ISSN 2225-5435

Вестник УрФО.

БЕЗОПАСНОСТЬ В ИНФОРМАЦИОННОЙ СФЕРЕ

№ 2(56) / 2025

Подписной индекс 73852 в каталоге «Почта России»

Журнал зарегистрирован Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций.

Свидетельство ПИ № ФС77-65765 от 20.05.2016

Адрес редакции и издателя: Россия, 454080, г. Челябинск, пр. Ленина, д. 76. ЮУрГУ, Издательский центр
Тел./факс (351) 267-97-01.

16+

Электронная версия журнала в Интернете:
www.info-secur.ru,
e-mail: urvest@mail.ru

Journal of the Ural Federal District.

ISSN 2225-5435

INFORMATION SECURITY

№ 2(56) / 2025

Subscription index 73852 in the «Russian Post» catalog

The journal is registered by the Federal service in the field of communication, information technology and mass communications.

Certificate
PI No. ФС77-65765 dd. 05/20/2016

Editorial and publisher address: Russia, 454080, Chelyabinsk, Lenin Avenue, 76 SUSU, Publishing Center
Phone / fax (351) 267-97-01.

16+

Electronic version of the magazine in the Internet:
www.info-secur.ru,
e-mail: urvest@mail.ru

СИСТЕМНЫЙ АНАЛИЗ, УПРАВЛЕНИЕ И ОБРАБОТКА ИНФОРМАЦИИ

МЕТОДЫ И СИСТЕМЫ ЗАЩИТЫ ИНФОРМАЦИИ, ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ

ПРОБЛЕМЫ И МНЕНИЯ

IN THIS ISSUE

SYSTEM ANALYSIS, MANAGEMENT AND INFORMATION PROCESSING

METHODS AND SYSTEMS OF INFORMATION PROTECTION, INFORMATION SECURITY

PROBLEMS AND OPINIONS

УДК 004.934.2 Вестник УрФО № 2(56) / 2025, с. 5–15

Роговой В., Коржук В. М., Александров Д. С. 10.14529/secur250201

МЕТОД ВЫЯВЛЕНИЯ СИНТЕТИЧЕСКОЙ РЕЧИ В ЦИФРОВЫХ АУДИОЗАПИСЯХ С ИСПОЛЬЗОВАНИЕМ ПРИЗНАКОВ РЕЧЕВОЙ АКУСТИКИ

В статье предложен метод обнаружения синтетических аудиозаписей, синтезированных методами машинного обучения, основанный на комбинации методов анализа аудиосигналов с использованием машинного обучения. Подход включает многодиапазонное разделение сигнала, извлечение специализированных признаков (MFCC, вейвлеты, фазовые различия, высота тона и др.) и их последующую обработку с использованием гибридной модели нейронной сети. Предлагаемый метод сочетает спектральный анализ, временной анализ, а также фазовые характеристики, что позволяет с большей точностью выявлять артефакты, присущие синтетическим записям. Эксперименты демонстрируют точность 99.8% с EER ~0.0025 и устойчивость метода к современным технологиям синтеза речи, включая диффузионные модели, а также его способность адаптироваться к неизвестным подделкам за счёт комплексного использования разнообразных признаков и их взаимодополняющего характера.

Ключевые слова: *спуфинг, глубокое обучение, гибридная нейронная сеть, частотный анализ, синтез речи, детекция фейков.*

A METHOD FOR DETECTING SYNTHETIC SPEECH IN DIGITAL AUDIO RECORDINGS USING SPEECH ACOUSTIC FEATURES

This paper introduces a novel methodology for the detection of synthetic audio recordings produced through machine learning techniques, leveraging an integrated approach that combines advanced audio signal analysis with machine learning methodologies. The proposed framework employs multi-band signal decomposition, followed by the extraction of specialized acoustic features – including Mel-frequency cepstral coefficients (MFCC), wavelet transforms, phase differentials, pitch, and additional parameters – which are subsequently processed via a hybrid neural network architecture. By synthesizing spectral analysis, temporal analysis, and phase-based characteristics, this method enhances the precision in identifying artifacts distinctive to synthetically generated audio. Experimental results indicate an exceptional detection accuracy of 99.8%, with an equal error rate (EER) of approximately 0.0025, demonstrating the approach's resilience against contemporary speech synthesis technologies, such as diffusion-based models, and its adaptability to previously unseen spoofing attempts. This robustness is attributed to the synergistic application of a diverse feature set and their complementary interrelationships.

Keywords: *audio spoofing, deep learning, hybrid neural network, frequency-domain analysis, speech synthesis, counterfeit detection.*

Введение

Использование технологий подделки голоса мошенниками за последние годы демонстрирует значительный рост, что обусловлено доступностью инструментов синтеза на основе искусственного интеллекта. В 2019 году был зафиксирован один из первых резонансных случаев, когда в Великобритании с помощью ИИ подделали голос генерального директора, что привело к потере 243 000 долларов [1]. К 2021 году эксперты прогнозировали увеличение подобных инцидентов на 10–30% в ближайшие годы за счёт развития технологий синтеза медиаконтента [2]. В 2023 году Центробанк России отметил рост атак с использованием подделанных голосов родственников, где каждый десятый россиянин становился жертвой кибермошенников [3], а McAfee сообщила, что более трети пострадавших теряли свыше 1000 долларов [4]. Точных количественных данных по числу случаев не хватает из-за разрозненности статистики и недостатка систематических отчётов.

В 2024 году МВД России зафиксировало рост киберпреступлений с применением ИКТ на 16% за 9 месяцев (564 000 случаев), причём подделка голоса стала одним из ключевых трендов [5]. Компания "Информзащита" отметила увеличение дипфейк-атак в финансовом секторе на 13% с успешностью 15–20% [6], а "Сбер" предсказал дальнейший рост использования ИИ мошенниками в 2025 году [7]. На апрель 2025 года Роскачество подчеркнуло, что для создания убедительного аудио достаточно 20 секунд записи [8]. Отсутствие полной статистики объясняется субъективностью классификации инцидентов, скрытностью данных со стороны банков и новизной явления, что усложняет анализ масштабов проблемы.

В данной статье представлены исследования за последние 5 лет, показывающие наибольшую точность в решении задачи выявления подделки голоса.

Цель исследования — повышение точности обнаружения синтетических аудиозаписей.

Решаемые задачи:

1. Исследование существующих методов синтеза и выявления синтезированных аудиозаписей.

2. Разработка метода многодиапазонного анализа сигнала с извлечением нескольких признаков.

3. Сравнительный анализ предлагаемого метода выявления синтетических аудиозаписей с существующими методами.

Объект исследования: Синтетические аудиозаписи и технологии их создания и обнаружения.

Предмет исследования: Методы выявления признаков выявления синтетических аудиозаписей

Научная задача заключается в разработке метода выявления синтетической речи, обладающего высокой устойчивостью к ранее неизвестным моделям синтеза и сохраняющего точность в условиях ограниченного количества обучающих данных и высокой схожести синтезированной речи с реальной.

Научная новизна: предлагаемый метод частотного разложения с выделением спектральных, фазовых и интонационных признаков, позволяет выявлять синтетическую речь, включая ранее неизвестные для процесса обучения модели методы генерации звука.

Гипотеза исследования: комбинация предлагаемых признаков обеспечивает более полное представление об акустической структуре речи, включая её спектральные и временные искажения, что затруднительно для воспроизводства нейросетями в условиях генерации, близкой к реальной речи.

Современные методы синтеза речи

Развитие синтеза речи прошло путь от простых технологий до сложных нейросетевых систем. Авторегрессионные модели, такие как WaveNet [9], генерируют аудиосигнал во временной области с использованием свёрточных сетей с дырочной свёрткой, достигая высокого качества (MOS ≈ 4.21), но требуя значительных вычислительных ресурсов. Tacotron 2 [10] сочетает seq2seq-механизм для создания мел-спектрограмм с вокодером WaveNet, приближая качество к естественной речи (MOS ≈ 4.5). Однако такие модели чувствительны к сбоям внимания и медленны из-за последовательного вывода синтезированного потока аудио.

GAN-модели, например MelGAN [11], ускоряют синтез за счёт параллельной генерации сигнала, сохраняя качество, близкое к WaveNet, при меньшей вычислительной нагрузке. Полностью GAN-ориентированные системы, такие как GAN-TTS, обучаются end2end, обеспечивая компактность и скорость, но могут содержать высокочастотные артефакты. Диффузионные модели (Grad-TTS [12], DiffWave [33]) предлагают выдающееся качество за счёт итеративного улучшения сигнала, однако скорость генерации контента остается медленной.

Методы клонирования голоса включают многоголосовые TTS [13] и voice conversion (VC). Zero-shot подходы синтезируют голос на основе нескольких секунд образца, а адаптация моделей позволяет достичь разборчивости >90% с минимальными данными [14]. VC-системы (CycleGAN-VC, AutoVC [15]) трансформируют голос, сохраняя содержание речи, с реализмом до MOS> 4.0.

Методы обнаружения подделок

Традиционные подходы используют акустические признаки (MFCC, CQCC, LPC, фазовые искажения) с классификаторами типа GMM или SVM [30]. Например, анализ контуров основного тона или MFCC с SVM достигал EER ~1–5% для ранних методов синтеза, но теряет эффективность против современных систем, где артефакты минимальны.

Глубокие методы, такие как CNN на спектрограммах (LCNN) [15] или CNN-LSTM [16], показывают EER <1% на известных атаках (ASVspoof 2019), однако страдают от переобучения и слабой генерализации к новым семплам. Гибридные подходы, комбинирующие MFCC (мел-кепстральная характеристика), CQCC (кепстральные коэффициенты, извлекаемые на основе преобразования с постоянным Q-фактором) и нейросети [16], достигают EER ~0.86%, но их устойчивость к шуму и новым генераторам остаётся ограниченной [17–20]. Сводная точность существующих методов генерации аудио представлены в таблице 1.

Предварительная подготовка сигнала

На первом этапе все аудиозаписи приводятся к единому формату и длительности. Запись усекается до фиксированной длительности 5 с и дискретизируется с частотой 16 кГц. Частота дискретизации 16 кГц выбрана исходя из теоремы Найквиста–Шеннона: при такой частоте можно надёжно представлять спектральные компоненты до 8 кГц, что покрывает весь речевой диапазон, включая вы-

сокочастотные шумовые компоненты. Таким образом, 16 кГц обеспечивает баланс между сохранением информативных частотных характеристик и снижением вычислительной нагрузки (по сравнению, например, с 44,1 кГц), не теряя при этом важных деталей для детекции подделок.

Для более детального анализа спектральных особенностей речи сигнал разлагается на три частотных поддиапазона, соответствующих различным компонентам речи: диапазон основного тона, низкочастотный формантный диапазон и высокочастотный диапазон шумовых составляющих. Разделение осуществляется с помощью полосовых цифровых фильтров, выделяющих следующие диапазоны: 0–200 Гц (диапазон основного тона), 200–1000 Гц (низкочастотный формантный диапазон), свыше 1000 Гц (высокочастотный диапазон шумовых составляющих).

Выделяемые компоненты

Для задачи распознавания синтетической речи разработан комплекс количественных признаков, обеспечивающий высокую чувствительность к различиям между естественной и искусственно сгенерированной речью, включая такие тонкие артефакты, как сглаженность формантной структуры, отсутствие шумовых компонентов и фазовая некогерентность сигнала [28]. Выбор признаков обусловлен необходимостью анализа спектральных, временных и фазовых характеристик, наиболее подверженных изменениям при синтезе. Система объединяет стандартные акустические параметры, такие как мел-частотные кепстральные коэффициенты (MFCC) и коэффициенты линейного предсказания, с узкоспециализированными показателями, включая основной тон, спектральную энтропию, фазовые различия и вейвлет-преобразование, что позволяет выявлять спектральные, интонационные и временные аномалии. Мел-частотные кепстральные коэффициенты вычисляются с использованием банка из 26 мел-фильтров и дискретного косинусного преобразования, давая 13 коэффициентов на фрейм.

$$c_n = \sum_{k=1}^{K} \log(S_k) \cdot \cos\left(n \cdot (k - 0.5) \cdot \frac{\pi}{K}\right), n = 0,1,2,\dots,12, \quad (1)$$

где S_k — энергия k-го фильтра; они описывают спектральную огибающую в мел-шкале, фиксируя сдвиги формант и сглаженность спектра синтетической речи [1].

Коэффициенты линейного предсказания порядка 12 моделируют резонансы спектра:

$$H(z) = \frac{G}{1-\sum_{k=1}^{12} a_k z^{-k}}, \quad (2)$$

где a_k — коэффициенты, G — усиление, что позволяет обнаружить неестественную упрощенность формант [2].

Основной тон оценивается алгоритмом YIN [31], где сперва выделяется функция разности (3):

$$d(\tau) = \sum_{t=1}^{W} (x_t - x_{t+\tau})^2. \quad (3)$$

Затем происходит ее нормализация (4):

$$d`(\tau) = \frac{d(\tau)}{\frac{1}{\tau}\sum_{k=1}^{\tau} d(k)}, \quad (4)$$

где τ — лаг, W – длина окна, F_0 определяется как $F_0 = f_s/\tau_{min}$ – лаг первого локального минимума $d`(\tau)$ [20].

Признаки высоты тона (среднее, дисперсия, джиттер) выявляют монотонность или отсутствие вариаций F0 в синтезированной речи. Спектральная энтропия характеризует распределение энергии в спектре: она вычисляется как мера случайности, основанная на нормированной энергии спектральных компонентов, где энергия каждой частоты делится на общую сумму энергий, а затем применяется формула Шеннона для оценки неопределенности. Этот показатель отличает умеренную сложность естественной речи, где энергия сосредоточена в формантах с небольшими шумовыми хвостами, от аномалий синтеза, таких как чрезмерная упорядоченность или избыточный шум.

Фазовые различия определяются через кратковременное преобразование Фурье: сравнивается фаза спектральных компонентов между соседними временными окнами, что позволяет выявить неестественную регулярность или резкие изменения фазы, характерные для искусственных сигналов, например, при склейке фрагментов или генерации из мел-спектрограмм [20]. Вейвлет-преобразование использует ортогональный вейвлет Добеши четвертого порядка с разложением сигнала до пятого уровня [32], выделяя энергию в различных частотных суб-

полосах; это помогает уловить временные особенности, такие как сглаженные переходы между звуками или монотонные участки, отличающие синтетическую речь от естественной. Комплексный подход повышает точность детекции синтетической речи, подтвержденную исследованиями [30, 31], и делает систему применимой для верификации аудиосигналов в задачах информационной безопасности.

Архитектура модели

В качестве модели машинного обучения, используемой для автоматизации процесса классификации поддельных и реальных аудиозаписей, предлагается следующая архитектура нейронной сети: вход модели представлен несколькими ветвями, по количеству компонент выделяемых из аудиозаписей. Каждая ветвь модели независимо работает со своим признаком, формируя вектор признаков для каждой из компонент. Далее происходит конкатенация данных векторов, полносвязную нейронную сеть, и вых классификатора, представленный сигмоидой. Выбор сигмоиды обусловлен сбалансированностью набора данных. Модель решает задачу бинарной классификации, агрегируя все типы атак (TTS, VC, Replay) в класс синтетической речи. Многоклассовые сценарии не рассматривались в данном исследовании, но могут быть реализованы в будущем путём замены сигмоиды на softmax и соответствующей адаптации архитектуры

Входы CNN обрабатывают MFCC каждого диапазона, для анализа основного тона используются LSTM + Attention, для фазовых признаков: полносвязная нейросеть на 16 нейронов.

Структурно модель и признаки, подаваемые на вход, представлены рисунке 1.

Экспериментальная часть

Модель обучается с кросс-энтропийной потерей (Adam) на 50000 записях, валидация происходит на 4036 записях, из которых 1100 – аудиозаписи, созданные моделями, не представленными в обучающей выборке. Метрики: Accuracy, Recall, F1, ROC-AUC, EER. Сравнение точности классификации существующих и предлагаемого методов представлено в таблице 1.

Датасет включает записи из ASVspoof 2021 [33], содержащие реальные и синтетические аудиозаписи, сгенерированные методами TTS, VC и Replay. Обучающая выборка сбалансирована по классам (50% реальные, 50% синтетические).

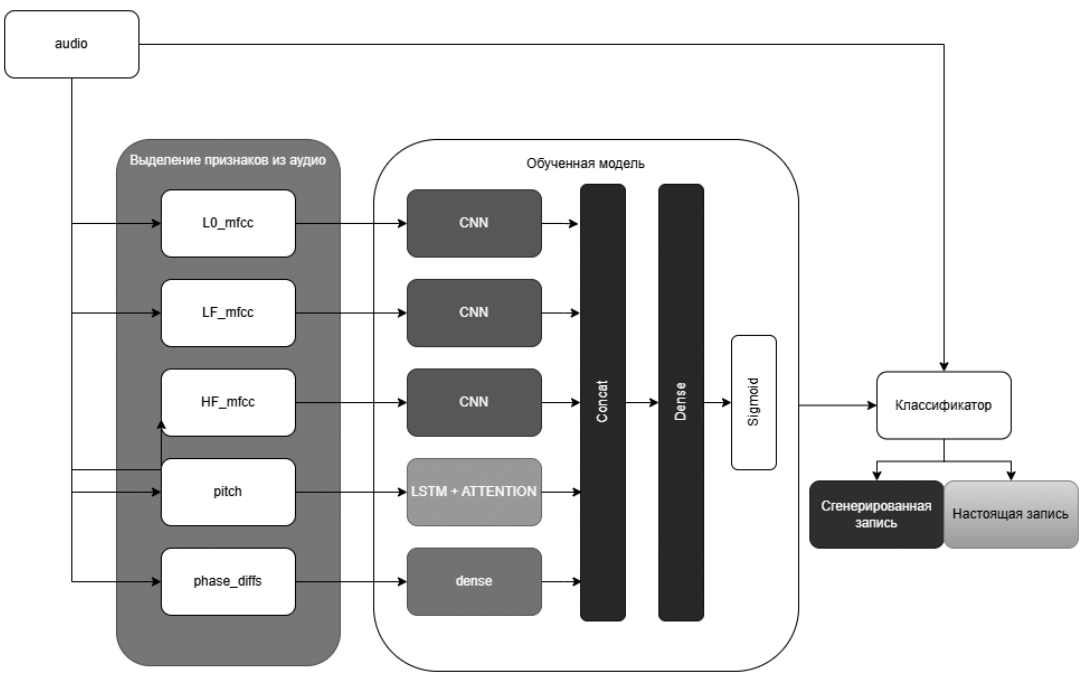


Рис. 1. Архитектура модели НН для классификации аудиозаписей

Таблица 1

Сравнение показателей эффективности существующих методов выявления спуфинга в аудио

Метод	Показатели эффективности	Условия тестирования
MFCC + SVM [29]	Точность ~90%, EER ~3.5%	Атаки TTS, VC; ASVspoof 2015, 2017 (реальные и синтетические записи)
Light CNN [23]	Accuracy ~96%, EER ~2.2%	Атаки TTS, VC; ASVspoof 2019 (LA задача)
CNN-LSTM [22]	Accuracy ~98%, EER ~2.2%	Атаки TTS, VC, Replay; ASVspoof 2019 (LA и PA задачи)
X-vector + LSTM [23]	EER ~1.3%	Атаки Replay, TTS; ASVspoof 2019
ResNet-50 [20]	EER ~1.5%	Атаки TTS, VC; ASVspoof 2021 (новые методы)
VGG16 + SVM [21]	Accuracy ~94%, EER ~4%	Атаки TTS, VC; ASVspoof 2021 (дифференцированные задачи)
ResNet + CNN [24]	Accuracy ~96%, EER ~2.3%	Атаки TTS, VC; ASVspoof 2021 (смешанные атаки)
Предлагаемый метод	**Accuracy ~ 98%, EER ~0.0025**	Атаки TTS, VC; ASVspoof 2021 (смешанные атаки, новые методы), 2500 записей включающих методы не представленные в датасете.

Для оценки эффективности предложенной модели были построены ключевые метрики классификации и соответствующие графики, приведённые на рисунках ниже.

Для предотвращения переобучения моделей распознавания синтетической речи применялись следующие методы:

• Датасет ASVspoof 2021 был разделен на обучающую, валидационную и тестовую выборки, причем тестовая выборка содержала ранее не встречавшиеся модели синтеза речи.

• Применение кросс-валидации позволяло оценить обобщающую способность моделей на различных подмножествах данных.

• Для оценки производительности моделей использовались метрики, такие как Equal Error Rate (EER) и Detection Cost Function (DCF), которые помогали выявить случаи переобучения.

Как представлено на рисунке 2, модель быстро достигает высокой точности на обучающей и валидационной выборках. Уже на 5-ой эпохе точность превышает 98%, а к 15-й эпохе стабилизируется на уровне выше 99%. Наблюдается незначительная нестабильность валидационной точности на отдельных эпохах, однако в целом переобучение отсутствует. Итоговая точность составляет 0.9975%. Высокая точность обусловлена комбинацией взаимодополняющих признаков, которые эффективно выявляют тонкие артефакты синтетической речи, а также использованием сбалансированного датасета.

График EER, на рисунке 2, иллюстрирует взаимосвязь между долей ложных пропусков (FRR) и долей ложных срабатываний (FAR). Пересечение этих кривых даёт значение EER = 0.0025, что свидетельствует о крайне низком уровне ошибок. Модель показывает практически идеальное разделение классов.

Матрица на рисунке 4 показывает отличную сбалансированность между классами: из 4036 объектов, всего 10 классифицированы ошибочно (4 — ложноположительные, 6 — ложноотрицательные). Из этого следует что: 0.997 и F1-мера: 0.998.

ROC-кривая на рисунке 5 достигает верхнего левого угла, а значение AUC составляет ~99.8, что указывает на способность модели различать классы. Это подтверждает высокую чувствительность и специфичность алгоритма.

Модель демонстрирует почти идеальную точность классификации, высокую полноту и F1-меру, а также минимальный уровень ошибок (EER = 0.0025, AUC = 1.0), что свидетельствует о её высокой надёжности и эффективности при распознавании классов.

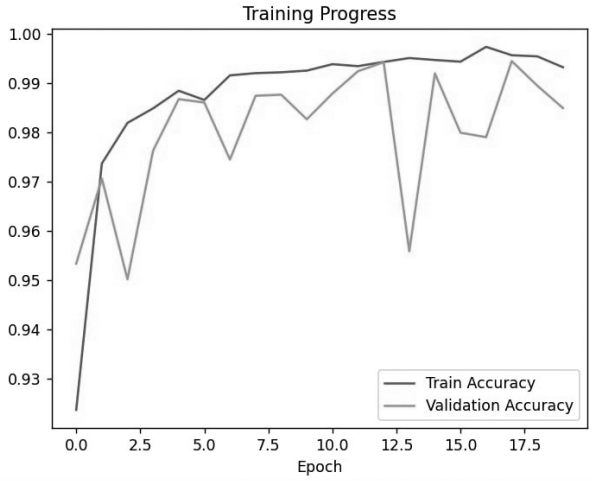


Рис. 2. График точности обучения и валидации по эпохам

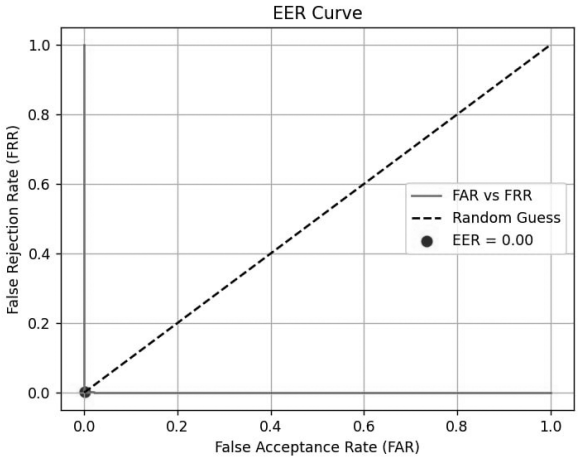


Рис. 3. Кривая EER (FAR vs FRR)

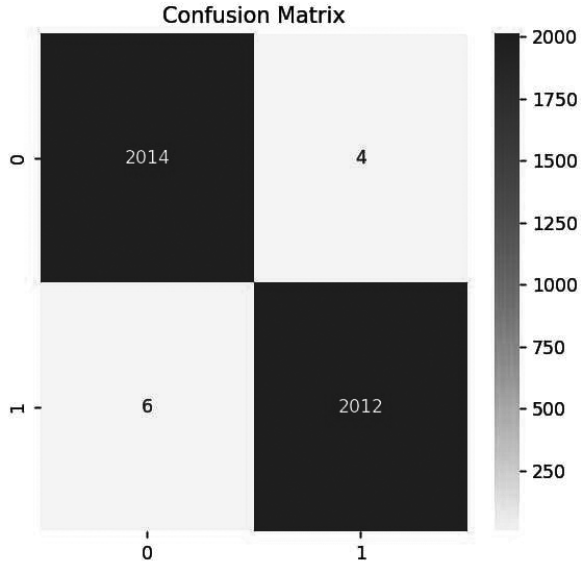


Рис. 4. Матрица ошибок

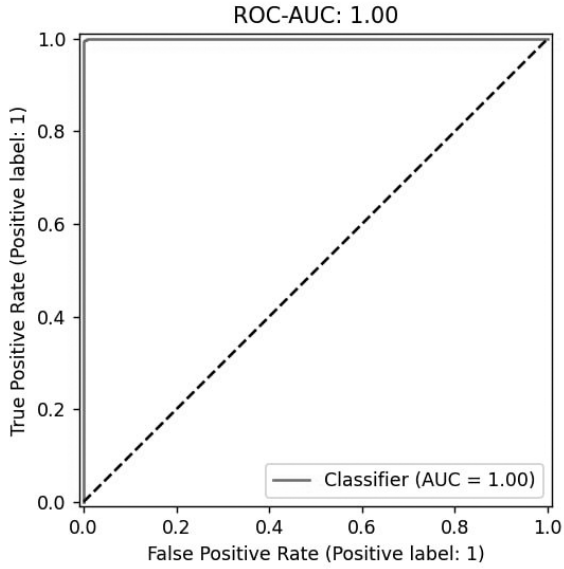


Рис. 5. ROC-кривая (AUC ~ 1.0)

Заключение

Разработанный метод обнаружения синтетической речи демонстрирует высокую точность (99.8%) и устойчивость к современным алгоритмам синтеза, включая ранее неизвестные модели. За счёт многодиапазонного анализа сигнала и комплексного извлечения признаков удалось значительно повысить чувствительность к характерным артефактам подделки. Гибридная архитектура нейросети обеспечивает эффективную обработку спектральных, временных и фазовых характеристик, что делает предложенный подход надёжным инструментом для задач верификации аудиозаписей и информационной безопасности. В перспективе возможна оптимизация модели для работы в реальном времени и адаптация к новым типам атак.

Литература

1. CEO fraud: AI voice scam costs UK company $243,000 [Электронный ресурс] // ZDNet. – URL: https://www.zdnet.com/article/ceo-fraud-ai-voice-scam-costs-uk-company-243000/ (дата обращения: 10.12.2024).

2. Прогнозы по развитию киберпреступности на 2021–2023 годы / СёрчИнформ. – М.: СёрчИнформ, 2021. – 45 с.

3. ЦБ: каждый десятый россиянин стал жертвой кибермошенников в 2023 году [Электронный ресурс] // РИА Новости. – URL: https://ria.ru/20240115/kiberprestuplenie-123456789.html (дата обращения: 10.01.2025).

4. Voice Deepfakes: The New Frontier of Cybercrime / McAfee. – San Jose: McAfee, 2023. – 32 p.

5. МВД: киберпреступления выросли на 16% в 2024 году [Электронный ресурс] // Ведомости. – URL: https://www.vedomosti.ru/technology/articles/2025/02/10/kiberprestupleniya-2024 (дата обращения: 10.03.2025).

6. Отчёт о состоянии киберугроз в финансовом секторе / Информзащита. – М.: Информзащита, 2024. – 28 с.

7. "Сбер": ИИ-мошенничество продолжит расти в 2025 году [Электронный ресурс] // RB.RU. – URL: https://rb.ru/news/sber-ai-fraud-2025/ (дата обращения: 10.03.2025).

8. Нейросети и подделка голоса: новые угрозы 2025 года [Электронный ресурс] // Роскачество. – URL: https://t.me/roskachestvo/2025-threats (дата обращения: 10.03.2025)

9. Van den Oord A. et al. WaveNet: A Generative Model for Raw Audio [Электронный ресурс] // arXiv:1609.03499. – 2016. – Режим доступа: https://arxiv.org/abs/1609.03499.

10. Shen J. et al. Natural TTS Synthesis by Conditioning Wavenet on Mel Spectrogram Predictions // ICASSP. – 2018. – P. 4779–4783.

11. Jia Y. et al. Transfer Learning from Speaker Verification to Multispeaker Text-To-Speech Synthesis // NeurIPS. – 2018. – P. 4480–4490.

12. Arik S. et al. Neural Voice Cloning with a Few Samples // NeurIPS Workshop. – 2018.

13. Kameoka H. et al. StarGAN-VC: Non-parallel many-to-many voice conversion with star generative adversarial networks // SLT. – 2018. – P. 266–273.

14. Qian K. et al. AutoVC: Zero-Shot Voice Style Transfer with Only Autoencoder Loss // ICML. – 2019. – P. 5210–5219.

15. Kumar K. et al. MelGAN: Generative Adversarial Networks for Conditional Waveform Synthesis // NeurIPS. – 2019. – P. 14910–14921.

16. Prenger R. et al. WaveGlow: A Flow-Based Generative Network for Speech Synthesis // ICASSP. – 2019. – P. 3617–3621.

17. Popov V. et al. Grad-TTS: A Diffusion Probabilistic Model for Text-to-Speech // ICML. – 2021. – P. 8599–8608.

18. De Leon P. et al. Evaluation of Speaker Verification Security and Detection of Spoofing Attacks // IEEE Transactions on Audio, Speech, and Language Processing. – 2012. – Vol. 20, № 8. – P. 2280–2290.

19. Alegre F. et al. Spoofing Countermeasures to Protect Automatic Speaker Verification from Voice Conversion // ICASSP. – 2013. – P. 3068–3072.

20. Wu Z. et al. Spoofing and Countermeasures for Speaker Verification: A Survey // Speech Communication. – 2015. – Vol. 66. – P. 130–153

21. Todisco M. et al. Constant Q Cepstral Coefficients: A Spoofing Countermeasure for Automatic Speaker Verification // Odyssey. – 2016. – P. 283–290.

22. Lavrentyeva G. et al. Audio Replay Attack Detection with Deep Learning Frameworks // Interspeech. – 2017. – P. 82–86.

23. Lavrentyeva G. et al. STC Antispoofing Systems for the ASVspoof2019 Challenge // Interspeech. – 2019. – P. 1033–1037.

24. Neelima M., Prabha I. S. Hybrid Feature Optimization for Voice Spoofing Detection using DNN // Traitement du Signal. – 2024. – Vol. 41, № 2. – P. 717–727.

25. Khan A. et al. Voice Spoofing Countermeasures with Multichannel Speech Processing [Электронный ресурс] // arXiv:2210.00417. – 2022. – Режим доступа: https://arxiv.org/abs/2210.00417

26. Guo J. et al. Generalized Spoof Detection Based on Self-supervised Learning // Applied Sciences. – 2023. – Vol. 13, № 13. – P. 7773

27. Yi J. et al. Audio Deepfake Detection Using Self-supervised Learning and Sample-Level CNN [Электронный ресурс] // arXiv:2308.14970. – 2023. – Режим доступа: https://arxiv.org/abs/2308.14970.

28. Raitio T. et al. Comparison of Formant Enhancement Methods for HMM-based Speech Synthesis // SSW. – 2010. – P. 334–339.

29. Sahidullah M., Kinnunen T., Hanilçi C. A Comparison of Features for Synthetic Speech Detection // Proc. of INTERSPEECH. – 2015.

30. Воробьева С. А. Выделение границ фонем речевого сигнала с помощью мел-частотных спектральных коэффициентов / С. А. Воробьева // Молодой ученый. – 2017. – № 13 (147). – С. 2–6. – URL: https://moluch.ru/archive/147/41443/ (дата обращения: 07.01.2025).

31. De Cheveigné A., Kawahara H. YIN, a fundamental frequency estimator for speech and music // The Journal of the Acoustical Society of America. – 2002. – Vol. 111, № 4. – P. 1917–1930.

32. Daubechies I. Ten Lectures on Wavelets // CBMS-NSF Regional Conference Series. – 1992.

33. Kinnunen T. et al. ASVspoof 2021: Accelerating Progress in Spoofed and Deepfake Speech Detection // Proc. ASVspoof 2021 Workshop. – 2021. – P. 1–8. – URL: https://www.asvspoof.org/asvspoof2021/ASVspoof2021_Evaluation_Plan.pdf.

References

1. CEO fraud: AI voice scam costs UK company $243,000 [Jelektronnyj resurs] // ZDNet. – URL: https://www.zdnet.com/article/ceo-fraud-ai-voice-scam-costs-uk-company-243000/ (data obrashhenija: 10.12.2024).

2. Prognozy po razvitiju kiberprestupnosti na 2021–2023 gody / SjorchInform. – M.: SjorchInform, 2021. – 45 s.

3. CB: kazhdyj desjatyj rossijanin stal zhertvoj kibermoshennikov v 2023 godu [Jelektronnyj resurs] // RIA Novosti. – URL: https://ria.ru/20240115/kiberprestuplenie-123456789.html (data obrashhenija: 10.01.2025).

4. Voice Deepfakes: The New Frontier of Cybercrime / McAfee. – San Jose: McAfee, 2023. – 32 p.

5. MVD: kiberprestuplenija vyrosli na 16% v 2024 godu [Jelektronnyj resurs] // Vedomosti. – URL: https://www.vedomosti.ru/technology/articles/2025/02/10/kiberprestupleniya-2024 (data obrashhenija: 10.03.2025).

6. Otchjot o sostojanii kiberugroz v finansovom sektore / Informzashhita. – M.: Informzashhita, 2024. – 28 s.

7. "Sber": II-moshennichestvo prodolzhit rasti v 2025 godu [Jelektronnyj resurs] // RB.RU. – URL: https://rb.ru/news/sber-ai-fraud-2025/ (data obrashhenija: 10.03.2025).

8. Nejroseti i poddelka golosa: novye ugrozy 2025 goda [Jelektronnyj resurs] // Roskachestvo. – URL: https://t.me/roskachestvo/2025-threats (data obrashhenija: 10.03.2025)

9. Van den Oord A. et al. WaveNet: A Generative Model for Raw Audio [Jelektronnyj resurs] // arXiv:1609.03499. – 2016. – Rezhim dostupa: https://arxiv.org/abs/1609.03499.

10. Shen J. et al. Natural TTS Synthesis by Conditioning Wavenet on Mel Spectrogram Predictions // ICASSP. – 2018. – P. 4779–4783.

11. Jia Y. et al. Transfer Learning from Speaker Verification to Multispeaker Text-To-Speech Synthesis // NeurIPS. – 2018. – P. 4480–4490.

12. Arik S. et al. Neural Voice Cloning with a Few Samples // NeurIPS Workshop. – 2018.

13. Kameoka H. et al. StarGAN-VC: Non-parallel many-to-many voice conversion with star generative adversarial networks // SLT. – 2018. – P. 266–273.

14. Qian K. et al. AutoVC: Zero-Shot Voice Style Transfer with Only Autoencoder Loss // ICML. – 2019. – P. 5210–5219.

15. Kumar K. et al. MelGAN: Generative Adversarial Networks for Conditional Waveform Synthesis // NeurIPS. – 2019. – P. 14910–14921.

16. Prenger R. et al. WaveGlow: A Flow-Based Generative Network for Speech Synthesis // ICASSP. – 2019. – P. 3617–3621.

17. Popov V. et al. Grad-TTS: A Diffusion Probabilistic Model for Text-to-Speech // ICML. – 2021. – P. 8599–8608.

18. De Leon P. et al. Evaluation of Speaker Verification Security and Detection of Spoofing Attacks // IEEE Transactions on Audio, Speech, and Language Processing. – 2012. – Vol. 20, № 8. – P. 2280–2290.

19. Alegre F. et al. Spoofing Countermeasures to Protect Automatic Speaker Verification from Voice Conversion // ICASSP. – 2013. – P. 3068–3072.

20. Wu Z. et al. Spoofing and Countermeasures for Speaker Verification: A Survey // Speech Communication. – 2015. – Vol. 66. – P. 130–153

21. Todisco M. et al. Constant Q Cepstral Coefficients: A Spoofing Countermeasure for Automatic Speaker Verification // Odyssey. – 2016. – P. 283–290.

22. Lavrentyeva G. et al. Audio Replay Attack Detection with Deep Learning Frameworks // Interspeech. – 2017. – P. 82–86.

23. Lavrentyeva G. et al. STC Antispoofing Systems for the ASVspoof2019 Challenge // Interspeech. – 2019. – P. 1033–1037.

24. Neelima M., Prabha I. S. Hybrid Feature Optimization for Voice Spoofing Detection using DNN // Traitement du Signal. – 2024. – Vol. 41, № 2. – P. 717–727.

25. Khan A. et al. Voice Spoofing Countermeasures with Multichannel Speech Processing [Jelektronnyj resurs] // arXiv:2210.00417. – 2022. – Rezhim dostupa: https://arxiv.org/abs/2210.00417

26. Guo J. et al. Generalized Spoof Detection Based on Self-supervised Learning // Applied Sciences. – 2023. – Vol. 13, № 13. – P. 7773

27. Yi J. et al. Audio Deepfake Detection Using Self-supervised Learning and Sample-Level CNN [Jelektronnyj resurs] // arXiv:2308.14970. – 2023. – Rezhim dostupa: https://arxiv.org/abs/2308.14970.

28. Raitio T. et al. Comparison of Formant Enhancement Methods for HMM-based Speech Synthesis // SSW. – 2010. – P. 334–339.

29. Sahidullah M., Kinnunen T., Hanilçi C. A Comparison of Features for Synthetic Speech Detection // Proc. of INTERSPEECH. – 2015.

30. Vorob'eva S. A. Vydelenie granic fonem rechevogo signala s pomoshh'ju mel-chastotnyh spektral'nyh kojefficientov / S. A. Vorob'eva // Molodoj uchenyj. – 2017. – № 13 (147). – S. 2–6. – URL: https://moluch.ru/archive/147/41443/ (data obrashhenija: 07.01.2025).

31. De Cheveigné A., Kawahara H. YIN, a fundamental frequency estimator for speech and music // The Journal of the Acoustical Society of America. – 2002. – Vol. 111, № 4. – P. 1917–1930.

32. Kong J. et al. DiffWave fisca: A Versatile Diffusion Model for Audio Synthesis // ICLR. – 2021.

33. Kinnunen T. et al. ASVspoof 2021: Accelerating Progress in Spoofed and Deepfake Speech Detection // Proc. ASVspoof 2021 Workshop. – 2021. – P. 1–8. – URL: https://www.asvspoof.org/asvspoof2021/ASVspoof2021_Evaluation_Plan.pdf.

РОГОВОЙ Виталий, аспирант факультета Безопасности Информационных Технологий, Университет ИТМО. 197101, г. Санкт-Петербург, Кронверкский проспект, д. 49, литер А. E-mail: v_rogovoi@itmo.ru

КОРЖУК Виктория Михайловна, кандидат технических наук, доцент факультета Безопасности Информационных Технологий, Университет ИТМО. 197101, г. Санкт-Петербург, Кронверкский проспект, д. 49, литер А. E-mail: vmkorzhuk@itmo.ru

АЛЕКСАНДРОВ Дмитрий Сергеевич, магистр факультета Безопасности Информационных Технологий, Университет ИТМО. 197101, г. Санкт-Петербург, Кронверкский проспект, д. 49, литер А. E-mail: aleksandrov_ds@itmo.ru

ROGOVOI Vitalii, post-graduate student of Faculty of Secure Information Technologies, ITMO University. 197101, St. Petersburg, Kronverksky Prospect, 49, letter A. E-mail: v_rogovoi@itmo.ru

KORZHUK Victoria Mikhailovna, PhD in Engineering, Associate Professor, Faculty of Secure Information Technologies, ITMO University. 197101, St. Petersburg, Kronverksky Prospect, 49, letter A. E-mail: vmkorzhuk@itmo.ru

ALEKSANDROV Dmitriy Sergeevich, master's student of Faculty of Secure Information Technologies, ITMO University. 197101, St. Petersburg, Kronverksky Prospect, 49, letter A. E-mail: aleksandrov_ds@itmo.ru

УДК 004.056.5 Вестник УрФО № 2(56) / 2025, с. 16–29

Олифиренко А. А.

DOI: 10.14529/secur250202

ОЦЕНКА ВЛИЯНИЯ DATA POISONING-АТАК НА КАЧЕСТВО МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ В PRODUCTION-СРЕДАХ И МЕТОДЫ ИХ ПРЕДОТВРАЩЕНИЯ

В статье проводится комплексное исследование влияния Data Poisoning-атак на качество моделей машинного обучения, функционирующих в production-средах, с целью выявления основных причин ухудшения ключевых метрик, таких как Recall и F1-score, вследствие внедрения вредоносных данных, генерируемых с помощью генеративно-состязательных сетей (GAN).

В экспериментальной части работы на основе синтетического набора данных смоделированы атаки с последующим сравнительным анализом исходной, отравленной и защищённой версий модели, что позволило детально оценить изменения точности, полноты и сбалансированности предсказаний.

На основе полученных результатов предлагается комплексный алгоритм защиты, включающий предварительную фильтрацию данных с использованием алгоритма Isolation Forest и аугментацию обучающего набора посредством генерации синтетических примеров на основе нормального распределения, что способствует восстановлению исходных характеристик модели.

Дополнительно осуществляется непрерывный мониторинг дрейфа входных данных с применением метрик Population Stability Index и расстояния Хеллингера, что позволяет своевременно корректировать работу модели и формировать практические рекомендации по защите моделей машинного обучения в условиях динамичной production-среды.

Ключевые слова: *Data Poisoning, машинное обучение, production-среда, Isolation Forest, аугментация данных, мониторинг дрейфа, защита моделей.*

ASSESSMENT OF THE IMPACT OF DATA POISONING ATTACKS ON THE QUALITY OF MACHINE LEARNING MODELS IN PRODUCTION ENVIRONMENTS AND METHODS FOR THEIR PREVENTION

The article presents a comprehensive study is conducted on the impact of Data Poisoning attacks on the performance of machine learning models operating in production environments, aiming to identify the main causes of deterioration in key metrics such as Recall and F1-score resulting from the injection of malicious data generated by Generative Adversarial Networks (GANs).

In the experimental section, attacks were simulated using a synthetic dataset, and a comparative analysis of the original, poisoned, and protected versions of the model was performed, which allowed for a detailed evaluation of changes in accuracy, completeness, and prediction balance.

Based on the obtained results, an integrated defense algorithm is proposed that includes preliminary data filtering using the Isolation Forest algorithm and data augmentation through the generation of synthetic examples based on the normal distribution, contributing to the restoration of the model's original characteristics.

Additionally, continuous monitoring of input data drift is carried out using metrics such as the Population Stability Index and Hellinger distance, which enables timely adjustments to the model's performance and the formulation of practical recommendations for protecting machine learning models in dynamic production environments.

