

Труды XXX научной конференции по радиофизике

**СЕКЦИЯ
«ИНФОРМАЦИОННЫЕ СИСТЕМЫ.
СРЕДСТВА, ТЕХНОЛОГИИ, БЕЗОПАСНОСТЬ»**

Председатель – Л.Ю. Ротков, секретарь – А.А. Рябов.
Нижегородский государственный университет им. Н.И. Лобачевского.

УЧЕТ РЕЛЬЕФА МЕСТНОСТИ ПРИ ВЫБОРЕ ПОЗИЦИЙ ПРИЕМНЫХ ПУНКТОВ УГЛОМЕРНОЙ РАДИОТЕХНИЧЕСКОЙ СИСТЕМЫ ПАССИВНОЙ ЛОКАЦИИ

И.Н. Карельский, Л.Ю. Ротков, С.А. Дема

ННГУ им. Н.И. Лобачевского

Современные угломерные радиотехнические системы пассивной локации, состоящие из двух и более разнесенных в пространстве приемных пеленгаторных пунктов (ПП) могут быть использованы в качестве обнаружителя и определителя координат радиоизлучающих малоразмерных БПЛА – дронов различного назначения, несущих угрозу охраняемому объекту (ОО).

В этом случае ПП должны быть наилучшим образом размещены относительно ОО в директивно выделенной полосе размещения (ВПР) с целью получения максимальной рабочей зоны системы (РЗС) на угрожаемых направлениях и ожидаемых высотах полета дронов.

Размеры и форма ВПР зависят от требуемой площади для размещения одного ПП, наличия источников первичного электропитания, возможностей по реализации системы связи, электромагнитной совместимости с другими РЭС, допустимыми ошибками местоопределения дрона источника радиоизлучения (ИРИ) и др.

Целью работы является установление координат ПП в ВПР, обеспечивающих максимальную площадь РЗС на угрожаемых направлениях полета FPV - дронов с учетом рельефа местности в районе применения системы и ограничений на размещение ПП.

РЗС образуется в результате совместного перекрытия некоторой области пространства зонами обнаружения отдельных пеленгаторов (ЗОП). Конфигурация этих зон определяется максимальными дальностями обнаружения ИРИ ($R_{обн}$) на различных направлениях их приема по азимуту при заданной высоте полета дрона.

Поскольку полет дрона происходит на малых и предельно малых высотах, а излучение сигнала происходит в УКВ диапазоне, то на распространение сигнала и его обнаружение существенное влияние оказывает рельеф местности. Поэтому дальность $R_{обн}$ выбирается как минимальное значение из трех предварительно рассчитанных дальностей: – дальности прямой видимости ИРИ ($R_{пв}$); – дальности до препятствия ($R_{пр}$), находящегося на заданном азимуте по линии дальности прямой видимости и заслоняющего прием сигнала из-за наличия угла закрытия в вертикальной плоскости (ε_3); – энергетической дальности приема сигналов ИРИ ($R_{эд}$).

Указанные дальности можно вычислить по формулам:

$R_{пв} = 4,12(h_a^{1/2} + h_{ири}^{1/2})$ [1], где h_a – абсолютная высота энергетического центра антенны ПП; $h_{ири}$ – абсолютная высота полета ИРИ;

$R_{пр}$ – выбирается при условии, что угол визирования ИРИ в вертикальной плоскости ($\varepsilon_{ири}$) для заданной высоты полета меньше угла закрытия, рассчитываемого по формуле $\varepsilon_3 = (h_{пр} - h_a)57,3/R_{пр}$ [2], где $h_{пр}$ – абсолютная высота препятствия;

$R_{эд} = 10^{(P_{TX} + G_{TX} + G_{RX} - L_{TX} - L_{other} - S_{RX} - M_{link} - 20 \log_{10}(f) - 32,44)/20}$, [3] где: P_{TX} – мощность передатчика FPV – дрона (от 25 мВт до 10 Вт и более); G_{TX} – коэффициент усиления передающей антенны дрона (около 3 dB); G_{RX} – коэффициент усиления

приемной антенны пеленгатора (от 5 до 16 dB); L_{TX} (dB) – потери в антенне и кабеле передатчика; L_{other} (dB) – дополнительные потери при распространении и приеме сигнала; S_{RX} – чувствительность приёмника, (обычно -90 ... -100) dBm; M_{link} – запас устойчивой радиосвязи (рекомендуется 10-15 dB); f – несущая частота сигнала, например, 2,4 или 5,8 ГГц.

