательное обучение базовых моделей, где каждая следующая модель стремится минимизировать ошибки предыдущей.[5] В качестве базовых моделей используются деревья решений, которые строятся по листьям: модель находит лист с максимальным значением ошибки и выбирает его для дальнейшего роста дерева, это позволяет более точно и эффективно улучшать модель.

Пусть композиция из N моделей:

$$\mathrm{F}_N\left(x\right)=\sum_{n=1}^{N} f_n\left(x\right),$$

где $f_n\left(x\right)$ – базовая модель.

На N-ном шаге ищется такая модель $b_N\left(x\right)$, которая минимизирует квадратичное отклонение от сдвига s_i, то есть уменьшит ошибку предыдущей модели:

$$f_N(x) = \arg\min_{b \in A} \sum_{i=1}^{l} \left(f(x_i) - s_i \right)^2,$$

где s_i – сдвиг на конкретной модели, $f(x_i)$ – композиция предыдущих моделей. При этом сдвиг s_i показывает, как нужно скорректировать предыдущую модель, чтобы уменьшить ошибку.

Далее композиция моделей обновляется по правилу:

$$\mathrm{F}_N(x) = \mathrm{F}_{N-1}(x) + \eta f_N(x),$$

где $\mathrm{F}_{N-1}(x)$ – композиция моделей после N-1 шагов, η – скорость обучения.

Таким образом, на каждом шаге текущая композиция моделей обновляется, это позволяет постепенно снижать ошибку. После окончания обучения получается итоговая сложная модель, хорошо отражающая зависимости в данных, её можно использовать для работы с новой информацией, в случае рассматриваемой системы – для классификации новых звонков.

VI. Заключение

В ходе работы был проведён анализ методов осуществления атак с использованием голосового фишинга, методов защиты от таких атак, а также предложена архитектура системы, позволяющая защищать абонента во время звонка.

Главное оружие злоумышленников во время атаки – социальная инженерия, однако именно технические средства позволяют обходить модули защиты и дозваниваться до абонентов. Основной тенденцией в атаках является переход в мессенджеры, поскольку уровень защиты у этого канала связи значительно ниже, чем у обычных звонков.

Большинство методов защиты направлены на предотвращение фишингового звонка до того, как он дойдёт до абонента, для этого используются антифрод-системы, антиспам-системы, метки безопасности звонка. Основными тенденциями в сфере защиты являются использование ИИ для анализа технических характеристик и аудиопотока звонка, а также комбинирование банковских и телефонных антифрод-систем.

Благодарность

Выражаю благодарность своему научному руководителю Обухову Александру Валерьевичу за ценные советы при планировании исследования и рекомендации по оформлению статьи.

Список литературы

[1] Психология фишинга: как мошенники используют психологические приемы для обмана пользователей и как им противостоять. Комсомольская правда URL: https://www.kp.ru/daily/27607/4934207/ (дата обращения: 18.06.2025).

[2] Газизулин Р.М., Хартьян Д.Ю. Сравнительный анализ моделей распознавания русскоязычной речи на примере телефонных звонков. Тюмень: 2024.

[3] GigaAM: the family of open-source acoustic models for speech processing // GitHub URL: https://github.com/salute-developers/GigaAM?tab=readme-ov-file (дата обращения: 10.05.2025).

[4] Минаев В.А., Симонов А.В. Сравнение моделей-трансформеров BERT при выявлении деструктивного контента в социальных медиа // Информация и безопасность. 2022. С. 341-348.

[5] Градиентный бустинг // Яндекс Образование URL: https://education.yandex.ru/handbook/ml/article/gradientnyj-busting (дата обращения: 25.06.2025).