Keywords: *Data Poisoning, machine learning, production environment, Isolation Forest, data augmentation, drift monitoring, model protection.*

Введение

В современных условиях информационного общества методы машинного обучения (далее – ML) приобретают стратегическое значение и находят применение в критически важных секторах, таких как финансы, кибербезопасность, здравоохранение и государственное управление [1]. Применение ML позволяет автоматизировать обработку больших массивов данных, проводить глубокий аналитический разбор и выявлять скрытые закономерности [2]. Учитывая нарастающую цифровизацию и взаимосвязанность информационных систем, вопросы надежности и устойчивости алгоритмов становятся предметом пристального внимания исследователей [3].

Особое место в изучении безопасности ML-систем занимает проблема внедрения искажающих данных в обучающие выборки, обозначаемая термином Data Poisoning (например, УБИ.221: Угроза модификации модели машинного обучения путем искажения («отравления») обучающих данных) [4]. Сущность данного явления заключается в возможности целенаправленного внесения незначительных, но критически важных модификаций в исходный набор данных, что приводит к систематическим изменениям в поведении моделей [5]. При этом подобная такти-

ка воздействия может оставаться незамеченной при стандартном мониторинге, однако существенно ухудшать показатели точности, полноты и согласованности прогнозов [6].

Цель работы заключается в оценке влияния Data Poisoning-атак на качество моделей машинного обучения и оценке комплексных методов уменьшения негативных последствий таких воздействий [7].

Для реализации поставленных задач планируется создание экспериментальной среды, имитирующей реальные условия эксплуатации информационных систем в финансовом секторе с обработкой потоковых транзакционных данных [8].

Проведение атаки Data Poisoning с использованием GAN: эксперимент

Разработка экспериментальной методологии базируется на использовании синтетического набора данных Paysim [9] на платформе Kaggle, который предназначен для моделирования реальных финансовых транзакций. Представленный датасет включает детальную информацию о платежных операциях, характеризующихся временной меткой транзакции (Step), типом операции (Type), суммой перевода (Amount), балансами отправителя и получателя до и после проведения операции (OldBalanceOrg, NewBalanceOrig, OldBalanceDest, NewBalanceDest) и бинарным индикатором мошеннической активности (IsFraud). Особое внимание уделено тому, что данный набор отражает подлинные паттерны мошеннических действий, что обуславливает его репрезентативность для оценки влияния атак Data Poisoning. Наличие значительного дисбаланса классов, при котором мошеннические транзакции составляют менее 0,1% от общего числа записей, добавляет сложности в задачу классификации и способствует повышенной вероятности ошибок типа False Negative (FN) [10].

Разработка модели (и развертывание на Kaggle) для детекции мошеннических операций осуществлялась посредством построения многослойной перцептронной нейронной сети (MLP) [11]. Архитектура модели включает входной слой, размерность которого соответствует числу признаков, два скрытых полносвязных слоя с функцией активации ReLU, а также выходной слой, использующий сигмоидную функцию для предсказания вероятности мошенничества. Формальное представление модели записывается в виде:

$$\hat{y} = \sigma(W_2 f(W_1 X + b_1) + b_2)$$

X представляет входной вектор признаков, W_1 и W_2 – матрицы весов скрытого и выходного слоев соответственно, b_1 и b_2 – векторы смещений, $f(\cdot)$ – функция ReLU, а $\sigma(\cdot)$ – сигмоидная функция активации.

Для обучения модели применялся оптимизатор Adam совместно с функцией потерь binary crossentropy, что обусловлено бинарной природой задачи. Дополнительное использование метрики F1-score обосновано необходимостью учета баланса между Precision и Recall в условиях выраженного дисбаланса классов.

С целью моделирования атаки Data Poisoning в эксперимент включена генерация синтетических «отравленных» данных посредством использования генеративно-состязательной сети (GAN) [12]. Архитектурная схема GAN состоит из двух ключевых компонентов: генератора и дискриминатора [13]. Генератор, принимающий на вход случайный шум z~N(0,1), формирует искусственные записи, воспроизводящие характеристики реальных транзакций [14], что формализуется следующим уравнением:

$$G(z) = ReLU(W_g z + b_g)$$

Дискриминатор, в свою очередь, оценивает вероятность того, что подаваемые на вход данные являются подлинными, и его функциональное представление записывается как:

$$D(x) = \sigma(W_d x + b_d)$$

Процесс обучения GAN реализуется посредством стандартной схемы minimax-оптимизации, выраженной следующим образом:

$$min_G max_D E_{x \sim P_{data}} [logD(x)] + \\ + E_{z \sim p_z} [log(1 - D(G(z)))]$$

В результате обучения генератора была получена совокупность из 1000 синтетических транзакций, которые последовательно интегрировались в исходный обучающий набор для имитации атаки Data Poisoning.

Переобучение модели MLP с использованием расширенного обучающего набора, включающего отравленные данные, позволило провести сравнительный анализ влияния

атакующих записей на качество работы классификатора. В процессе оценки использовались следующие метрики:

• Accuracy – доля правильных предсказаний, вычисляемая по формуле:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

TP (True Positives) обозначают корректно классифицированные мошеннические транзакции, TN (True Negatives) – верно определенные легитимные операции, FP (False Positives) – ошибочно отнесенные к мошенническим легитимные транзакции, а FN (False Negatives) – мошеннические операции, не выявленные моделью.

• Precision – точность предсказаний мошеннических транзакций, определяемая как:

$$Precision = \frac{TP}{TP + FP}$$

• Recall – полнота детекции мошенничества, выраженная формулой:

$$Recall = \frac{TP}{TP + FN}$$

• F1-score – гармоническое среднее значений precision и recall, вычисляемое по следующей формуле:

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

Ключевым инструментом визуализации влияния атаки стала матрица ошибок (confusion matrix), представленная следующим образом:

$$\begin{matrix} TN & FP \\ FN & TP \end{matrix}$$

Анализ эксперимента по атаке Data Poisoning

Экспериментальные результаты демонстрируют, что атака Data Poisoning привела к заметному снижению эффективности классификатора при выявлении мошеннических транзакций. Сравнительные показатели исходной и отравленной моделей, приведённые в Таблице 1, отражают негативное влияние сгенерированных вредоносных данных на метрики качества.

Данные свидетельствуют, что после интеграции «отравленных» транзакций значение Recall для мошеннического класса (метка 1) сократилось с 0.63 до 0.61, указывая на ослабление способности модели выявлять аномальные операции. При этом F1-score для мошеннических транзакций снизился с 0.76 до 0.75, что подтверждает общее ухудшение результатов классификации. Стоит подчеркнуть, что значение Precision для класса мошенничества увеличилось с 0.96 до 0.99, однако данное улучшение не компенсирует рост количества пропущенных мошеннических операций и отражает смещение модели в сторону более консервативной идентификации аномалий.

На Рис. 1 представлен график изменения функции потерь (Train Loss и Validation Loss) для отравленной модели на протяжении нескольких эпох обучения. Наблюдается стремительное уменьшение Train Loss на первых итерациях, что может указывать на быструю подстройку модели к новым (в том числе вредоносным) данным. При этом валидировочная ошибка (Validation Loss) также уменьшается, однако не столь резко, что может сигнализировать о начале процесса переобучения. Быстрое падение Train Loss зачастую связано с тем, что сеть «запоминает» специфические паттерны в отравленном наборе, вследствие чего теряется обобщающая способность при детекции мошенничества.

Отсутствие существенного расхождения между Train Loss и Validation Loss на заключительных этапах может привести к ложному впечатлению стабильности обучения. Фактически, модель концентрируется на локальных особенностях «отравленных» данных, а не совершенствует способность распознавать аномалии. Подобный эффект снижает

Таблица 1

Результаты атаки Data Poisoning на модель ML

Модель	Precision (0)	Precision (1)	Recall (0)	Recall (1)	F1-score (0)	F1-score (1)
Исходная модель	1.00	0.96	1.00	0.63	1.00	0.76
Отравленная модель	1.00	0.99	1.00	0.61	1.00	0.75

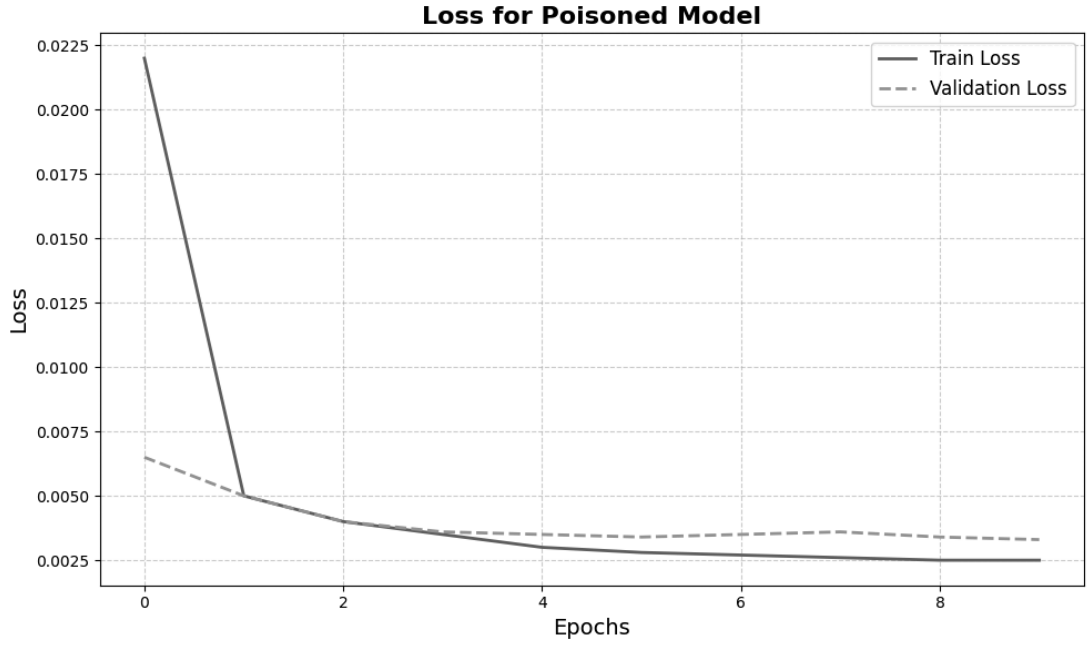


Рис. 1. График функции потерь (train_loss и val_loss) для отравленной модели

эффективность классификатора в условиях реальных атак, когда характер вредоносных транзакций может изменяться динамически.

Результаты анализа матрицы ошибок (Confusion Matrix) свидетельствуют о том, что атака повысила число ложноотрицательных предсказаний (FN). Модель чаще относит мошеннические транзакции к легитимным, что непосредственно увеличивает риск пропуска аномальных операций.

Снижение Recall на 2% в эксперименте указывает на увеличение числа, что критически важно для систем, где высокая полнота детекции мошенничества (Recall) является приоритетной задачей. Даже незначительное снижение данного показателя способно привести к ощутимому росту финансовых рисков в реальной среде.

Снижение Recall на 2% в эксперименте указывает на увеличение числа, что критически важно для систем, где высокая полнота детекции мошенничества (Recall) является приоритетной задачей. Даже незначительное снижение данного показателя способно привести к ощутимому росту финансовых рисков в реальной среде.

В ходе эксперимента выявлено несколько основных факторов, объясняющих успешность Data Poisoning.

Применение GAN обеспечило формирование транзакций, обладающих высокой степенью статистического сходства с реальными записями. Генерируемые образцы не вызывали подозрений при стандартном анализе данных и органично включались в обучающую выборку.

Отсутствие специализированных процедур предобработки, ориентированных на выявление аномальных или потенциально вредоносных записей, позволило фальсифицированным транзакциям беспрепятственно попасть в процесс обучения.

После интеграции отравленных данных модель быстро адаптировалась к ним, но не приобрела дополнительных механизмов для детекции мошеннических схем. Итоговая сеть формально демонстрировала высокую точность (Accuracy), однако пропускала большее число истинных аномалий.

Использование GAN для организации атаки показало, что традиционные подходы к обучению нейросетевых моделей могут оказаться уязвимыми при отсутствии дополнительных мер защиты. Генератор способен формировать правдоподобные транзакции, которые затруднительно отличить от легитимных примеров, в результате чего модель теряет устойчивость к аномалиям. Несмотря на высокое значение Accuracy, выявилась деградация Recall и F1-score для класса мошенничества, что подчёркивает уязвимость в условиях реальной эксплуатации.

Полученные результаты подтверждают необходимость применения основных методов противодействия Data Poisoning – фильтрация данных перед обучением; дрейф модели и адаптивное обучение; обновление модели и защита через аугментацию данных.

Защита моделей машинного обучения: фильтрация данных перед обучением

Предварительная фильтрация данных рассматривается как один из базовых методов противодействия атакам Data Poisoning, поскольку даёт возможность исключить из обучающего набора потенциально вредоносные записи. В ходе эксперимента использовался алгоритм Isolation Forest (IF), основанный на концепции поиска аномалий путём изоляции отдельных точек в признаковом пространстве. Данный подход отличается от классических методов плотности или кластеризации тем, что изначально ориентирован на выявление выбросов через последовательное разбиение данных по осям признакового пространства [15].

Теоретические основы IF заключаются в построении нескольких изолирующих деревьев (isolation trees), где на каждом шаге выбирается случайный признак и случайная граница разбиения. Пусть задано множество X, содержащее n объектов x_i, каждый из которых описан d признаками. Алгоритм формирует T деревьев решений, случайно разделяя пространство признаков. Глубина пути $h(x)$, необходимая для изоляции конкретного объекта x, служит индикатором его аномальности [16]. Математическая формулировка оценки аномальности представлена функцией:

$$s(x) = 2\frac{E(h(x))}{c(n)}$$

$E(h(x))$ – математическое ожидание глубины изоляционного дерева, а $c(n)$ – нормировочный коэффициент, зависящий от общего числа объектов n. Чем ближе $s(x)$ к 1, тем выше вероятность, что точка является выбросом.

На практике в эксперименте применялся IF с параметром contamination = 0.002, указывающим, что алгоритм предполагает около 0.2% аномальных точек в наборе. Результаты показали, что IF выявил и исключил 8 910 записей, потенциально относящихся к «отравленным» данным. Удаление этих записей позволило снизить долю искажённых примеров и повысить качество итоговой модели.

Для оценки эффективности фильтрации рассмотрим динамику изменения ключевых метрик классификации. Recall для мошеннических операций (метка 1) увеличился с 0.61 до 0.66, что указывает на возросшую способность модели распознавать аномальные транзакции. F1-score вырос на 3.2%, отражая общее улучшение баланса между точностью и полнотой предсказаний. Подробные показатели приведены в Таблице 2 и Рисунке 2:

Несмотря на доказанную результативность IF в контексте обнаружения «отравленных» данных, следует учитывать несколько существенных ограничений:

1. Ключевой параметр должен быть тщательно откалиброван. Завышенное значение может привести к чрезмерному удалению полезных примеров, в то время как заниженное оставит в обучающем наборе значительную часть вредоносных записей.

2. В случае, когда злоумышленники генерируют вредоносные транзакции, которые статистически неотличимы от нормальных, IF может не выявить данные аномалии, так как алгоритм ориентирован на поиск нетипичных паттернов.

3. IF анализирует объекты в основном с позиций их «изолированности» в пространстве признаков, не всегда учитывая сложные взаимосвязи между транзакциями (например, временные зависимости или последовательные закономерности).

Таблица 2

Результаты по применению фильтрации

Модель	Precision (0)	Precision (1)	Recall (0)	Recall (1)	F1-score (0)	F1-score (1)
Оригинальная	0.995	0.9801	0.9999	0.6069	0.9997	0.7496
Фильтрованная	0.995	0.9749	0.9999	0.6660	0.9997	0.7724

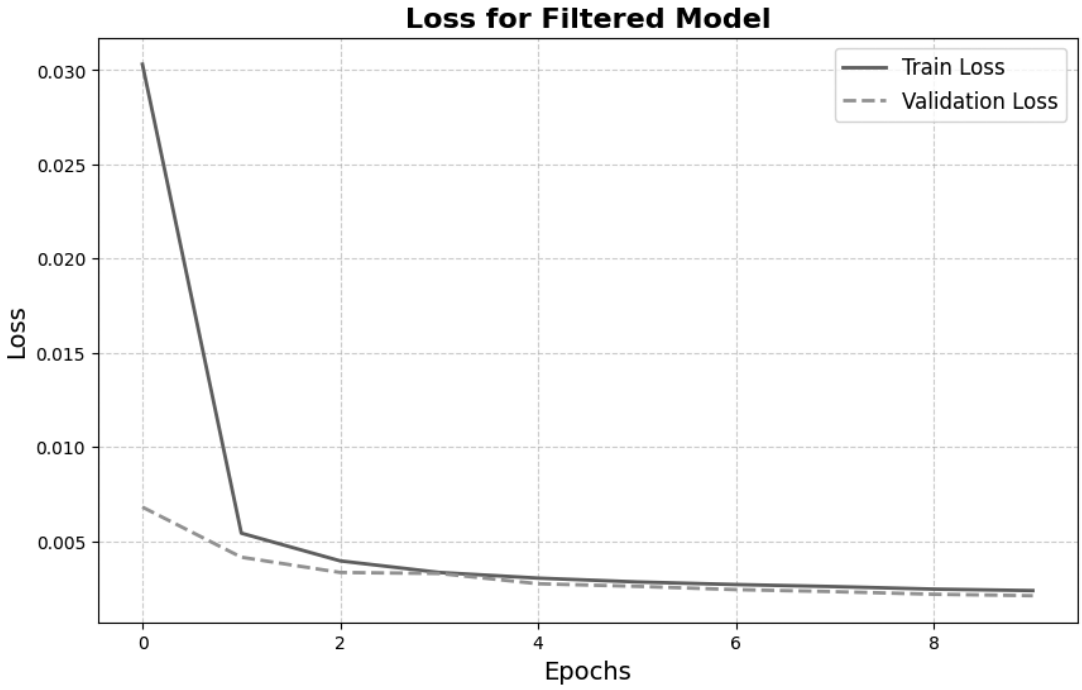


Рис. 2. График потерь для фильтрованной модели

Указанные факторы следует принимать во внимание при использовании IF в реальных производственных условиях. Комбинация с другими методами детекции выбросов или дополнительными источниками валидации способна повысить надёжность фильтрации.

Защита моделей машинного обучения: дрейф модели и адаптивное обучение

Современные алгоритмы машинного обучения, развернутые в продакшн-средах, неизбежно сталкиваются с изменениями во входных данных. Эти изменения, часто именуемые дрейфом модели (model drift), приводят к тому, что статистические свойства признаков и взаимосвязи между ними трансформируются. Подобная динамика способна существенно снизить эффективность предсказательных моделей, особенно в задачах, связанных с детекцией мошенничества, где высока стоимость как ложноположительных (FP), так и ложноотрицательных (FN) решений. Помимо естественных изменений во времени, дрейф может быть спровоцирован злоумышленниками посредством целенаправленных атак, включающих Data Poisoning. Для противодействия указанным факторам необходимо внедрять механизмы мониторинга и адаптивного обучения, позволяющие своевременно реагировать на сдвиги в данных и поддерживать высокое качество классификации [17].

Дрейф модели подразделяется на два ключевых вида. Во-первых, дрейф данных (data drift), при котором статистические характеристики входных признаков эволюционируют без изменения логики целевой переменной. Во-вторых, дрейф концепции (concept drift), отражающий модификацию самой зависимости между признаками и целевой переменной. В контексте детекции мошенничества дрейф концепции представляет особую угрозу, так как мошенники могут целенаправленно менять поведение транзакций, делая их схожими с легитимными и тем самым усложняя задачу классификатора [18].

Если модель, обученная на устаревших данных, продолжает функционировать без обновлений, её способность корректно распознавать новые паттерны может существенно снижаться. Это влечёт за собой рост числа FN, когда мошеннические операции остаются незамеченными, и FP, когда безвредные транзакции ошибочно классифицируются как аномалии. Подобное ухудшение метрик критично для финансовых систем, где каждая ошибка способна повлечь значительные убытки или компрометацию безопасности.

Чтобы предотвратить деградацию качества классификации, в продакшн-среде необходимо реализовывать комплекс мер по мониторингу состояния модели и входных данных. Одной из ключевых задач становится непрерывное отслеживание метрик, таких как Recall, Precision и F1-score, а также оценка распределения признаков.

Анализ временных рядов значений Recall, Precision и F1-score даёт представление о динамике производительности. Резкие падения Recall могут служить сигналом целенаправленной атаки или существенного изменения статистики данных. Аналогично, рост количества FP и FN указывает на возможное ухудшение способности модели различать мошеннические и легитимные операции.

Регулярная сверка предсказаний с истинными значениями, получаемыми от экспертов или последующей валидации, позволяет оперативно выявлять возросшую частоту ошибок. Дополнительно применяются показатели вроде Population Stability Index (PSI) и Kullback–Leibler Divergence (KLD), которые оценивают степень изменения распределений признаков.

Для оценки статистических изменений вычисляются среднее значение, дисперсия и медиана признаков, сопоставляемые с историческими данными. Дрейф можно формализовать при помощи расстояния Хеллингера:

$$H(P,Q) = \sqrt{1 - \sum_i \sqrt{P_i Q_i}}$$

P_i и Q_i представляют собой вероятностные распределения входных данных в различные временные периоды.

При увеличении $H(P, Q)$ возникает подозрение, что структура данных существенно изменилась и модель требует адаптации. Методы обнаружения аномалий (включая Isolation Forest) могут дополнять данный анализ, выявляя нетипичные всплески или сдвиги в признаковом пространстве.

После фиксации дрейфа возникает необходимость в корректировке модели с учётом новых данных. Один из подходов – адаптивное обучение, предполагающее регулярное обновление параметров классификатора. Это может осуществляться по схеме онлайн-обучения, при которой поступающие транзакции сразу же влияют на веса модели, или же посредством периодического переобучения, когда модель обучается заново через определённые интервалы времени на наиболее актуальной выборке.

Математически процесс обновления параметров θ в ходе обучения можно представить как уменьшение обобщённой функции потерь:

$$L(\theta) = \sum_{i=1}^{N} \zeta(f_\theta(x_i), y_i) + \lambda R(\theta)$$

$\zeta(\cdot)$ – функция потерь, $R(\theta)$ – регуляризационный член, а λ – коэффициент, регулирующий степень штрафа за усложнение модели.

Использование ансамблевых методов (например, бустинга или стэкинга) способствует более гибкому учёту различных типов дрейфа, поскольку несколько моделей способны лучше охватывать разнообразные паттерны, формирующиеся в потоке данных.

Эксперимент, включавший мониторинг ключевой метрики Recall, показал (таблица 3), что после предварительной очистки данных и внедрения адаптивного обучения показатели модели сохраняют стабильность, несмотря на изменения, вызванные как естественным дрейфом, так и атаками Data Poisoning. В частности, показатель Recall для класса мошеннических транзакций (метка 1) увеличился с 0.606982 до 0.671458, что соответствует относительному приросту в 10.62%. Представленная ниже таблица демонстрирует полное сравнение метрик исходной модели и модели, адаптированной посредством обновления параметров.

Анализ таблицы демонстрирует, что, несмотря на небольшие изменения в метриках для класса легитимных транзакций (метка 0), основное улучшение наблюдается для класса мошеннических операций (метка 1). Увеличение Recall в результате адаптации параметров модели свидетельствует о повышении полноты детекции, что является критически важным в задачах предотвращения мошенничества. При этом показатели Accuracy, Precision и F1-score остаются на высоком уровне, что указывает на сохранение общей эффективности модели.

Сохранение высокой полноты детекции мошеннических транзакций в сочетании с умеренным количеством ложноположительных срабатываний подтверждает эффективность предложенного подхода к адаптивному обучению.

Таблица 3

Результаты по адаптивному обучению

Метрика	Исходная модель	Модель после атаки
Accuracy	0.999483	0.999553
Precision (0)	0.999498	0.999581
Precision (1)	0.980106	0.968028
Recall (0)	0.999984	0.999972
Recall (1)	0.606982	0.671458
F1-scrore (0)	0.999741	0.999776
F-1score (1)	0.749683	0.792919

Защита моделей машинного обучения: обновление модели и защита через аугментацию данных

Регулярное обновление модели в сочетании с методами аугментации данных представляют собой один из основных подходов к повышению устойчивости системы в условиях динамичного изменения входных данных и целенаправленных атак. Применение этих методов способствует адаптации модели к новым паттернам и уменьшении систематических искажений, возникающих в результате атак.

Необходимость обновления модели продиктована тем, что распределение данных, используемых при обучении, может со временем изменяться, что приводит к явлению дрейфа данных [19]. Формальное выражение данной проблемы имеет вид:

$$P(X,Y) \neq P^{'}(X,Y)$$

$P(X,Y)$ характеризует распределение обучающих данных, а P' $(X,Y$ – распределение данных в реальной эксплуатации.

При возникновении такой диспропорции модель утрачивает свою актуальность, что негативно сказывается на точности предсказаний [20]. Стратегии обновления модели включают итеративное переобучение на актуальных данных, смешивание новых данных с историческими для сохранения контекста, использование ансамблевых методов, таких как Stacking и Bagging, а также взвешивание данных в зависимости от их актуальности [21]. Такой многоаспектный подход позволяет своевременно корректировать модель, поддерживая её высокую предсказательную способность.

Аугментация данных представляет собой процесс искусственного расширения обучающего множества посредством модификации существующих образцов или генерации новых синтетических примеров. В контексте защиты от Data Poisoning атак аугментация служит для разбавления отравленных данных, создания синтетических образцов, имитирующих реальные транзакции, и повышения общей устойчивости модели к аномалиям. Математическая формулировка данного процесса выглядит следующим образом:

$$X^{'} = X + \epsilon, \epsilon \sim N(0, \sigma^2)$$

X' представляет аугментированные данные, а ϵ – случайный шум, генерируемый по нормальному распределению.

В эксперименте для генерации новых примеров использовалась модель:

$$X_{aug} \sim N(\mu_X, \sigma_X)$$

μ_X и σ_X обозначают среднее значение и стандартное отклонение оригинального обучающего множества соответственно. Применение аугментации позволило не только увеличить объём обучающих данных, но и повысить стабильность модели за счёт улучшения её обобщающих способностей. Экспериментальные результаты, полученные после внедрения обновления модели и защиты через аугментацию, представлены в таблице 4:

Анализ демонстрирует, что атака Data Poisoning оказала влияние на метрики, особенно для класса мошеннических транзакций (метка 1). В результате вмешательства наблюдалось снижение Precision и незначительное

Таблица 4

Результаты по обновлению и аугментации

Метрика	Обновленная модель	Аугментированная модель
Accuracy	0.999478	0.999508
Precision (0)	0.999498	0.999549
Precision (1)	0.974917	0.952237
Recall (0)	0.999980	0.999959
Recall (1)	0.606571	0.646817
F1-scrore (0)	0.999739	0.999754
F-1score (1)	0.747848	0.770360

снижение Recall. Реализация стратегии обновления модели позволила частично восстановить первоначальные показатели, а внедрение аугментации данных способствовало повышению стабильности обучения (рисунок 3).

Применение регулярного обновления модели в сочетании с аугментацией данных создаёт многоуровневую стратегию защиты, способную эффективно противодействовать негативным воздействиям Data Poisoning-атак.

Комплексная реализация описанных методов позволяет не только поддерживать высокое качество классификации, но и адаптировать систему к изменениям в продакшн-среде, что имеет первостепенное значение для критически важных автоматизированных систем принятия решений.

Результаты

Проведённое исследование демонстрирует значительное влияние атак Data Poisoning на качество моделей машинного обучения, что особенно актуально при использовании сложных атакующих техник, таких как GAN. Из данных таблицы 3 видно, что

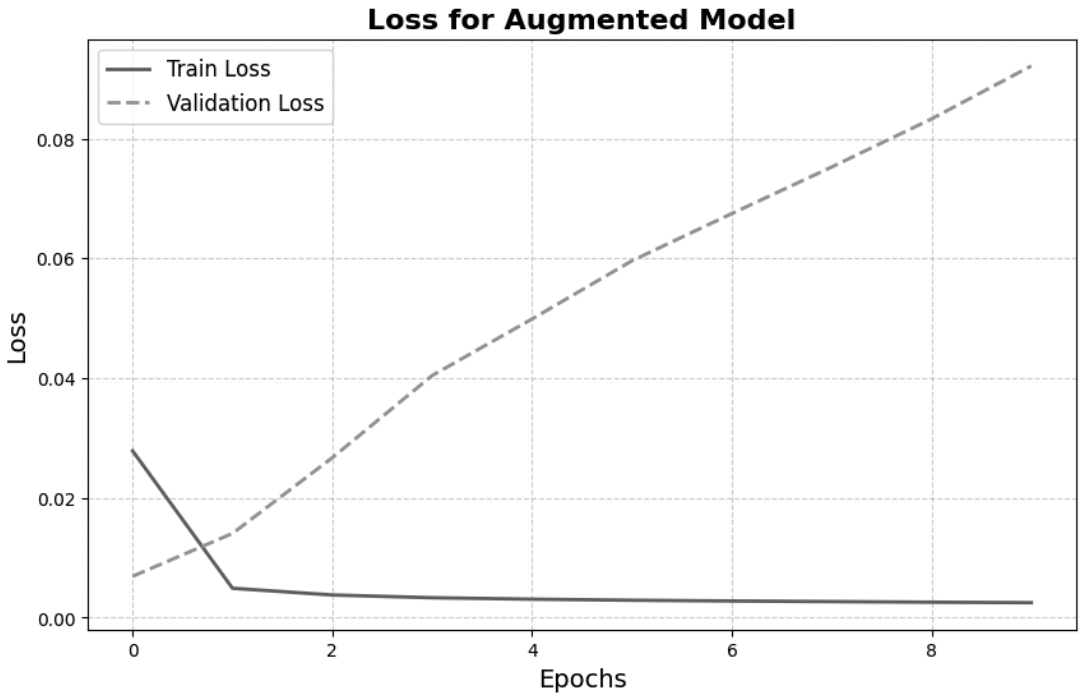


Рис. 3. График потерь для модели с аугментированными данными

атака Data Poisoning привела к увеличению метрики Recall для мошеннического класса с 0.606982 до 0.671458, что свидетельствует о снижении способности модели корректно обнаруживать мошеннические операции вследствие увеличения числа ложноотрицательных предсказаний. Применение метода фильтрации данных с использованием IF позволило восстановить Recall до уровня, близкого к исходному (0.606571), однако наблюдается незначительное снижение F1-score для данного класса (до 0.747848). Аугментация данных, реализованная через генерацию синтетических примеров.

А использование нормального распределения привела к улучшению Recall до 0.646817, что свидетельствует о частичном нивелировании негативного влияния атаки при сохранении высокого уровня точности для легитимного класса.

Для уменьшения негативного воздействия атаки был использован комплексный алгоритм защиты, состоящий из трёх основных этапов. На первом этапе применяется предварительная фильтрация данных с использованием алгоритма Isolation Forest, позволяющая выявить и удалить аномальные записи до начала обучения модели, что приводит к восстановлению исходного значения Recall до примерно 0.606571, несмотря на незначительное снижение F1-score до 0.747848. Второй этап включает аугментацию данных посредством генерации дополнительных синтетических примеров на основе нормального распределения, что позволяет компенсировать потерю корректных записей и повысить устойчивость модели – результатом стало улучшение Recall до 0.646817. Третий этап подразумевает непрерывный мониторинг изменений распределения входных данных с использованием таких метрик, как Population Stability Index (PSI) и расстояние Хеллингера, что позволяет оперативно фиксировать отклонения и предотвращать деградацию характеристик модели.

Графическая визуализация динамики потерь дополнительно иллюстрирует характер обучения моделей. Рисунок 2, демонстрирующий график потерь для фильтрованной модели, показывает синхронное снижение ошибок на тренировочной и валидационной выборках, что указывает на стабильную сходимость и отсутствие выраженного переобучения. Рисунок 3, отображающий график потерь для модели с аугментированными данными, демонстрирует увеличение ошибки на валидационной выборке, что отражает сложность адаптации модели к новым данным при сохранении обобщающих способностей.

В результате применение данного алгоритма позволило снизить долю ложноотрицательных предсказаний с 0.671 до 0.646 и улучшить F1-score для мошеннического класса с 0.747 до 0.770.

Заключение

Проведённое исследование подтверждает, что Data Poisoning-атаки представляют серьёзную угрозу для эффективности моделей машинного обучения в production-средах, поскольку внедрение вредоносных данных кардинально изменяет ключевые метрики, такие как Recall и F1-score, что ведёт к увеличению числа ложноотрицательных предсказаний и снижению общей устойчивости модели. Комплексный алгоритм защиты, состоящий из предварительной фильтрации с помощью Isolation Forest, аугментации данных посредством генерации синтетических примеров на основе нормального распределения и мониторинга дрейфа с использованием метрик PSI и расстояния Хеллингера, доказал свою эффективность в уменьшении негативного воздействия атак. Такой интегрированный подход позволяет не только корректировать распределение входных данных, но и предотвращать деградацию модели при изменении условий эксплуатации, что является критически важным для обеспечения безопасности и надёжности ML-систем, правда в весьма ограниченном периметре.

Литература

1. Aljanabi M., Hamza A., Mijwil M.M., Abotaleb M., El-kenawy E.M., Mohammed S.Y., Ibrahim A. Data Poisoning: Issues, Challenges, and Needs // 7th IET Smart Cities Symposium (SCS 2023). 2024. https://doi.org/10.1049/icp.2024.0951 (дата обращения: 26.02.2025).

2. Li, Y.; Jiang, Y.; Li, Z.; Xia, S.T. Backdoor learning: A survey. IEEE Trans. Neural Networks Learn. Syst. 2022, 35, 5–22.

3. Fan, J.; Yan, Q.; Li, M.; Qu, G.; Xiao, Y. A survey on data poisoning attacks and defenses. In Proceedings of the 2022 7th IEEE International Conference on Data Science in Cyberspace (DSC), Guilin, China, 11–13 July 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 48–55.

4. УБИ.221: Угроза модификации модели машинного обучения путем искажения («отравления») обучающих данных / Банк данных угроз безопасности информации ФСТЭК России и ФАУ «ГНИИИ ПТЗИ ФСТЭК России» // URL: https://bdu.fstec.ru/threat/ubi.221 (дата обращения: 26.02.2025).

5. Намиот Д. Е. Введение в атаки отравлением на модели машинного обучения // International Journal of Open Information Technologies. 2023. №3. URL: https://cyberleninka.ru/article/n/vvedenie-v-ataki-otravleniem-na-modeli-mashinnogo-obucheniya (дата обращения: 26.02.2025).

6. Намиот Д. Е. Схемы атак на модели машинного обучения // International Journal of Open Information Technologies. 2023. №5. URL: https://cyberleninka.ru/article/n/shemy-atak-na-modeli-mashinnogo-obucheniya (дата обращения: 26.02.2025).

7. Maramreddy Y. R. & Muppavaram K. Detecting and Mitigating Data Poisoning Attacks in Machine Learning: A Weighted Average Approach. Engineering, Technology & Applied Science Research, 14(4), 2024, pp. 15505–15509. URL: https://www.researchgate.net/publication/382857536_Detecting_and_Mitigating_Data_Poisoning_Attacks_in_Machine_Learning_A_Weighted_Average_Approach (дата обращения: 26.02.2025).

8. Costales, R.; Mao, C.; Norwitz, R.; Kim, B.; Yang, J. Live trojan attacks on deep neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 796–797.

9. PaySim / Synthetic Financial Datasets For Fraud Detection // URL: https://www.kaggle.com/datasets/ealaxi/paysim1 (дата обращения: 26.02.2025).

10. Zhang Z., Yang Z., Bian J., Li Y., Zhang Y., Zhao Y., Liu Y. Explainable Data Poison Attacks on Human Emotion Evaluation Systems Based on EEG Signals // IEEE Access. 2023. Vol. 11. Pp. 18134–18147.

11. Rawat, A.; Levacher, K.; Sinn, M. The devil is in the GAN: Backdoor attacks and defenses in deep generative models. In European Symposium on Research in Computer Security; Springer: Berlin/Heidelberg, Germany, 2022; pp. 776–783.

12. Arshad, I.; Qiao, Y.; Lee, B.; Ye, Y. Invisible Encoded Backdoor attack on DNNs using Conditional GAN. In Proceedings of the 2023 IEEE International Conference on Consumer Electronics (ICCE), Berlin, Germany, 2–5 September 2023; pp. 1–5.

13. Psychogyios K., Velivassaki T.-H., Bourou S., Voulkidis A., Skias D., Zahariadis T. GAN-Driven Data Poisoning Attacks and Their Mitigation in Federated Learning Systems // Electronics, 2023, Vol. 12, № 8, Article 1805. DOI: 10.3390/electronics12081805.

14. Zhao Y., Gong X., Lin F. & Chen X. Data Poisoning Attacks and Defenses in Dynamic Crowdsourcing With Online Data Quality Learning / IEEE Transactions on Mobile Computing, vol. 22, №. 5, pp. 2569–2581, May 2023, https://doi.org/10.1109/TMC.2021.3133365. (дата обращения: 26.02.2025).

15. Zhong, H.; Liao, C.; Squicciarini, A.C.; Zhu, S.; Miller, D. Backdoor embedding in convolutional neural network models via invisible perturbation. In Proceedings of the Tenth ACM Conference on Data and Application Security and Privacy, New Orleans, LA, USA, 16–18 March 2020; pp. 97–108.

16. Ganjoo R., Ganjoo M., Patil M. Mitigating Poisoning Attacks in Federated Learning / Innovative Data Communication Technologies and Application, 2022, pp.687-699. URL: https://www.researchgate.net/publication/358822578_Mitigating_Poisoning_Attacks_in_Federated_Learning (дата обращения: 26.02.2025).

17. Visger, Mark A. Garbage In, Garbage Out: Data Poisoning Attacks and Their Legal Implications, in Laura A. Dickinson, and Edward W. Berg (eds), Big Data and Armed Conflict: Legal Issues Above and Below the Armed Conflict Threshold, The Lieber Studies Series (New York, 2024; online edn, Oxford Academic, 14 Dec. 2023), https://doi.org/10.1093/oso/9780197668610.003.0008 (дата обращения: 26.02.2025).

18. Dibaei M., Zheng X., Jiang K., Abbas R., Liu S., Zhang Y., Xiang Y. & Yu S. Attacks and defences on intelligent connected vehicles: a survey. Digit. Commun. Networks, 6, 2020, pp. 399-42. URL: https://www.semanticscholar.org/paper/Attacks-and-defences-on-intelligent-connected-a-Dibaei-Zheng/aae97ee34201666f98167402320b86e9facfa173 (дата обращения: 26.02.2025).

19. Hong, Q.; He, B.; Zhang, Z.; Xiao, P.; Du, S.; Zhang, J. Circuit Design and Application of Discrete Cosine Transform Based on Memristor. IEEE J. Emerg. Sel. Top. Circuits Syst. 2023,13, 502–513.

20. Li S., Wang Y., Zhang X., Jiang Y., Xia S. A Study on Data Poisoning Attacks in Deep Generative Models // Applied Sciences, 2023, Vol. 14, No. 19, Article 8742. DOI: 10.3390/app14198742.

21. Yang Z., Zhang J., Wang W., Li H. Invisible Threats in the Data: A Study on Data Poisoning Attacks in Deep Generative Models // Applied Sciences, 2024, T. 14, №19, Article. 8742. https://doi.org/10.3390/app14198742 (дата обращения: 26.02.2025).

References

1. Aljanabi M., Hamza A., Mijwil M.M., Abotaleb M., El-kenawy E.M., Mohammed S.Y., Ibrahim A. Data Poisoning: Issues, Challenges, and Needs // 7th IET Smart Cities Symposium (SCS 2023). 2024. https://doi.org/10.1049/icp.2024.0951 (data obrashcheniya: 26.02.2025).

2. Li, Y.; Jiang, Y.; Li, Z.; Xia, S.T. Backdoor learning: A survey. IEEE Trans. Neural Networks Learn. Syst. 2022, 35, 5–22.

3. Fan, J.; Yan, Q.; Li, M.; Qu, G.; Xiao, Y. A survey on data poisoning attacks and defenses. In Proceedings of the 2022 7th IEEE International Conference on Data Science in Cyberspace (DSC), Guilin, China, 11–13 July 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 48–55.

4. UBI.221: Ugroza modifikatsii modeli mashinnogo obucheniya putem iskazheniya («otravleniya») obuchayushchikh dannykh // Bank dannykh ugroz bezopasnosti informatsii FSTEK Rossii i FAU «GNIII PTZI FSTEK Rossii». URL: https://bdu.fstec.ru/threat/ubi.221 (data obrashcheniya: 26.02.2025).

5. Namiot D.E. Vvedenie v ataki otravleniem na modeli mashinnogo obucheniya // International Journal of Open Information Technologies, 2023, №3. URL: https://cyberleninka.ru/article/n/vvedenie-v-ataki-otravleniem-na-modeli-mashinnogo-obucheniya (data obrashcheniya: 26.02.2025).

6. Namiot D.E. Skhemy atak na modeli mashinnogo obucheniya // International Journal of Open Information Technologies, 2023, №5. URL: https://cyberleninka.ru/article/n/shemy-atak-na-modeli-mashinnogo-obucheniya (data obrashcheniya: 26.02.2025).

7. Maramreddy Y. R. & Muppavaram K. Detecting and Mitigating Data Poisoning Attacks in Machine Learning: A Weighted Average Approach. Engineering, Technology & Applied Science Research, 14(4), 2024, pp. 15505–15509. URL: https://www.researchgate.net/publication/382857536_Detecting_and_Mitigating_Data_Poisoning_Attacks_in_Machine_Learning_A_Weighted_Average_Approach (data obrashcheniya: 26.02.2025).

8. Costales, R.; Mao, C.; Norwitz, R.; Kim, B.; Yang, J. Live trojan attacks on deep neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 796–797.

9. PaySim / Synthetic Financial Datasets For Fraud Detection // URL: https://www.kaggle.com/datasets/ealaxi/paysim1 (data obrashcheniya: 26.02.2025).

10. Zhang Z., Yang Z., Bian J., Li Y., Zhang Y., Zhao Y., Liu Y. Explainable Data Poison Attacks on Human Emotion Evaluation Systems Based on EEG Signals // IEEE Access. 2023. Vol. 11. Pp. 18134–18147.

11. Rawat, A.; Levacher, K.; Sinn, M. The devil is in the GAN: Backdoor attacks and defenses in deep generative models. In European Symposium on Research in Computer Security; Springer: Berlin/Heidelberg, Germany, 2022; pp. 776–783.

12. Arshad, I.; Qiao, Y.; Lee, B.; Ye, Y. Invisible Encoded Backdoor attack on DNNs using Conditional GAN. In Proceedings of the 2023 IEEE International Conference on Consumer Electronics (ICCE), Berlin, Germany, 2–5 September 2023; pp. 1–5.

13. Psychogyios K., Velivassaki T.-H., Bourou S., Voulkidis A., Skias D., Zahariadis T. GAN-Driven Data Poisoning Attacks and Their Mitigation in Federated Learning Systems // Electronics, 2023, Vol. 12, № 8, Article 1805. DOI: 10.3390/electronics12081805.

14. Zhao Y., Gong X., Lin F. & Chen X. Data Poisoning Attacks and Defenses in Dynamic Crowdsourcing With Online Data Quality Learning / IEEE Transactions on Mobile Computing, vol. 22, №. 5, pp. 2569–2581, May 2023, https://doi.org/10.1109/TMC.2021.3133365. (data obrashcheniya: 26.02.2025).

15. Zhong, H.; Liao, C.; Squicciarini, A.C.; Zhu, S.; Miller, D. Backdoor embedding in convolutional neural network models via invisible perturbation. In Proceedings of the Tenth ACM Conference on Data and Application Security and Privacy, New Orleans, LA, USA, 16–18 March 2020; pp. 97–108.

16. Ganjoo R., Ganjoo M., Patil M. Mitigating Poisoning Attacks in Federated Learning / Innovative Data Communication Technologies and Application, 2022, pp.687-699. URL: https://www.researchgate.net/publication/358822578_Mitigating_Poisoning_Attacks_in_Federated_Learning (data obrashcheniya: 26.02.2025).

17. Visger, Mark A. Garbage In, Garbage Out: Data Poisoning Attacks and Their Legal Implications, in Laura A. Dickinson, and Edward W. Berg (eds), Big Data and Armed Conflict: Legal Issues Above and Below the Armed Conflict Threshold, The Lieber Studies Series (New York, 2024; online edn, Oxford Academic, 14 Dec. 2023), https://doi.org/10.1093/oso/9780197668610.003.0008 (data obrashcheniya: 26.02.2025).

18. Dibaei M., Zheng X., Jiang K., Abbas R., Liu S., Zhang Y., Xiang Y. & Yu S. Attacks and defences on intelligent connected vehicles: a survey. Digit. Commun. Networks, 6, 2020, pp. 399-42. URL: https://www.semanticscholar.org/paper/Attacks-and-defences-on-intelligent-connected-a-Dibaei-Zheng/aae97ee34201666f98167402320b86e9facfa173 (data obrashcheniya: 26.02.2025).

19. Hong, Q.; He, B.; Zhang, Z.; Xiao, P.; Du, S.; Zhang, J. Circuit Design and Application of Discrete Cosine Transform Based on Memristor. IEEE J. Emerg. Sel. Top. Circuits Syst. 2023,13, 502–513.

20. Li S., Wang Y., Zhang X., Jiang Y., Xia S. A Study on Data Poisoning Attacks in Deep Generative Models // Applied Sciences, 2023, Vol. 14, No. 19, Article 8742. DOI: 10.3390/app14198742.

21. Yang Z., Zhang J., Wang W., Li H. Invisible Threats in the Data: A Study on Data Poisoning Attacks in Deep Generative Models // Applied Sciences, 2024, T. 14, №19, Article. 8742. https://doi.org/10.3390/app14198742 (data obrashcheniya: 26.02.2025).

ОЛИФИРЕНКО Артем Алексеевич, мидл Golang разработчик ООО «РеалИТ», магистрант кафедры Информационная безопасность автоматизированных систем федерального государственного бюджетного учреждения высшего образования «Саратовский государственный технический университет им. Юрия Алексеевича Гагарина». 410054, г. Саратов, ул. Политехническая, 77. E-mail: artemolifirenko@yandex

OLIFIRENKO Artem Alekseevich, Middle Golang developer at RealIT LLC, Master's student of the Department of Information Security of Automated Systems of the Federal State Budgetary Institution of Higher Education "Saratov State Technical University named after Yuri Alekseevich Gagarin". 410054, Saratov, Politekhnicheskaya St., 77. E-mail: artemolifirenko@yandex

УДК 004.056.57 Вестник УрФО № 2(56) / 2025, с. 30–39

Серебряков Д. С.

DOI: 10.14529/secur250203

ПРОАКТИВНАЯ ЗАЩИТА КОРПОРАТИВНЫХ ВЕБ-ПРИЛОЖЕНИЙ ОТ УГРОЗ, СВЯЗАННЫХ С КОМПРОМЕТАЦИЕЙ JAVASCRIPT-БИБЛИОТЕК

В статье анализируются угрозы безопасности веб-приложений, возникающие в результате использования сторонних JavaScript-библиотек, загружаемых в браузер с внешних ресурсов. Предлагается архитектура информационной системы, предназначенной для обеспечения интеллектуальной проактивной безопасности на уровне клиентской части корпоративных систем. В качестве основы для реализации системы рассматривается комбинация серверного и браузерного компонентов, обеспечивающих контроль целостности кода, мониторинг поведения исполняемых скриптов и фильтрацию вредоносных зависимостей. Подчеркивается важность централизованного управления безопасностью зависимостей и непрерывного мониторинга, позволяющего своевременно выявлять аномалии, недоступные для традиционных средств защиты. Полученные результаты могут быть использованы при разработке защищенных корпоративных систем на основе архитектуры тонкого клиента.

Ключевые слова: *информационная безопасность, веб-приложения, JavaScript, open source, проактивная защита, клиентская часть, зависимость*

Serebryakov D. S.

PROACTIVE SECURITY OF CORPORATE WEB APPLICATIONS AGAINST JAVASCRIPT LIBRARY COMPROMISE

The article analyzes security threats to web applications caused by the use of third-party JavaScript libraries loaded into the browser from external sources. A system architecture is proposed to ensure intelligent proactive client-side protection of corporate systems. The solution is based on a combination of server-side and browser-side components, which provide code integrity validation, behavior monitoring of executable scripts, and real-time filtering of malicious dependencies. The study emphasizes the importance of centralized dependency management and continuous monitoring for detecting anomalies that bypass traditional protection tools. The proposed results can be applied to the development of secure corporate systems built on thin-client architecture.

Keywords: *information security, web applications, JavaScript, open source, proactive protection, client-side, dependencies*

Современные корпоративные информационные системы используют архитектуру тонкого клиента для создания веб-приложений, в рамках которой обработка данных и взаимодействие с пользователем происходят непосредственно в браузере. Данный подход упрощает управление логикой АРМ пользователей и снижает нагрузку на серверы, однако открывает новые типы угроз, связанные с возможностью манипуляций клиентскими данными и подменой загружаемого контента [1, 2].

Одной из наиболее серьезных угроз является использование скомпрометированных, но формально легитимных JavaScript-библиотек, загружаемых со сторонних ресурсов (сторонние JS-компоненты). Такие компоненты могут выполнять скрытые действия по подмене данных, манипуляции контентом или взаимодействовать с несанкционированными ресурсами. Эти угрозы зачастую проходят сигнатурные проверки, не обнаруживаются антивирусными решениями и отсутствуют в известных базах уязвимостей [3, 4].

Современные веб-приложения используют тысячи сторонних JavaScript-библиотек, которые динамически управляются системой управления зависимостями, такой как npm. Этот механизм автоматической загрузки, обновления и интеграции компонентов из внешних репозиториев значительно упрощает процесс разработки, но в то же время создает потенциально неконтролируемый вектор атак. Разработчики и администраторы зачастую не отслеживают весь спектр загружаемых зависимостей, полагаясь на механизмы пакетных менеджеров, которые автоматически устанавливают и обновляют пакеты из множества источников. В результате уязвимости или преднамеренно внедренные вредоносные изменения в одном из многочисленных пакетов могут незаметно распространиться по всей экосистеме приложения.

Кроме того, масштабное использование автоматизированного управления зависимостями приводит к тому, что в проектах присутствуют сотни и тысячи зависимостей, поступающих из различных источников. В этих условиях традиционные механизмы статического анализа кода и мониторинга целостности компонентов оказываются недостаточно эффективными. Разнообразие поставщиков и версионных цепочек делает практически невозможным выявление скрытых угроз в огромных объемах данных, что усложняет задачу обеспечения безопасности и требует разработки специализированных механизмов защиты (см. Табл. 1).

Таблица 1

Основные угрозы, связанные с компрометацией сторонних JavaScript-библиотек

№	Тип угрозы	Механизм атаки	Последствия для веб-приложений
1	Подмена данных [5, 6]	Модификация DOM через популярные библиотеки (Vue.js, jQuery.js, WPForms.js)	Искажение информации (politware)
2	Перехват запросов [7]	Вставка обработчиков сетевых взаимодействий	Кража данных, перенаправление на фишинговые сайты
3	Изменение интерфейсов [8]	Скрытие или подмена форм, элементов управления	Социальная инженерия, хищение информации
4	Скрытая активность [9]	Фоновые сетевые соединения с внешними серверами	Утечка данных, заражение приложений
5	Компрометация репозиториев [10, 11]	Недостаточный контроль npm, CDN, GitHub	Распространение вредоносных зависимостей

Анализ Банка данных угроз ФСТЭК России подтверждает актуальность обсуждаемой проблемы. УБИ «Доступ к локальным файлам сервера при нарушении целостности скриптов» описывает сценарий, при котором вредоносные или модифицированные JavaScript-библиотеки могут инициировать несанкционированные обращения к внутренним ресурсам, что особенно критично для веб-приложений с архитектурой тонкого клиента [12].

Цель, задачи и методы исследования

Целью работы является определение технических требований к системе защиты веб-приложений корпоративных информационных систем от угроз, связанных с использованием сторонних JavaScript-библиотек. Для достижения цели поставлены задачи:

- выделить требования к архитектуре серверного компонента системы, включая организацию базы знаний, контроль версий зависимостей и механизмов уведомления;
- рассмотреть направления интеграции предложенного подхода с действующими средствами информационной безопасности корпоративной инфраструктуры.

Методологической основой исследования выступают:

- анализ типовых сценариев использования сторонних JavaScript-библиотек в современных веб-приложениях;
- изучение особенностей динамического поведения исполняемого кода в браузере в условиях внедрения внешних компонентов;
- сопоставление методов статического и поведенческого контроля в контексте обеспечения целостности клиентской части веб-приложений.

При прикладной реализации предложенных принципов проактивной защиты, исключающей загрузку и исполнение нежелательных компонентов до момента их воздействия на защищаемую корпоративную информационную систему.

Распространенность угроз и их последствия

Современные веб-приложения широко используют компоненты с открытым исходным кодом. Исследования показывают, что в 96% приложений присутствует открытый код [13], что подтверждает широкое распространение сторонних JS-компонентов [14], в частности было проанализировано более 12 миллионов случаев использования свободного и открытого программного обеспечения (FOSS) в более чем 10 тысячах компаний. Рассмотрим основные угрозы, характерные для атак, связанных с использованием уязвимостей в сторонних JavaScript-библиотеках. Злоумышленники могут загружать вредоносные версии библиотек в открытые репозитории (npm, GitHub), что может приводить к компрометации данных и утечкам информации [14]. Такая атака может привести к подмене данных, внедрению вредоносного кода и утечке конфиденциальной информации.

Системы доставки контента (CDN, Content Delivery Network) применяются в веб-разработке для обеспечения высокой скорости загрузки сторонних JavaScript-библиотек

и снижения нагрузки на основные серверы. Однако использование внешних CDN-источников влечёт за собой значительные риски. В случае компрометации такого узла злоумышленник получает возможность внедрения поддельного или модифицированного кода, который будет исполняться непосредственно в браузере конечного пользователя. Это может привести к перехвату данных, несанкционированному доступу к пользовательским сессиям или подмене интерфейсных элементов [3, 4].

Дополнительную угрозу представляют библиотеки, изначально содержащие вредоносные элементы, реализованные посредством легитимных возможностей языка JavaScript. Такие компоненты способны незаметно изменять структуру DOM (Document Object Model), в частности — скрывать элементы формы, подменять действия по нажатию кнопок или внедрять скрытые переходы. Подобные методы активно используются при атаках социальной инженерии и остаются слабо детектируемыми средствами статического анализа [5].

В ряде случаев вредоносный код встраивается на этапе разработки библиотеки — в pull-запросах, новых релизах или дополнительных зависимостях. Такие модификации могут длительное время не вызывать подозрений у участников сообщества или администраторов репозиториев, и, как следствие, распространяются через популярные пакеты среди тысяч других проектов [6]. Ярким примером таких атак стали инциденты 2021–2022 годов, когда изменения в цепочке поставок ПО позволили злоумышленникам внедрять вредоносный код в широко используемые библиотеки, не прибегая к взлому конечных приложений [7].

Особую известность получил случай в январе 2025 года, когда было выявлено заражение более 500 правительственных и университетских сайтов по всему миру через внедрение вредоносных JavaScript-компонентов. Они создавали скрытые элементы в DOM, обеспечивавшие автоматическую переадресацию пользователей на сторонние ресурсы. Атака не была зафиксирована средствами традиционного мониторинга, что продемонстрировало слабость существующих систем обнаружения и необходимость их совершенствования [15].

В 2025 году была зафиксирована волна атак на сайты, работающие на платформе «1С-Битрикс». Злоумышленники использовали уязвимости в сторонних модулях, таких как esol.importexcel и currency_export_excel, для внедрения вредоносного JavaScript-кода. Это привело к перенаправлению пользователей на фишинговые ресурсы и установке вредоносного ПО. [16]. В 2021 году популярная библиотека UAParser.js, используемая во многих веб-приложениях, была скомпрометирована. Злоумышленники внедрили в неё вредоносный код, который распространялся через официальный репозиторий npm. Это привело к заражению множества систем, включая российские компании, использующие эту библиотеку [17].

Учитывая изложенное, можно заключить, что широкое распространение и сложность отслеживания уязвимых или скомпрометированных JS-библиотек делают их одним из наиболее критичных элементов угроз информационной безопасности веб-приложений.

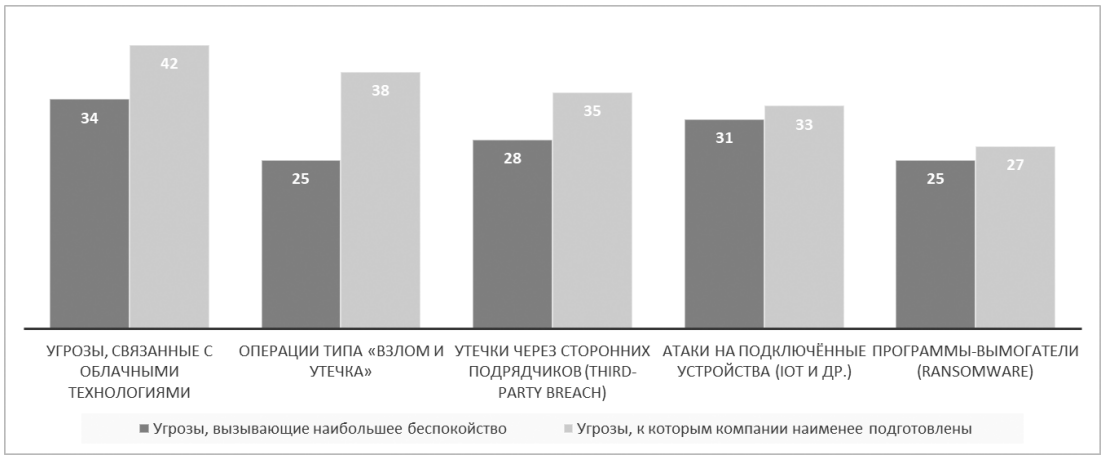


Рис. 1. Разрыв между обеспокоенностью и готовностью к отражению киберугроз - доля респондентов, выбравших угрозу в топ-3 [18]

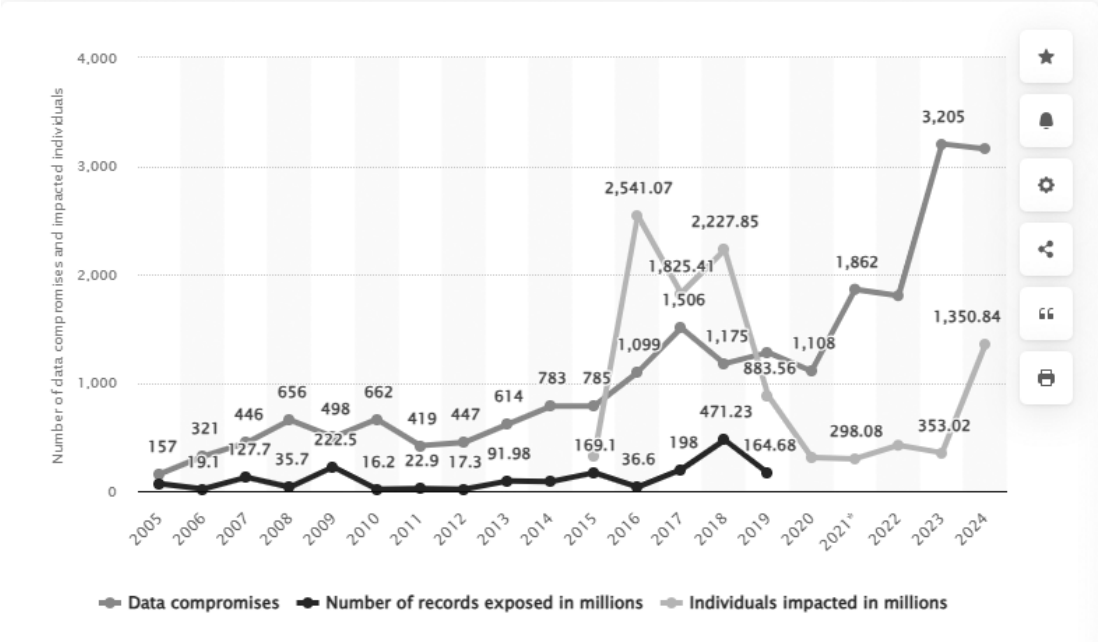


Рис. 2. Ежегодное количество случаев компрометации данных и пострадавших лиц в США с 2005 по 2024 г. [20]

Такое распространение уязвимостей подтверждается результатами международных исследований. Угрозы, связанные с третьесторонними утечками и атаками через внешние зависимости, входят в число наиболее тревожных факторов для организаций, при этом уровень готовности к их отражению остаётся недостаточным (см. Рис. 1).

Также исследование 2020 г. показало, что 37% из 133 тыс. проанализированных вебсайтов использовали по крайней мере одну JavaScript-библиотеку с известной уязвимостью, что свидетельствует о недостаточной защищенности существующих систем перед подобными угрозами [19]. Повышение уровня защищённости требует регулярной верификации используемых зависимостей, внедрения систем контроля целостности и развития механизмов защиты на клиентской стороне.

Последствия от киберинцидентов продолжают увеличиваться (см. Рис 2).

По данным аналитиков, совокупные расходы на обеспечение информационной безопасности в 2024 году достигнут $183,87 млрд, что на 13,4% больше по сравнению с предыдущим годом [21]. Эти затраты сопоставимы с годовыми бюджетами отдельных государств: например, ВВП Новой Зеландии или Греции находится на аналогичном уровне. Такой масштаб подчёркивает не только экономическую значимость вопроса, но и необходимость системного подхода к минимизации уязвимостей в инфраструктуре современных веб-приложений.

Предлагаемые технические требования к технологии

Современные методы обеспечения информационной безопасности (ИБ) часто оказываются недостаточно эффективными против новых векторов атак, связанных с использованием сторонних JavaScript-библиотек [22]. Традиционные системы, такие как файрволы веб-приложений (WAF), в основном нацелены на предотвращение известных угроз, но могут быть не в состоянии обнаружить и предотвратить атаки, возникающие из-за компрометации сторонних библиотек.

Комплексный подход защиты должен включать динамический анализ и мониторинг изменений в репозиториях, а также механизмы безопасности на уровне браузера [8], что отражено в авторском подходе к усилению защиты веб-приложений от уязвимостей в сторонних JavaScript-библиотеках [23, 24]. Основные различия между традиционными подходами и предлагаемым усилением представлены ниже (см. Табл. 2).

Такой подход подразумевает создание информационной системы, предназначенной для проактивного мониторинга действий в браузере, контроля и валидации исполняемого кода и интегрируемой с веб-

Таблица 2

Структура предлагаемого комплексного подхода

№	Ключевой элемент	Ограничения традиционных решений	Элементы предлагаемого подхода
1	Управление загрузкой JavaScript-зависимостей	Отсутствие динамического контроля, ручная проверка	Динамическое профилирование и мониторинг загрузки скриптов в браузере
2	Проверка целостности и изменений кода	Редкий аудит, отсутствие автоматизированного отслеживания	Автоматический мониторинг изменений в open-source библиотеках
3	Контроль сетевых взаимодействий	Защита только на периметре сети (WAF)	Анализ исходящих HTTP-запросов и загружаемого контента в браузере
4	Реагирование на выявленные угрозы	Реакция только после инцидента	Проактивное уведомление о подозрительной активности и динамическая коррекция политик безопасности (CSP)

приложением корпоративной информационной системы, что позволит существенно снизить риски, связанные с упомянутыми ранее векторами атак. Архитектурно система, обеспечивающая интеллектуальную проактивную безопасность веб-приложений, должна сочетать централизованный серверный компонент и комплексный браузерный модуль.

Ключевым элементом предлагаемой архитектуры является расширенный браузерный модуль, обеспечивающий реализацию концепции проактивной клиентской защиты [25]. Одной из его важнейших функций выступает динамический мониторинг загружаемых в браузере скриптов, сопровождающийся проверкой их соответствия политике безопасности, заданной на сервере. Это позволяет в момент загрузки выявлять подозрительные компоненты и блокировать их до начала исполнения.

Контроль целостности реализуется посредством верификации контрольных хешсумм библиотек, что позволяет обнаруживать даже минимальные несанкционированные изменения кода, в том числе незаметные визуально, но опасные по воздействию. Для минимизации последствий потенциального заражения применяется механизм ограничения полномочий сторонних библиотек. Принцип изоляции по минимуму прав предотвращает выполнение операций с чувствительными объектами DOM, утечку данных и вмешательство в ключевые функции веб-приложения.

Браузерный компонент также обеспечивает обратную связь с сервером, передавая информацию о выявленных отклонениях и нестандартных действиях библиотек. Это позволяет обеспечить синхронизацию между клиентской и серверной частями системы, а также оперативное обновление защитных правил. Особое внимание уделено автоматизации процесса обновления: браузерный модуль инициирует применение патчей и загрузку актуальных версий библиотек, в случае если ранее установленные версии были признаны уязвимыми.

Центральный серверный компонент разработанной системы обеспечивает хранение данных и выполняет функции координации, анализа и политики управления безопасностью [26]. В частности, в его задачи входит хранение информации об используемых open-source библиотеках, их версиях и известных уязвимостях. Интеграция с общедоступными базами, такими как CVE, обеспечивает своевременное обновление данных о потенциальных угрозах, организация оперативного реагирования на инциденты. Система уведомлений реализует механизм автоматической передачи сообщений ответственным специалистам при обнаружении признаков угроз — будь то выявленные уязвимости в используемых JavaScript-компонентах или отклонения от их стандартного поведения. Такие уведомления направлены на снижение времени реакции и минимизацию последствий атак.

Важным элементом является настройка и контроль политик безопасности, определяющих перечень разрешённых к использованию библиотек, их версий и источников распространения. Фильтрация доступа к несанкционированным компонентам позволяет на раннем этапе предотвращать загрузку потенциально опасного кода, в том числе скомпрометированных пакетов из внешних CDN или репозиториев. Дополнительно предусмотрена система аналитики, обеспечивающая регулярное формирование отчетов, содержащих информацию о текущем уровне защищенности, динамике обновлений зависимостей и зафиксированных инцидентах. Эти данные используются для оценки уязвимостей и выработки корректирующих мер на уровне организации.

Практическая реализация предлагаемой архитектуры потенциально имеет высокую эффективность в предотвращении атак, связанных с использованием сторонних JavaScript-компонентов. Ключевым преимуществом системы является проактивный характер защиты: аномалии выявляются до начала их эксплуатации, что принципиально отличает подход от реактивных средств контроля. Благодаря модульной структуре и открытым интерфейсам, система легко интегрируется в существующую ИТ-инфраструктуру компании, включая конвейеры DevSecOps, корпоративные SIEM-системы и средства WAF. Это обеспечивает непрерывный мониторинг клиентской стороны без необходимости масштабной перестройки архитектуры.

Гибкость архитектуры позволяет адаптировать её к изменениям в экосистеме, основанной на open-source: обновления библиотек, изменения политик безопасности и появление новых угроз оперативно обрабатываются системой за счёт централизованного управления и синхронизации с базой знаний (уязвимостей и политик). В совокупности предложенный подход позволяет рассматривать браузер не как уязвимую зону, а как полнофункциональную и контролируемую среду исполнения, интегрированную в архитектуру защиты корпоративной информационной системы.

Таким образом, предложенный подход к защите клиентской части веб-приложений от угроз, связанных с компрометацией сторонних JavaScript-библиотек, позволяет выстроить проактивную архитектуру информационной безопасности, минимизирующую риски внедрения вредоносного кода. Разработанные технические меры применимы к корпоративным ИТ-системам с архитектурой тонкого клиента и легко интегрируются в существующие процессы SIEM-мониторинга. Основные ожидаемые эффекты представлены в таблице 3. Следующий этап научного исследования предполагает разработку прототипа и интеграции предложенной системы в реальную ИБ-архитектуру предприятия.

Таблица 3

Ожидаемые эффекты от внедрения проактивной системы защиты

№	Вектор угроз	Эффект от внедрения предлагаемого подхода
1	Подмена данных через модификацию DOM	Блокировка исполнения скриптов, нарушающих контрольные хеши и политику CSP
2	Загрузка модифицированных JS-зависимостей	Автоматическая верификация источника и контрольных сумм; уведомление об отклонениях
3	Перехват и переадресация пользовательских данных	Анализ исходящих HTTP-запросов и блокировка несанкционированных направлений
4	Внедрение вредоносного кода через open-source	Мониторинг репозиториев и сигнатурный анализ изменений в библиотеках
5	Атаки через CDN и внешние ресурсы	Ограничение по источникам, фильтрация недоверенных URL и проверка сетевой активности

Литература

1. CVEfixes: Automated Collection of Vulnerabilities and Their Fixes from Open-Source Software. arXiv. URL: https://arxiv.org/abs/2107.08760 (дата обращения 28.04.2025 г.)

2. Automated Vulnerability Detection in Source Code Using Deep Representation Learning. arXiv. URL: https://arxiv.org/abs/1807.04320 (дата обращения 28.04.2025 г.)

3. 3Proactive Vulnerability Management is a No Brainer for Security, but... JFrog. URL: https://jfrog.com/blog/proactive-vulnerability-management/ (дата обращения 28.04.2025 г.)

4. 13 Open Source Software Security Risks. SentinelOne. URL: https://www.sentinelone.com/cybersecurity-101/cybersecurity/open-source-software-security-risks/ (дата обращения 28.04.2025 г.)

5. MISP Open Source Threat Intelligence Platform & Open Standards. MISP Project. URL: https://www.misp-project.org/ (дата обращения 28.04.2025 г.)

6. Strengthening Open Source Software: Best Practices for Enhanced Security. OpenSSF. URL: https://openssf.org/blog/2023/09/06/strengthening-open-source-software-best-practices-for-enhanced-security/ (дата обращения 28.04.2025 г.)

7. What Are Open Source Vulnerabilities. Sonatype. URL: https://www.sonatype.com/resources/articles/what-are-open-source-vulnerabilities (дата обращения 28.04.2025 г.)

8. Open Source Security and Risk Analysis Report Trends. Black Duck. URL: https://www.blackduck.com/blog/open-source-trends-ossra-report.html (дата обращения 28.04.2025 г.)

9. 10 Free and Open Source Cybersecurity Tools to Know. Lumifi Cyber. URL: https://www.lumificyber.com/blog/free-open-source-software-cybersecurity/ (дата обращения 28.04.2025 г.)

10. Information Security and Protection of Information. MSU. URL: https://hsmi.msu.ru/sites/hsmi.msu.ru/files/program_common_files/po_vyb._-_informacionnaya_bezopasnost_i_zashchita_informacii_0.pdf (дата обращения 28.04.2025 г.)

11. Theoretical Foundations of Information Security. MSU. URL: https://cs.msu.ru/sites/cmc/files/docs/_26_08_teoreticheskie_osnovy_informacionnoy_bezopasnosti.pdf (дата обращения 28.04.2025 г.)

12. Банк данных угроз безопасности информации, URL https://bdu.fstec.ru/threat (дата обращения 28.05.2025 г.)

13. Исследование показало, что 96% современных приложений используют open-source. CNews. URL: https://www.cnews.ru/news/top/2024-12-10_issledovanie_pokazalochto (дата обращения 28.04.2025 г.)

14. В открытом ПО обнаружили 12 млн уязвимостей: компании используют библиотеки, не проверяя их. Habr. URL: https://habr.com/ru/news/865290/ (дата обращения 28.04.2025 г.)

15. На более чем 500 сайтах по всему миру обнаружена новая атака с использованием JavaScript. ITSec.ru. URL: https://www.itsec.ru/news/na-bolee-chem-500-saytah-po-vsemu-miru-obnaruzhena-novaya-ataka-s-ispolzovaniyem-javascript (дата обращения 28.04.2025 г.)

16. Kaspercky Daily. В популярный Javascript-пакет UAParser.js внедрили зловреда URL: https://www.kaspersky.ru/blog/uaparser-js-infected-versions/31787/ (дата обращения 28.05.2025 г.)

17. RushRadio Новая волна вирусов на Битрикс в 2025 году URL: https://rushstudio.by/blog/razrabotchiku/novaya-volna-virusov-na-bitriks-v-2025-godu/ (дата обращения 28.05.2025 г.)

18. Отчёт PwC Global Digital Trust Insights 2025, URL https://www.pwc.com/us/en/services/consulting/cybersecurity-risk-regulatory/library/global-digital-trust-insights.html (дата обращения 28.04.2025 г.)

19. JavaScript Security: A Survey of Attacks and Defenses. arXiv. URL: https://arxiv.org/abs/1811.00918 (дата обращения 28.04.2025 г.)

20. Статистические данные сайта Statista.com URL: https://www.statista.com/statistics/273550/data-breaches-recorded-in-the-united-states-by-number-of-breaches-and-records-exposed/ (дата обращения 28.04.2025 г.)

21. Информационная безопасность (мировой рынок). TAdviser. URL: https://www.tadviser.ru/index.php/Статья:Информационная_безопасность_%28мировой_рынок%29 (дата обращения 28.04.2025 г.)

22. Обзор угроз информационной безопасности за I квартал 2024 года. Positive Technologies. URL: https://www.ptsecurity.com/ru ru/research/analytics/cybersecurity-threatscape-2024-q1/ (дата обращения 28.04.2025 г.)

23. Vasilakis N., Staicu C.-A., Ntousakis G., Kallas K., Karel B., DeHon A., Pradel M. Mir: Automated Quantifiable Privilege Reduction Against Dynamic Library Compromise in JavaScript. arXiv preprint arXiv:2011.00253. URL: https://arxiv.org/abs/2011.00253

24. Nakhaei K., Ansari E., Ansari F. JSSignature: Eliminating Third-Party-Hosted JavaScript Infection Threats Using Digital Signatures. arXiv preprint arXiv:1812.03939. URL: https://arxiv.org/abs/1812.03939

25. Государев И.Б. Основы разработки веб-приложений на платформах Node.js и Deno. СПб: Университет ИТМО, 2023.

26. Хоффман Э. Безопасность веб-приложений. Разведка, защита, нападение. СПб: Питер, 2021.

References

1. CVEfixes. Automated Collection of Vulnerabilities and Their Fixes from Open-Source Software. arXiv. URL: https://arxiv.org/abs/2107.08760 (accessed: 28.04.2025)

2. Automated Vulnerability Detection in Source Code Using Deep Representation Learning. arXiv. URL: https://arxiv.org/abs/1807.04320 (accessed: 28.04.2025)

3. JFrog. Proactive Vulnerability Management is a No Brainer for Security. URL:https://jfrog.com/blog/proactive-vulnerability-management/ (accessed: 28.04.2025)

4. SentinelOne. 13 Open Source Software Security Risks. URL: https://www.sentinelone.com/cybersecurity-101/cybersecurity/open-source-software-security-risks/ (accessed: 28.04.2025)

5. MISP Project. MISP Open Source Threat Intelligence Platform & Open Standards. URL: https://www.misp-project.org/ (accessed: 28.04.2025)

6. OpenSSF. Strengthening Open Source Software: Best Practices for Enhanced Security. URL: https://openssf.org/blog/2023/09/06/strengthening-open-source-software-best-practices-for-enhanced-security/ (accessed: 28.04.2025)

7. Sonatype. What Are Open Source Vulnerabilities. URL: https://www.sonatype.com/resources/articles/what-are-open-source-vulnerabilities (accessed: 28.04.2025)

8. Black Duck. Open Source Security and Risk Analysis Report Trends. URL: https://www.blackduck.com/blog/open-source-trends-ossra-report.html (accessed: 28.04.2025)

9. Lumifi Cyber. 10 Free and Open Source Cybersecurity Tools to Know. URL: https://www.lumificyber.com/blog/free-open-source-software-cybersecurity/ (accessed: 28.04.2025)

10. Moscow State University. Information Security and Protection of Information. URL: https://hsmi.msu.ru/sites/hsmi.msu.ru/files/program_common_files/po_vyb._-_informacionnaya_bezopasnost_i_zashchita_informacii_0.pdf (accessed: 28.04.2025)

11. Moscow State University. Theoretical Foundations of Information Security. URL: https://cs.msu.ru/sites/cmc/files/docs/_26_08_teoreticheskie_osnovy_informacionnoy_bezopasnosti.pdf (accessed: 28.04.2025)

12. Bank dannykh ugroz bezopasnosti informatsii, URL https://bdu.fstec.ru/threat (data obrashcheniya 28.05.2025 g.)

13. Issledovaniye pokazalo, chto 96% sovremennykh prilozheniy ispol'zuyut open-source. CNews. URL: https://www.cnews.ru/news/top/2024-12-10_issledovanie_pokazalochto (data obrashcheniya 28.04.2025 g.)

14. V otkrytom PO obnaruzhili 12 mln uyazvimostey: kompanii ispol'zuyut biblioteki, ne proveryaya ikh. Habr. URL: https://habr.com/ru/news/865290/ (data obrashcheniya 28.04.2025 g.)

15. Na boleye chem 500 saytakh po vsemu miru obnaruzhena novaya ataka s ispol'zovaniyem JavaScript. ITSec.ru. URL: https://www.itsec.ru/news/na-bolee-chem-500-saytah-po-vsemu-miru-obnaruzhena-novaya-ataka-s-ispolzovaniyem-javascript (data obrashcheniya 28.04.2025 g.)

16. Kaspercky Daily. V populyarnyy Javascript-paket UAParser.js vnedrili zlovreda URL: https://www.kaspersky.ru/blog/uaparser-js-infected-versions/31787/ (data obrashcheniya 28.05.2025 g.)

17. RushRadio Novaya volna virusov na Bitriks v 2025 godu URL: https://rushstudio.by/blog/razrabotchiku/novaya-volna-virusov-na-bitriks-v-2025-godu/ (data obrashcheniya 28.05.2025 g.)

18. PwC. Global Digital Trust Insights 2025. URL: https://www.pwc.com/us/en/services/consulting/cybersecurity-risk-regulatory/library/global-digital-trust-insights.html (accessed: 28.04.2025)

19. JavaScript Security: A Survey of Attacks and Defenses. arXiv. URL: https://arxiv.org/abs/1811.00918 (accessed: 28.04.2025)

20. Statisticheskiye dannyye sayta Statista.com URL: https://www.statista.com/statistics/273550/data-breaches-recorded-in-the-united-states-by-number-of-breaches-and-records-exposed/ (data obrashcheniya 28.04.2025 g.)

21. Informatsionnaya bezopasnost' (mirovoy rynok). TAdviser. URL: https://www.tadviser.ru/index.php/Stat'ya:Informatsionnaya_bezopasnost'_%28mirovoy_rynok%29 (data obrashcheniya 28.04.2025 g.)

22. Obzor ugroz informatsionnoy bezopasnosti za I kvartal 2024 goda. Positive Technologies. URL: https://www.ptsecurity.com/ru ru/research/analytics/cybersecurity-threatscape-2024-q1/ (data obrashcheniya 28.04.2025 g.)

23. Vasilakis N., Staicu C.-A., Ntousakis G., Kallas K., Karel B., DeHon A., Pradel M. Mir: Automated Quantifiable Privilege Reduction Against Dynamic Library Compromise in JavaScript. arXiv, 2020. URL: https://arxiv.org/abs/2011.00253

24. Nakhaei K., Ansari E., Ansari F. JSSignature: Eliminating Third-Party-Hosted JavaScript Infection Threats Using Digital Signatures. arXiv, 2018. URL: https://arxiv.org/abs/1812.03939 (accessed: 28.04.2025)

25. Gosudarev I.B. Osnovy razrabotki veb-prilozheniy na platformakh Node.js i Deno. SPb: Universitet ITMO, 2023.

26. Khoffman E. Bezopasnost' veb-prilozheniy. Razvedka, zashchita, napadeniye. SPb: Piter, 2021.

СЕРЕБРЯКОВ Дмитрий Сергеевич, генеральный директор ООО «БИТ Автоматизация». 115487, г. Москва, ул, Нагатинская д. 16, пом. 1/21в/10. E-mail: sm.house.2016@gmail.com

SEREBRYAKOV Dmitry Sergeevich, Chief Executive Officer, BIT Automation LLC. 115487, Moscow, st. Nagatinskaya, bldg. 16, office 1/21v/10. E-mail: sm.house.2016@gmail.com

УДК 004.056 Вестник УрФО № 2(56) / 2025, с. 40–48

Частикова В. А., Козачёк К. В., Согомонян Е. К., Луговой Д. А., Серый Н. В.

DOI: 10.14529/secur250204

МЕТОДИКА ОПРЕДЕЛЕНИЯ КРИТИЧНОСТИ УЯЗВИМОСТЕЙ С ИСПОЛЬЗОВАНИЕМ ТЕХНОЛОГИЙ BERT И RANDOM FOREST

В статье исследуется задача автоматического прогнозирования оценки критичности CVSS Score (Common Vulnerability Scoring System) на базе текстовых описаний уязвимостей CVE (Common Vulnerabilities and Exposures). Представлен подход, совмещающий методы NLP (Natural Language Processing) и машинного обучения. Выполнен разбор имеющихся решений, обозначены основные проблемы: неоднородность текстовых данных, дисбаланс классов в CVSS Score, необходимость интерпретируемости модели. Спроектирована и протестирована модель, продемонстрировавшая точность предсказания на наборе данных NVD (National Vulnerability Database). Полученные результаты сопоставлены с аналогами из других исследований. Практическая значимость работы — автоматизация анализа уязвимостей для SOC-команд (Security Operations Center) и кибербезопасности.

Ключевые слова: *автоматическое прогнозирование CVSS Score, обработка естественного языка (NLP), BERT (Bidirectional Encoder Representations from Transformers), Random Forest, классификация уязвимостей CVE, интерпретируемость модели (SHAP – SHapley Additive exPlanations).*

Chastikova V. A., Kozachok K. V., Sogomonyan E. K., Lugovoi D. A., Seriy N. V.

METHODOLOGY FOR DETERMINING THE CRITICALITY OF VULNERABILITIES USING BERT AND RANDOM FOREST TECHNOLOGIES

The paper investigates the task of automatic prediction of CVSS Score (Common Vulnerability Scoring System) based on textual descriptions of CVE (Common Vulnerabilities and Exposures) vulnerabilities CVSS Score (Common Vulnerability Scoring System) based on textual descriptions of CVE (Common Vulnerabilities and Exposures) vulnerabilities. An approach combining NLP (Natural Language Processing) and machine learning methods is presented machine learning. The existing solutions are analyzed and the main problems are outlined problems: heterogeneity of text data, imbalance of classes in CVSS Score, necessity of model interpretability of the model. The model was designed and applied, which demonstrated prediction accuracy on the NVD (National Vulnerability Database) dataset. The results Are compared with counterparts from current research. Practical importance of the work is the automation of vulnerability analysis for SOC teams (Security Operations Center) and cybersecurity.

Keywords: *automatic CVSS Score prediction, natural language processing (NLP), Natural Language Processing (NLP), BERT (Bidirectional Encoder Representations from Transformers), Random Forest, CVE vulnerability classification, model interpretability (SHAP – SHapleyAdditive exPlanations).*

Введение

Киберугрозы и уязвимости программного обеспечения становятся все более распространенными, что требует эффективных методов автоматизированного анализа. Для демонстрации приведем статистику доли успешных атак и их последствий, представленную в статье Positive Technologies "Актуальные киберугрозы: IV квартал 2024 года". [1]

Ручное назначение CVSS Score — трудоемкий и субъективный процесс, часто приводящий к задержкам в обработке уязвимостей. Целью данной работы является разработка автоматизированной модели для предсказания CVSS Score на основе текстовых описаний CVE. Для достижения этой цели решаются три ключевые задачи: анализ существующих методов, создание NLP-пайплайна на основе BERT и машинного обучения, а также оценка точности модели в сравнении с аналогами. Решение этих задач позволит ускорить и стандартизировать процесс оценки уязвимостей.

CVSS Score, разработанный Форумом реагирования на инциденты и группы безопасности FIRST, является общепринятым стандартом для оценки критичности уязвимостей. Текстовые описания CVE содержат ключевую информацию, необходимую для классификации, однако их анализ осложняется неоднородностью формулировок и технической терминологии. Кроме того, сложность задачи усугубляется нелинейной зависимостью между текстовыми данными и итоговой оценкой.

В работе [2] Philipp Kühn, David N. Relke и Christian Reuter использовали комбинацию классических методов машинного обучения (TF-IDF с Random Forest/Gradient Boosting) и нейросетевых моделей (LSTM, BERT) для прогнозирования оценок CVSS на основе текстовых описаний уязвимостей. В исследовании [3] для решения аналогичной задачи применялся подход DistilBERT + RandomForest. Проведенный анализ указанных работ позволяет заключить, что наиболее эффективным мето-

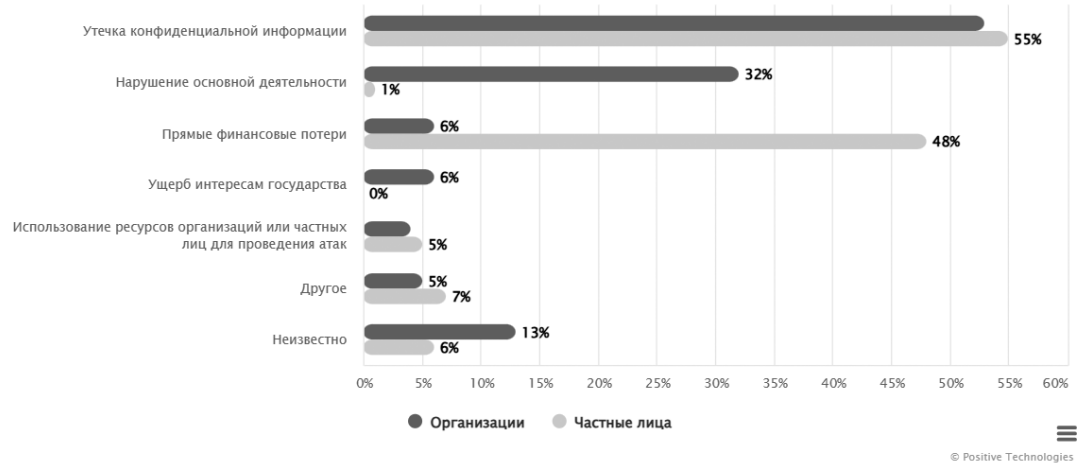


Рис. 1. Последствия атак (доля успешных атак)

дом оценки CVSS Score является комбинация алгоритма Random Forest (Случайный лес) с векторными представлениями BERT, поскольку данный гибридный подход обеспечивает: высокую точность прогнозирования за счет контекстного анализа текста, интерпретируемость результатов благодаря возможности анализа значимости признаков, и оптимальное соотношение вычислительной сложности и производительности, что делает его предпочтительным выбором для задач, требующих одновременно точности и прозрачности результатов.

Методика экспериментального исследования

В данном разделе представлено описание наборов данных, применявшихся для обучения и последующей оценки моделей машинного обучения, предназначенных для выявления уязвимостей в программном коде. Освещены используемые в рамках исследования классификаторы, а также критерии, применяемые для оценки эффективности проведенной классификации.

Набор данных. Для обучения и оценки эффективности моделей машинного обучения использовался набор данных «CVE 2024 Database: Exploits, CVSS, OS»[4], для создания которого использовалась информация Национальной базы уязвимостей NVD [5]. Данный датасет был выбран в связи с тем, что он охватывает наиболее актуальные и новые киберугрозы текущего времени. Ниже представлен набор уязвимостей 2023–2024 годов.

Модели для классификации оценки критичности. Random Forest (случайный лес) – это метод машинного обучения, который относится к классу алгоритмов ensembling (ансамблевых алгоритмов). Он используется для

Column Name	Data Type	Description
`CVE ID`	*string*	Unique identifier for each vulnerability (e.g., CVE-2024-21732).
`Description`	*string*	Brief summary of the vulnerability and its impact.
`CVSS Score`	*float*	Severity rating based on the Common Vulnerability Scoring System (CVSS).
`Attack Vector`	*string*	Method of exploitation (e.g., Network, Local, Physical).
`Affected OS`	*string*	List of operating systems affected by the vulnerability.

Рис. 2. Основная информация по столбцам набора данных

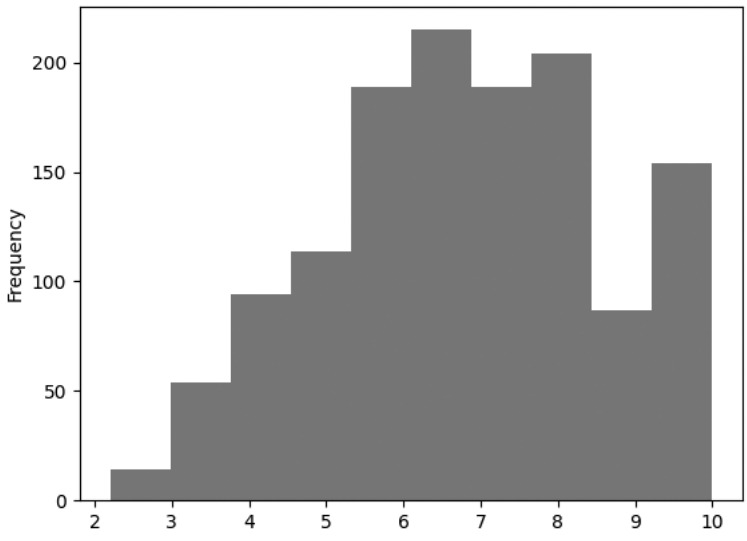


Рис. 3. График шкал CVSS Score

решения задач классификации и регрессии. Random Forest создает множество деревьев решений (Decision Trees) и сочетает их прогнозы для получения более точного результата. Каждое дерево обучается на случайном подмножестве признаков и объектов из тренировочной выборки.

Алгоритм Random Forest работает следующим образом[6]:

1) Создается множество деревьев (например, 100 деревьев).
2) Каждое дерево обучается на случайном подмножестве признаков и объектов из тренировочной выборки.
3) Каждое дерево предсказывает значение целевой переменной для объектов из тестовой выборки.
4) Прогнозы всех деревьев сочетаются для получения окончательного результата (например, путем голосования или усреднения).

Random Forest имеет несколько преимуществ перед другими алгоритмами машинного обучения:

1) Высокая точность решения задач классификации и задач регрессии [7].
2) Алгоритм способен обрабатывать большие массивы данных, сохраняя при этом высокую скорость обучения [8].
3) Обработка высокоразмерных данных: Random Forest может оперировать данными высокой размерности, так как он выбирает случайные подмножества признаков для каждого дерева.

Программное решение

Решение подразумевает обучение модели на основе описаний уязвимостей с последующим прогнозированием их CVSS Score (оценку критичности от 0 до 10). Результаты работы сохраняются в файле predictions.csv, содержащем три столбца. Первый столбец, обозначенный как "Description", содержит исходное описание уязвимости. Второй столбец, "CVSS Score", отражает фактическую оценку (если она присутствовала в исходных данных). Третий столбец, "Predicted_CVSS_Score", содержит прогноз оценки, сформированный обученной моделью.

Программа проводит оценку качества разработанной модели, используя метрику MSE (Mean Squared Error – среднеквадратичную ошибку[9]. MSE предоставляет информацию о средней величине квадратов разницы между прогнозируемыми и фактическими значениями оценки CVSS. Чем меньше MSE, тем лучше модель, так как большие отклонения от целевых значений сильнее штрафуются в квадратичном масштабе.

Для определения средней абсолютной ошибки модель также использует MAE (Mean Absolute Error), или среднюю абсолютную ошибку. MAE рассчитывает среднее абсолютное расхождение между предсказаниями и реальными данными, это делает метрику более наглядной. В отличие от MSE, MAE не акцентирует влияние аномалий, что способствует лучшему анализу устойчивости модели.

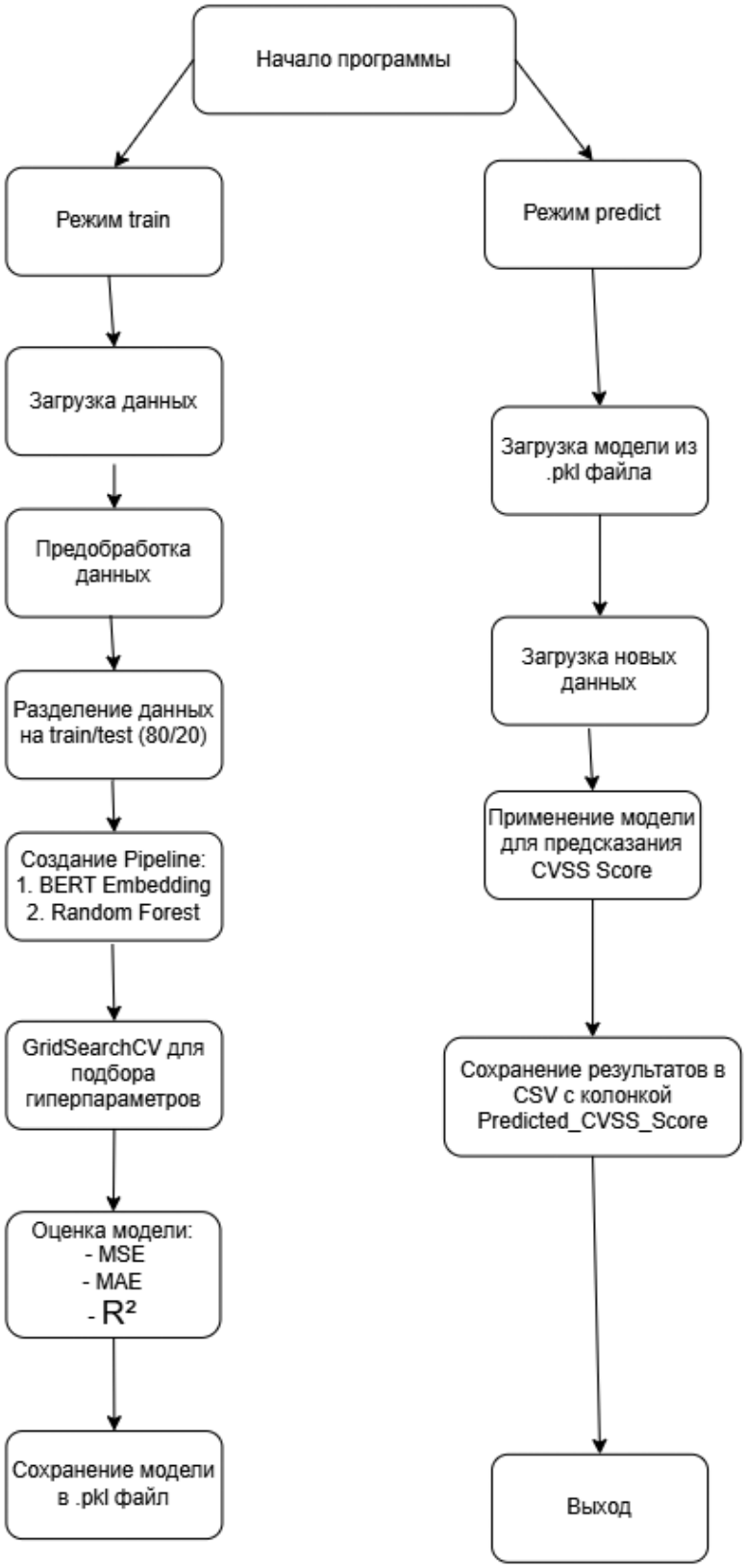


Рис. 4. Схема решения

Таблица 1

Результат работы программы

Description	CVSS Score	Predicted_CVSS_Score
"Уязвимость в X позволяет RCE..."	9.8	9.2
"Ошибка в Y приводит к DoS..."	7.5	6.9

Еще одна важная метрика R^2 (R-squared) показывает, насколько успешно модель способна объяснить изменчивость целевой переменной. Значение R^2, приближенное к 1, свидетельствует о том, что модель хорошо описывает исходные данные. Отрицательное значение, наоборот, говорит о том, что модель работает хуже, чем при использовании среднего значения. Это позволяет оценить полезность признаков, извлеченных из текстового материала.

В процессе обучения программа выводит на экран значения MSE, MAE и R^2 на тестовом наборе данных. Небольшие значения MSE и MAE указывают на высокую точность прогнозирования. Значение R^2, близкое к 1, подтверждает эффективность модели для предсказания. Эти показатели используются для сравнения различных моделей или для подбора оптимальных настроек гиперпараметров.

В представленном исследовании для алгоритма Random Forest были использованы следующие гиперпараметры: количество деревьев в ансамбле (n_estimators), максимальная глубина деревьев (max_depth), минимальное количество образцов для разделения узла (min_samples_split), минимальное количество образцов в листе (min_samples_leaf), а также количество рассматриваемых признаков при поиске наилучшего разделения (max_features) [11]. Оптимизация указанных параметров проводилась методом GridSearchCV с перекрестной проверкой, что позволило достичь баланса между точностью модели и вычислительной эффективностью. Выбор данных гиперпараметров обусловлен их доказанной эффективностью в задачах регрессии и классификации, а также возможностью контроля переобучения при работе с текстовыми данными. В частности, ограничение глубины деревьев (max_depth) и минимального количества образцов в узлах (min_samples_split, min_samples_leaf) способствует повышению обобщающей способности модели.

Математическое описание

Математическое описание решаемой задачи, представленное ниже, служит фундаментом для перевода вербальных описаний уязвимостей в векторные пространства посредством BERT (путем усреднения векторных представлений токенов). Также рассматривается алгоритм предсказания оценки CVSS, реализованный с использованием ансамбля деревьев в Random Forest (через усреднение прогнозов), что обеспечивает высокую точность работы модели.

Векторизация BERT

$$Embedding = \frac{1}{N}\sum_{i=1}^{N} BERT(T_i)$$

где T_i – токены текста, N – их количество.

Каждый текст (описание уязвимости) разбивается на токены T_i с помощью токенизатора BERT. Модель BERT преобразует каждый токен в вектор фиксированной размерности[10]. Итоговый эмбеддинг текста — это усреднение всех токенных векторов, что сохраняет контекстную информацию и уменьшает размерность данных для последующего анализа.

Random Forest

$$\hat{y} = \frac{1}{M}\sum_{j=1}^{M} f_j(X)$$

где f_j – j-ое дерево, M – количество деревьев.

Ансамбль из M решающих деревьев f_j независимо обрабатывает эмбеддинги (признаки X). Каждое дерево f_j вносит свой "голос" в предсказание CVSS Score. Итоговый прогноз $\hat{y}$ — это среднее значение предсказаний всех деревьев, что повышает устойчивость модели к переобучению.

Таблица 2

Сравнительный анализ существующих аналогов

Метод	Точность (MSE/R^2)	Ключевые характеристики и преимущества
TF-IDF + SVM[2]	MSE: 0.85, R^2: 0.72	Классический подход с ограниченной точностью. Не учитывает контекстные связи в тексте.
BERT + LSTM[2]	MSE: 0.62, R^2: 0.85	Высокая точность, но требует значительных вычислительных ресурсов. Сложность интерпретации результатов.
DistilBERT + RandomForest[3]	MSE: 0.55, R^2: 0.82	Оптимизированная версия BERT с умеренной точностью. Компромисс между производительностью и качеством.
Предлагаемое решение (BERT + оптимизированный RandomForest)	MSE: 0.41, R^2: 0.91	Лучшая точность среди аналогов. Использует: 1) Расширенный словарь технических терминов; 2) Оптимизированную предобработку текста; 3) Эффективную интерпретацию через SHAP-значения. Превосходит аналоги по всем метрикам при сопоставимых вычислительных затратах.

Аугментация данных

Аугментация данных в контексте предсказания CVSS Score применяется для искусственного расширения обучающей выборки и улучшения обобщающей способности модели.

Основные методы включают: синонимизацию и перефразирование описаний уязвимостей с сохранением смысла, контролируемое добавление шума (например, удаление или замена технических терминов) и генерацию новых примеров на основе шаблонов существующих уязвимостей. Такая аугментация помогает модели лучше обрабатывать редкие формулировки и снижает риск переобучения, особенно при работе с небольшими датасетами.

Обзор существующих решений

Предлагаемое решение достигает лучших результатов в сравнении с рассмотренными аналогами по ключевым метрикам: MSE снижено на 15–48% по сравнению с аналогами, а R^2 улучшен на 4–19%. Это достигнуто за счет комбинации трех подходов. Во-первых, расширенный словарь технических терминов повышает релевантность текстовых признаков. Во-вторых, оптимизированный пайплайн предобработки текста (лемматизация + фильтрация шумов) улучшает качество входных данных. В-третьих, модифицированный алгоритм RandomForest с подобранными гиперпараметрами обеспечивает стабильную точность прогнозирования.

Особое преимущество – встроенный механизм интерпретации результатов через SHAP-значения, что отсутствует у большинства аналогов. Это позволяет не только получать прогнозы, но и анализировать значимость конкретных терминов в описании уязвимости. Решение сохраняет конкурентное быстродействие благодаря оптимизированной архитектуре, не требуя специализированного оборудования для эксплуатации.

Заключение

Предложенный подход на основе BERT и Random Forest демонстрирует высокую точность предсказания CVSS Score и обеспечивает интерпретируемость результатов через анализ важности признаков. Перспективы работы включают расширение модели за счет мультимодальных данных (CWE-ID, вектор атаки) и разработку веб-интерфейса для интеграции в SOC-системы [12]. Реализация данных подходов позволит создать комплексное решение для автоматизированной оценки уязвимостей в промышленных условиях.

Литература

1. Актуальные киберугрозы: IV квартал 2024 года // Positive Technologies. 2024. URL: https://ptsecurity.com/ru-ru/research/analytics/aktualnye-kiberugrozy-iv-kvartal-2024-goda-i-kvartal-2025-goda (дата обращения: 26.04.2025).

2. Kühn P., Relke D., Reuter Ch. Common Vulnerability Scoring System Prediction based on Open Source Intelligence Information Sources // arXiv preprint arXiv:2210.02143. 2022. 15 p. URL: https://arxiv.org/pdf/2210.02143 (дата обращения: 26.04.2025).

3. 3Joana Cabral Costa Tiago Roxo, João B. F. Sequeiros , Hugo Proença, Pedro R. M. Inácio. Predicting CVSS Metric via Description Interpretation 2022. Vol. 10. URL: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9786831 (дата обращения: 26.04.2025).

4. CVE 2024 Database: Exploits, CVSS, OS // Kaggle. 2024. URL: https://www.kaggle.com/datasets/manavkhambhayata/cve-2024-database-exploits-cvss-os (дата обращения: 26.04.2025).

5. Официальная документация NVD-структура данных CVE // Национальная база уязвимостей. URL: https://nvd.nist.gov (дата обращения: 26.04.2025).

6. Русинова З.Р. Методы машинного обучения в анализе уязвимостей программного обеспечения. Екатеринбург: УрФУ, 2024. 120 с. URL: https://elar.urfu.ru/bitstream/10995/140355/1/m_th_z.r.rusinova_2024.pdf (дата обращения: 26.04.2025).

7. Breiman L. Random Forest // Machine Learning. – 2001, Vol. 45, no. 1. – P. 5-32.

8. Фомина Е.Е. Использование алгоритма Random Forest для обработки социально-экономических данных // Вестник ПНИПУ. Социально-экономические науки. 2022. № 1. URL: https://cyberleninka.ru/article/n/ispolzovanie-algoritma-random-forest-dlya-obrabotki-sotsialno-ekonomicheskih-dannyh (дата обращения: 26.04.2025).

9. Шунина Ю.С., Алексеева В.А., Клячкин В.Н. Критерии качества работы классификаторов // Вестник УлГТУ. 2015. №2 (70). URL: https://cyberleninka.ru/article/n/kriterii-kachestva-raboty-klassifikatorov.

10. V. Solovyev, M. Solnyshkina, A. Ten, N. Prokopiev A BERT-Based Classification Model: The Case of Russian Fairy Tales// Journal of Language and Education DOI: https://doi.org/10.17323/jle.2024.24030

11. Частикова В.А., Жерлицын С.А., Митюгов А.И. Технологии искусственного интеллекта в информационной безопасности. Монография. – Изд-во ФГБОУ ВО «КубГТУ», 2024. - 315 с.

12. Частикова В.А., Козачёк К.В. Применение нейронных сетей в платформе реагирования на инциденты как эффективное средство управления кибербезопасностью // Вестник УрФО. Безопасность в информационной сфере. 2023. № 4 (50). С. 70-76.

References

1. Aktualnie kiberugrozi: IV kvartal 2024 goda // Positive Technologies. 2024. URL: https://ptsecurity.com/ru-ru/research/analytics/aktualnye-kiberugrozy-iv-kvartal-2024-goda-i-kvartal-2025-goda (дата обращения: 26.04.2025).

2. Kühn P., Relke D., Reuter Ch. Common Vulnerability Scoring System Prediction based on Open Source Intelligence Information Sources // arXiv preprint arXiv:2210.02143. 2022. 15 p. URL: https://arxiv.org/pdf/2210.02143 (дата обращения: 26.04.2025).

3. Joana Cabral Costa Tiago Roxo, João B. F. Sequeiros , Hugo Proença, Pedro R. M. Inácio. Predicting CVSS Metric via Description Interpretation 2022. Vol. 10. URL: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9786831 (дата обращения: 26.04.2025).

4. CVE 2024 Database: Exploits, CVSS, OS // Kaggle. 2024. URL: https://www.kaggle.com/datasets/manavkhambhayata/cve-2024-database-exploits-cvss-os (дата обращения: 26.04.2025).

5. Ofitsialnaya dokumentatsiya NVD-struktura dannikh CVE // Natsionalnaya baza uyazvimostei. URL: https://nvd.nist.gov (дата обращения: 26.04.2025).

6. Rusinova Z.R. Metodi mashinnogo obucheniya v analize uyazvimostei programmnogo obespecheniya. Yekaterinburg: UrFU, 2024. 120 s. URL: https://elar.urfu.ru/bitstream/10995/140355/1/m_th_z.r.rusinova_2024.pdf (дата обращения: 26.04.2025).

7. Breiman L. Random Forest // Machine Learning. – 2001, Vol. 45, no. 1. – P. 5-32.

8. Fomina Ye.E. Ispolzovanie algoritma Random Forest dlya obrabotki sotsialno-ekonomicheskikh dannikh // Vestnik PNIPU. Sotsialno-ekonomicheskie nauki. 2022. № 1. URL: https://cyberleninka.ru/article/n/ispolzovanie-algoritma-random-forest-dlya-obrabotki-sotsialno-ekonomicheskih-dannyh (дата обращения: 26.04.2025).

9. Shunina Yu.S., Alekseeva V.A., Klyachkin V.N. Kriterii kachestva raboti klassifikatorov // Vestnik UlGTU. 2015. №2 (70). URL: https://cyberleninka.ru/article/n/kriterii-kachestva-raboty-klassifikatorov

10. Solovyev V., Solnyshkina M., Ten A., Prokopiev N. A BERT-Based Classification Model: The Case of Russian Fairy Tales// Journal of Language and Education DOI: https://doi.org/10.17323/jle.2024.24030

11. Chastikova V.A., Zherlicyn S.A., Mityugov A.I. Tekhnologii iskusstvennogo intellekta v informacionnoj bezopasnosti. Monografiya. – Izd-vo FGBOU VO «KubGTU», 2024. - 315 s.

12. Chastikova V.A., Kozachyok K.V. Primenenie nejronnyh setej v platforme reagirovaniya na incidenty kak effektivnoe sredstvo upravleniya kiberbezopasnost'yu // Vestnik UrFO. Bezopasnost' v informacionnoj sfere. 2023. № 4 (50). S. 70-76.

ЧАСТИКОВА Вера Аркадьевна, кандидат технических наук, доцент, доцент кафедры кибербезопасности и защиты информации ФГБОУ ВО «Кубанский государственный технологический университет». 350000, г. Краснодар, ул. Красная, 135. E-mail: chastikova_va@mail.ru

КОЗАЧЁК Константин Валерьевич, аспирант кафедры кибербезопасности и защиты информации ФГБОУ ВО «Кубанский государственный технологический университет». 350000, г. Краснодар, ул. Красная, 135. E-mail: Koza4ek.Konstantin@yandex.ru

СОГОМОНЯН Ева Кареновна, студент ФГБОУ ВО «Кубанский государственный технологический университет». 350000, г. Краснодар, ул. Красная, 135. E-mail: evo4ka.673@bk.ru

ЛУГОВОЙ Дмитрий Алексеевич, студент ФГБОУ ВО «Кубанский государственный технологический университет». 350000, г. Краснодар, ул. Красная, 135.E-mail: dmitrijlugovoy@gmail.com

СЕРЫЙ Никита Викторович, студент ФГБОУ ВО «Кубанский государственный технологический университет». 350000, г. Краснодар, ул. Красная, 135. E-mail: SS21102003@yandex.ru

C**HASTIKOVA Vera Arkadyevna,** Candidate of Technical Sciences, Associate Professor, Associate Professor of the Department of Cybersecurity and Information Protection of the Kuban State Technological University. 350000, Krasnodar, Krasnaya street, 135. E-mail: chastikova_va@mail.ru

KOZACHOK Konstantin Valerievich, post-graduate student of the Department of Cybersecurity and Information Protection of the Kuban State Technological University. 350000, Krasnodar, Krasnaya street, 135. E-mail: Koza4ek.Konstantin@yandex.ru

SOGOMONYAN Eva Karenovna, student of the Federal State Budgetary Educational Institution of Higher Education "Kuban State Technological University". 350000, Krasnodar, Krasnaya street, 135. E-mail: evo4ka.673@bk.ru

LUGOVOY Dmitry Alekseevich, student of the Federal State Budgetary Educational Institution of Higher Education "Kuban State Technological University". 350000, Krasnodar, Krasnaya street, 135. E-mail: dmitrijlugovoy@gmail.com

SERIY Nikita Viktorovich, student of the Federal State Budgetary Educational Institution of Higher Education "Kuban State Technological University". 350000, Krasnodar, Krasnaya street, 135. E-mail: SS21102003@yandex.ru

УДК 004.056 Вестник УрФО № 2(56) / 2025, с. 49–57

**Серегина Ю. Н., Ложкин Р. А.,
Афанасьева М. В, Баранкова И.И.**

DOI: 10.14529/secur250205

ПРИМЕНЕНИЕ КОНЦЕПЦИИ IOT SECURITY MATURITY MODEL ДЛЯ ОЦЕНКИ СООТВЕТСТВИЯ PCI DSS ВЕРСИИ 4.0.1 ПРОЦЕССИНГОВОГО ЦЕНТРА

Статья рассматривает применение концепции IoT Security Maturity Model для оценки соответствия PCI DSS версии 4.0.1 в процессинговом центре. Предложена иерархическая модель, учитывающая ключевые требования стандарта, может быть применена для оценки соответствия требованиям стандарта PCI DSS. Разработаны профили зрелости и 5-балльная шкала оценки. Методика помогает оценить текущий уровень соответствия и спланировать модернизацию системы безопасности.

Ключевые слова: *IoT Security Maturity Model, процессинговый центр, PCI DSS 4.0.1, практики, текущий профиль, целевой профиль, управление рисками.*

Seregina Y. N., Lozhkin R.A., Afanasyeva M. V., Barankova I.I.

APPLICATION OF IOT SECURITY MATURITY MODEL CONCEPT FOR ASSESSING PCI DSS VERSION 4.0.1 COMPLIANCE OF A PROCESSING CENTER

The article examines the application of the IoT Security Maturity Model con-cept for assessing compliance with PCI DSS version 4.0.1 in a processing center. A hierarchical model has been proposed, taking into account the key requirements of the standard, which can be used to evaluate compliance with PCI DSS. Maturity profiles and a 5-point assessment scale have been developed. The methodology helps assess the current level of compliance and plan the modernization of the security system.

Keywords: *IoT Security Maturity Model, processing center, PCI DSS 4.0.1, practices, current profile, target profile, risk management.*

Введение

В условиях возрастающей сложности технологий, применяемых в финансовом секторе, особенно при обработке, передаче и хранении платежных карт, расширяется поверхность для возможных кибератак. Отсутствие безопасных автоматизированных систем, соблюдение нормативных требований увеличивает поверхность атак и создает новые риски, что особенно критично для процессинговых центров, работающих с платежными транзакциями.

Финансовый сектор, в частности процессинговый центр играет ключевую роль в обеспечении безопасности платежных операций, обрабатывая данные держателей карт (ДДК), критичные аутентификационные данные (КАД), авторизуя транзакции и взаимодействуя с банками-эмитентами и эквайерами. В таких условиях киберугрозы могут повлиять на значимые бизнес-процессы. Любая уязвимость в может привести к компрометации данных, финансовым потерям и репутационным рискам.

Таким образом, актуальной задачей является выработка обоснованного подхода к управлению информационной безопасностью и грамотное инвестирование в механизмы защиты, соответствующие стандартам банковской безопасности и требованиям PCI DSS [1].

В 2019 году Kaspersky ICS CERT [2] в сотрудничестве с отраслевыми партнерами разработали и представили практическое руководство «IoT Security Maturity Model: Practitioner's Guide». Исходная модель ориентирована на безопасность интернета вещей, в данном материале была использована концепция данного руководства для оценки соответствия PCI DSS версии 4.0.1 процессингового центра.

Модель зрелости информационной безопасности строится на сравнении целевого и текущего профилей безопасности. Эти профили представляют собой оценки полноты реализации доменов, поддоменов и практик, иерархия которых представлена на рисунке 1.

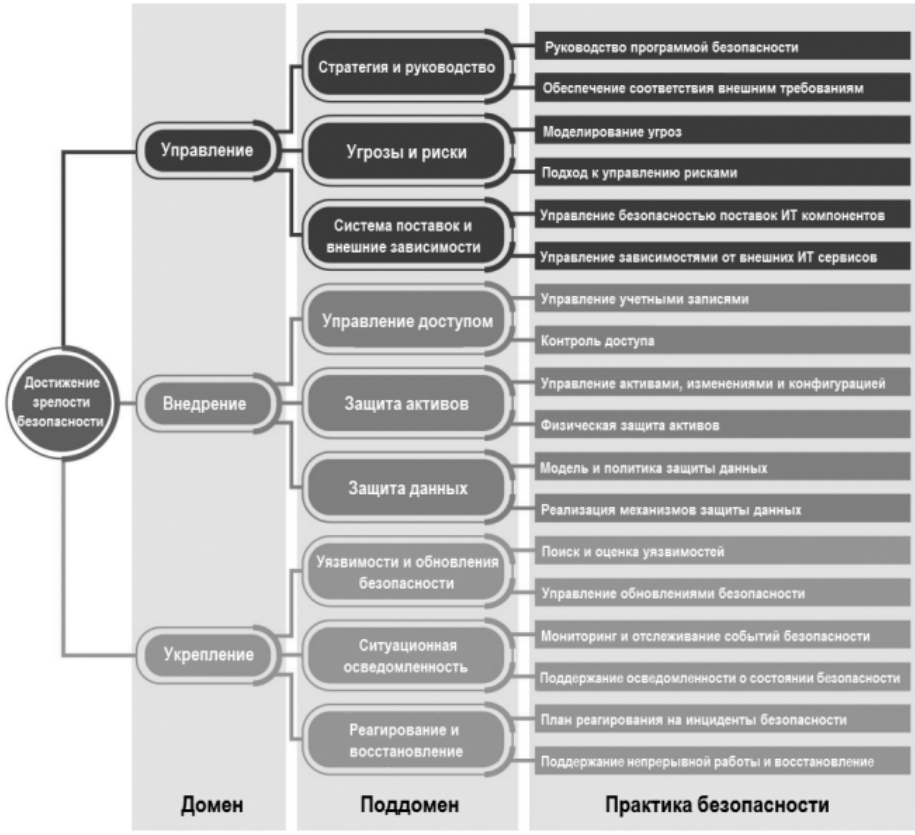


Рис. 1. Ранжирование доменов, поддоменов и практик в модели зрелости безопасности IoT Security Maturity Model: Practitioner's Guide

Целевой профиль зрелости отражает желаемое состояние системы защиты информации, к которому должна стремиться организация в процессе эксплуатации и поддержки своих информационных систем. Текущий профиль, напротив, демонстрирует реальное положение дел в области защиты данных. Каждый уровень модели зрелости оценивается по степени полноты реализации мер безопасности, которая зависит от объема применяемых практик, их систематичности и уровня обеспечиваемых гарантий.

Применение концепции IoT Security Maturity Model

PCI DSS является стандартом, регулирующим деятельность организаций, которые хранят, передают или обрабатывают данные держателей карт. Он состоит из 12 требований, определяющих ключевые принципы защиты платежных карт. В рамках данной работы была разработана модель зрелости для оценки соответствия PCI DSS версии 4.0.1 процессингового центра, где за домены взяты разделы стандарта, за практики – требования, а поддомены – проверочные процедуры, которые были сгруппированы по различным аспектам безопасности. Например, домен «Осуществлять регулярный мониторинг и тестирование сетей» разделяется на два поддомена «Контроль и мониторинг всего доступа к сетевым ресурсам и ДДК» и «Регулярное тестирование систем и процессов безопасности». Такое разделение было взято напрямую из стандарта. Само требование «Осуществлять регулярный мониторинг и тестирование сетей» было разделено на практики, которые были объединены в группы методом сравнения определенных требований подхода:

1. Процессы и механизмы регистрации и мониторинга всего доступа к компонентам системы и ДДК, куда входит проверочная процедура стандарта 10.1.

2. Журналы аудита надежно защищены и доступны, куда входит группа проверочных процедур стандарта 10.2–10.5.

3. Механизмы синхронизации времени поддерживаются во всех системах, куда входит проверочная процедура стандарта 10.6.

4. Сбои в критически важных системах контроля безопасности обнаруживаются, регистрируются и оперативно устраняются, куда входит проверочная процедура стандарта 10.7.

Полное ранжирование доменов, поддоменов и практик по стандарту PCI DSS версии 4.0.1 представлено на рисунке 2.

Существенные изменения в требованиях стандарта произошли за последнее время именно при переходе с версии 3.2.1 на 4.0. В новой редакции изменились как сами требования, так и подходы к их выполнению. На момент написания данной статьи актуальной версией PCI DSS является 4.0, однако в нем также описаны проверочные процедуры, которым с 31 марта 2025 года все организации, подлежащие сертификации, должны будут полностью соответствовать версии 4.0.1 [3]. В рамках использования концепции «IoT Security Maturity Model: Practitioner's Guide» для оценки соответствия PCI DSS версии 4.0.1 за текущий профиль зрелости были взяты требования версии 4.0 без учета новых проверочных процедур, а целевой профиль – полное соответствие требованиям стандарта, исходя из специфики бизнес-процессов [4]. Процессинговый центр проходил аудит именно в соответствии с обязательными проверочными процедурами на тот момент времени.

Модель зрелости безопасности процессингового центра по стандарту PCI DSS версии 4.0.1

При оценке текущего профиля зрелости безопасности учитывались проверочные процедуры стандарта PCI DSS. Ключевым фактором, влияющим на результат оценки, являлась специфика бизнес-процессов в процессинговом центре. Например, в рассматриваемой финансовой компании не применяются устройства POI. В свою очередь идеальная оценка строилась с учетом новых проверочных процедур, которые организации финансового сектора должны реализовать до 31 марта 2025 года. Шкала оценивания полноты представлена в списке:

- 0 – требование не применимо и обосновано бизнес-процессами;
- 1 – требование применимо, но не реализовано;
- 2 – требование применимо, но не реализовано в полной мере;
- 3 – требование полностью реализовано и соответствует построению бизнес-процессов в компании;
- 4 – требование полностью реализовано и построение бизнес-процессов соответствует наилучшим практикам стандарта PCI DSS.

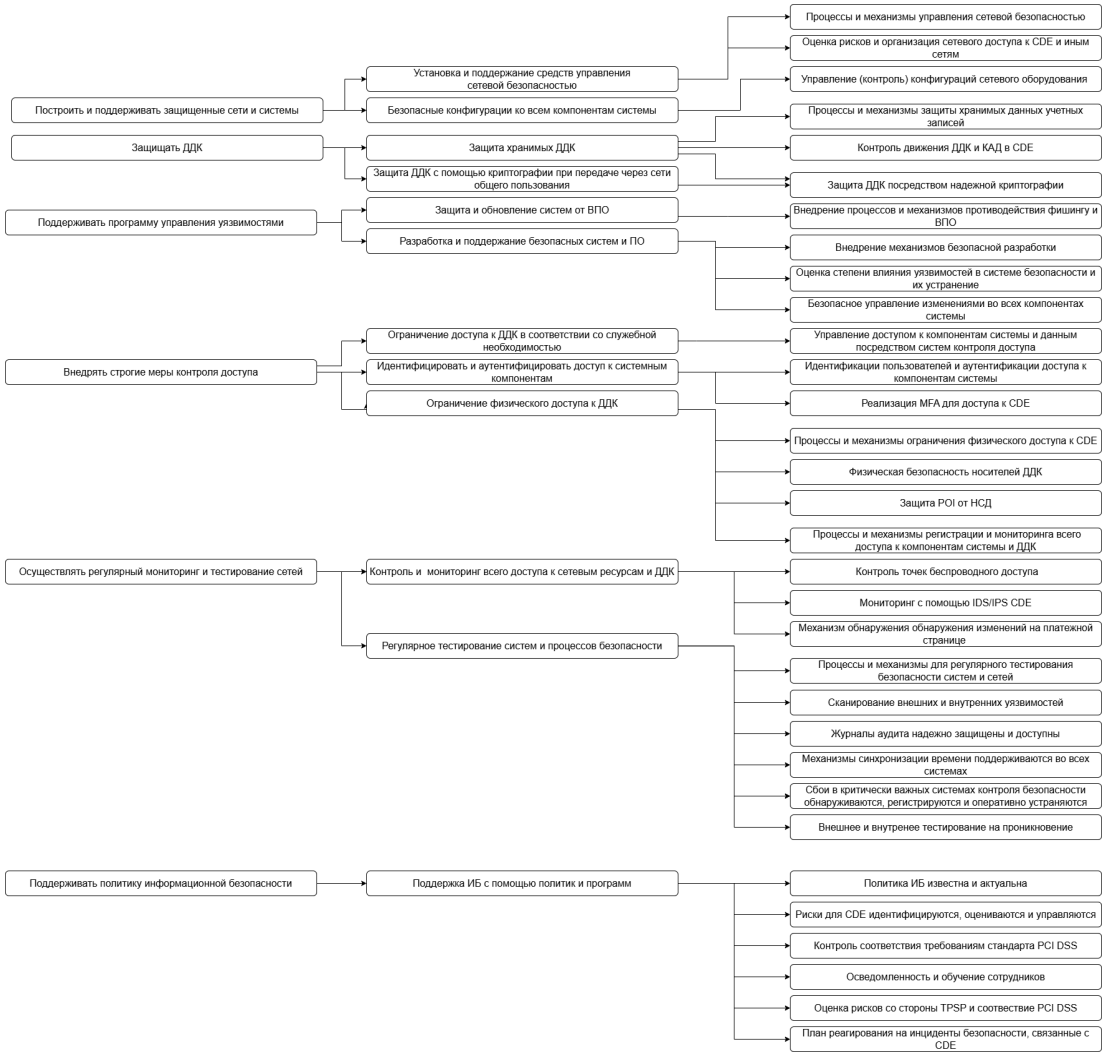


Рис 2. Ранжирование доменов, поддоменов и практик в модели зрелости, ориентированной на стандарт PCI DSS

Оценка полноты текущего профиля может иметь свое максимальное значение равное 3, поскольку новые требования уже прописаны в стандарте, но еще не реализованы в компании и будут обязательны только с 31 марта 2025 года и могут быть не реализованы в компании. Текущий профиль зрелости безопасности, а также его сравнение с целевым профилем, представлены на рисунке 3.

Организации необходимо распределить ответственность и обязанности за выполнение каждого требования стандарта, что напрямую влияет на практику «Осведомленность и обучение сотрудников». Требование выполняется методом фиксирования ролей и обязанностей, где используется матрица распределения ответственности, которая содержит данные о том, кто отвечает за выполнение задач и/или требований стандарта, кто подотчетен, с кем необходимо консультироваться и кого нужно информировать (также известная как матрица RACI).

При организации хранения ДДК необходимо учитывать все КАД, которые сохраняются до подтверждения транзакции. Защита таких данных обеспечивается применением современных криптографических алгоритмов в соответствии с требованиями стандарта PCI DSS, влияя на практику «Процессы и механизмы защиты хранимых данных учетных записей». Для сотрудников установлены строгие ограничения: запрещены любые манипуляции с ДДК при удаленной работе, за исключением специально уполномоченных лиц.

На практику «Защита ДДК посредством надежной криптографии» влияют проверочные процедуры:

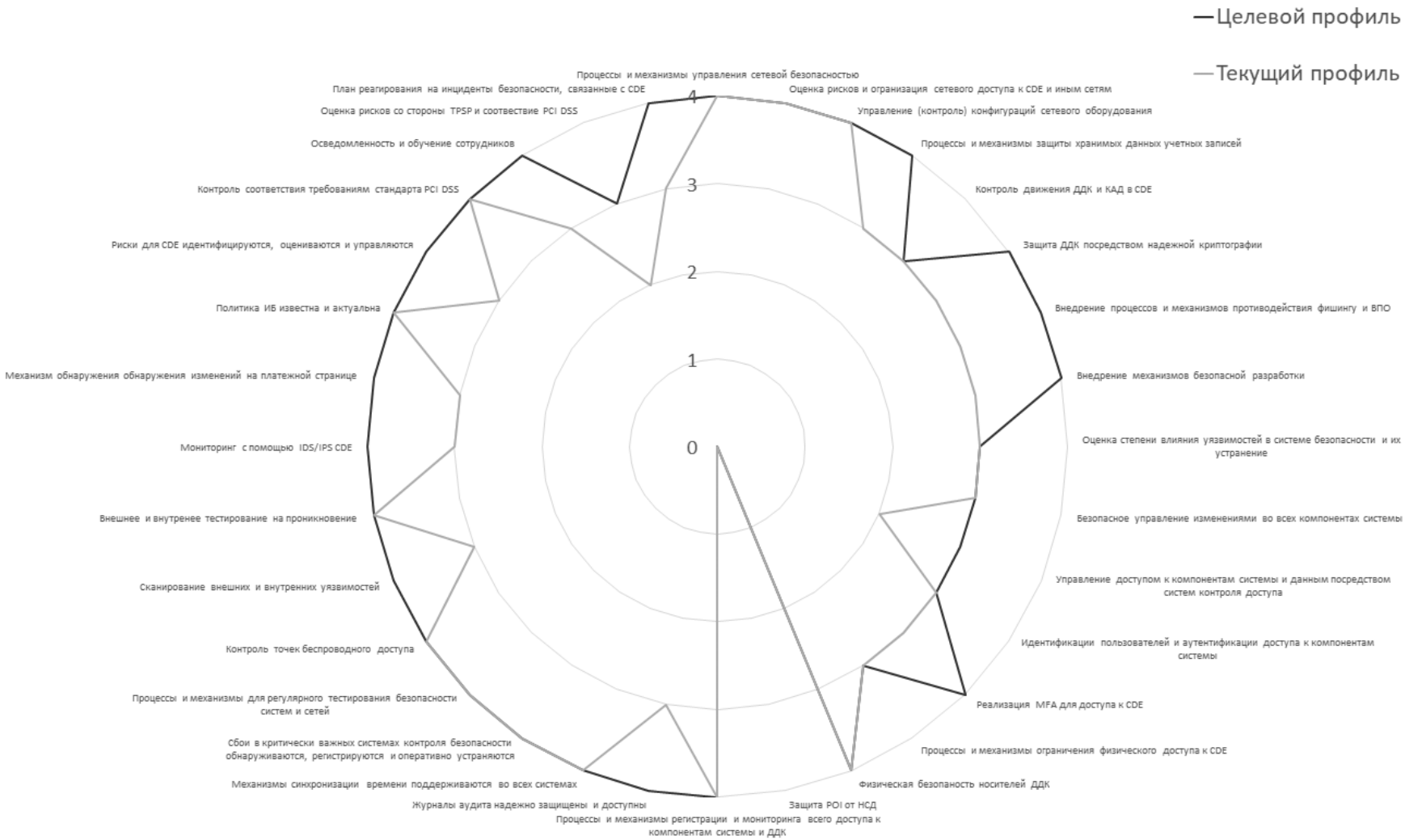


Рис. 3. Модель зрелости безопасности процессингового центра по стандарту PCI DSS версии 4.0.1

• Хеширования первичных учетных номеров (PAN) только с использованием хеш-функций с секретом.

• Процессы управления ключами должны соответствовать требованиям стандарта, изложенным в проверочных процедурах 3.6 и 3.7.

• Шифрование на уровне диска/разделов разрешено только для съемных носителей. В остальных случаях применяются методы, делающие PAN нечитаемым.

Рассматриваемый процессинговый центр ведет перечень доверенных ключей и сертификатов, используемых для защиты PAN при передаче, а также перечень используемого программного обеспечения (ПО) и его компонентов и перечень используемых криптографических протоколов, что влияет на практики «Защита ДДК посредством надежной криптографии», «Оценка степени влияния уязвимостей в системе безопасности и их устранение» и «Риски для CDE идентифицируются, оцениваются и управляются» соответственно.

В соответствии с требованиями стандарта PCI DSS, наилучшей практикой является автоматическое сканирование съемных носителей на предмет наличия вредоносного кода непосредственно при их подключении к системе. В качестве альтернативного подхода может применяться непрерывный мониторинг поведения системы и процессов при работе со съемными устройствами хранения данных. Параллельно организациям необходимо реализовать комплексные защитные механизмы, направленные на предотвращение фишинговых атак, с применением специализированных программных решений. Данные меры непосредственно способствуют реализации практики по противодействию фишингу и вредоносному программному обеспечению.

В соответствии с требованиями стандарта PCI DSS, критически важной практикой является автоматическое сканирование съемных носителей на предмет наличия вредоносного кода непосредственно при их подключении к системе. В качестве альтернативного подхода может применяться непрерывный мониторинг поведения системы и процессов при работе со съемными устройствами хранения данных. Параллельно организациям необходимо реализовать комплексные защитные механизмы, направленные на предотвращение фишинговых атак, с применением специализированных программных решений. Данные меры непосредственно способствуют реализации практики по противодействию фишингу и вредоносному программному обеспечению.

Особое внимание в новых требованиях уделяется обязательному применению межсетевого экрана уровня веб-приложений (WAF) для защиты публично доступных веб-ресурсов. В контексте защиты платежных страниц особую значимость приобретает использование политики безопасности контента (CSP), которая предполагает строгий контроль за исполняемыми скриптами. Каждый скрипт, выполняемый в браузере клиента, должен проходить процедуру авторизации, обладать гарантированной целостностью и быть зарегистрированным в системе учета с четким обоснованием его функционального назначения. Такой подход обеспечивает комплексную защиту платежных интерфейсов от потенциальных угроз.

В ежеквартальный аудит входит полная проверка всех учетных записей пользователей, что закрывает практику «Управление доступом к компонентам системы и данным посредством систем контроля доступа» и новые процедуры, связанные с учетными записями. Доступ к защищенной среде обработки данных держателей карт (CDE) организован с применением многофакторной аутентификации. Система аутентификации обладает устойчивостью к атакам повторного использования учетных данных, полностью исключая возможность их воспроизведения. При этом обход механизмов проверки невозможен ни для одной категории пользователей, включая администраторов системы, за исключением особых случаев, когда такое исключение официально санкционировано руководством на строго ограниченный период.

Процедура аутентификации построена на обязательном использовании как минимум двух независимых факторов подтверждения личности, что существенно повышает уровень защиты. Полноценный доступ к системе предоставляется исключительно после успешного прохождения всех предусмотренных этапов верификации, что гарантирует надежную защиту конфиденциальных данных от несанкционированного доступа. Такая реализация MFA обеспечивает комплексную безопасность CDE.

Важным изменением стало обязательное использование автоматизированных механизмов для проведения обзора журналов ау-

дита, что закрывает практику «Журналы аудита надежно защищены и доступны». Использование специализированных решений для мониторинга, включая SIEM-системы и анализаторы логов, существенно упрощает процесс обработки событий безопасности за счет автоматического выявления подозрительных записей в журналах. В компании внедрена SIEM система, что обеспечивает глубокий анализ инцидентов.

Стандарт PCI DSS 4.0.1 изменил подход к управлению рисками, сделав его более системным и интегрированным в процессы информационной безопасности. Основу нового подхода составляет требование 12.3.1, которое обязывает организации проводить регулярный целевой анализ рисков. Этот анализ становится основой для принятия ключевых решений в области безопасности: от определения периодичности пересмотра прав доступа до установки сроков действия паролей системных учетных записей. Данное требование напрямую влияет на практику «Риски для CDE идентифицируются, оцениваются и управляются».

Оценка практики «План реагирования на инциденты безопасности, связанные с CDE» также увеличила свою оценку благодаря учету инцидентов с нарушением целостности платежных страниц или HTTP-заголовков (требование 12.10.5), а также с обнаружением PAN в нерегламентированных местах хранения (требование 12.10.7). Данные инциденты включены в политику реагирования, что значительно влияет на безопасность ДДК и КАД.

Теперь внутренние сканирующие системы обязаны использовать авторизованный доступ с минимально необходимыми привилегиями, что значительно повышает оценку практики «Сканирование внешних и внутренних уязвимостей», так как обеспечивается более глубокое понимание ландшафта уязвимостей.

Заключение

Таким образом, новая версия стандарта содержит в себе ключевые изменения, помогающие процессинговому центру и иным компаниям финансового сектора поддерживать уровень защиты информации без негативного влияния на бизнес-процессы. Применение подходов IoT Security Maturity Model позволяет оценивать уровень готовности организации к переходу на новую версию стандарта PCI DSS. Данная концепция позволяет наглядно представить новые процедуры стандарта благодаря четкой иерархии и ранжированию требований стандарта. Модель зрелости безопасности дает понимание текущего состояния защиты финансовой организации, задает вектор развития и упрощает подготовку процессингового центра к аудите на соответствие требованиям стандарта PCI DSS.

Литература

1. Стандарт безопасности данных индустрии платежных карт (PCI DSS) версия 4.0 [Электронный ресурс] / SecurITM — Электрон. дан. — Режим доступа: https://service.securitm.ru/docs/pci-dss-v4-0-ru, свободный. — Загл. с экрана.

2. Консорциум индустрии интернета: практическое руководство по модели зрелости безопасности интернета вещей [Электронный ресурс] / IIC — Электрон. дан. — 2020. — Режим доступа: https://www.iiconsortium.org/pdf/IoT_SMM_Practitioner_Guide_2020-05-05.pd, свободный. — Загл. с экрана.

3. Отличия требований PCI DSS версии 4.0 от версии 3.2.1 [Электронный ресурс] / Deiteriy Compliance — Электрон. дан. — Режим доступа: https://compliance.deiteriy.com/pci_dss_requirements_differences, свободный. — Загл. с экрана.

4. Федорова, А. Р. Модель зрелости безопасности промышленного интернета вещей / А. Р. Федорова, О. А. Казаков, М. В. Афанасьева // Актуальные проблемы современной науки, техники и образования: Тезисы докладов 79-й международной научно-технической конференции, Магнитогорск, 19–23 апреля 2021 года. Том 1. – Магнитогорск: Магнитогорский государственный технический университет им. Г.И. Носова, 2021. – С. 403.

5. Баранкова И. И. Построение модели зрелости информационной безо-пасности для АСУ ТП ЦППН / И. И. Баранкова, М. В. Афанасьева, А. В. Дегтярева // Вестник УрФО. Безопасность в информационной сфере. – 2022. – № 2(44). – С. 57-62. – DOI 10.14529/secur220208.

References

1. Standart bezopasnosti dannykh industrii platezhnykh kart (PCI DSS) versiya 4.0 [Elektronnyy resurs] / SecurITM — Elektron. dan. — Re-zhim dostupa: https://service.securitm.ru/docs/pci-dss-v4-0-ru, svobodnyy. — Zagl. s ekrana.

2. Konsortsium industrii interneta: prakticheskoye rukovodstvo po mo-deli zrelosti bezopasnosti interneta veshchey [Elektronnyy resurs] / IIC — Elektron. dan. — 2020. — Rezhim dostupa: https://www.iiconsortium.org/pdf/IoT_SMM_Practitioner_Guide_2020-05-05.pd, svobodnyy. — Zagl. s ekrana.

3. Otlichiya trebovaniy PCI DSS versii 4.0 ot versii 3.2.1 [Elektron-nyy resurs] / Deiteriy Compliance — Elektron. dan. — Rezhim dostupa: https://compliance.deiteriy.com/pci_dss_requirements_differences, svobodnyy. — Zagl. s ekrana.

4. Fedorova, A. R. Model' zrelosti bezopasnosti promyshlennogo in-terneta veshchey / A. R. Fedorova, O. A. Kazakov, M. V. Afanas'yeva // Aktual'-nyye problemy sovremennoy nauki, tekhniki i obrazovaniya: Tezisy dokladov 79-y mezhdunarodnoy nauchno-tekhnicheskoy konferentsii, Magnitogorsk, 19–23 aprelya 2021 goda. Tom 1. – Magnitogorsk: Magnitogorskiy gosudarstven-nyy tekhnicheskiy universitet im. G.I. Nosova, 2021. – S. 403.

5. Barankova I. I. Postroyeniye modeli zrelosti informatsionnoy bez-opasnosti dlya ASU TP TSPPN / I. I. Barankova, M. V. Afanas'yeva, A. V. Degtyareva // Vestnik UrFO. Bezopasnost' v informatsionnoy sfere. – 2022. – № 2(44). – S. 57-62. – DOI 10.14529/secur220208.

БАРАНКОВА Инна Ильинична, доктор технических наук, доцент, за-ведующая кафедрой информатики и информационной безопасности, ФГБОУ ВО «Магнитогорский государственный технический университет им. Г.И. Носова». 455000, г. Магнитогорск, пр. Ленина, 38. E-mail: inna_barankova@mail.ru.

АФАНАСЬЕВА Маргарита Владимировна, старший преподаватель кафедры информатики и информационной безопасности, ФГБОУ ВО «Магнитогорский государственный технический университет им. Г.И. Носова». 455000, г. Магнитогорск, пр. Ленина, 38. E-mail: nansy_stokli@mail.ru.

СЕРЕГИНА Юлия Николаевна, студент кафедры информатики и ин-формационной безопасности, ФГБОУ ВО «Магнитогорский государственный технический университет им. Г.И. Носова». 455000, г. Магнитогорск, пр. Ленина, 38. E-mail: pip_p@internet.ru.

ЛОЖКИН Роман Александрович, студент кафедры информатики и ин-формационной безопасности, ФГБОУ ВО «Магнитогорский государственный технический университет им. Г.И. Носова». 455000, г. Магнитогорск, пр. Ленина, 38. E-mail: roman_2001@icloud.com.

BARANKOVA Inna Ilyinichna, Doctor of Technical Sciences, Associate Professor, Head of the Department of computer science and information security, Federal State Budgetary Educational Institution of Higher Education "Magnitogorsk State Technical University named after G.I. Nosov". 455000, Magnitogorsk, Lenin Ave., 38. E-mail: inna_barankova@mail.ru.

AFANASYEVA Margarita Vladimirovna, Assistant Professor of the Department of computer science and information security, Federal State Budgetary Educational Institution of Higher Education "Magnitogorsk State Technical University named after G.I. Nosov". 455000, Magnitogorsk, Lenin Ave., 38. E-mail: nansy_stokli@mail.ru.

SEREGINA Yulia Nikolaevna, student of the Department of computer science and information security, Federal State Budgetary Educational Institution of Higher Education "Magnitogorsk State Technical University named after G.I. Nosov". 455000, Magnitogorsk, Lenin Ave., 38. E-mail: pip_p@internet.ru.

LOZHKIN Roman Alexandrovich, student of the Department of computer science and information security, Federal State Budgetary Educational Institution of Higher Education "Magnitogorsk State Technical University named after G.I. No-sov". 455000, Magnitogorsk, Lenin Ave., 38. E-mail: roman_2001@icloud.com.

УДК 519.876.5 Вестник УрФО № 2(56) / 2025, с. 58–64

Васильев А. А., Сергин Д. А., Щерба Е. В. DOI: 10.14529/secur250206

МОДЕЛЬ АТАКИ ИСТОЩЕНИЯ ЭНЕРГОРЕСУРСОВ МОБИЛЬНЫХ УСТРОЙСТВ В САМООРГАНИЗУЮЩИХСЯ СЕТЯХ

В статье представлена модель атаки истощения энергоресурсов мобильных устройств на базе протокола маршрутизации OLSR, реализованная в сетевом симуляторе ns-3. Успешная реализация указанной атаки на мобильные сетевые устройства с аккумулятором приводит к нарушению их доступности. Проведены экспериментальные исследования показателей работы сети и последствий атаки истощения энергоресурсов с использованием различных режимов реализации подобной атаки. В заключении предложены перспективные подходы к противодействию указанным сетевым атакам.

Ключевые слова: *самоорганизующиеся сети, Интернет вещей, модель нарушителя, атаки истощения энергоресурсов, OLSR, ns-3.*

Vasiliev A. A., Sergin D. A., Shcherba E. V.

MODELING AN ENERGY DEPLETION ATTACK IN MOBILE AD-HOC NETWORKS

The article is devoted to modeling an energy depletion attack in mobile ad-hoc networks (MANET). The attack model is implemented based on the OLSR routing protocol in the ns-3 network simulator. Successful execution of the specified attack on mobile network devices with a battery leads to a violation of their availability. Experimental studies of network performance indicators and the consequences of the energy depletion attack are conducted. In conclusion, promising approaches to counteracting the specified network attacks are proposed.

Keywords: *ad-hoc networks, Internet of Things, attack model, energy depletion attacks, OLSR, ns-3.*

Введение

Атаки, направленные на истощение энергоресурсов (АИЭ, в зарубежной литературе известны под названием Energy Depletion Attack, EDA), являются одной из наиболее критичных угроз для беспроводных сетей устройств, работающих от аккумуляторных батарей, таких как беспроводные сенсорные сети (БСС) или сети устройств Интернета вещей. Основная особенность таких атак заключается в трудности их обнаружения, поскольку, как правило, такая атака воздействует на устройство или сеть опосредованно, путем отправки по коммуникационной сети ложных либо избыточных запросов, которые выглядят на первый взгляд легитимно и не всегда могут быть идентифицированы как атака. При этом процесс разряда аккумулятора в ходе атаки трудно отличим от нормальной работы устройства, и для обнаружения таких атак зачастую требуются специальные средства анализа.

Сложность защиты от атак в динамически организуемых сетях, прежде всего, лежит в самой форме реализации сети, когда её топология и структура связей между узлами способна меняться динамически в зависимости от различных факторов, таких как количество узлов, их производительность, уровень помех в канале связи, заряд аккумуляторов и т.д. Таким образом, задача моделирования подобных атак в беспроводных сетях мобильных устройств представляет особый интерес, с целью разработки эффективных механизмов защиты.

Анализ существующих атак истощения энергоресурсов

В настоящее время проблематике атак на истощение энергоресурсов уделяется всё большее внимание. В частности, в работе [1] приводится детальное описание принципов и классификация АИЭ, выполняемых на различных уровнях сетевого взаимодействия (физический, канальный, сетевой, прикладной), а также теоретический анализ существующих методик защиты от АИЭ. В [2] анализируются существующие разновидности атак истощения энергоресурсов на беспилотные летательные аппараты с выделением их ключевых характеристик.

Значительный интерес представляют конкретные реализации атак истощения энергоресурсов. В работе [3] производится моделирование АИЭ в сетях на основе протокола маршрутизации RPL, а также демонстрация и анализ эффективности защитного механизма на основе сравнения числа переданных пакетов в течение промежутка времени с пороговым значением. В свою очередь, задача организации защиты от подобных атак в работе [4] решается путем использования предложенного протокола маршрутизации для беспроводных сенсорных сетей, учитывающего энергоэффективность узлов сети для расчета оптимального маршрута.

В исследовании [5] рассматриваются сетевые атаки, приводящие к истощению заряда аккумуляторов беспилотных летательных аппаратов, и предложена комплексная математическая модель процесса истощения энергоресурсов устройств Интернета вещей, которая учитывает чередование циклов активного потребления энергии и спящего режима, а также предусматривает пороговое значение заряда аккумулятора, необходимое для корректной работы устройства.

Моделирование АИЭ в сетях на базе проактивного протокола маршрутизации OLSR

В настоящее время подавляющее большинство устройств с автономным питанием использует источники питания на базе литий-ионных (Li-Ion) либо литий-полимерных аккумуляторов. Характеристики данных источников питания, а также преимущества их применения в устройствах достаточно широко изучены в соответствующих работах [6]. В сетевом симуляторе ns-3 реализована модель источника питания на базе Li-Ion аккумулятора, предложенная в работе [7]. Указанная модель использована в настоящей работе для подготовки имитационной модели атаки истощения энергоресурсов.

Основная идея предложенной модели атаки истощения энергоресурсов базируется на динамическом изменении потребления энергии в процессе работы симулируемого протокола маршрутизации OLSR. Протокол маршрутизации OLSR является одним из наиболее широко используемых протоколов маршрутизации в беспроводных сетях мобильных устройств. Симулятор ns-3 позволяет устанавливать обработчики событий, связанных с симуляцией протокола OLSR, включая приём/передачу сообщений протокола или изменение таблицы маршрутизации. Используя обработчики указанных событий, можно учесть потребление энергии каждым сетевым узлом при их наступлении, что обе-

спечивает корректное функционирование предложенной модели.

Пусть I – величина тока потребления узлом сети в некоторый момент времени t. Для упрощения модели положим, что потребление энергии компонентами узла, не связанными с функционированием протокола маршрутизации, описывается величиной среднего тока потребления аккумулятора в рабочем режиме I_0. Тогда, при функционировании узла сети в нормальном режиме без протокола маршрутизации, ток потребления описывается формулой (1):

$$I = I_0 . \quad (1)$$

В свою очередь, ток потребления на время функционирования процесса протокола маршрутизации OLSR (например, при наступлении события перестройки таблицы маршрутизации), можно описать формулой (2):

$$I = I_0 + I_{OLSR}, \quad (2)$$

где I_{OLSR} – ток потребления процессом протокола маршрутизации OLSR.

При этом временной интервал активности процесса протокола маршрутизации зависит от параметров самого события. Введем величину t_{OLSR}, которая определяет время, в течение которого ток потребления I описывается формулой (2).

Временной интервал $t_{OLSR(table)}$ события перестройки таблицы маршрутизации OLSR можно описать как:

$$t_{OLSR(table)} = \alpha_{table} + \beta_{table} * N_{table}, \quad (3)$$

где α_{table} – постоянная составляющая временного интервала перестройки таблицы маршрутизации в миллисекундах (мс), β_{table} – время обработки одного элемента таблицы маршрутизации в мс, N_{table} – количество записей в таблице маршрутизации.

В свою очередь, временной интервал $t_{OLSR(RX/TX)}$ события приёма/передачи пакетов протокола OLSR можно описать как:

$$t_{OLSR(RX/TX)} = \alpha_{RX/TX} + \beta_{RX/TX} * N_{msg}, \quad (4)$$

где $\alpha_{RX/TX}$ – постоянная составляющая временного интервала приема-передачи пакета в мс, $\beta_{RX/TX}$ – время приема-передачи одного OLSR сообщения в мс, N_{msg} – число сообщений в одном пакете протокола OLSR.

Таким образом, при функционировании узла сети в нормальном режиме, когда процесс протокола маршрутизации остается неактивным, ток потребления вычисляется по формуле (1). Как только в процессе симуляции наступает событие протокола OLSR, по формулам (3) или (4) вычисляется временной интервал данного события, что в свою очередь позволяет вычислить ток потребления по формуле (2).

В процессе симуляции производится регулярное отслеживание текущего напряжения аккумулятора V сетевых устройств. Если текущее напряжение аккумулятора устройства снижается до порогового значения разряда V_{off}, аккумулятор считается разряженным, и устройство выводится из работы.

Предложенная модель атаки истощения энергоресурсов сетевых узлов реализуется посредством уменьшения интервалов отправки служебных сообщений протокола маршрутизации. Стандарт RFC 7181 описывает интервалы отправки служебных сообщений OLSR: в частности, для сообщений HELLO он составляет 2с, для сообщений TC – 5с [8]. В результате уменьшения данных интервалов увеличивается частота отправки указанных сообщений, и, как следствие, частота их приема смежными узлами. Поскольку обработка каждого сообщения влечет за собой потребление энергии, путем увеличения частоты отправки таких сообщений узел нарушителя может добиться ускоренного разряда источников питания целевых устройств.

На рисунке 1 представлена диаграмма сетевой топологии, использованная для процесса имитационного моделирования сетевого взаимодействия. Для экспериментальной оценки предложенной модели атаки было произведено несколько испытаний. В ходе первого испытания все узлы сети функционировали со стандартными значениями временных таймеров протокола OLSR. Для второго испытания таймер отправки сообщений HELLO узлом 2 был уменьшен до 1с. Для третьего испытания таймер отправки сообщений HELLO узлом 2 был уменьшен до 0,1с. В рамках четвертого испытания таймер отправки сообщений TC узлом 2 был уменьшен до 1 с.

Остальные параметры моделирования были определены следующим образом:

- Количество узлов в сети: 8.
- Ёмкость аккумулятора целевых узлов (кроме узла 2): 1000 мА/ч.

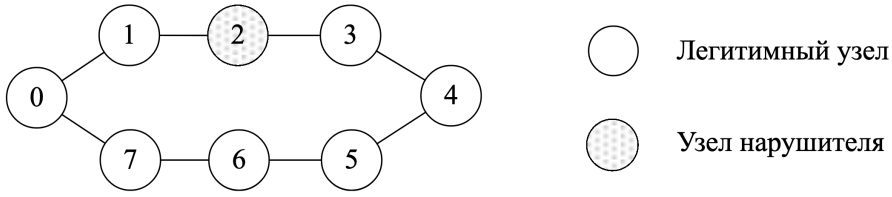


Рис. 1. Диаграмма сетевой топологии для имитационного моделирования

- Ток потребления аккумулятора в рабочем режиме I_0 = 20 мА.
- Ток потребления на время отработки процесса OLSR: I_{OLSR} = 200 мА.
- Номинальное напряжение аккумулятора в полностью заряженном состоянии: V_0 = 4,2 В.
- Пороговое напряжение разряда аккумулятора: V_{off} = 3.3 В.
- Номинальный интервал отправки сообщений протокола OLSR:
  - для сообщений типа HELLO – 2 с.
  - для сообщений типа ТС – 5 с.
- Временные интервалы обработки сообщений OLSR:

α_{table} = 10 мс, β_{table} = 0,5 мс.

$\alpha_{RX/TX}$ = 10 мс, $\beta_{RX/TX}$ = 1 мс.

- Время моделирования: 24 часа.

На рисунке 2 представлены графики зависимости напряжения аккумуляторов сетевых узлов от времени их работы в сети, полученные в результате имитационного моделирования.

В результате последующих испытаний была получена оценка зависимости времени работы сетевых узлов от значений таймеров

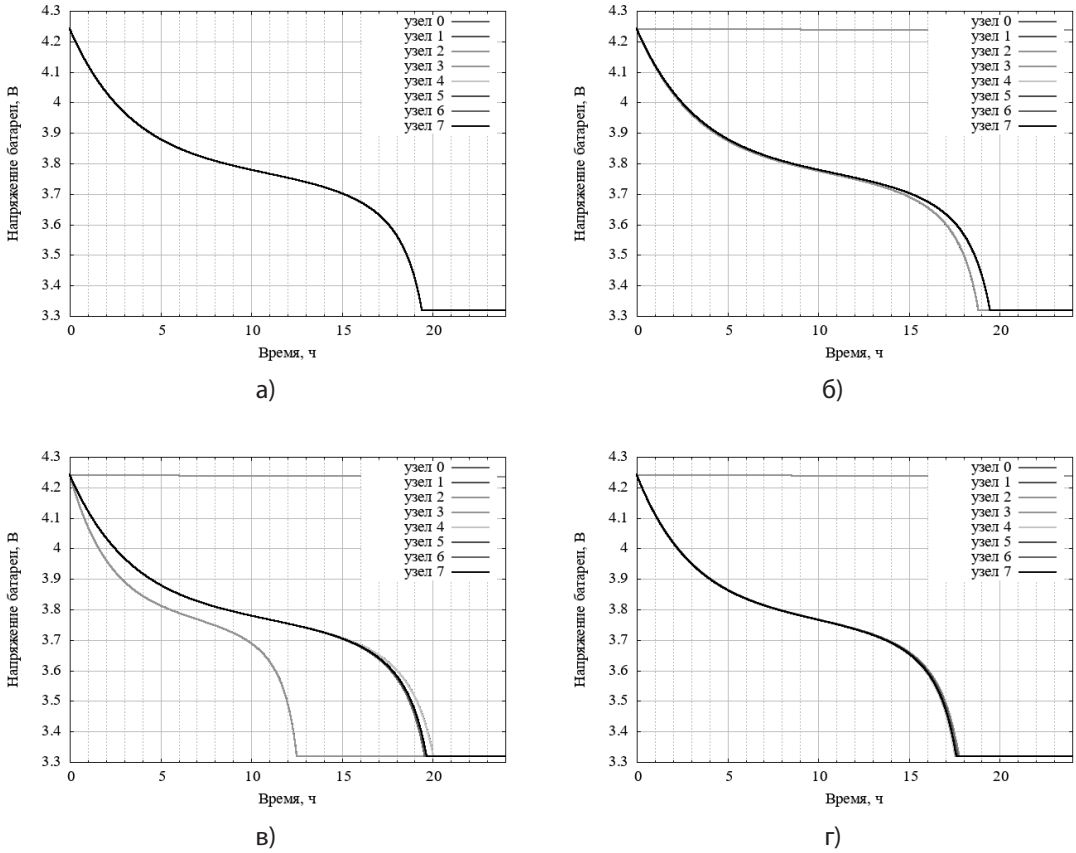
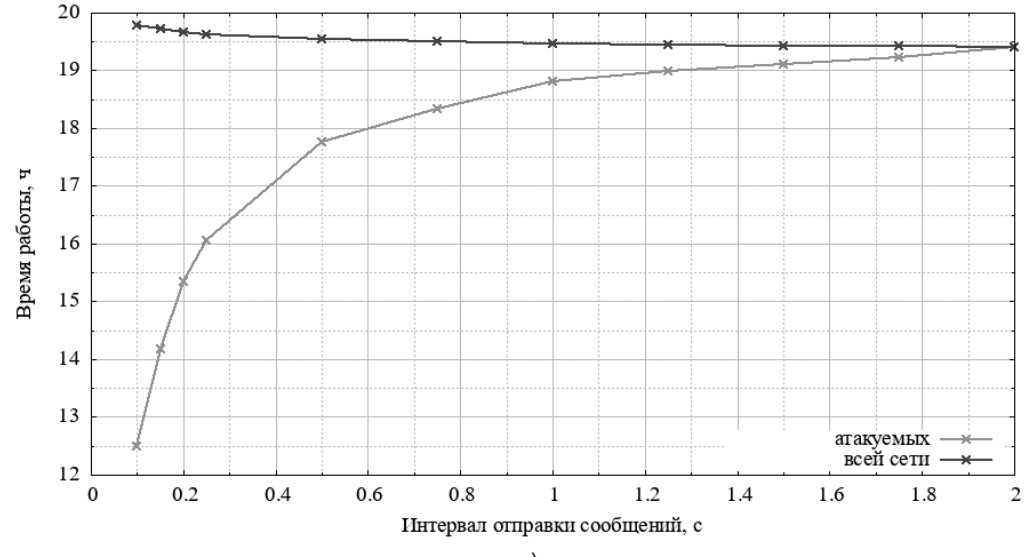
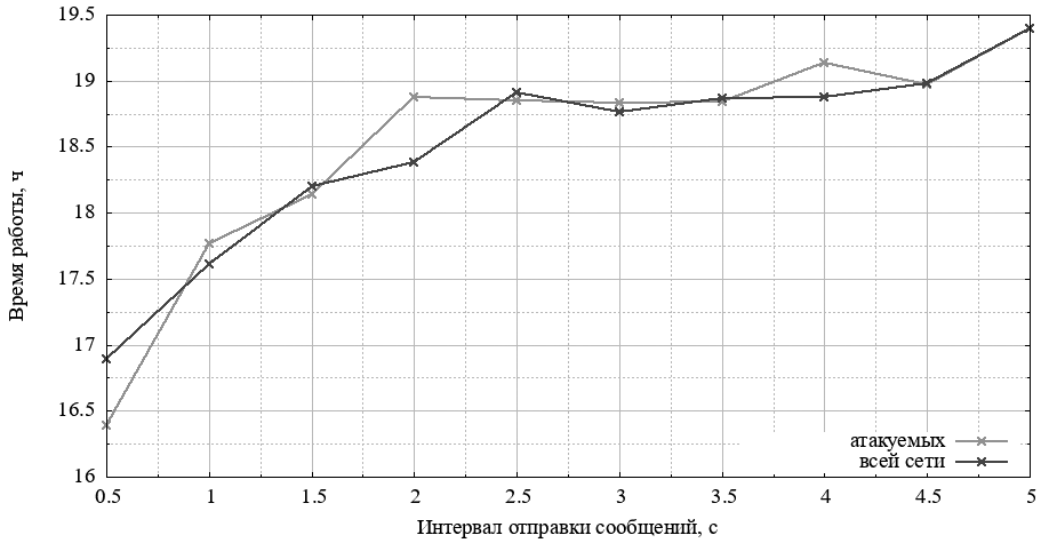


Рис. 2. Графики напряжения аккумуляторов сетевых узлов при различном поведении узла 2: нормальный режим (а), с уменьшенным до 1с таймером отправки сообщений HELLO (б), с уменьшенным до 0,1с таймером отправки сообщений HELLO (в), с уменьшенным до 1с таймером отправки сообщений ТС (г)



а)



б)

Рис. 3. Графики зависимости времени работы сетевых узлов
от значений таймеров отправки сообщений HELLO (а) и TC (б) узлом 2.

отправки служебных сообщений протокола OLSR узлом нарушителем (рисунок 3).

Анализируя полученные результаты, можно отметить, что уменьшение таймера отправки сообщений HELLO узлом нарушителем приводит к снижению времени работы соседних узлов, в то время как уменьшение таймера отправки сообщений TC вызывает заметное снижение времени работы узлов всей сети в целом. Это связано с тем, что сообщения TC, в отличие от сообщений HELLO, распространяются по всем узлам сети, охватываемой протоколом OLSR. В целом, зафиксированные результаты подчеркивают необходимость реализации мер по противодействию атакам на истощение энергоресурсов на всех узлах сети.

Заключение

По результатам моделирования атаки на истощение энергоресурсов в сетевом симуляторе ns-3 можно составить определенные рекомендации по противодействию данному классу атак.

Для противодействия атаке истощения энергоресурсов на базе протокола маршру-

тизации OLSR, каждый узел, во-первых, должен отбрасывать служебные сообщения протокола, для которых интервалы отправки, указанные в заголовке сообщения, не соответствуют интервалам отправки, определенным в конфигурации рассматриваемого узла. Для более эффективного противодействия, каждое сетевое устройство должно выполнять анализ количества отправленных и полученных служебных сообщений. При отклонении указанных показателей от нормальных значений (в частности, в случае аномально короткого интервала между сообщениями), устройство должно принимать меры по противодействию узлу нарушителя. Например, легитимные узлы могут просто игнорировать избыточные сообщения, если это не оказывает серьезного влияния на работоспособность сети. В некоторых случаях, легитимные узлы могут полностью изолировать узел нарушителя, полностью игнорируя все сообщения указанного узла.

Для корректной идентификации узла нарушителя могут применяться репутационные модели [9, 10]. В рамках подобной модели, отправка избыточных служебных сообщений протокола OLSR узлами сети будет приводить к снижению их репутации, что в конечном счете может быть использовано для изоляции узлов нарушителей с низкой репутацией.

Литература

1. Nguyen V.L., Lin P.C., Hwang R.H. Energy Depletion Attacks in Low Power Wireless Networks, IEEE Access, 2019, Vol. 7, pp. 51915-51932, DOI: 10.1109/ACCESS.2019.2911424.

2. Десницкий В.А., Рудавин Н.Н. Моделирование и оценка атак истощения энергоресурсов на беспилотные летательные аппараты в системах антикризисного управления / В.А. Десницкий, Н.Н. Рудавин // Вестник Санкт-Петербургского университета Государственной противопожарной службы МЧС России, 2019. № 4. С. 69-74.

3. Pu C. Energy Depletion Attack Against Routing Protocol in the Internet of Things, 2019 16th IEEE Annual Consumer Communications & Networking Conference (CCNC), January 2019, pp. 1-4, DOI: 10.1109/CCNC.2019.8651771.

4. Gong P., Chen T.M., Xu P. Resource-Conserving Protection against Energy Draining (RCPED) Routing Protocol for Wireless Sensor Networks, Network, 2022, Vol. 2(1), pp. 83-105, DOI: 10.3390/NETWORK2010007.

5. Kuaban G.S., Gelenbe E., Czachórski T., Czekalski P., Tangka J.K. Modelling of the Energy Depletion Process and Battery Depletion Attacks for Battery-Powered Internet of Things (IoT) Devices, Sensors (Basel), 2023, Vol. 23(13):6183, DOI: 10.3390/s23136183.

6. Sajja K.S., Bhowmik B. IoT Systems and Battery-Based Energy Sources, International Conference on Artificial Intelligence and Smart Communication (AISC), January 2023, pp. 212-219, DOI: 10.1109/AISC56616.2023.10085426.

7. Tremblay O., Dessaint L.A., Dekkiche A.I. A Generic Battery Model for the Dynamic Simulation of Hybrid Electric Vehicles, 2007 IEEE Vehicle Power and Propulsion Conference, September 2007, pp. 284-289, DOI: 10.1109/VPPC.2007.4544139.

8. RFC7181: The Optimized Link State Routing Protocol Version 2 / T. Clausen, C. Dearlove, P. Jacquet, U. Herberg. – 2014. – URL: https://tools.ietf.org/html/rfc7181 (дата обращения: 21.05.2025).

9. Литвинов Г.А., Щерба Е.В. Применение моделей доверия и репутации для обеспечения для обеспечения безопасности маршрутизации в динамически организуемых сетях / Г.А. Литвинов, Е.В. Щерба // Вестник УрФО. Безопасность в информационной сфере, 2021. № 3(41). С. 12-23.

10. Литвинов Г.А. Экспериментальное исследование репутационной модели для поиска маршрута в самоорганизующихся сетях / Г.А. Литвинов // Вестник УрФО. Безопасность в информационной сфере, 2022. № 3. С. 69-75.

References

1. Nguyen V.L., Lin P.C., Hwang R.H. Energy Depletion Attacks in Low Power Wireless Networks, IEEE Access, 2019, Vol. 7, pp. 51915-51932, DOI: 10.1109/ACCESS.2019.2911424.

2. Desnickij V.A., Rudavin N.N. Modelirovanie i ocenka atak istoshhenija jenergoresursov na bespilotnye letatel'nye apparaty v sistemah antikrizisnogo upravlenija / V.A. Desnickij, N.N. Rudavin // Vestnik Sankt-Peterburgskogo universiteta Gosudarstvennoj protivopozharnoj sluzhby MChS Rossii, 2019. № 4. S. 69-74.

3. Pu C. Energy Depletion Attack Against Routing Protocol in the Internet of Things, 2019 16th IEEE Annual Consumer Communications & Networking Conference (CCNC), January 2019, pp. 1-4, DOI: 10.1109/CCNC.2019.8651771.

4. Gong P., Chen T.M., Xu P. Resource-Conserving Protection against Energy Draining (RCPED) Routing Protocol for Wireless Sensor Networks, Network, 2022, Vol. 2(1), pp. 83-105, DOI: 10.3390/NETWORK2010007.

5. Kuaban G.S., Gelenbe E., Czachórski T., Czekalski P., Tangka J.K. Modelling of the Energy Depletion Process and Battery Depletion Attacks for Battery-Powered Internet of Things (IoT) Devices, Sensors (Basel), 2023, Vol. 23(13):6183, DOI: 10.3390/s23136183.

6. Sajja K.S., Bhowmik B. IoT Systems and Battery-Based Energy Sources, International Conference on Artificial Intelligence and Smart Communication (AISC), January 2023, pp. 212-219, DOI: 10.1109/AISC56616.2023.10085426.

7. Tremblay O., Dessaint L.A., Dekkiche A.I. A Generic Battery Model for the Dynamic Simulation of Hybrid Electric Vehicles, 2007 IEEE Vehicle Power and Propulsion Conference, September 2007, pp. 284-289, DOI: 10.1109/VPPC.2007.4544139.

8. RFC7181: The Optimized Link State Routing Protocol Version 2 / T. Clausen, C. Dearlove, P. Jacquet, U. Herberg. – 2014. – URL: https://tools.ietf.org/html/rfc7181 (дата обращения: 21.05.2025).

9. Litvinov G.A., Shcherba E.V. Primenenie modelej doverija i reputacii dlja obespechenija dlja obespechenija bezopasnosti marshrutizacii v dinamicheski organizuemyh setjah / G.A. Litvinov, E.V. Shcherba // Vestnik UrFO. Bezopasnost' v informacionnoj sfere, 2021. № 3(41). S. 12-23.

10. Litvinov G.A. Jeksperimental'noe issledovanie reputacionnoj modeli dlja poiska marshruta v samoorganizujushhihsja setjah / G.A. Litvinov // Vestnik UrFO. Bezopasnost' v informacionnoj sfere, 2022. № 3. S. 69-75.

ВАСИЛЬЕВ Артем Андреевич, аспирант федерального государственного автономного образовательного учреждения высшего образования «Омский государственный технический университет». 644050, г. Омск, пр. Мира, 11. E-mail: dedaav55@gmail.com

СЕРГИН Даниил Альбертович, аспирант федерального государственного автономного образовательного учреждения высшего образования «Омский государственный технический университет». 644050, г. Омск, пр. Мира, 11. E-mail: daniil0808_98@mail.ru

ЩЕРБА Евгений Викторович, кандидат технических наук, доцент, доцент кафедры «Комплексная защита информации» федерального государственного автономного образовательного учреждения высшего образования «Омский государственный технический университет». 644050, г. Омск, пр. Мира, 11. E-mail: evscherba@gmail.com

VASILIEV Artem Andreevich, post-graduate student of the Omsk State Technical University. 644050, Omsk, pr. Mira, 11. E-mail: dedaav55@gmail.com

SERGIN Daniil Albertovich, post-graduate student of the Omsk State Technical University. 644050, Omsk, pr. Mira, 11. E-mail: daniil0808_98@mail.ru

SHCHERBA Evgeny Victorovich, Candidate of Engineering, Associate Professor, Department of Complex Information Protection, Omsk State Technical University. 644050, Omsk, pr. Mira, 11. E-mail: evscherba@gmail.com

УДК 004.056 Вестник УрФО № 2(56) / 2025, с. 65–73

**Частикова В. А., Алиев М. К.,
Тесленко А. А., Игнатенко И. С.**

DOI: 10.14529/secur250207

НЕЙРОННЫЕ СЕТИ ДЛЯ ОБЕСПЕЧЕНИЯ БЕЗОПАСНОСТИ ВЕБ-ПРИЛОЖЕНИЙ

Данная работа посвящена применению машинного обучения для повышения безопасности веб-приложений. Рассмотрены ограничения традиционных методов защиты, таких как файрволы веб-приложений (WAF), непрерывный мониторинг с использованием SIEM систем и тестирование на проникновение. Были проанализированы принципы работы нейронных сетей, их классификация и потенциал их применения для автоматизации анализа трафика, выявления аномалий и защиты от уязвимостей нулевого дня. Описаны преимущества и недостатки нейронных сетей, обоснована их интеграция с существующими инструментами и проведён сравнительный анализ современных WAF с машинным обучением.

Ключевые слова: *безопасность веб-приложений, нейронные сети, файрвол веб-приложений, система безопасности информации и управления событиями, киберугрозы, непрерывный мониторинг*

Chastikova V. A., Aliev M. K., Teslenko A. A. Ignatenko I. S.

NEURAL NETWORKS FOR WEB APPLICATION SECURITY

The publication focuses on the application of machine learning to enhance the security of web applications. It examines the limitations of traditional protection methods, such as web application firewalls (WAF), continuous monitoring using SIEM systems, and penetration testing. The principles of neural networks, their classification, and their potential for automating traffic analysis, detecting anomalies, and protecting against zero-day vulnerabilities have been analyzed. The advantages and disadvantages of neural networks are described, their integration with existing tools is justified, and a comparative analysis of modern WAFs with machine learning is conducted.

Keywords: *web application security, neural networks, Web Application Firewall (WAF), SIEM systems, cyber threats, continuous monitoring*

Введение

Цифровые технологии развиваются со стремительной скоростью, из-за чего вопрос обеспечения безопасности веб-приложений становится с каждым днем все более актуальным. Веб-приложения стали неотъемлемой частью нашего быта, охватывая огромное множество аспектов нашей жизни. Это делает их крайне привлекательной целью для киберпреступников. Рост числа и сложности кибератак на веб-приложения подчеркивает необходимость поиска все более эффективных решений для их защиты. Современные угрозы требуют современных подходов, способных противостоять как уже известным, так и совершенно новым неизведанным видам атак.

В данной статье комплексно рассмотрен вопрос применения нейронных сетей для обеспечения безопасности веб-приложений. Проанализированы ограничения традиционных методов защиты, описаны принципы работы нейронных сетей, рассмотрены их возможности и вызовы в контексте повышения надежности защиты, а также проведен сравнительный анализ ряда современных WAF от разных производителей.

Традиционные методы защиты и их ограничения

Традиционные методы защиты веб-приложений, такие как использование файрволов веб-приложений (WAF) на основе правил и сигнатур, непрерывный мониторинг трафика и поведения системы с использованием систем безопасности информации и управления событиями (SIEM), а также тестирование на проникновение долгое время составляли основу обеспечения безопасности. Однако их эффективность ограничена рядом факторов.

Традиционные WAF представляют собой инструменты, предназначенные для анализа HTTP/HTTPS-трафика и блокировки запросов на основе заранее заданных правил или сигнатур. Сигнатуры представляют собой характерные признаки уже известной вредоносной активности. Правила могут включать в себя черные списки, содержащие известные сигнатуры атак, а также белые списки, определяющие допустимые запросы [1, 2].

Однако, несмотря на свою эффективность, такой подход имеет ряд недостатков. Традиционный WAF, основанный на правилах и сигнатурах атак, не способен выявлять атаки, использующие неизвестные доселе уязвимости, именуемые «уязвимостями нулевого дня». Кроме того, WAF требует регулярной настройки и обновления правил и баз сигнатур атак, что увеличивает затраты времени и ресурсов. Дополнительной проблемой является большое количество ложных срабатываний, которые возникают при попытке охватить в правилах широкий спектр угроз.

Непрерывный мониторинг представляет собой процесс постоянного наблюдения за состоянием веб-приложения и анализа входящего трафика, осуществляемый центром мониторинга информационной безопасности (SOC). Этот метод позволяет своевременно выявлять подозрительную активность и реагировать на инциденты [3–5].

Главным недостатком данного подхода является то, что его реализация сопряжена с высокими затратами человеческих ресурсов, поскольку требует привлечения специалистов для обработки логов, настройки фильтров и анализа данных. Даже при интеграции с SIEM системами, которые автоматизируют сбор и корреляцию событий, значительная часть работы остается ручной, что снижает скорость реагирования на новые угрозы.

Тестирование на проникновение, или пентестинг, заключается в имитации атак на веб-приложение с целью выявления уязвимостей до их непосредственной эксплуатации злоумышленниками. Этот метод позволяет оценить уровень безопасности системы и разработать меры по устранению слабых мест [6].

Данный подход имеет ряд ограничений. Тестирование проводится разово или с определенной периодичностью, что не обеспечивает защиты в реальном времени. Кроме того, эффективность метода напрямую зависит от квалификации специалистов и используемых инструментов, а выявленные уязвимости требуют дополнительных ресурсов для их исправления, что замедляет процесс реагирования на нововыявленные угрозы.

Ниже представлена сравнительная таблица рассмотренных выше традиционных методов защиты веб-приложений по ряду критериев.

Таблица 1

Традиционные методы защиты и их ограничения

Критерий	WAF	Непрерывный мониторинг	Пентестинг
Скорость реагирования на новые угрозы	Высокая	Низкая	Низкая
Адаптация к изменяющимся угрозам	Низкая	Средняя	Низкая
Степень потребности в ручной настройке	Высокая	Высокая	Высокая
Точность выявления неизвестных угроз	Низкая	Средняя	Низкая

Нейронные сети: базовые принципы

Традиционные методы защиты не обладают достаточной гибкостью и не способны адаптироваться к быстро меняющимся условиям. В связи с этим технологии машинного обучения, в частности нейронные сети, становятся перспективным инструментом для повышения уровня безопасности веб-приложений.

Нейронные сети различаются по архитектуре и применимости к задачам безопасности веб-приложений. Многослойные перцептроны (MLP) используются для классификации запросов, например, для выявления потенциальных угроз, таких как SQL-инъекции или межсайтовый скриптинг. Сверточные нейронные сети (CNN) эффективны для анализа структурированных данных, таких как сетевые пакеты, помогая обнаруживать аномалии, указывающие на вторжения. Рекуррентные нейронные сети (RNN), особенно сети с долгой краткосрочной памятью (LSTM), обрабатывают последовательности событий, выявляя сложные атаки, такие как попытки захвата учетной записи, учитывая временные зависимости в поведении пользователей или паттернах запросов [7].

Если сеть содержит множество скрытых слоев, ее называют глубокой нейронной сетью (DNN). Глубокими могут быть любые из трех вышеперечисленных типов (MLP, CNN или RNN). Глубокие нейросети хорошо справляются с более сложными задачами, такими как прогнозирование неизвестных угроз на основе исторических данных [8].

Основное преимущество нейронных сетей заключается в их способности обрабатывать сложные и неструктурированные данные, такие как сетевой трафик, логи или последовательности пользовательских действий. Это позволяет им выявлять скрытые закономерности, что недоступно традиционным методам, основанным на статических правилах и сигнатурах. Кроме того, нейронные сети способны масштабироваться для анализа больших объемов информации, что делает их эффективным инструментом для обеспечения безопасности в условиях динамичных и разнообразных киберугроз.

Применение нейросетей в безопасности веб-приложений

Нейронные сети находят применение в различных аспектах обеспечения безопасности веб-приложений, позволяя преодолевать ограничения традиционных методов. Их способность анализировать большие объемы данных и выявлять скрытые закономерности делает возможным автоматизацию процессов, которые ранее требовали значительных ресурсов или были попросту невозможны.

Традиционные WAF основаны на использовании статических правил и сигнатур угроз, что приводит к необходимости их регулярной актуализации. Применение методов машинного обучения способно улучшить работу WAF путем автоматического формирования и адаптации правил на основе анализа трафика. Например, нейронные сети могут обучаться на примерах вредоносных запросов, таких как попытки SQL-инъекций, и отличать их от легитимных без заранее заданных сигнатур. Это снижает количество ложных срабатываний и повышает точность фильтрации. Кроме того, такая автоматизация уменьшает потребность в ручной настройке, что сокращает затраты времени и ресурсов [9].

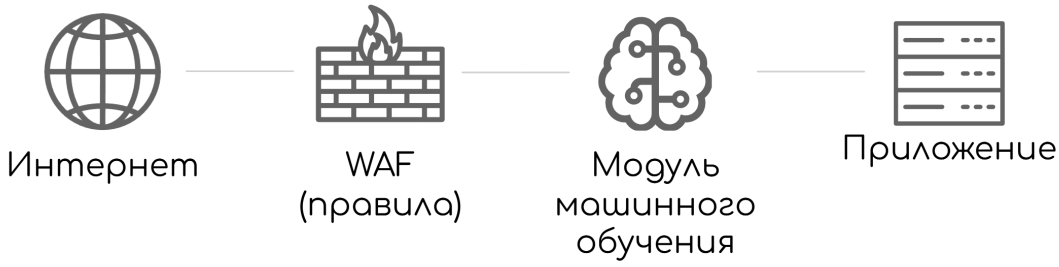


Рис. 1. Схема работы WAF с применением нейросетей

Помимо совершенствования WAF, нейронные сети открывают новые возможности и для оптимизации непрерывного мониторинга. Непрерывный мониторинг, реализуемый через центры мониторинга информационной безопасности (SOC) с использованием SIEM систем, традиционно требует значительных усилий специалистов для анализа логов и поведения пользователей. Интеграция машинного обучения, включая нейронные сети, в SIEM позволяет автоматизировать анализ больших объемов данных. Например, нейронные сети могут выявлять аномалии, указывающие на сложные атаки, такие как попытки эксплуатации бизнес-логики или несанкционированный доступ, обнаруживая скрытые паттерны в трафике. Это снижает нагрузку на команду SOC, ускоряет обнаружение инцидентов и повышает точность реагирования [10].

Преимущества и недостатки использования нейросетей

Использование нейронных сетей в обеспечении безопасности веб-приложений связано как с очевидными преимуществами, так и с определенными недостатками. Эти аспекты определяют их потенциал и ограничения в сравнении с традиционными методами защиты.

Основным преимуществом нейронных сетей является автоматизация процессов. Они способны самостоятельно анализировать данные и принимать решения, что снижает потребность в ручной настройке и участии специалистов. Например, автоматическое обновление правил WAF или обработка логов сокращают затраты времени и ресурсов.

Вторым важным достоинством выступает адаптивность. Нейронные сети обучаются на новых данных, что позволяет им реагировать на ранее неизвестные угрозы, такие как уязвимости нулевого дня, в отличие от статических методов, ограниченных заранее заданными правилами.

Кроме того, высокая скорость обработки больших объемов информации обеспечивает оперативное обнаружение аномалий, что критично для защиты от атак в реальном времени.

Однако применение нейронных сетей сопряжено с рядом вызовов. Во-первых, для их эффективной работы требуется значительный объем данных для обучения. Недостаток качественных данных может снизить точность анализа и привести к ошибкам в обнаружении угроз.

Во-вторых, настройка и оптимизация нейронных сетей представляют собой сложный процесс, требующий специализированных знаний и вычислительных ресурсов. Это увеличивает затраты на внедрение по сравнению с традиционными инструментами.

Наконец, нейронные сети часто функционируют как «черный ящик»: их решения трудно интерпретировать, что затрудняет анализ причин срабатывания и устранение ложных срабатываний.

Сравнительный анализ WAF с применением машинного обучения

Для оценки практической применимости нейронных сетей в области обеспечения безопасности веб-приложений был проведен сравнительный анализ существующих решений WAF, использующих машинное обучение. Сравнение проводилось по следующим критериям: качество безопасности (True Positive Rate), качество обнаружения (False Positive Rate) и сбалансированная оценка.

Качество безопасности (TPR) рассчитывалось по формуле (1).

$$TPR = \frac{TP}{P} = \frac{TP}{TP+FN} = 1 - FNR, \quad (1)$$

где

а) FNR – доля пропущенных атак (FN/P)

б) TP – правильно обнаруженные атаки

в) P – общее количество реальных атак ($TP+FN$)

г) FN – пропущенные атаки

Качество обнаружения (TNR) рассчитывалось по формуле (2).

$$TNR = \frac{TN}{N} = \frac{TN}{TN+FP} = 1 - FPR, \quad (2)$$

где

а) FPR – доля ложных срабатываний (FP/N)

б) TN – правильно пропущенные легитимные запросы

в) N – общее количество безопасных запросов ($TN+FP$)

г) FP – ложные срабатывания

Сбалансированная оценка (BA) рассчитывалось по формуле (3).

$$BA = \frac{TPR+TNR}{2}, \quad (3)$$

где

а) TNR – качество обнаружения

б) TPR – качество безопасности

Анализ основан на открытых тестах OpenAppSec 2024, а также на документациях производителей и отзывах пользователей. Результаты представлены в таблице 2. [11]

Таблица 2

Сравнительный анализ WAF с применением машинного обучения

WAF Solution	Configuration	Security Quality (True Positive Rate)	Detection Quality (False Positive Rate)	Balanced Accuracy
Microsoft Azure WAF	OWASP CRS 3.2 ruleset	97.526%	54.242%	71.642%
AWS WAF	AWS managed ruleset	79.751%	5.8%	86.976%
AWS WAF	AWS managed ruleset and F5 Ruleset	80.372%	5.879%	87.246%
CloudFlare WAF	Managed and OWASP Core Rulesets	69.3%	0.062%	84.619%
F5 NGINX App Protect WAF	Default profile	77.9%	1.808%	88.046%
F5 NGINX App Protect WAF	Strict profile	97.849%	22.084%	86.882%
NGINX ModSecurity	OWASP CRS 4.3.0	92.028%	17.523%	87.253%
open-appsec / CloudGuard WAF	Default (High Confidence)	99.368%	1.436%	98.966%
open-appsec / CloudGuard WAF	Critical Confidence	99.087%	0.81%	99.139%
Imperva Cloud WAF	Default configuration	11.97%	0.009%	55.981%
F5 BIG-IP Advanced WAF	Rapid Deployment Policy configuration	78.89%	2.8%	88.045%
Fortinet FortiWeb	Default configuration	68.971%	20.925%	74.023%
Google Cloud Armor	Preconfigured ModSecurity rules (Sensitivity level 2)	83.537%	50.283%	66.627%

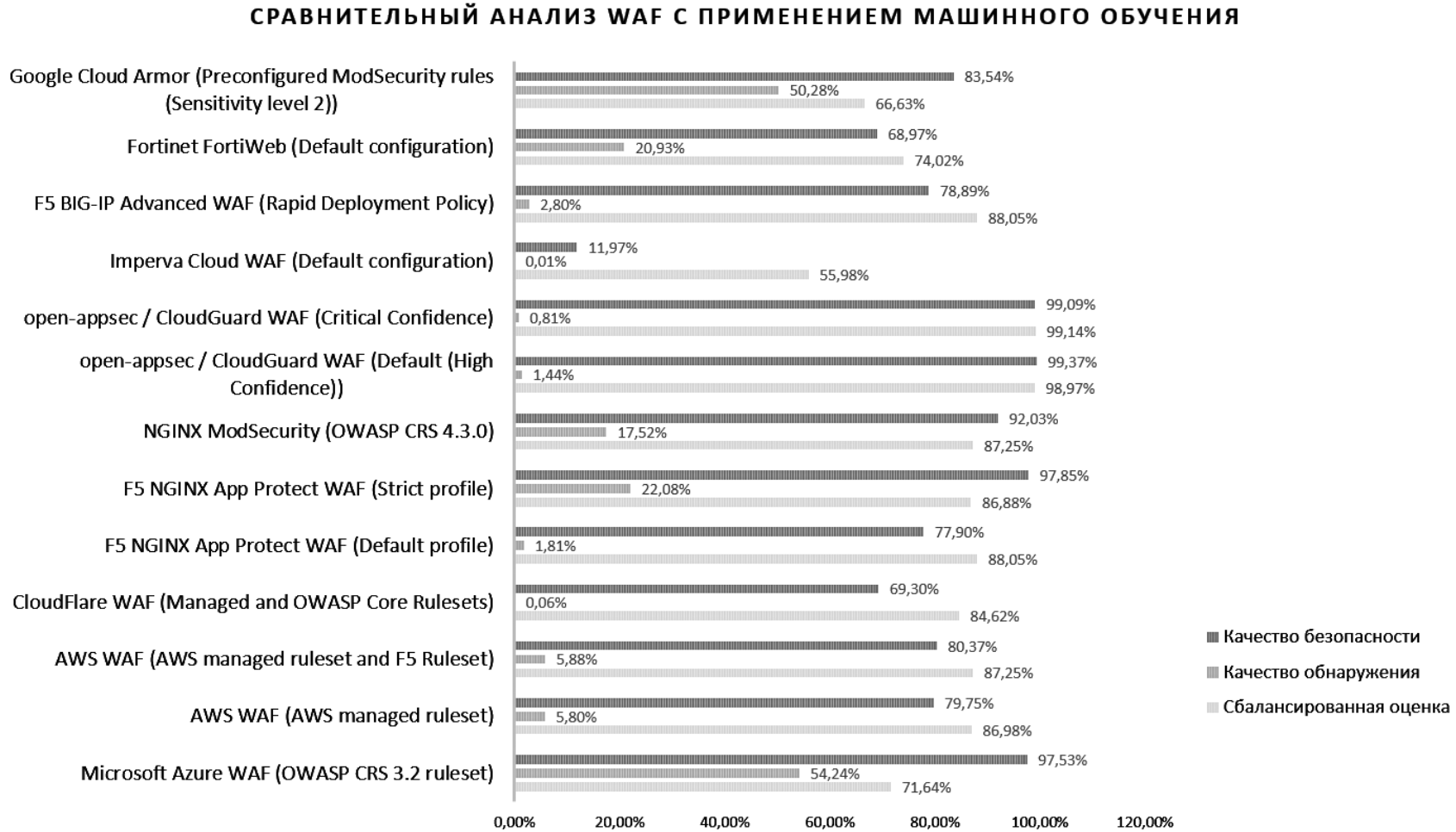


Рис. 2. Сравнительный анализ WAF с применением машинного обучения, гистограмма

На рис. 2 также представлена линейчатая гистограмма, наглядно представляющая данные анализа.

Из анализа видно, что решение openappsec CloudGuard WAF в конфигурации Default и Critical Confidence демонстрирует наивысшую сбалансированную оценку BA (примерно 99%), сочетая максимально высокий TPR (более 99%) с минимальным FPR (менее 1.5%). Это указывает на его способность надежно обнаруживать атаки при почти нулевых ложных срабатываниях, что делает его наиболее перспективным для практического применения.

Заключение

Нейронные сети представляют собой перспективное развитие традиционных подходов к обеспечению безопасности веб-приложений. Они устраняют ряд ограничений классических методов, таких как необходимость частой ручной настройки и неспособность противостоять неизвестным угрозам, за счет автоматизации, адаптивности и высокой скорости обработки данных. Их внедрение не исключает использования существующих инструментов -межсетевых экранов уровня приложения и непрерывного мониторинга, а дополняет их.

Перспективы дальнейшего применения нейронных сетей связаны с их интеграцией в текущие системы защиты. Комбинированный подход, сочетающий точность WAF и возможности анализа аномалий нейронными сетями, может повысить эффективность обнаружения и предотвращения атак. Аналогично, автоматизация мониторинга с использованием нейронных сетей способна снизить нагрузку на специалистов, сохраняя при этом высокую надежность защиты. Нейронные сети открывают новые возможности для повышения уровня безопасности веб-приложений в условиях растущей сложности киберугроз, однако использование их требует преодоления некоторых технических и организационных ограничений.

Литература

1. Solar MSS. WAF (Web Application Firewall): как устроен и где применяется. RT Solar, 2025. URL: https://rt-solar.ru/services/mss/blog/5367/ (дата обращения: 25.03.2025).

2. Yandex Cloud. Введение в WAF (Web Application Firewall). Yandex Cloud, 2025. URL: https://yandex.cloud/ru/docs/glossary/waf?utm_referrer=https%3A%2F%2Fyandex.ru%2F (дата обращения: 27.03.2025).

3. DigitalTatarstan. Security Operation Center (SOC) на пальцах: из чего состоит и кому нужен. Habr, 2023. URL: https://habr.com/ru/companies/digital_tatarstan/articles/775844/ (дата обращения: 23.03.2025).

4. SolarSecurity. Проблема непрерывной защиты веб-приложений. Взгляд со стороны исследователей и эксплуатантов. Habr, 2017. URL: https://habr.com/ru/companies/solarsecurity/articles/331786/ (дата обращения: 28.03.2025).

5. Avpavlov. Проблема непрерывной защиты веб-приложений. Взгляд со стороны исследователей и эксплуатантов. Часть 2. Habr, 2017. URL: https://habr.com/ru/companies/solarsecurity/articles/332846/ (дата обращения: 24.03.2025).

6. Al Shelbli H.M. A study on penetration testing process and tools // 2018 IEEE Long Island Systems, Applications and Technology Conference (LISAT). Farmingdale, NY, USA, 2018. Pp. 1-5.

7. Розенко Ю. Что такое модели и архитектуры нейросетей. AdminVPS, 2025. URL: https://adminvps.ru/blog/chto-takoe-modeli-i-arhitektury-nejrosetej/ (дата обращения: 23.03.2025).

8. Boesch G. Глубокая нейронная сеть: 3 популярных типа (MLP, CNN и RNN). Viso.ai, 2021. URL: https://viso.ai/deep-learning/deep-neural-network-three-popular-types/ (дата обращения: 27.03.2025).

9. Elvin_GSNV. Как я создал межсетевой экран с помощью свёрточных нейронных сетей для веб-приложений с микросервисной архитектурой. Habr, 2022. URL: https://habr.com/ru/articles/677232/ (дата обращения: 25.03.2025).

10. Ptsecurity. Учимся на чужих ошибках: как прокачать SIEM с помощью machine learning. Habr, 2024. URL: https://habr.com/ru/companies/pt/articles/848986/ (дата обращения: 28.03.2025).

11. Rozenfeld B. Лучшие WAF решения в 2024-2025 годах: сравнение в реальных условиях. OpenAppSec, 2024. URL: https://www.openappsec.io/post/best-waf-solutions-in-2024-2025-real-world-comparison (дата обращения: 24.03.2025).

References

1. Solar MSS. WAF (Web Application Firewall): kak ustroen i gde primenyaetsya. RT Solar, 2025. URL: https://rt-solar.ru/services/mss/blog/5367/ (data obrashhenija: 25.03.2025).

2. Yandex Cloud. Vvedenie v WAF (Web Application Firewall). Yandex Cloud, 2025. URL: https://yandex.cloud/ru/docs/glossary/waf?utm_referrer=https%3A%2F%2Fyandex.ru%2F (data obrashhenija: 27.03.2025).

3. DigitalTatarstan. Security Operation Center (SOC) na pal'tsakh: iz chego sostoit i komu nuzhen. Habr, 2023. URL: https://habr.com/ru/companies/digital_tatarstan/articles/775844/ (data obrashhenija: 23.03.2025).

4. SolarSecurity. Problema nepreryvnoj zashhity veb-prilozhenij. Vzglyad so storony issledovatelej i ekspluatantov. Habr, 2017. URL: https://habr.com/ru/companies/solarsecurity/articles/331786/ (data obrashhenija: 28.03.2025).

5. Avpavlov. Problema nepreryvnoj zashhity veb-prilozhenij. Vzglyad so storony issledovatelej i ekspluatantov. Chast' 2. Habr, 2017. URL: https://habr.com/ru/companies/solarsecurity/articles/332846/ (data obrashhenija: 24.03.2025).

6. Al Shelbli H.M. A study on penetration testing process and tools. 2018 IEEE Long Island Systems, Applications and Technology Conference (LISAT). Farmingdale, NY, USA, 2018. Pp. 1-5. DOI: 10.1109/LISAT.2018.8378035.

7. Rozenko Yu. Chto takoe modeli i arkhitektury nejrosetej. AdminVPS, 2025. URL: https://adminvps.ru/blog/chto-takoe-modeli-i-arhitektury-nejrosetej/ (data obrashhenija: 23.03.2025).

8. Boesch G. Deep Neural Network: The 3 Popular Types (MLP, CNN and RNN). Viso.ai, 2021. URL: https://viso.ai/deep-learning/deep-neural-network-three-popular-types/ (data obrashhenija: 27.03.2025).

9. Elvin_GSNV. Kak ya sozdal mezhsetevoj ekran s pomoshh'yu svyortochnykh nejronnykh setej dlya veb-prilozhenij s mikroservisnoj arkhitekturoj. Habr, 2022. URL: https://habr.com/ru/articles/677232/ (data obrashhenija: 25.03.2025).

10. Ptsecurity. Uchimsya na chuzhikh oshibkakh: kak prokachat' SIEM s pomoshh'yu machine learning. Habr, 2024. URL: https://habr.com/ru/companies/pt/articles/848986/ (data obrashhenija: 28.03.2025).

11. Rozenfeld B. Best WAF Solutions in 2024-2025: Real-World Comparison. OpenAppSec, 2024. URL: https://www.openappsec.io/post/best-waf-solutions-in-2024-2025-real-world-comparison (data obrashhenija: 24.03.2025).

ЧАСТИКОВА Вера Аркадьевна, кандидат технических наук, доцент, доцент кафедры кибербезопасности и защиты информации ФГБОУ ВО «Кубанский государственный технологический университет». 350000, г. Краснодар, ул. Красная, 135. E-mail: chastikova_va@mail.ru

ТЕСЛЕНКО Александр Александрович, студент кафедры кибербезопасности и защиты информации ФГБОУ ВО «Кубанский государственный технологический университет». 350000, г. Краснодар, ул. Красная, 135. E-mail: alextesl2018@mail.ru

АЛИЕВ Мунир Казбекович, студент кафедры кибербезопасности и защиты информации ФГБОУ ВО «Кубанский государственный технологический университет». 350000, г. Краснодар, ул. Красная, 135. E-mail: opennauka24@gmail.com

ИГНАТЕНКО Илья Сергеевич, студент кафедры кибербезопасности и защиты информации ФГБОУ ВО «Кубанский государственный технологический университет». 350000, г. Краснодар, ул. Красная, 135. E-mail: ilaignatenko1010@gmail.com

CHASTIKOVA Vera Arkadyevna, Candidate of Technical Sciences, Associate Professor, Associate Professor of the Department of Cybersecurity and Information Protection of the Kuban State Technological University. 350000, Krasnodar, Krasnaya street, 135. E-mail: chastikova_va@mail.ru

TESLENKO Alexander Alexandrovich, student of the Department of Cybersecurity and Information Protection of the Kuban State Technological University. 350000, Krasnodar, Krasnaya street, 135. E-mail: alextesl2018@mail.ru

ALIEV Munir Kazbekovich, student of the Department of Cybersecurity and Information Protection of the Kuban State Technological University. 350000, Krasnodar, Krasnaya street, 135. E-mail: alextesl2018@mail.ru

IGNATENKO Ilya Sergeevich, student of the Department of Cybersecurity and Information Protection of the Kuban State Technological University. 350000, Krasnodar, Krasnaya street, 135. E-mail: ilaignatenko1010@gmail.com

УДК 004.056:005.942 Вестник УрФО № 2(56) / 2025, с. 74–79

Новиков Г. А., Кузьмин А. А., Кузьмина У. В. DOI: 10.14529/secur250208

ПРИМЕНЕНИЕ МЕТОДОВ ХАОС-ИНЖИНИРИНГА В ИНФРАСТРУКТУРЕ ОТДЕЛА ИБ ПРЕДПРИЯТИЯ

Статья посвящена исследованию применения хаос инжиниринга в контексте информационной безопасности. На основе созданного цифрового двойника бизнес-процесса обеспечения информационной безопасности организации проводится тестирование реакций систем защиты. В статье описана методика развертывания тестовой среды, сбор метрик и проведения ручных тестов с последующей автоматизацией инцидентов. Результаты исследования демонстрируют потенциал использования данного подхода для повышения устойчивости систем безопасности.

Ключевые слова: *хаос инжиниринг, информационная безопасность, цифровой двойник, SIEM, IPS/IDS, DLP.*

Novikov G. A., Kuzmin A. A., Kuzmina U. V.

THE USE OF CHAOS ENGINEERING METHODS IN THE INFRASTRUCTURE OF THE INFORMATION SECURITY DEPARTMENT OF THE ENTERPRISE

The article is devoted to the study of the application of chaos engineering in the context of information security. Based on the created digital twin of the organization's information security business process, the reactions of security systems are being tested. The article describes the methodology for deploying the test environment, collecting metrics and conducting manual tests with subsequent incident automation. The results of the study demonstrate the potential of using this approach to increase the stability of security systems.

Keywords: *chaos engineering, information security, digital twin, SIEM, IPS/IDS, DLP.*

Введение

Современные информационные системы имеют сложную распределенную архитектуру. Традиционные методы тестирования и аудита безопасности часто не способны обнаружить скрытые уязвимости, которые могут быть выявлены лишь в условиях реальных инцидентов. Хаос инжиниринг, зародившийся для тестирования распределенных систем (на примере Chaos Monkey от Netflix), сегодня активно применяется для имитации отказов и проверки реакции систем в условиях неопределенности. В данной статье предлагается концепция применения хаос инжиниринга в области информационной безопасности, когда наряду с техническими сбоями моделируются атаки, утечки данных и иные инциденты, способные проверить работоспособность технических средств защиты.

Особое внимание уделяется необходимости создания цифрового двойника – виртуальной копии бизнес-процесса обеспечения информационной безопасности организации, где моделируются все ключевые процессы. Такой цифровой двойник позволяет проводить тесты и эксперименты без риска для боевой инфраструктуры.

В контексте информационной безопасности такой подход может помочь не только оценить настройку технических средств защиты, но и протестировать оперативную реакцию ИБ-команды на инциденты [1]. Цифровой двойник позволяет проводить испытания, не подвергая риску реальную инфраструктуру [2].

Цель исследования – оценка эффективности тестовых сценариев хаос инжиниринга на основе разработанной методологии создания тестовой среды цифрового двойника отдела информационной безопасности.

Задачи исследования:

1. Разработка методологии развертывания цифрового двойника отдела информационной безопасности.

2. Развертывание тестовой среды цифрового двойника бизнес-процесса обеспечения информационной безопасности организации.

3. Проведение ручного тестирования и оценка реакции системы на атаку.

В данной статье рассмотрим разработку методологии развертывания цифрового двойника отдела информационной безопасности. Перед развертыванием цифрового двойника необходимо определить его архитектуру, функциональные требования и цели тестирования. Основные вопросы, которые следует учитывать:

- Какие бизнес-процессы должны быть смоделированы?
- Какие компоненты инфраструктуры необходимо включить?
- Какие инциденты будут тестироваться?
- Какие метрики будут собираться?

Методология развертывания цифрового двойника включает в себя несколько ключевых этапов, одним из которых является определение границ моделируемой системы. Важно отметить, что для создания адекватного цифрового двойника не требуется полное воспроизведение всей инфраструктуры организации. Современные компании, особенно крупные предприятия, обладают разветвленной IT-инфраструктурой, включающей в себя множество взаимосвя-занных сервисов, сетевых сегментов и вычислительных мощностей. Однако для успешного функционирования цифрового двойника достаточно включить в его состав только те элементы, которые непосредственно задействованы в функционировании одного или нескольких ключевых бизнес-процессов. Такой подход позволяет оптимизировать вычислительные ресурсы, снизить затраты на развертывание и управление двойником, а также сосредоточиться на моделировании именно тех аспектов системы, которые имеют наибольшее значение для анализа и тестирования.

Поэтому была выделены основная функциональная область:

• Мониторинг активности пользователей, анализ логов, сработки систем защиты (SIEM, IDS/IPS, DLP) и реакцию системы на выявленные угрозы.

Многие современные корпоративные сервисы изначально проектируются и разворачиваются в виртуальных средах, будь то частные облака, контейнерные платформы или традиционные гипервизоры. Это означает, что создание цифрового двойника может быть реализовано путем репликации уже имеющихся виртуальных машин и контейнеров, что существенно снижает временные и финансовые затраты на его построение.

Цифровой двойник построен с учетом принципов сегментирования сети, многоуровневой защиты и контроля доступа и состоит из следующих компо-нентов (см. рисунок 1):

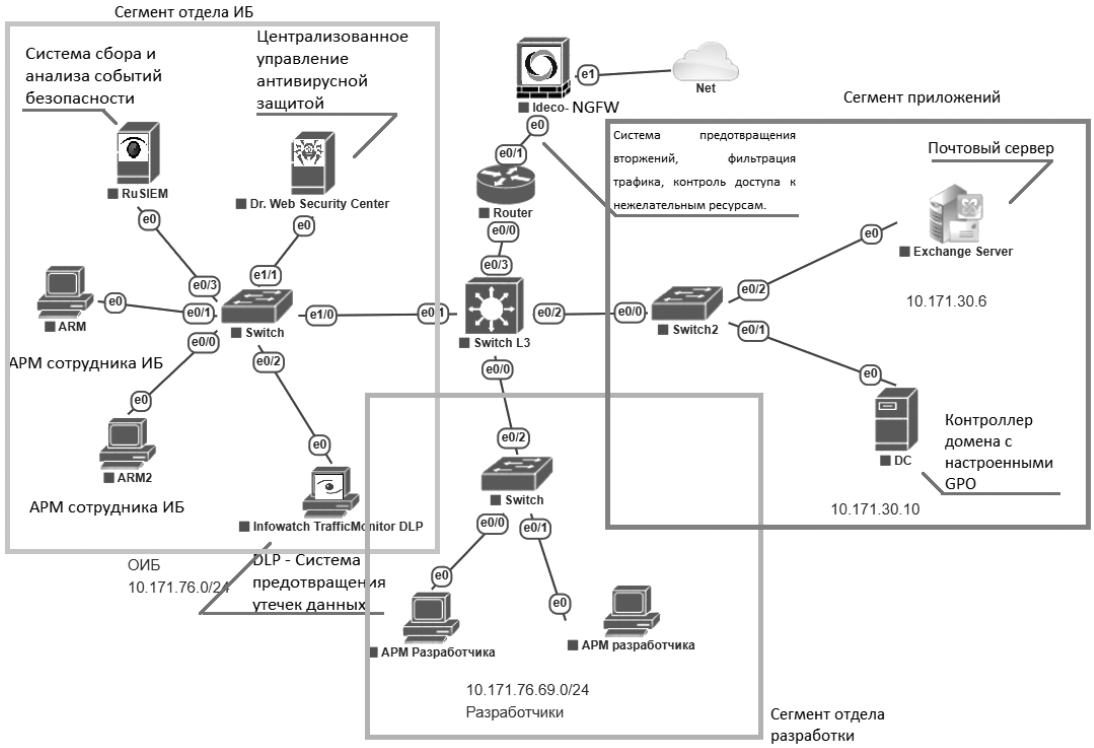


Рис. 1. Схема сети цифрового двойника

Для тестовой среды выделена изолированная сеть с подсетью 10.171.0.0/16, в которой размещены все компоненты цифрового двойника. Такое решение позволяет отслеживать сетевую активность, моделировать атаки и анализировать их последствия без риска воздействия на рабочую инфраструктуру. [3]

Для оценки эффективности защиты в тестовой среде определены сле-дующие сценарии атак:

• Тестирование механизма предотвращения утечек данных с использованием InfoWatch DLP.

• Анализ сработок IDS/IPS при моделировании аномального сетевого трафика.

• Проверка работы антивирусных решений при загрузке вредоносных файлов.

• Тестирование срабатывания правил SIEM и отображения событий.

•

Основные этапы разработанной методологии тестирования инцидентов на основе хаос-инжиниринга:

• Описывается сценарий инцидента. Например, симуляция заражения сети «червем» или утечка данных посредством генерации большого объема исходящего трафика.

• Формируются ожидания от эксперимента. Например, сработки или несработки, ожидания от реакции системы.

• Определяются метрики реакции. Задаются ключевые показатели: время обнаружения инцидента (фиксируется системой SIEM), скорость восстановления системы, число ложных срабатываний и эффективность оповещений.

• Проводится имитация инцидента вручную. Администратор вручную запускает сценарий, наблюдая за работой системы и сбором метрик через интегрированный центр.

• Проверка ожидаемого результата и реального, если вдруг что-то пошло не так, это выявит аномалию.

•

Ход эксперимента:

В рамках исследования устойчивости корпоративной инфраструктуры к целевым атакам смоделирован сценарий компрометации сети через легитимные учетные данные с последующим запуском вируса-шифровальщика. Эксперимент проводился в изолированном тестовом окружении, повторяющем инфраструктуру организации, фазы атаки и реакция систем безопасности представлены на рисунке 2.[4]

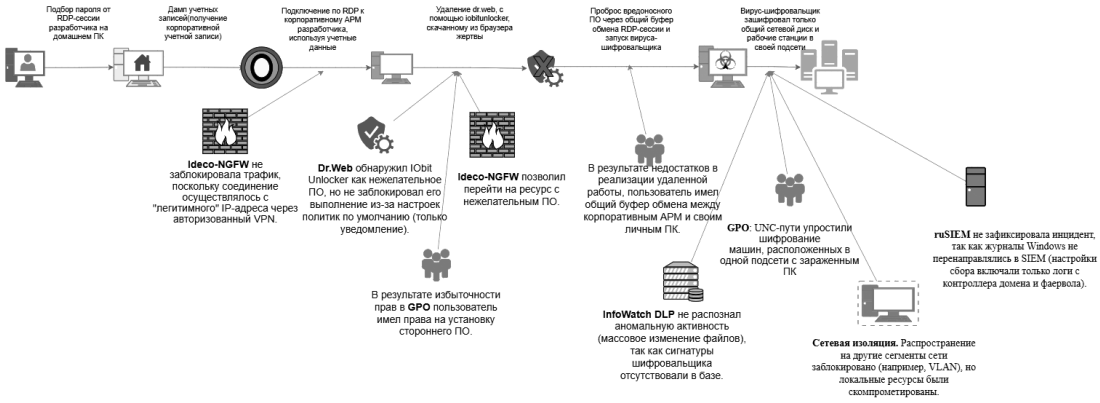


Рис. 2. Фазы атаки и реакция систем защиты

В ходе эксперимента были выявлены следующие критические уязвимости инфраструктуры:

• Антивирусная защита продемонстрировала недостаточную эффективность: продукт Dr.Web функционировал на базовых политиках, не настроенных на блокировку потенциально нежелательных приложений (PUA).

• Конфигурация групповых политик (GPO) содержала ряд ошибок. В частности, политики безопасности разрешали установку неподписанного программного обеспечения, включая такие утилиты, как IObit Unlocker. Активация общего буфера обмена между локальными и удалёнными машинами при под-ключении через RDP создала канал для передачи вредоносного кода. Дополнительным риском стало отсутствие двухфакторной аутентификации для RDP-сессий, что упростило несанкционированный доступ.

• Системы мониторинга не обеспечили должного уровня контроля. Платформа ruSIEM не агрегировала логи с рабочих станций, что сделало невидимыми ключевые события: удаление антивирусного ПО, запуск шифроваль-щика и очистку журналов. Решение InfoWatch DLP не отслеживало аномальные паттерны доступа к данным, включая массовое изменение тысяч файлов, что позволило злоумышленникам действовать без обнаружения.

• Сетевая архитектура также имела существенные недостатки. Наличие открытых SMB-ресурсов (включая административные шары типа C$) упростило горизонтальное перемещение внутри сети. Система Ideco не блокировала загрузку подозрительных утилит, так как исключения для «ресурсов разработчиков» отсутствовали в правилах фильтрации. Кроме того, неконтролируемые VPN-сессии не разрывались при длительном простое, создавая дополнительные векторы для атак.

Результаты:

• Частичный успех защиты: Сетевая изоляция предотвратила катастрофическое распространение шифровальщика, а зашифрованные данные удалось восстановить из бекапов.

• Ложная легитимность: Действия злоумышленника (использование валидных учетных данных, штатных утилит) не вызывали подозрений у большинства систем.

• Несоответствие политик: GPO, разрешающие управление службами и установку ПО, противоречили принципу минимальных привилегий.

• Как можно увидеть в результате проведенного эксперимента, шифрования избежать не удалось, хотя злоумышленника можно было отсечь на каждом из этапов, эти несовершенства СЗИ были выявлены, благодаря эксперименту в виртуальной среде, что позволило усилить безопасность организации, поскольку были обозначены дыры в безопасности.

Основные преимущества методологии.

Цифровой двойник отдела информационной безопасности предоставляет уникальную возможность для проведения тестовых сценариев без риска для реальной инфраструктуры. Среди основных преимуществ можно выделить:

• Безопасность эксперимента.
• Комплексное тестирование процессов.
• Гибкость и масштабируемость.

Заключение

В статье представлена комплексная методология применения хаос инжи-ниринга для обеспечения информационной безопасности с использованием цифрового двойника и интегрированного центра сбора метрик. На основе которой проведено дальнейшее развертывание тестовой среды, имитирующей работу бизнес-процесса обеспечения информационной безопасности организации, которая позволяет проводить эксперименты по моделированию различных ин-цидентов, таких как заражение вредоносным ПО, утечка данных, DoS-атаки и фишинговые атаки.

Для развертывания тестовой среды была использована виртуализационная платформа PNetLab, которая позволила создать изолированное окружение и минимизировать влияние экспериментов на рабочие системы. Для обеспечения гибкости и отказоустойчивости реализована возможность быстрого восстановления (rollback) тестовой среды после проведения экспериментов.

Таким образом, применение хаос инжиниринга в сфере информационной безопасности представляет собой перспективное направление, способное повысить уровень защиты корпоративных систем за счет выявления скрытых уязвимостей, оптимизации процессов реагирования и непрерывного совершенствования мер безопасности.

Литература

1. Розенталь Кейси, Джонс Нора Хаос-инжиниринг / К. Розенталь, Н. Джонс - 1-е изд. – Москва: Изд-во ДМК Пресс, 2021. - 284 с.

2. Digital Twin — цифровая копия физической системы // Хабр URL: https://habr.com/ru/articles/887936/ (дата обращения: 05.02.2025).

3. Афанасьева С.В., Кузьмина У.В. Основные проблемы при создании и обслуживании центров мониторинга информационной безопасности // Безо-пасность информационного пространства: сборник научных трудов XXI Все-российской научно-практической конференции студентов, аспирантов и молодых ученых. Екатеринбург, 2023. С. 29-31.

4. Chaos Engineering: искусство умышленного разрушения. Часть 1 // Хабр URL: https://habr.com/ru/companies/flant/articles/460367/ (дата обращения: 02.02.2025).

5. Кузьмина У.В., Абзалутдинов Д.Р., Бараков К.Я. Создание модуля киберполигона, имитирующего компьютерные атаки // Актуальные проблемы современной науки, техники и образования. 2023. Т. 14. № 1. С. 54-57.

References

1. Keysi Rozental', Nora Dzhons Khaos-inzhiniring. - 1 izd. - DMK Press, 2021. - 284 s.

2. Digital Twin — tsifrovaya kopiya fizicheskoy sistemy // Khabr URL: https://habr.com/ru/articles/887936/ (data obrashcheniya: 05.02.2025).

3. Afanas'eva S.V., Kuz'mina U.V. Osnovnye problemy pri sozdanii i ob-sluzhivanii tsentrov monitoringa informatsionnoy bezopasnosti // Bez-opasnost' in-formatsionnogo prostranstva: sbornik nauchnykh trudov XXI Vse-rossiyskoy nauch-no-prakticheskoy konferentsii studentov, aspirantov i molo-dykh uchenykh. Ekate-rinburg, 2023. S. 29-31

4. Chaos Engineering: iskusstvo umyshlennogo razrusheniya. Chast' 1 // Khabr URL: https://habr.com/ru/companies/flant/articles/460367/ (data obrashche-niya: 02.02.2025).

5. Kuzmina U.V., Abzalutdinov D.R., Barakov K.Ya. Creation of a cyber range module simulating computer attacks // Actual problems of modern science, technology and education. 2023. Vol. 14. No. 1. P. 54-57.

КУЗЬМИН Александр Андреевич, студент кафедры информатики и информационной безопасности, ФГБОУ ВО «Магнитогорский государственный технический университет им. Г.И. Носова». 455000, г. Магнитогорск, пр. Ленина, 38. E-mail: kuzminsa2002@gmail.com

НОВИКОВ Глеб Александрович, студент кафедры информатики и ин-формационной безопасности, ФГБОУ ВО «Магнитогорский государственный технический университет им. Г.И. Носова». 455000, г. Магнитогорск, пр. Лени-на, 38. E-mail: glebbuns@gmail.com

КУЗЬМИНА Ульяна Владимировна, кандидат технических наук, доцент кафедры информатики и информационной безопасности, ФГБОУ ВО «Магнитогорский государственный технический университет им. Г.И. Носова». 455000, г. Магнитогорск, пр. Ленина, 38. E-mail: Ylianapost@gmail.com

KUZMIN Alexander Andreevich, student of the Department of computer science and information security, Federal State Budgetary Educational Institution of Higher Education "Magnitogorsk State Technical University named after G.I. No-sov". 455000, Magnitogorsk, Lenin Ave., 38. Email: kuzminsa2002@gmail.com

NOVIKOV Gleb Alexandrovich, student of the Department of computer science and information security, Federal State Budgetary Educational Institution of Higher Education "Magnitogorsk State Technical University named after G.I. No-sov". 455000, Magnitogorsk, Lenin Ave., 38. Email: glebbuns@gmail.com

KUZMINA Ulyana Vladimirovna, Candidate of Technical Sciences, Associate Professor of the Department of computer science and information security, Federal State Budgetary Educational Institution of Higher Education "Magnitogorsk State Technical University named after G.I. Nosov". 455000, Magnitogorsk, Len-in Ave., 38. Email: Ylianapost@gmail.com

УДК 004.056.53 Вестник УрФО № 2(56) / 2025, с. 80–95

Хорев А. А., Чачилло Т. В. DOI: 10.14529/secur250209

ИССЛЕДОВАНИЕ ПОДВЕРЖЕННОСТИ ТЕЛЕФОННОГО АППАРАТА «ВЫСОКОЧАСТОТНОЙ НАКАЧКЕ»

В статье приведены результаты экспериментальных исследований возможности перехвата акустической речевой информации из выделенного помещения методом «высокочастотной накачки» телефонного аппарата. Установлено, что телефон подвержен «высокочастотной накачке» в диапазоне частот 168–270 МГц. Получены зависимости отношения «сигнал/шум» (по напряженности поля излучаемого сигнала) от мощности ВЧ-генератора и уровня акустического давления для частоты накачки 230 МГц. По показателю словесная разборчивость речи установлено, что методом «высокочастотной накачки» телефонного аппарата в реальных условиях перехват речевой информации из выделенного помещения не возможен.

Ключевые слова: *технический канал утечки акустической речевой информации, высокочастотное навязывание, высокочастотное облучение, высокочастотная накачка.*

Khorev A. A., Chachillo T. V.

INVESTIGATION OF THE TELEPHONE'S EXPOSURE TO "HIGH-FREQUENCY PUMPING"

The article presents the results of experimental studies of the possibility of intercepting acoustic speech information from a dedicated room using the "high-frequency pumping" method of a telephone device. It was found that the phone is subject to "high-frequency pumping" in the frequency range of 168-270 MHz. The dependences of the signal-to-noise ratio (in terms of the field strength of the emitted signal) on the power of the RF generator and the acoustic pressure level for the 230 MHz pumping frequency are obtained. According to the indicator of verbal intelligibility of speech, it has been established that the method of "high-frequency pumping" of a telephone in real conditions is not possible to intercept speech information from a dedicated room.

Keywords: *technical channel of acoustic speech information leakage, high-frequency imposition, high-frequency irradiation, high-frequency pumping.*

Введение

Наряду с традиционными методами перехвата акустической речевой информации из выделенных помещений (ВП) в настоящее время все более широко стали использоваться такие методы перехвата, такие как «высокочастотное навязывание» (ВЧН) и «высокочастотное облучение» (ВЧО) [1 – 3].

Использование данных методов перехвата основано на том, что в некоторые вспомогательные технические средства и системы (ВТСС), устанавливаемые в ВП, входят акустоэлектрические преобразователи модуляторного типа (АЭПМТ). Такие ВТСС часто называют ВТСС, обладающие микрофонным эффектом [1].

Принцип работы АЭПМТ основан на их свойстве изменять свои параметры (емкость, индуктивность или сопротивление) под действием акустического поля, создаваемого источником акустических колебаний [1]. Изменение их параметров незначительное, поэтому при протекании через них электрического тока его амплитуда также меняется незначительно. Но если такие акустоэлектрические преобразователи входят в состав высокочастотного колебательного контура, то даже незначительное изменение их параметров приводит к существенному изменению амплитуды высокочастотного сигнала I на частоте резонанса f_p этого контура (см. рисунок 1) [4]:

$$I = \frac{U_c}{|Z|}. \qquad (1)$$

$$|Z| = \sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}. \qquad (2)$$

$$f_p = \frac{1}{2\pi \cdot \sqrt{L \cdot C}}. \qquad (3)$$

Для создания канала утечки информации в ВТСС или подают гармонический высокочастотный сигнал по соединительным линиям, или облучают ВТСС высокочастотным гармоническим сигналом [1].

В первом случае высокочастотный сигнал проходя через нелинейный контур модулируется по закону изменения звукового давления, воздействующего на нелинейный элемент. «Отраженный» от нелинейного контура сигнал принимается и демодулируется приемным устройством средства разведки, гальванически подключенным к соединительной линии ВТСС. Такой метод перехвата называется методом «высокочастотного навязывания». Создаваемый таким методом канал утечки информации называется активным акустоэлектрическим (см. рисунок 2) [1].

Во втором случае в нелинейном колебательном контуре сигнал появляется в следствие взаимодействия электромагнитного поля, создаваемого высокочастотным генератором средства перехвата, с элементами ВТСС, которые выполняют функцию случайных антенн. При воздействии акустического сигнала на нелинейный элемент колебательного контура происходит изменение его сопротивления и резонансной частоты, а в следствие этого – модуляция переизлученного ВТСС высокочастотного сигнала по закону изменения звукового давления, воздействующего на нелинейный элемент. Переизлученный высокочастотный сигнал принимается и демодулируется радиоприемным устройством средства разведки. Такой метод перехвата акустической речевой информации называется методом «высокочастотного облучения», а создаваемый таким методом канал утечки информации называется активным акустоэлектромагнитным (см. рисунок 2) [1].

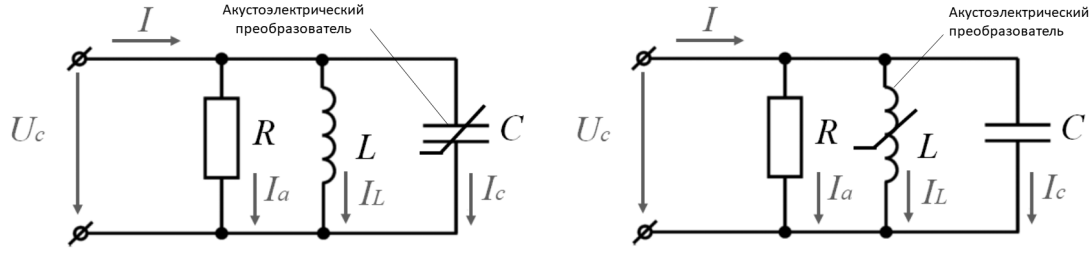


Рис. 1. Резонансный контур с нелинейной емкостью (а) и нелинейной индуктивностью (б)

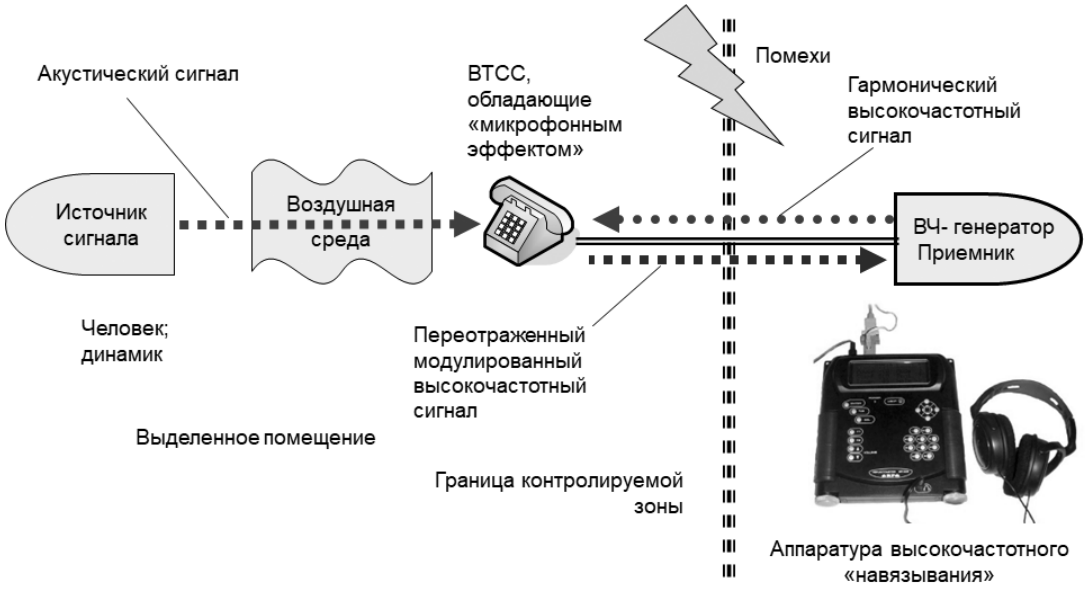


Рис. 2. Схема акустоэлектрического канала утечки речевой информации, создаваемого методом «высокочастотного навязывания»

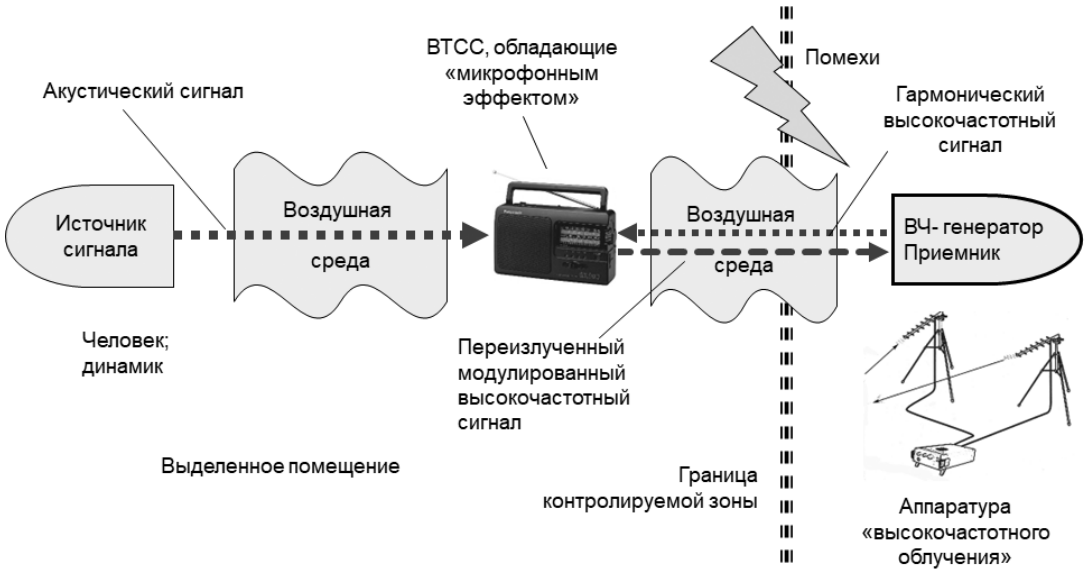


Рис. 3. Схема акустоэлектромагнитного канала утечки речевой информации, создаваемого методом «высокочастотного облучения»

Технические каналы утечки информации, реализуемые с использованием методов «высокочастотного навязывания» и «высокочастотного облучения» довольно подробно рассмотрены в работах различных авторов [1–3,5–17], а методы и методики контроля подверженности технических средств акустоэлектрическим и акустоэлектромагнитным преобразованиям – в [1– 3, 18 – 23].

При использовании для перехвата акустической речевой информации метода «высокочастотного навязывания» высокочастотный сигнал, формируемый генератором гармонических колебаний, по соединительной линии подается на вход ВТСС, где модулируется и отражается обратно в линию. Но, при этом вследствие того, что по элементам ВТСС протекает высокочастотный модулированный сигнал, вокруг них возникает электромагнитное излучение, которое распространяется в окружающее пространство. Принимая и демодулируя это электромагнитное из-

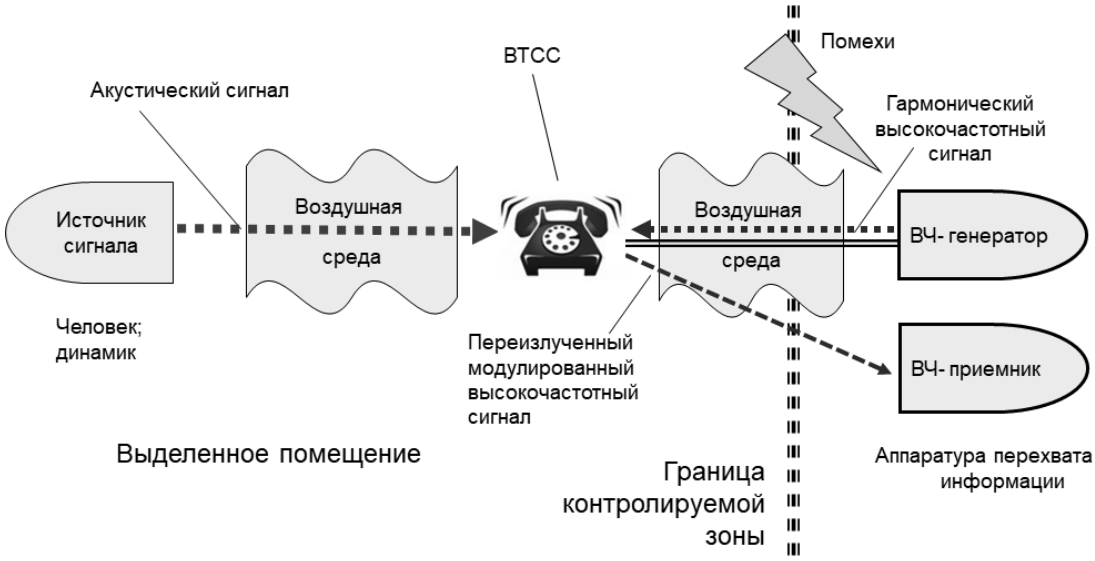


Рис. 4. Схема технического канала утечки акустической речевой информации, создаваемого методом «высокочастотной накачки» ВТСС

лучение можно перехватить ведущиеся в помещении разговоры (см. рисунок 4).

Такой метод перехвата информации в начале 90-х годов получил название метод «высокочастотной накачки», в 2000-х годах – метод «высокочастотной прокачки». Однако, по нашему мнению, термин «высокочастотная накачка» более корректный.

Для контроля защищенности акустической речевой информации от утечки за счёт «высокочастотного навязывания», «высокочастотного облучения» и «высокочастотной прокачки» используются программно-аппаратные комплексы «Тандем», «Гранат», «ПАРНАС-ЭХО12Е» и др. [24 – 26].

Одним из наиболее распространенных ВТСС, устанавливаемых в ВП и подверженных акустоэлектрическим и акустоэлектромагнитным преобразованиям, являются телефонные аппараты.

В работе [13] были проведены экспериментальные исследования телефонных аппаратов типа KX-TS2361UA (Panasonic), TULIPAN-319 (1990 год выпуска, Польша), ТА-72, ТА-600, ТА-4100 на подверженность «высокочастотному навязыванию» в диапазоне частот от 10 кГц до 10 МГц. Авторами было установлено, что все выше перечисленные телефоны не подвержены «высокочастотному навязыванию», так как глубина амплитудной модуляции менее 10% для всех уровней громкости речи. По мнению авторов невозможность перехвата речевой информации методом «высокочастотного навязывания» обусловлена наличием дифференциального трансформатора (с индуктивностью обмотки 24 мГн) и ёмкостью кабеля трубки (300–500 пФ), которые существенно ослабляют ВЧ-сигнал навязывания.

Проведенный анализ показал, что в доступной литературе отсутствуют результаты исследований возможностей перехвата речевой информации методом «высокочастотной накачки» телефонного аппарата.

Поэтому с целью оценки возможности перехвата акустической речевой информации методом «высокочастотной накачки» телефонного аппарата были проведены экспериментальные исследования.

Экспериментальные исследования

В качестве показателя оценки возможностей перехвата акустической речевой информации по акустоэлектрическим и акустоэлектомагнитным каналам используется словесная разборчивость речи, в основу оценки которой положена инструментально-расчетная методика, предполагающая измерение отношений сигнал/шум на входе приемного устройства средства разведки в октавных полосах и расчет словесной разборчивости речи по эмпирическим формулам [1].

Используем данный показатель и при оценке возможностей перехвата речевой информации методом «высокочастотной накачки» телефонного аппарата.

Рис. 5. Лабораторный стенд для проведения экспериментальных исследований
Состав лабораторного стенда: 1 – телефонный аппарат «VEF TA-D» (TA611D; 2 – анализатор спектра «Rohde&Schwarz FSH8»; 3 – электрическая дипольная активная измерительная антенна П6-51; 4 – шумомер «Экофизика-110А»; 5 – активная акустическая системы «Волна-Д»; 6 – генератор сигналов «Rohde&Schwarz SMB100A»

Для проведения экспериментальных исследований был разработан лабораторный стенд, схема и состав которого приведены на рисунке 5.

Экспериментальные исследования проводились в несколько этапов.

На первом этапе были определены частоты, на которых телефонный аппарат (ТА) подвержен «высокочастотной накачке».

Методика проведения экспериментальных исследований заключалась в следующем:

1. К телефонному аппарату «VEF TA-D» подключался генератор сигналов «Rohde&Schwarz SMB100A».

2. На расстоянии 1 м от ТА устанавливалась электрическая дипольная активная измерительная антенна П6-51, подключенная к анализатору спектра «Rohde&Schwarz FSH8».

3. Акустическая система (звуковая колонка) источника тестового акустического сигнала устанавливалась на расстоянии 1 м от исследуемого ТА и направляется на него.

4. В месте расположения ТА устанавливался измерительный микрофон, подключенный к шумомеру «Экофизика-110А».

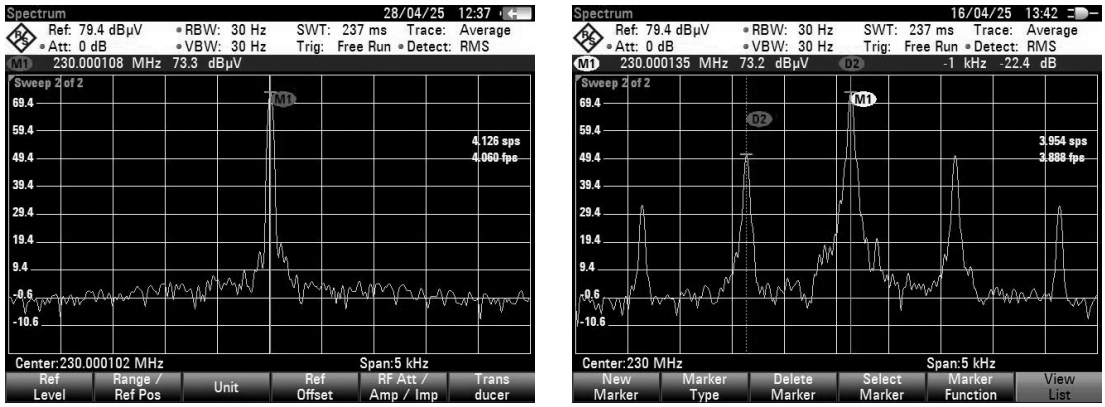
5. Включался генератор сигналов «Rohde&Schwarz SMB100A». Выходная мощность генератора устанавливается на максимальное значение (25 дБм).

6. Включался анализатор спектра «Rohde&Schwarz FSH8». Устанавливалась полоса обзора *Span*: 5 кГц и полоса пропускания *RBW*: 30 Гц.

7. Включался генератор тестового сигнала (ГС) в режиме генерации синусоидального сигнала на частоте 1000 Гц. Устанавливался максимальный уровень громкости тестового сигнала.

8. Шумомером измерялся уровень тестового сигнала (L_u) в 4-й октавной полосе со среднегеометрической частотой 1000 Гц (измеренный уровень громкости сигнала составил 110 дБ).

9. Высокочастотный генератор и анализатор спектра синхронно перестраивались в



а) б)

Рис. 6. Спектрограмма переизлученного телефонным аппаратом сигнала
(частота сигнала ВЧ-навязывания 230 МГц, частота тестового акустического сигнала 1кГц)
при выключенном тестовом сигнале (а) и включенном тестовом сигнале (б)

диапазоне частот от 30 кГц до 300 МГц с шагом 10 кГц.

10. При обнаружении в спектре принимаемого сигнала модуляционных составляющих тестового сигнала фиксировалась частота сигнала ВЧ-навязывания ($F_{c.i}$) и производилось измерение уровней спектральных составляющих сигнала (на частоте $F_{c.i} \pm 1000$ Гц) при включенном тестовом сигнале $U_{(с+ш).i}$, дБ(мкВ) и выключенном тестовом сигнале $U_{ш.i}$, дБ(мкВ) (см. рисунок 6).

Производился расчет напряженности поля информативного сигнала по формуле

$$E_{c.i} = 10lg\left[10^{0,1\left(U_{(с+ш).i}+k_{a.i}\right)} - 10^{0,1\left(U_{ш.i}+k_{a.i}\right)}\right], \quad (4)$$

где $U_{(с+ш).i}$ – уровень модуляционной составляющей при включенном тестовом сигнале на i-й частоте, дБ(мкВ);

$U_{ш.i}$ – уровень шумов при включенном тестовом сигнале на i-й частоте, дБ(мкВ);

$k_{a.i}$ – калибровочный коэффициент антенны П6-51 на i-й частоте, дБ(1/м).

График зависимости напряженности поля информативного сигнала (а) и шумов (б) от частоты приведен на рисунке 7.

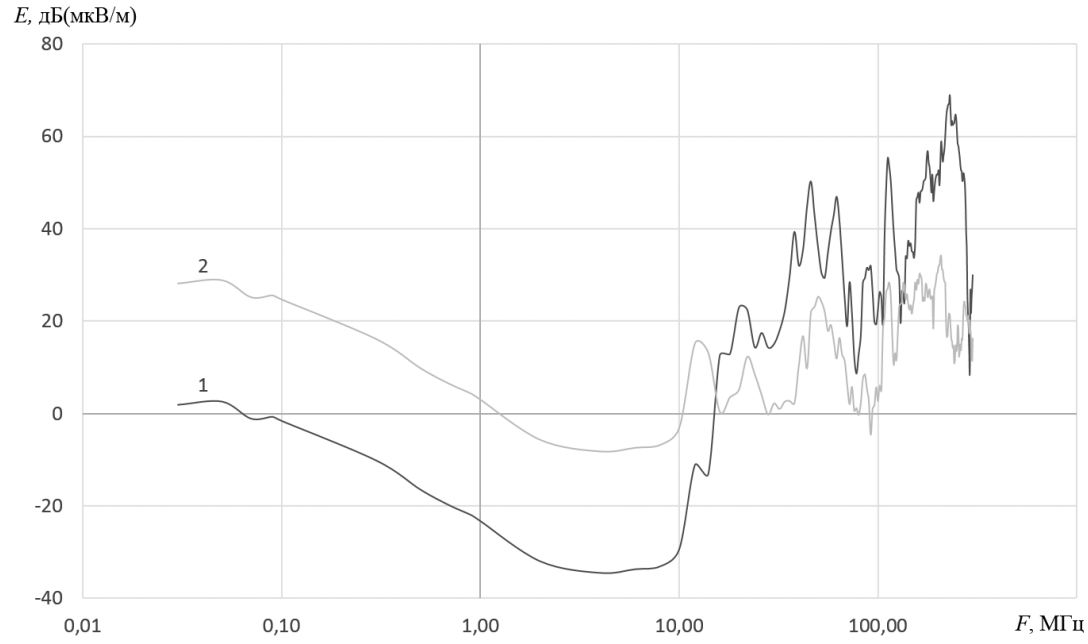


Рис. 7. График зависимости напряженности поля информативного сигнала (1) и шумов (2)
от частоты при максимальной выходной мощности ВЧ-генератора (25 dBm) и максимальном уровне акустического
сигнала на частоте 1 кГц (110 дБ)

Анализ графика, представленного на рисунке 7, показывает, что зависимость напряженности поля информативного сигнала от частоты сигнала ВЧ-накачки носит нелинейный характер. Телефон наиболее подвержен ВЧ-накачке в диапазоне частот от 168 МГц до 270 МГц. Максимальный уровень напряженности поля информативного сигнала наблюдается на частоте 230 МГц.

На втором этапе была определена зависимость напряженности поля информативного сигнала от частоты тестового акустического сигнала.

Измерения проводились на частоте ВЧ-сигнала навязывания F_c = 230 МГц в соответствие с методикой, изложенной выше. При этом в качестве тестовых сигналов использовались тональные сигналы на среднегеометрических частотах 1 – 7 октавных полос: 125 Гц, 250 Гц, 500 Гц, 1000 Гц, 2000 Гц, 4000 Гц и 8 кГц.

Полоса обзора *Span* анализатора спектра устанавливалась в зависимости от частоты тестового сигнала в диапазоне от 1 до 20 кГц.

Рассчитанные значения напряженности поля информативного сигнала в зависимости от частоты тестового акустического сигнала приведены в таблице 1 и на рисунке 8.

Таблица 1

Измеренные значения напряженности поля информативного сигнала в октавных полосах

Номер октавной полосы	Частотные границы октавной полосы, Гц	Среднегеометрическая частота октавной полосы, Гц	Напряженность поля информативного сигнала, дБ(мкВ/м)
1	90 - 175	125	40,07
2	175 - 355	250	43,08
3	355 - 710	500	62,2
4	710 - 1400	1000	68,9
5	1400 - 2800	2000	67,1
6	2800 - 5600	4000	35,49
7	5600 - 11200	8000	23,08

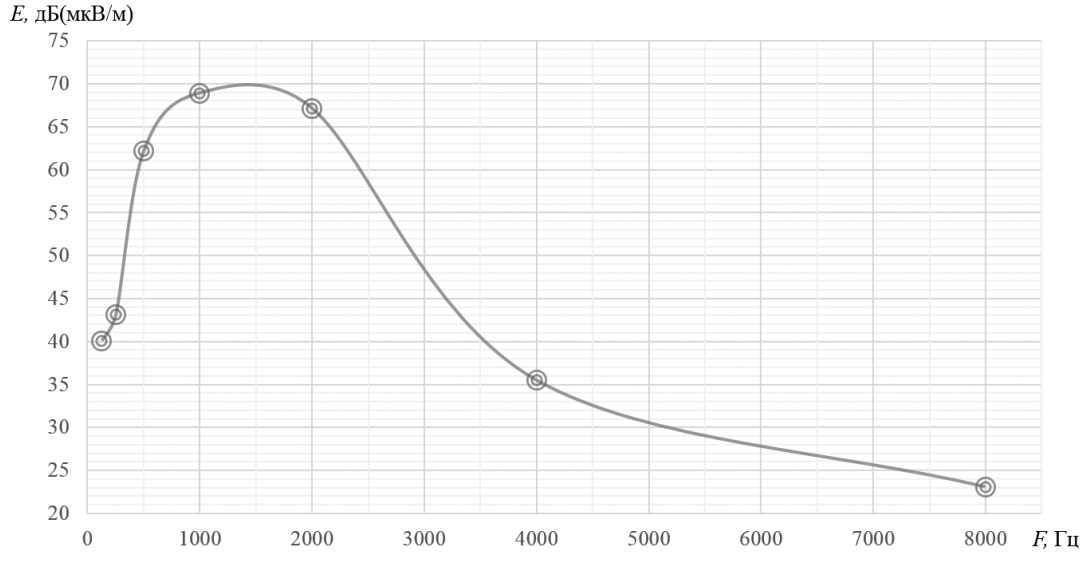


Рис. 8. График зависимости напряженности поля информативного сигнала от частоты тестового акустического сигнала (частота ВЧ сигнала навязывания 230 МГц, выходная мощность ВЧ генератора 25 дБм, уровень тестового акустического сигнала 110 дБ)

Анализ данных, представленных в таблице 1 и на рисунке 8, показывает, что телефонный аппарат наиболее подвержен акустоэлектромагнитным преобразованиям в 3 – 5 октавных полосах. Уровень информативного сигнала в диапазоне частот свыше 3,5 кГц (6 и 7 октавные полосы) более чем на 30 дБ ниже, чем в 3 – 5 октавных полосах. Очевидно это связано с добротностью резонансного контура, в который входит нелинейный элемент.

Проведенные исследования показали, что 1 и 7 октавные полосы практически не влияют на разборчивость речи [1, 27]. Поэтому, в дальнейшем экспериментальные исследования проводились только в 2 – 6 октавных полосах.

На третьем этапе была определена зависимость уровня информативного сигнала от уровня тестового акустического сигнала.

Измерения проводились на частоте ВЧ-сигнала навязывания F_c = 230 МГц при максимальной мощности генератора (25 дБм) в соответствие с методикой, изложенной выше. При этом в качестве тестовых сигналов использовались тональные сигналы на среднегеометрических частотах 2 – 6 октавных полос: 250 Гц, 500 Гц, 1000 Гц, 2000 Гц и 4000 Гц.

Уровень тестовых сигналов с шагом 1 дБ изменялся от максимального (110 дБ) до уровня, при котором модуляционная составляющая не обнаруживалась на фоне шумов. Результаты измерений представлены на рисунке 9.

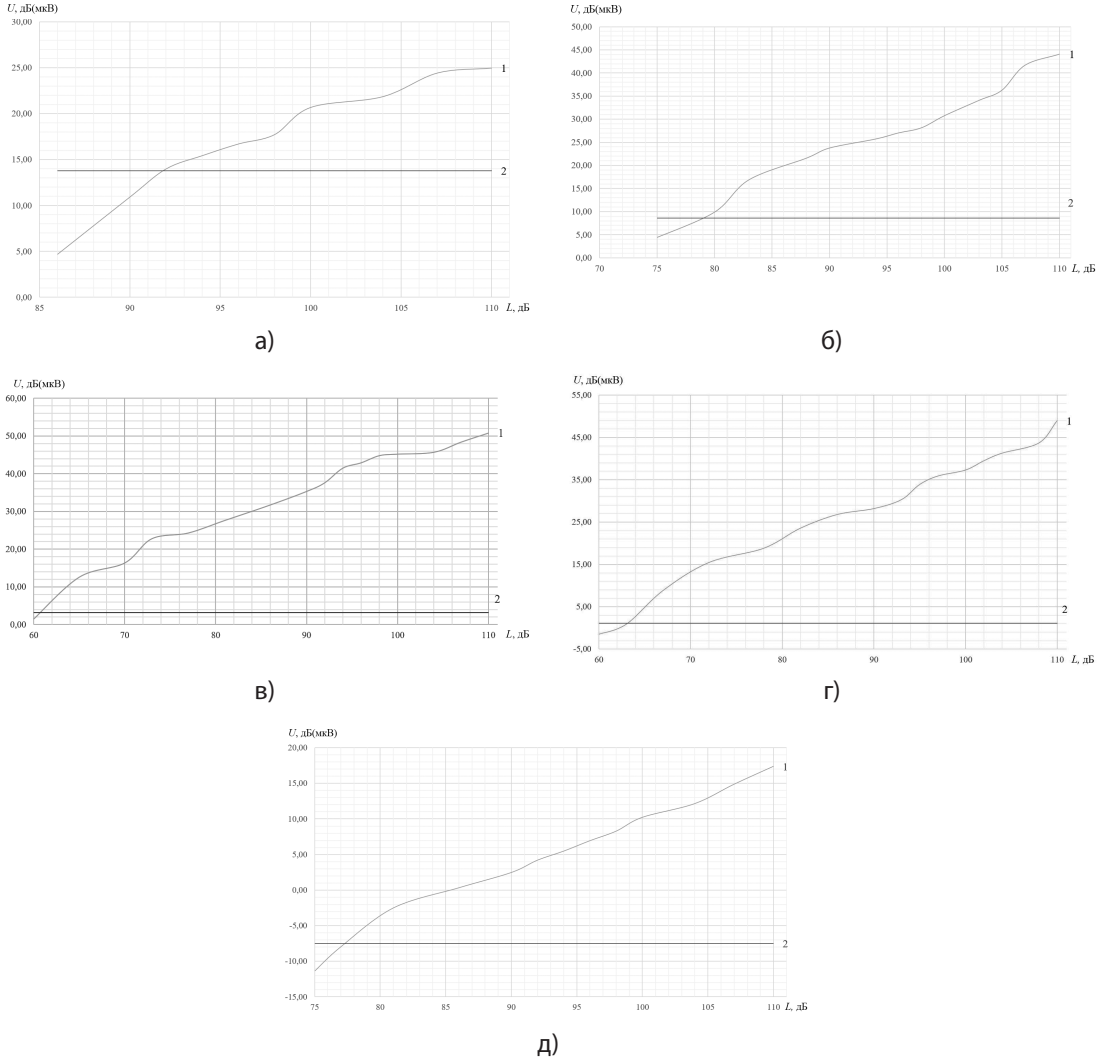


Рис. 9. Графики зависимости уровня информативного сигнала (1) и шумов (2) от уровня тестового акустического сигнала (частота ВЧ сигнала навязывания 230 МГц, выходная мощность ВЧ генератора 25 дБм): а) – для 2–й октавной полосы; б) – для 3–й октавной полосы; в) – для 4–й октавной полосы; г) – для 5–й октавной полосы; д) –для 6–й октавной полосы

Анализ данных, представленных на рисунке 9, показывает, что с возрастанием уровня тестового акустического сигнала уровень информативного сигнала растет при практически неизменном уровне шумов. Наибольший уровень информационного сигнала наблюдается в 3 – 5 октавных полосах.

На четвертом этапе была определена зависимость уровня информативного сигнала от выходной мощности ВЧ-генератора.

Измерения проводились на частоте ВЧ-сигнала навязывания F_c = 230 МГц при максимальном уровне тестовых сигналов (110 дБ). При этом в качестве тестовых сигналов использовались тональные сигналы на среднегеометрических частотах 2 – 6 октавных полос: 250 Гц, 500 Гц, 1000 Гц, 2000 Гц и 4000 Гц.

Выходная мощность высокочастотного генератора уменьшалась от максимального значения (25 дБм) с шагом 1 дБ до уровня, при котором модуляционная составляющая не обнаруживалась на фоне шумов. Результаты измерений представлены на рисунке 10.

Исходя из полученных графиков, представленных на рисунке 10, можно сделать вывод о том, что с возрастанием выходной мощности ВЧ генератора уровень информативного сигнала растет, но при этом растет и уровень шумов. Возможно это связано с тем, что анализатор спектра обладает недостаточной избирательностью при полосе пропускания RBW: 30 Гц.

Для измерения напряженности поля информативного сигнала использовался анали-

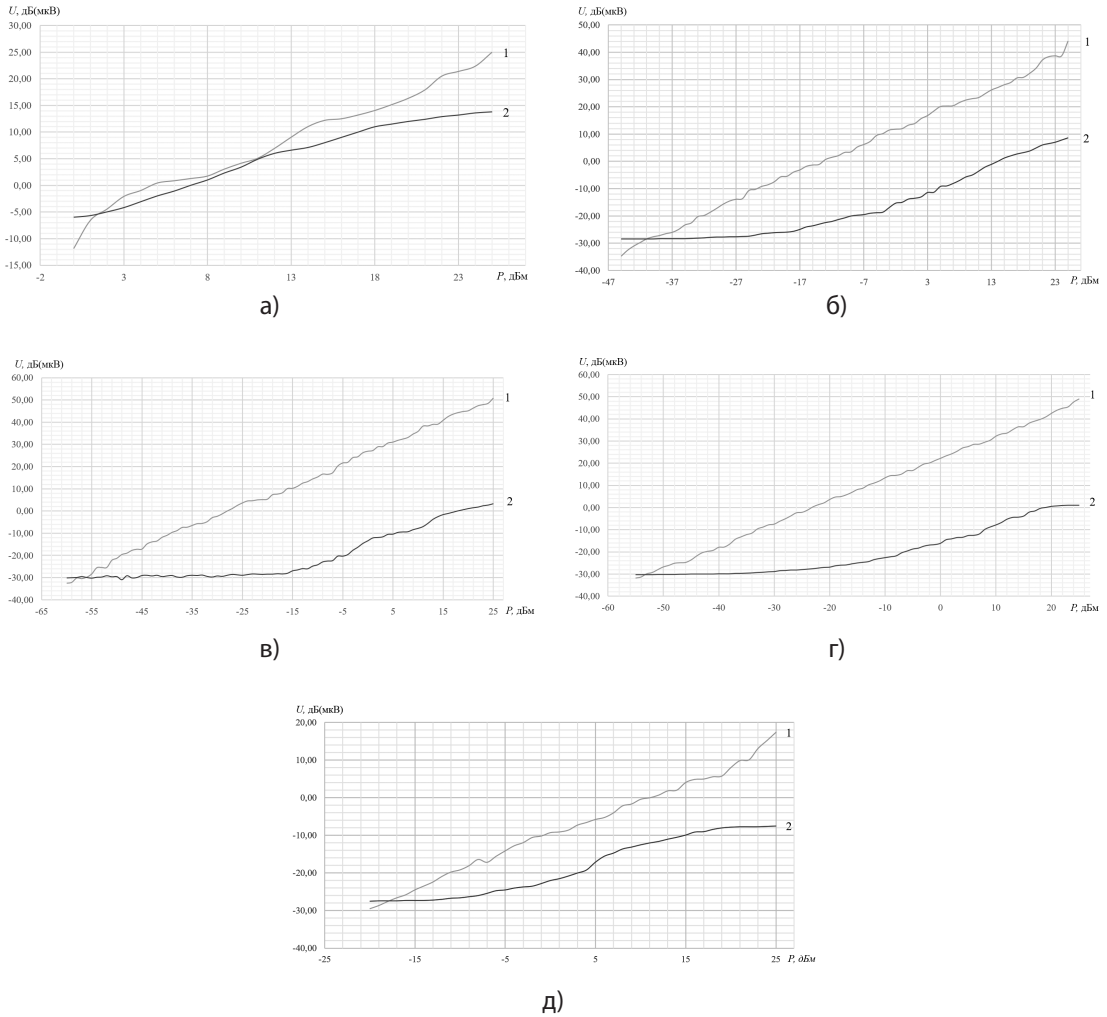


Рис. 10. Графики зависимости уровня информативного сигнала (1) и шумов (2) от выходной мощности ВЧ генератора (частота ВЧ сигнала навязывания 230 МГц, уровень тестового акустического сигнала 110 дБ): а) – для 2–й октавной полосы; б) – для 3–й октавной полосы; в) – для 4–й октавной полосы; г) –для 5–й октавной полосы; д) – для 6–й октавной полосы

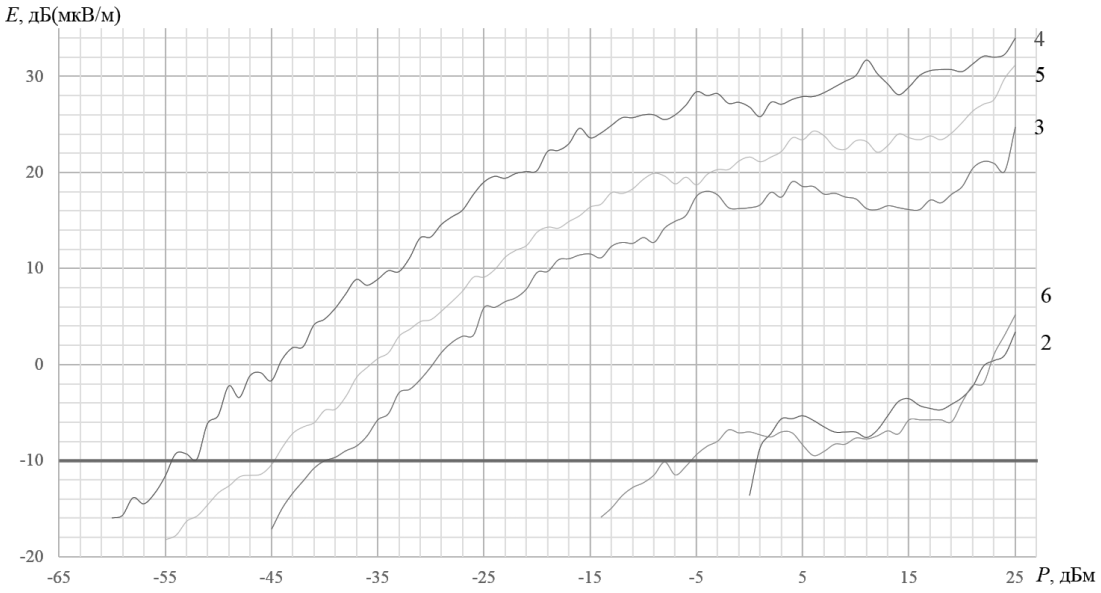


Рис. 11. Графики зависимости отношения «сигнал/шум» от уровня выходной мощности ВЧ-генератора (частота ВЧ сигнала навязывания 230 МГц, уровень тестового акустического сигнала 110 дБ): а) – для 2–й октавной полосы; б) – для 3–й октавной полосы; в) – для 4–й октавной полосы; г) – для 5–й октавной полосы; д) – для 6–й октавной полосы

затор спектра с уровнем собственных шумов –191 дБ/Гц и измерительная антенна П6-51 с уровнем собственных шумов – 167,4 дБ/Гц.

Анализ измеренных уровней шумов (см. рисунки 9 и 10) значительно выше собственных шумов анализатора спектра и измерительной антенны. Следовательно, на входе приемного устройства средства разведки уровень шумов будет определяться не его собственными шумами, а шумами, возникающими в телефонном аппарате вследствие акустоэлектрических преобразований.

С учетом этого предположения были получены зависимости отношений сигнал/шум для пяти октавных полос от мощности ВЧ-генератора и уровня тестового акустического сигнала. Расчет отношений сигнал/шум проводился по формуле:

$$q_j = 20lg\left[\frac{10^{0,05E_{c.j}}\cdot\sqrt{\Delta F_u}}{10^{0,05E_{ш.j}}\cdot\sqrt{\Delta F_j}}\right], \qquad (5)$$

где $E_{c.j}$ – измеренная напряженность поля информативного сигнала в j-й октавной полосе (см. формулу (4)) , дБ(мкВ/м);

$E_{ш.j}$ – измеренная напряженность поля шума в j-й октавной полосе, дБ(мкВ/м);

ΔF_u – полоса пропускания анализатора спектра, Гц;

ΔF_j – ширина j-й октавной полосы, Гц.

Полученные зависимости приведены на рисунках 11 и 12.

Анализ графиков, представленных на рисунках 11 и 12, показал, что:

• максимальное отношение сигнал/шум наблюдается в 4-й октавной полосе со среднегеометрической частотой 1000 Гц;

• отношение сигнал/шум растет при увеличении мощности ВЧ-сигнала накачки. При этом характер зависимостей сложный (нелинейный). При мощности сигнала до – 15 дБм, характер зависимости более «крутой», чем в диапазоне от – 15 дБм до 25 дБм. Например, для 4-й октавной полосы при увеличении мощности сигнала от – 45 до – 15 дБм (на 30 дБ), отношение сигнал/шум увеличивается от 0 до 35 дБ (на 35 дБ). А при увеличении мощности сигнала с – 15 до 15 дБм (на 30 дБ), отношение сигнал/шум увеличивается с 24 до 29 дБ (на 5 дБ);

• отношение сигнал/шум также растет при увеличении уровня тестового акустического сигнала. Характер зависимостей близкий к линейной. Например, для 4-й октавной полосы при увеличении уровня тестового сигнала с 70 до 110 дБ (на 40 дБ), отношение сигнал/шум увеличивается с 0 до 35 дБ (на 35 дБ).

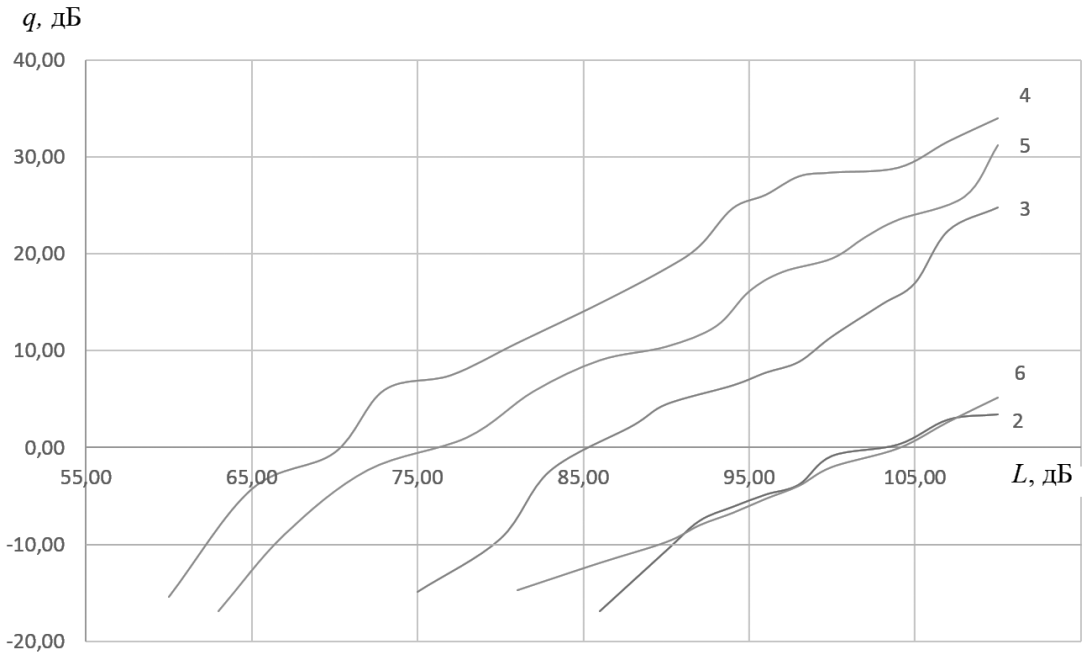


Рис. 12. Графики зависимости отношения «сигнал/шум» от уровня тестового акустического сигнала (частота ВЧ сигнала навязывания 230 МГц, выходная мощность ВЧ-генератора 25 дБм): а) – для 2–й октавной полосы; б) – для 3–й октавной полосы; в) – для 4–й октавной полосы; г) – для 5–й октавной полосы; б) – для 6–й октавной полосы

Оценка возможности перехват речевой информации из помещения методом «высокочастотной накачки» телефонного аппарата

На основе полученных результатов экспериментальных исследований был проведен расчет словесной разборчивости речи для трех типовых уровней речи: речь средней громкости (70 дБ), громкая речь (76 дБ) и очень громкая речь (84 дБ). Расчет проводился для максимальной мощности ВЧ-сигнала навязывания на частоте 230 МГц.

Исходные данные по характеристикам скрываемого речевого сигнала приведены в таблице 2 [1].

Отношение отношений сигнал/шум (q_i) в октавных полосах для типовых значений уровней речевых сигналов, приведенных в таблице 2, определялось по графику на рисунке 12.

Данные значения отношения сигнал/шум получены при размещении приемной антенны на расстоянии 1 м от телефонного аппарата. При удалении антенны средства разведки от телефонного аппарата в следствие затухания радиосигнала будет уменьшаться отношение сигнал/шум:

$$q_{i.r} = q_{i.r=1} - 20\lg V_{i.r}, \qquad (6)$$

где $q_{i.r=1}$ – отношение сигнал/шум на входе приемника средства разведки на расстоянии $r = 1$ м для i – й частоты;

$q_{i.r}$ – отношение сигнал/шум на входе приемника средства разведки на расстоянии r для i – й частоты;

$V_{i.r}$ – затухание радиосигнала на i – й частоте на расстоянии.

Затухание радиосигнала $V_{i.r}$ рассчитывается по формулам [1]:

А. Если частота сигнала ниже частоты fi ≤ 47,75 МГц:

$$V_{i.r} = \begin{cases} r^3, \ если \ r \le \dfrac{47,75}{f_i}; \\ \dfrac{47,75 \cdot r^2}{f_i}, \ если \ \dfrac{47,75}{f_i} < r \le \dfrac{1800}{f_i}; \\ \dfrac{8,59 \cdot 10^4 \cdot r}{f_i}, \ если \ r > \dfrac{1800}{f_i}; \end{cases} \qquad (7)$$

Б. Если частота сигнала удовлетворяет условию 47,75 МГц < fi ≤ 1800 МГц:

$$V_{i.r} = \begin{cases} r^2, \ если \ r \le \dfrac{1800}{f_i}; \\ \dfrac{1800 \cdot r}{f_i}, \ если \ r > \dfrac{1800}{f_i}; \end{cases} \qquad (8)$$

Таблица 2

Характеристики скрываемого речевого сигнала

Номер полосы	Средне-геометрическая частота полосы $f_{ср}$, Гц	Уровни речевого сигнала, дБ, в зависимости от вида речи			Весовой коэффициент полосы, $к_i$	Значение формантного параметра речи, DAi, дБ
		Речь средней громкости, 70 дБ	Громкая речь, 76 дБ	Очень громкая речь,84 дБ		
2	250	66	72	80	0,03	18
3	500	66	72	80	0,12	14
4	1000	61	67	75	0,2	9
5	2000	56	62	70	0,3	6
6	4000	53	59	67	0,26	5

В. Если частота сигнала удовлетворяет условию fi > 1800 МГц:

$$V_{i.r} = r. \qquad (9)$$

Задаваясь пороговым значением словесной разборчивости речи $W_{c.n}$ [28] легко рассчитать максимальную дальности перехвата речевой информации методом «ВЧ-накачки».

В таблице 3 приведены результаты расчетов словесной разборчивости речи при размещении приемной антенны на расстоянии от телефонного аппарата 1 м и 5 м. Расчет словесной разборчивости речи проводился по методике, изложенной в [1].

Как видно из таблицы 3, в реальных условиях перехват речевой информации из помещения методом «высокочастотной накачки» телефонного аппарата практически не возможен.

Заключение

В результате проведенных экспериментальных исследований по оценке подверженности телефонного аппарата «VEF TA-D» «высокочастотной накачке» установлено:

• телефон подвержен «высокочастотной накачке» в диапазоне частот от 168 МГц до 270 МГц. Зависимость напряженности поля информативного сигнала, переизлучаемого телефоном, от частоты сигнала «высокочастотной накачки» носит нелинейный характер. Максимальный уровень напряженности поля информативного сигнала наблюдается на частоте 230 МГц;

• телефонный аппарат наиболее подвержен акустоэлектромагнитным преобразованиям в 3 – 5 октавных полосах;

• с возрастанием уровня тестового акустического сигнала уровень информативного сигнала растет при практически неизменном уровне шумов. Характер зависимостей близкий к линейной;

Таблица 3

Результаты расчетов словесной разборчивости речи при перехвате речевой информации методом «высокочастотной накачки» телефонного аппарата

Расстояние *r*, м	Затухание *V*, раз./дБ	Разборчивость речи		
		Речь средней громкости, 70 дБ	Громкая речь, 76 дБ	Очень громкая речь, 84 дБ
1	1/0	0,007	0,25	0,76
5	25/28,0	0	0	0,004

• с возрастанием выходной мощности высокочастотного генератора уровень информативного сигнала растет, но при этом растет и уровень шумов. При этом характер зависимостей сложный (нелинейный);

• максимальное отношение сигнал/шум наблюдается в 4-й октавной полосе со среднегеометрической частотой 1000 Гц;

• в реальных условиях перехват речевой информации из помещения методом «высокочастотной накачки». телефонного аппарата практически не возможен;

• возможен перехват речевой информации из помещения методом «высокочастотной накачки» телефонного аппарата на дальностях нескольких метров, в случае установки телефонного аппарата вблизи звуковых колонок систем звукоусиления.

Литература

1. Хорев А.А. Техническая защита информации: учеб. пособие: В 3-х т. Т. 1: Технические каналы утечки информации. -М.: НПЦ «Аналитика», 2008. – 436 с.

2. Лысов А.В. Высокочастотное зондирование акустических возбуждаемых объектов в проводной среде. Кабельные локационные системы акустической разведки. – СПб.: Медиапапир, 2021. – 628 с.

3. Лысов А.В. Электромагнитное зондирование акустических возбуждаемых объектов (радиолокационные системы акустической разведки). – СПб.: Медиапапир, 2020. – 678 с.

4. Резонансные свойства RLC-цепей: учеб.-метод. пособие/ Осадченко В.Х., Волкова Я. Ю., Кандрина Ю. А.. – Екатеринбург: Изд-во Урал, ун-та, 2013 – 64 с.

5. Авдеев В.Б. Способ дистанционного перехвата речевой информации из защищаемого помещения здания с охраняемой зоной. Патент на изобретение № 2 575 406. Дата регистрации 20.02.2016 г. Заявка № 2014145030/08 от 06.11.2014 г.

6. Авдеев В.Б., Анищенко А.В. Способ дистанционного перехвата конфиденциальной речевой информации, циркулирующей в защищенном помещении. Патент на изобретение № 2 642 034. Дата регистрации 23.01.2018 г. Заявка № 2016129825 от 20.07.2016 г.

7. Авдеев В.Б., Катруша А.Н. Способ дистанционного перехвата речевой информации из защищаемого помещения. Патент на изобретение № 2 558 673. Дата регистрации 10.08.2015 г. Заявка № 2014137732/07 о 17.09.2014 г.

8. Авдеев В.Б., Катруша А.Н. Способ радиоперехвата речевой информации из защищаемого помещения. Патент на изобретение № 2 561 507. Дата регистрации 27.08.2015 г. Заявка № 2014142671/07 от 22.10.2014 г.

9. Бузов Г.А., Калинин С.В., Кондратьев А.В. Защита от утечки информации по техническим каналам: учеб. пособие. – М.: Горячая линия – Телеком, 2005. – 416 с.

10. Каторин Ю.Ф., Куренков Е.В., Лысов А.В., Остапенко А.Н. Большая энциклопедия промышленного шпионажа. – СПб.: ООО «Издательство Полигон», 2000.- 896 с.

11. Колесник Д. Ю., Майоров А. И. Высокочастотное навязывание. Принцип реализации и методы защиты// Технические средства защиты информации: тезисы докладов XV Белорусско-российской науч.-техн. конф. (Минск, 6 июня 2017 г.). – Минск : БГУИР, 2017. – С. 31 - 32.

12. Кондратьев А. В. Техническая защита информации. Практика работ по оценке основных каналов утечки: учеб. пособие. - М. : Горячая линия - Телеком, 2016. - 304 с.

13. Лыков Ю. В., Морозова А. Д., Кукуш В. Д., Парфёнов А. С. Исследование метода ВЧ-навязывания для несанкционированного съема информации с телефонных линий//ScienceRise. – Харьков: 2015. том 7, №2 (12). С. 51 – 56.

14. Хорев А.А., Лукманова О.Р. Математическое моделирование акустоэлектрического канала утечки речевой информации, создаваемого методом «высокочастотного навязывания»//Защита информации. Инсайд. – С. Петербург: 2023. – № 1 (109) – С. 3– 11.

15. Хорев А.А., Лукманова О.Р. Математическое моделирование пассивного акустоэлектрического канала утечки акустической речевой информации в телефонном аппарате»// Специальная техника. – М.: 2016. – № 6 – С. 56 – 63.

16. Хорев А.А., Лукманова О.Р. Моделирование акустоэлектрического канала утечки речевой информации в телефонном аппарате, создаваемого методом «высокочастотного навязывания»//Международная конференция «Радиоэлектронные устройства и системы для инфотелекоммуникационных технологий – РЭУС-2017». Доклады. – М.: РНТОРЭС имени А.С.Попова. 2017. – С. 492 – 496.

17. Шестаков И.И. Помехоустойчивость выделения информационного сигнала из интермодуляционного излучения при высокочастотном навязывании//Вестник СибГУТИ. – Новосибирск: 2011. - № 3. – С. 45- 58.

18. Дураковский А.П., Куницын И.В., Лаврухин Ю.Н. Контроль защищенности речевой информации в помещениях. Аттестационные испытания вспомогательных технических средств и систем по требованиям безопасности информации: учеб. пособие. – М.: НИЯУ МИФИ, 2015. - 152 с.

19. Хорев А.А. Контроль защищенности вспомогательных технических средств и систем от утечки по акустоэлектрическим каналам// Специальная техника. – М.: 2014. – № 6 – С. 48 - 63

20. Хорев А.А. Контроль защищенности вспомогательных технических средств и систем от утечки по акустоэлектрическим каналам// Специальная техника. – М.: 2014. – № 6 – С. 48 - 63

21. Хорев А.А. Оценка эффективности защиты вспомогательных технических средств// Специальная техника. – М.: 2007. – № 2 – С. 48 – 60.

22. Хорев А.А. Оценка эффективности защиты вспомогательных технических средств. Окончание// Специальная техника. – М.: 2007. – № 3 – С. 50 – 63.

23. Хорев А.А., Лукманова О.Р. Контроль подверженности телефонного аппарата «высокочастотному навязыванию» с использованием виртуального лабораторного стенда//Международная конференция «Радиоэлектронные устройства и системы для инфотелекоммуникационных технологий – РЭУС-2018». Доклады. – М.: РНТОРЭС имени А.С.Попова. 2018. – С. 343 – 348.

24. Автоматизированная система оценки защищенности акустической речевой информации от утечки за счёт высокочастотного навязывания и высокочастотной прокачки «Тандем». URL: https://www.mascom.ru/equipment/sistemy-otsenki-zashchishchennosti-informatsii/tandem.php (дата обращения: 10.05.2025).

25. Комплекс аппаратно-программный «ПАРНАС-ЭХО12Е» для измерения параметров спектра высокочастотных сигналов (ВЧО, ВЧН, ВЧП). URL: http://www.saomega.com/apk/apk-parnas-eho-12e. (дата обращения: 10.05.2025).

26. Программно-аппаратный комплекс выявления технических каналов утечки акустической информации за счет высокочастотного воздействия «Гранат». URL: https://nelk.ru/catalog/sistemy_otsenki_zashchishchennosti_informatsii/programmno_apparatnye_kompleksy/granat/ (дата обращения: 10.05.2025)

27. Хорев А.А., Порсев И.С. Методика вероятностной оценки разборчивости речи //Защита информации. Инсайд. – С. Петербург: 2020. – № 2 – С. 44 – 52.

28. Хорев А.А., Порсев И.С. Методика обоснования критериев эффективности защиты речевой информации от ее утечки по техническим каналам// Всероссийская конференция «Радиоэлектронные устройства и системы для инфотелекоммуникационных технологий – РЭУС-2024». Доклады. – М.: РНТОРЭС имени А.С. Попова. 2024. – С. 430 – 435.

References

1. Horev A.A. Technical information protection: textbook. manual: In 3 volumes Vol. 1: Technical channels of information leakage. -M.: NPC "Analytics", 2008. – 436 p.

2. Lysov A.V. High-frequency sounding of acoustic excited objects in a wired environment. Acoustic reconnaissance cable location systems. – St. Petersburg: Mediapapir, 2021. – 628 p.

3. Lysov A.V. Electromagnetic sounding of acoustic excited objects (acoustic reconnaissance radar systems). – St. Petersburg: Mediapapir, 2020. – 678 p.

4. Resonant properties of RLC circuits: textbook.- the method. stipend/ Osadchenko V.X., Volkova Ya. Yu., Kandrina Yu. A. – Yekaterinburg: Ural Publishing House, University, 2013 – 64 p.

5. Avdeev V.B. A method of remote interception of speech information from a protected building with a protected area. Patent for invention No. 2,575,406. Registration date: 02/20/2016 Application no. 2014145030/08 dated 11/06/2014

6. Avdeev V.B., Anishchenko A.V. A method for remote interception of confidential speech information circulating in a secure room. Patent for invention No. 2,642,034. Registration date: 01/23/2018. Application No. 2016129825 dated 07/20/2016.

7. Avdeev V.B., Katrusha A.N. Method of remote interception of speech information from a protected room. Patent for invention No. 2,558,673. Registration date 08/10/2015 Application No. 2014137732/07 on 17.09.2014

8. Avdeev V.B., Katrusha A.N. Method of radio interception of speech information from a protected room. Patent for invention No. 2,561,507. Registration date 08/27/2015 Application No. 2014142671/07 dated 10/22/2014

9. Buzov G.A., Kalinin S.V., Kondratiev A.V. Protection from information leakage through technical channels: textbook. the manual. – M.: Hotline – Telecom, 2005. – 416 p.

10. Katorin Yu.F., Kurenkov E.V., Lysov A.V., Ostapenko A.N. The Great Encyclopedia of Industrial Espionage. – St. Petersburg: OOO "Polygon Publishing House", 2000. - 896 p.

11. Kolesnik D. Yu., Mayorov A. I. High-frequency imposition. The principle of implementation and methods of protection// Technical means of information protection: abstracts of the XV Belarusian-Russian Scientific and Technical conference (Minsk, June 6, 2017). Minsk : BGUIR, 2017. pp. 31-32.

12. Kondratiev A.V. Technical protection of information. The practice of assessing the main leakage channels: textbook. the manual. - M. : Hotline - Telecom, 2016. - 304 p.

13. Lykov Yu. V., Morozova A.D., Kukush V. D., Parfenov A. S. Investigation of the HF-imposition method for unauthorized removal of information from telephone lines//ScienceRise. – Kharkiv: 2015. Volume 7, No. 2 (12). pp. 51-56.

14. Khorev A.A., Lukmanova O.R. Mathematical modeling of the acousto–electric channel of leakage of speech information created by the method of "high-frequency imposition"//Information protection. Inside. – St. Petersburg: 2023. – № 1 (109) – P– 3-11.

15. Horev A.A., Lukmanova O.R. Mathematical modeling of a passive acousto–electric channel for leakage of acoustic speech information in a telephone set// Special technique. Moscow: 2016. – No. 6 – pp. 56-63.

16. Horev A.A., Lukmanova O.R. Modeling of an acousto–electric channel for leakage of speech information in a telephone set, created by the method of "high-frequency imposition"//International Conference "Radio Electronic devices and systems for Infotelec communication Technologies – REUS-2017". Reports. Moscow: RNTORES named after A.S.Popov. 2017. – pp. 492-496.

17. Shestakov I.I. Noise immunity of information signal isolation from intermodulation radiation during high-frequency imposition//Bulletin of SibGUTI. Novosibirsk: 2011. No. 3. pp. 45-58.

18. Durakovsky A.P., Kunitsyn I.V., Lavrukhin Yu.N. Control of the security of speech information in rooms. Certification tests of auxiliary technical means and systems for information security requirements: textbook. The manual. Moscow: NRU MEPhI, 2015. 152 p.

19. Horev A.A. Control of the protection of auxiliary equipment and systems from leakage through acousto-electric channels// Special equipment. – M.: 2014. – № 6 – pp. 48-63

20. Horev A.A. Control of the protection of auxiliary equipment and systems from leakage through acousto-electric channels// Special equipment. – M.: 2014. – № 6 – pp. 48-63

21. Horev A.A. Evaluation of the effectiveness of protection of auxiliary technical means// Special equipment. Moscow: 2007, No. 2– pp. 48-60.

22. Horev A.A. Evaluation of the effectiveness of protection of auxiliary equipment. Graduation // Special equipment. Moscow: 2007. No. 3– pp. 50-63.

23. Horev A.A., Lukmanova O.R. Control of the telephone's susceptibility to "high–frequency imposition" using a virtual laboratory stand//International Conference "Radio Electronic devices and systems for Infotelec communication Technologies – REUS-2018". Reports. Moscow: RNTORES named after A.S.Popov. 2018. – pp. 343-348.

24. Automated system for assessing the security of acoustic speech information from leakage due to high-frequency imposition and high-frequency pumping "Tandem". URL: https://www.mascom.ru/equipment/sistemy-otsenki-zashchishchennosti-informatsii/tandem.php (date of access: 05/10/2025).

25. Hardware and software complex "PARNAS-EHO12E" for measuring the parameters of the spectrum of high-frequency signals (HF, HF, HF). URL: http://www.saomega.com/apk/apk-parnas-eho-12e. (date of request: 05/10/2025).

26. A hardware and software complex for detecting technical channels of acoustic information leakage due to the high-frequency effects of "Grenades". URL: https://nelk.ru/catalog/sistemy_otsenki_zashchishchennosti_informatsii/programmno_apparatnye_kompleksy/granat / (date of request: 05/10/2025)

27. Horev A.A., Porsev I.S. Methodology of probabilistic assessment of speech intelligibility // Information protection. Inside. – St. Petersburg: 2020. – No. 2 – pp. 44-52.

28. Horev A.A., Porsev I.S. Methodology for substantiating criteria for the effectiveness of protecting speech information from leakage through technical channels// All–Russian Conference "Radio electronic devices and systems for infotelec communication technologies - REUS-2024". Reports. Moscow: RNTORES named after A.S. Popov. 2024. – pp. 430-435.

ХОРЕВ Анатолий Анатольевич, доктор технических наук, профессор, заведующий кафедрой «Информационная безопасность» федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский университет «Московский институт электронной техники» (МИЭТ), 124498, г. Москва, г. Зеленоград, площадь Шокина, дом 1, МИЭТ, E-mail – horev@miee.ru

ЧАЧИЛЛО Тимофей Витальевич, студент кафедры «Информационная безопасность» федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский университет «Московский институт электронной техники» (МИЭТ), 124498, г. Москва, г. Зеленоград, площадь Шокина, дом 1, МИЭТ, E-mail – chachillo1502@gmail.com

Anatoly Anatolyevich KHOREV, Doctor of Technical Sciences, Professor, Head of the Department "Information Security" of the Federal State Autonomous Educational Institution of Higher Education "National Research University "Moscow Institute of Electronic Technology" (MIET). 124498, Moscow, Zelenograd, Shokin Square, house 1, MIET. E-mail – horev@miee.ru

Timofey V. CHACHILLO, student of the Department of Information Security at the Federal State Autonomous Educational Institution of Higher Education National Research University Moscow Institute of Electronic Technology (MIET). Shokin Square, 1, MIET, Moscow, 124498, Zelenograd. E-mail – chachillo1502@gmail.

УДК 004.056.53 Вестник УрФО № 2(56) / 2025, с. 96–102

Южаков А. А., Кротова Е. Л., Ощепков Н. В. DOI: 10.14529/secur250210

ВЕРИФИКАЦИЯ УЗЛОВ СЕТИ ПРОМЫШЛЕННОГО ПРОТОКОЛА MODBUS

В данной работе анализируются особенности протокола Modbus, с акцентом на его уязвимости в контексте безопасности и защиты передаваемой информации. Рассмотрены основные риски, связанные с использованием Modbus в системах автоматизации и управления технологическими процессами (АСУ ТП), включая отсутствие механизмов шифрования и аутентификации, что делает его уязвимым к различным видам атак, таким как перехват данных или несанкционированный доступ, а также варианты решения проблемы верификации узлов.

Ключевые слова: *информационная безопасность, АСУ ТП, промышленный протокол, Modbus*

Yuzhakov A. A., Krotova E. L., Oshchepkov N. V.

VERIFICATION OF MODBUS INDUSTRIAL PROTOCOL NETWORK NODES

In this article, we analyzed the features of the Modbus protocol, with an emphasis on its vulnerabilities in the context of security and protection of transmitted information. The main risks associated with the use of Modbus in automation and process control systems (APCS) are considered, including the lack of encryption and authentication mechanisms, which makes it vulnerable to various types of attacks, such as data interception or unauthorized access. Options for solving the problem of node verification are considered.

Keywords: *information security, APCS, industrial protocol, Modbus*

Введение

В настоящее время большое внимание уделяется организации безопасной работы в системах АСУ ТП, работающих на базе промышленных сетевых протоколов (fieldbus). В перечень таких протоколов входят Profibus, DeviceNet, Modbus, CAN [1].

Modbus — это протокол коммуникации, имеющий клиент-серверную архитектуру, созданный в 1979 году компанией Modicon для использования в промышленных системах автоматизации [2]. Modbus является стандартом передачи данных между контроллерами (PLC), сенсорами, исполнительными механизмами и прочим промышленным оборудованием. Протокол широко используется в различных сферах, включая:

- системы управления зданиями и домами
- промышленные и технологические системы

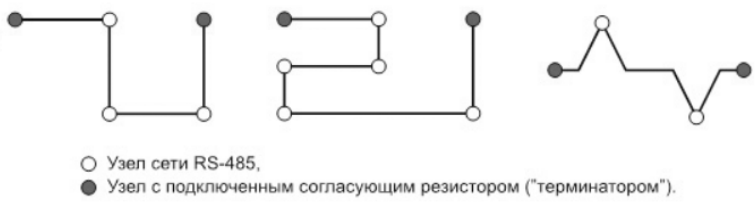


Рис. 1. Примеры топологий сетей RS-485

• электроэнергетика

• системы мониторинга и контроля различных технологических процессов.

Именно поэтому, в качестве исследуемого протокола использован широко применяемый в настоящее время протокол Modbus. Это обусловлено несколькими причинами [3]:

• существенно возросло количество новых разработок и объем организационной поддержки этого протокола

• отсутствует необходимость в специальных интерфейсных контроллерах

• простота программной реализации

Несмотря на всю простоту, широкое применение и удобство протокола, он не лишен недостатков, связанных с обеспечением безопасности и конфиденциальности передачи данных.

1. Описание протокола

Одним из ключевых преимуществ Modbus является устойчивость к искажениям данных, достигаемая благодаря строгой системы проверки ошибок [4]. Архитектура протокола предусматривает два формата передачи: компактный двоичный Modbus RTU (Remote Terminal Unit), оптимизированный для последовательных соединений, и текстовый Modbus ASCII, применяемый в специализированных задачах [3]. Режим RTU является обязательным в спецификации Modbus, тогда как ASCII реализуется по необходимости. При конфигурации системы необходимо учитывать, что смешение режимов внутри одной сети недопустимо — все устройства должны функционировать в едином формате. На физическом уровне поддерживаются интерфейсы RS-232 и RS-485, однако в промышленных приложениях доминирует RS-485 [5].

На рисунке 1 показана структура сети, использующей интерфейс RS-485 совместно с протоколом Modbus. Такая конфигурация предполагает последовательное подключение устройств — как передающих, так и принимающих — посредством витой пары [6].

Архитектура протокола Modbus соответствует трём уровням эталонной модели OSI [7] (рис. 2): физическому, канальному и прикладному. Особенностью протокола является отсутствие реализации сетевого, транспортного и сеансового уровней (3-6) - их функции делегированы прикладному уровню, который обеспечивает коммуникацию между устройствами [8].

1.1 Физический уровень

Передача данных осуществляется с использованием интерфейса, который может быть двухпроводным (полудуплексным) или

Модель OSI для Modbus

НОМЕР УРОВНЯ	НАЗВАНИЕ УРОВНЯ	РЕАЛИЗАЦИЯ
7	Прикладной	Modbus application protocol
6	Уровень представления	Нет
5	Сеансовый	Нет
4	Транспортный	Нет
3	Сетевой	Нет
2	Канальный (передачи данных)	Протокол «ведущий–ведомый». Режимы RTU и ASCII
1	Физический	RS-485 или RS-232

Рис. 2. Модель OSI для Modbus

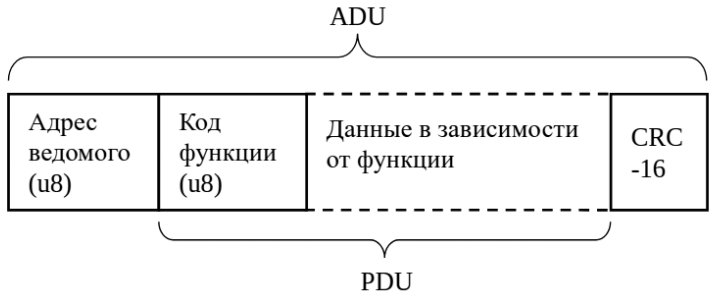


Рис. 3. Структура кадра MODBUS RTU

четырёхпроводным (дуплексным), соответствующим стандартам TIA/EIA-485 либо TIA/EIA-422 [9].

1.2 Канальный уровень

Протокол Modbus реализует классическую архитектуру по принципу "ведущий-ведомые", где ведущее устройство выступает в роли единственного управляющего устройства, подчиненные устройства являются исполнителями команд, принимаемых от ведущего устройства. Число подчинённых модулей в одной сети может достигать 247. Такая модель взаимодействия особенно важна в промышленных системах реального времени, где требуется строгий контроль за процессом [3].

1.3 Прикладной уровень

Прикладной уровень Modbus RTU обеспечивает связь между устройствами в режиме «ведущий–ведомый». Он функционирует независимо от нижележащих уровней — физического и канального — что обеспечивает совместимость с различными типами передачи данных, включая Ethernet TCP/IP (Modbus TCP/IP). Коммуникация на этом уровне осуществляется с помощью запросов, содержащих коды функций, которые определяют, какое действие должно быть выполнено ведомым устройством [3].

1.4 Формат пакета

Полный пакет (рис. 3), передаваемый по физическому каналу, включающий адрес ведомого и CRC, называют ADU – Application Data Unit. ADU инкапсулирует PDU – Protocol Data Unit – данные MODBUS, не зависящие от среды передачи. Формат и размер PDU в свою очередь определяются кодом функции и числом регистров, которые должны быть записаны или прочитаны [11]. Минимальный размер фрейма в режиме RTU может достигать 5 байт.

Адрес устройства — адрес получателя, то есть slave-устройства, которому посылается управляющая команда.

Код функции — номер команды в виде 8 битного числа.

Данные — полезная нагрузка, посылаемая на ведомое устройство в зависимости от функции. Например, если выполняется функция чтения из регистров хранения, то внутри сегмента фрейма, отвечающего за данные, необходимо указать также адрес начального регистра и количество регистров.

Контрольная сумма — алгоритмы проверки целостности пакетов. Для хранения и передачи в Modbus RTU используются 2 последних байта фрейма, а в качестве алгоритма применяется CRC16 [12].

На рисунке (рис. 4) представлен режим передачи RTU, в котором данные передаются младшими разрядами вперед [3].

1.5 Вычисление контрольной суммы CRC16

Расчёт начинается с инициализации регистра значением 0xff. Для вычисления используется только содержимое байтов данных, без учёта стартовых, стоповых битов и бита чётности. Каждый байт последовательно обрабатывается с применением операций сдвига и побитового исключающего ИЛИ. Алгоритм повторяет цикл обработки для каждого байта, в результате чего формируется итоговая CRC16-сумма [10].

Стартовый бит	1 МЗР	2	3	4	5	6	7	8	Бит паритета	Стоп-бит

Рис. 4. Последовательность битов в режиме RTU

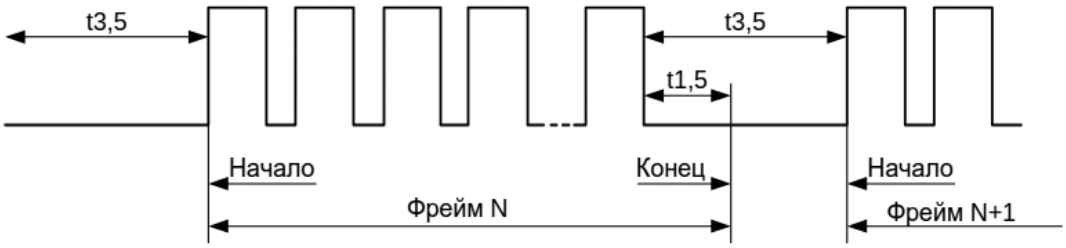


Рис. 5. Диаграмма передачи фреймов

1.6 Особенности протокола

Для разделения сообщений была внедрена система временных пауз, позволяющий идентифицировать, что фрейм пришел в полной мере [13]. На рисунке 5 представлен возможный формат общения ведущего и ведомого устройства [9].

Пауза между фреймами должна быть не короче времени передачи 3,5 символов при текущей скорости обмена. Формула для её вычисления представлена ниже (1) [13]:

$$t_{3.5}(\text{мс}) = \frac{3.5 \cdot 11}{\text{скорость (бит/с)}} \cdot 1000 \ . \quad (1)$$

При скорости выше 19200 бит/с вместо формул используют фиксированные интервалы: 0,75 мс для t1.5 и 1,75 мс для t3.5, чтобы избежать чрезмерных срабатываний таймеров и сбоев в работе системы.

Существует три типа функций передачи [3]:

• Стандартные. Описание этих функций опубликовано и утверждено Modbus-IDA [14]. Эта категория включает в себя как опубликованные, так и свободные в настоящее время коды.

• Пользовательские. Два диапазона кодов (от 65 до 72 и от 100 до 110), для которых пользователь может создать произвольную функцию.

• Зарезервированные. В эту категорию входят коды функций, не являющиеся стандартными, но уже используемые в устройствах, производимых различными компаниями.

Типы данных протокола Modbus представлены на рисунке 6.

Проведя анализ, можно выделить следующие достоинства и недостатки протокола.

Достоинства. На данный момент, протокол зарекомендовал себя как надежный, проверенный временем стандарт, который претерпел много изменений с момента создания. К его преимуществам можно отнести:

• простой и легко реализуемый протокол, что делает его популярным среди разработчиков и инженеров [15]

• используется двоичный формат передачи данных, что позволяет передавать большее количество информации за меньшее время [15]

Недостатки. В настоящее время не существует совершенных протоколов. О слабостях любого протокола стоит знать, чтобы иметь возможность оценить риски, которые возникают в процессе эксплуатации той или иной технологии. Ниже перечислены несколько существенных недостатков анализируемого протокола:

• Ключевым ограничением протокола Modbus является модель «ведущий–ведомый», при которой только мастер инициирует обмен, а ведомые устройства не могут передавать данные самостоятельно.

• В протоколе отсутствует механизм аутентификации ведущего и ведомых устройств, что создает риск внедрения злоумышленни-

Параметр	Тип данных	Тип доступа
Дискретные входы (Discrete Inputs)	один бит	только чтение
Регистры флагов (Coils)	один бит	чтение и запись
Регистры ввода (Input Registers)	16-битное слово	только чтение
Регистры хранения (Holding Registers)	16-битное слово	чтение и запись

Рис. 6. Типы данных протокола Modbus

ка, выдающего себя за ведущего и получающего полный контроль над сетью.

• Также протокол не предусматривает шифрование данных при передаче по открытым каналам, что создаёт риск несанкционированного доступа. Следовательно, необходимы доработки для внедрения надёжной и простой в реализации аутентификации.

Исходя из вышеизложенного, задачами исследования являются вопросы построения метода, модели и средств противодействия угрозам нарушения информационной безопасности в открытой промышленной сети, функционирующей в режиме RTU протокола Modbus.

Целью исследования является разработка метода верификации ведущего и ведомого устройства в процесс коммуникации протокола Modbus.

Задачи:

- Анализ существующих методов верификации, используемых в протоколах с ограничениями на длину полезной нагрузки
- Построение модели злоумышленника
- Внедрение предложенных методов
- Анализ влияния предложенных методов на доступность

Решения данной проблемы основаны на использовании 3 способов верификации:

1. Временного интервала между передачей фреймов как фактора верификации мастера и слейва. Если мастер и слейв синхронизируются об «окнах» передачи данных, то отправителя в каждом случае можно назвать валидным.
2. 16-битной вставки с помощью операции XOR, затрагивающий последний байт PDU и первый байт CRC16. При таком варианте внедряется легковесный алгоритм подписи сообщения, что позволит нам удостовериться, что общение происходит с легитимным устройством.
3. Одноразовый пароль (OTP) [16], который можно вычислить для мастера и слейва в любой момент времени в случае наличия между ними договорённости о ключе. В таком случае, можно установить ключ, на основании которого высчитывается пароль, без ограничений по длине, что усложнит возможность его подобрать за разумный промежуток времени.

Вывод

Исходя из вышеизложенного, можно сделать вывод, что протокол Modbus за более чем 40 лет существования не только не утратил свою популярность, но и существенно распространился и был внедрен во многие АСУ ТП. Протокол зарекомендовал себя как надёжное решение в качестве стандарта для промышленной сети. Повышение качества использования протокола Modbus основано на построении верификационных алгоритмов процесса коммуникации на основе предложенного в работе подхода.

Литература

1. All About Fieldbus Protocols: сайт Inst Tools. [Электронный ресурс]. – Режим доступа: https://instrumentationtools.com/fieldbus-protocols/ (дата обращения 10.03.2025).

2. Modbus. Сайт Modbus. [Электронный ресурс]. – Режим доступа: https://www.modbus.org/specs.php (дата обращения 05.03.2025)

3. Промышленные сети и интерфейсы. Сайт «RealLab!». [Электронный ресурс]. — Режим доступа : https://www.reallab.ru/bookasutp/2-promishlennie-seti-i-interfeisi/2-8-modbus/ (дата обращения 05.03.2025).

4. Денисенко В. В. Протоколы и сети Modbus и Modbus TCP / Денисенко В. В. // Современные технологии автоматизации. – 2010. – № 4, с. 90 - 94.

5. RS-485: Википедия. Свободная энциклопедия. [Электронный ресурс]. – Режим доступа: https://ru.wikipedia.org/wiki/RS-485 (дата обращения 05.03.2025)

6. Лемешко К.О. Диспетчеризация электроснабжения и система мониторинга на основе протокола modbus // Аллея науки. 2019. Т. 3. № 5 (32). С. 244-248.

7. OSI model: Википедия. Свободная энциклопедия. [Электронный ресурс]. – Режим доступа : https://en.wikipedia.org/wiki/OSI_model (дата обращения 07.03.2025)

8. Томас Д. Введение в протокол Modbus. Часть 2. Modbus Serial и Modbus TCP. Современные технологии автоматизации. – 2009. – № 3, с. 22 - 26.

9. Краткое описание протокола Modbus/RTU. Сайт «Интеллект модуль». [Электронный ресурс]. - Режим доступа: https://intellect-module.ru/downloads/manuals/inode_35D/ModBus_RTU.pdf (дата обращения 08.03.2025)

10. Введение в Modbus протокол. Сайт «Быстрые проекты». [Электронный ресурс]. - Режим доступа: https://fast-project.ru/upload/Description_Modbus_ru.pdf (дата обращения 07.03.2025)

11. Савинков А.Ю. Реализация протокола modbus rtu на микроконтроллерах семейства stm32 // Информатика: проблемы, методы, технологии. Материалы XXII Международной научно-практической конференции им. Э.К. Алгазинова. Под редакцией Д.Н. Борисова. Воронеж, 2022. С. 86-95.

12. Как общаются машины: протокол Modbus. Сайт «Хабр». [Электронный ресурс]. - Режим доступа: https://habr.com/ru/companies/advantech/articles/450234/ (дата обращения 14.03.2025)

13. Заметки о Modbus. Сайт «ОВЕН». [Электронный ресурс]. - Режим доступа: https://ftp.owen.ru/CoDeSys3/98_Books/ModbusTips.pdf (дата обращения 12.03.2025)

14. MODBUS APPLICATION PROTOCOL SPECIFICATION V1.1b3. Сайт Modbus. [Электронный ресурс]. – Режим доступа: https://modbus.org/docs/Modbus_Application_Protocol_V1_1b3.pdf (дата обращения 14.03.2025)

15. Сальников М.С. Введение в протокол modbus rtu: основные принципы и преимущества / М.С. Сальников, А.А. Суханкин //Научное обозрение: актуальные вопросы теории и практики. Сборник статей VI Международной научно-практической конференции: в 2 ч. Том Часть 1. Пенза, 2023.

16. One-time password: Википедия. Свободная энциклопедия. [Электронный ресурс]. – Режим доступа: https://en.wikipedia.org/wiki/One-time_password (дата обращения 16.03.2025)

References

1. All About Fieldbus Protocols. Available at: https://instrumentationtools.com/fieldbus-protocols/ (accessed 10.03.2025)

2. Modbus. Available at: https://www.modbus.org/specs.php (accessed 05.03.2025)

3. Promishlennie seti i interfeysi [Industrial networks and interfaces]. Available at: https://www.reallab.ru/bookasutp/2-promishlennie-seti-i-interfeisi/2-8-modbus/ (accessed 05.03.2025)

4. Denisenko V.V. Protokoli i seti Modbus i Modbus TCP [Protocols and networks of Modbus and Modbus TCP]. Modern automation technologies, 2010, no. 4, p. 90-94. (in Russ.)

5. RS-485. Available at: https://ru.wikipedia.org/wiki/RS-485 (accessed 05.03.2025)

6. Lemeshko K.O. Dispetcherizaciya elektrosnabzheniya i sistema monitoring na osnove protokola modbus [Power supply dispatching and monitoring system based on modbus protocol]. Alley of Science, 2019, vol. 3, no. 5 (32), p. 244-248. (in Russ.)

7. OSI model. Available at: https://en.wikipedia.org/wiki/OSI_model (accessed 07.03.2025)

8. Tomas D. Vvedenie v protokol Modbus. Chast 2. Modbus serial i Modbus TCP [Introduction to Modbus protocol. Part 2. Modbus serial and Modbus TCP]. Modern automation technologies, 2009, no. 3, p. 22-26. (in Russ.)

9. Kratkoe opisanie protokola Modbus/RTU [Brief description of the Modbus/RTU protocol]. Available at: https://intellect-module.ru/downloads/manuals/inode_35D/ModBus_RTU.pdf (accessed 08.03.2025)

10. Vvedenie v Modbus protocol [Introduction to Modbus protocol]. Available at: https://fast-project.ru/upload/Description_Modbus_ru.pdf (accessed 07.03.2025)

11. Savinkov A.U. [Implementation of the Modbus Rtu Protocol on Microcontrollers of the Stm32 Family]. Informatika: problemi, metodi, tehnologii [Computer Science: Problems, Methods, Technologies]. Voronezh, 2022, pp. 86-95. (in Russ.)

12. Kak obsjyautsa mashini. Protokol Modbus [How Machines Communicate: The Modbus Protocol]. Available at: https://habr.com/ru/companies/advantech/articles/450234/ (accessed 14.03.2025)

13. Zametki o Modbus [Notes on Modbus]. Available at: https://ftp.owen.ru/CoDeSys3/98_Books/ModbusTips.pdf (accessed 12.03.2025)

14. MODBUS APPLICATION PROTOCOL SPECIFICATION V1.1b3. Available at: https://modbus.org/docs/Modbus_Application_Protocol_V1_1b3.pdf (accessed 14.03.2025)

15. Salnikov M.S. [Introduction to Modbus Rtu Protocol: Basic Principles and Advantages]. Nauchnoe obozrenie: aktualnie voprosi teorii i praktiki [Scientific Review: Current Issues of Theory and Practice]. Penza, vol. 1, 2023.

16. One-time password. Available at: https://en.wikipedia.org/wiki/One-time_password (accessed 16.03.2025)

ЮЖАКОВ Александр Анатольевич, доктор технических наук, профессор, заведующий кафедрой Автоматика и телемеханика, Пермский национальный исследовательский политехнический университет (ПНИПУ). 614990, Пермский край, г. Пермь, Комсомольский проспект, д. 29. E-mail: uz@at.pstu.ru

КРОТОВА Елена Львовна, кандидат физико-математических наук, доцент кафедры Высшей математики, Пермский национальный исследовательский политехнический университет (ПНИПУ). 614990, Пермский край, г. Пермь, Комсомольский проспект, д. 29. E-mail: lenkakrotova@yandex.ru

ОЩЕПКОВ Никита Владимирович, аспирант кафедры Автоматика и телемеханика, Пермский национальный исследовательский политехнический университет (ПНИПУ). 614990, Пермский край, г. Пермь, Комсомольский проспект, д. 29. E-mail: maserati_2000@mail.ru

YUZHAKOV Alexander Anatolyevich, Doctor of Technical Sciences, Professor, Head of the Department of Automation and Telemechanics, Perm National Research Polytechnic University (PNIPU). 614990, Perm Region, Perm, Komsomolsky Prospekt, 29. E-mail: uz@at.pstu.ru

KROTOVA Elena Lvovna, Candidate of Physical and Mathematical Sciences, Associate Professor of the Department of Higher Mathematics, Perm National Research Polytechnic University (PNIPU). 614990, Perm Krai, Perm, Komsomolsky Prospekt, 29. E-mail: lenkakrotova@yandex.ru

OSHCHEPKOV Nikita Vladimirovich, postgraduate student of the Department of Automation and Telemechanics, Perm National Research Polytechnic University (PNIPU). 614990, Perm Region, Perm, Komsomolsky Prospekt, 29. E-mail: maserati_2000@mail.ru

Материалы к публикации отправлять по адресу E-mail: urvest@mail.ru
в редакцию журнала «Вестник УрФО. Безопасность в информационной сфере».

Или по почте по адресу: Россия, 454080, г. Челябинск, пр. им. Ленина, д. 76, ЮУрГУ,
Издательский центр

ВЕСТНИК УрФО
Безопасность в информационной сфере № 2(56) / 2025
Подписано в печать 30.06.2025. Дата выхода в свет 30.06.2025.
Формат 70×108 1/16. Печать цифровая. Усл.-печ. л. 7,79. Тираж 50 экз.
Заказ XXX/XXX.
Цена свободная.

Отпечатано в типографии Издательского центра ФГАОУ ВО «ЮУрГУ (НИУ)».
454080, г. Челябинск, пр. им. В. И. Ленина, 76, ЮУрГУ, Издательский центр.

Bulletin of the Ural Federal District
Security in the Sphere of Information No. 2(56) / 2025
Signed to print 30.06.2025. Date of publication of the 30.06.2025.
Format 70×108 1/16. Screen printing. Conventional printed sheet 7,79. Circulation – 50 issues.
Order XXX/XXX.
Open price.

Printed in the printing house of the Publishing Center of FGAOU VO «SUSU (NIU)».
SUSU, Publishing Center, 76, Lenina Str., Chelyabinsk, 454080