Расчеты показывают, что, обнаружение дрона, летающего на дальние расстояния (до 100 км), с излучаемой на частоте 5,8 ГГц мощностью $P_{TX}=1$ Вт, пеленгатором, имеющим приемную антенну в виде кольцевой антенной решетки [4] и чувствительность приемника -100 dBm, произойдет на дальности около 25 км. Дроны летающие на малые и средние дальности будут обнаруживаться на дальностях не превышающих 5-15 км в идеальных условиях и 2-5 км в городе/лесу.

С учетом энергетической дальности и известном среднеквадратическом отклонении в измерении угла β ИРИ (δ_β) (в радианах), можно определить желательное расстояние между двумя пеленгаторами – базу (d) по формуле: $d = R_{ад}^2 \delta_\beta / \sigma_{мп}$, при которой ошибка определения местоположения дрона ($\sigma_{мп}$) не будет превышать заданную величину. Так, например, если задана ошибка на линии траверза к базе $\sigma_{мп} = 200$ м и $R_{ад} = 5000$ м, то величина $d \approx 2000$ м. Это значение может быть задано как диаметр или как сторона квадрата ВПР с центром в середине ОО.

В общем случае для оценки точности определения местоположения несколькими пеленгаторами используют ковариационную матрицу ошибок [5]: $P = \sigma_\beta^2 \cdot (H^T H)^{-1}$, где H – матрица Якоби системы уравнений для измеряемых пеленгаторами пеленгов на ИРИ.

Если прямую пеленга на плоскости в прямоугольной СК описать уравнением $y_0 - y_i = \tan(\beta_i)(x_0 - x_i)$, где (x_i, y_i) – координаты i -го ПП, (x_0, y_0) – искомые координаты ИРИ, β_i – измеренный пеленг от i -го ПП на ИРИ, то элементы матрицы ошибок рассчитываются по формулам: $H_{i1} = \partial \beta_i / \partial x = (y_0 - y_i) / r_i^2$; $H_{i1} = \partial \beta_i / \partial y = (x_0 - x_i) / r_i^2$, где $r_i = \sqrt{(x_0 - x_i)^2 + (y_0 - y_i)^2}$ – расстояние от i -го ПП до ИРИ. Ошибка местоположения по осям СК: $\sigma_{мпх} = \sqrt{P_{11}}$, $\sigma_{мпу} = \sqrt{P_{22}}$; а радиальная ошибка: $\sigma_{мп} = (\sigma_{мпх}^2 + \sigma_{мпу}^2)^{0.5}$.

Алгоритм установления координат ПП в ВПР, предполагает поиск дискретного варианта размещения ПП на местности, при котором обеспечивается максимальная площадь РЗС при заданных ограничениях на ВПР, заданных сигнальных характеристиках ИРИ и заданной высоте полета дрона.

Указанная задача решается с помощью цифровой модели рельефа местности (ЦМР) района применения системы. Данные по рельефу могут быть получены из открытых источников, например, из Google Map. ЦМР позволяет визуализировать рельеф, рассчитать профили трасс ИРИ и оценить влияние возвышенностей на распространение сигнала.

Для учета рельефа местности земная поверхность в зоне применения разбивается на элементарные участки (ЭОП) – ячейки малого размера, соответствующие матрице высот, построенной по ЦМР. Профили местности представляют собой вертикальные разрезы земной поверхности плоскостью, проходящей через точки размещения антенн

ПП и ИРИ. При этом должны быть учтены все возвышенности антропогенного характера.

Варианты размещения (выбранные ЭОП) ПП в ВПР и местоположение ИРИ в РЗС можно представить в виде координатных матриц, содержащих $L_{пп} = m_{пп} * n_{пп}$ и $L_{ири} = p_{ири} * k_{ири}$ элементов (m, p и k, n по осям x и y соответственно). Для каждого возможного местонахождения ПП и заданного по координатам ИРИ пеленга производится пошаговый с интервалом, равным расстоянию между элементами матриц, расчет $R_{обн}$, с удалением ИРИ от ПП на каждом шаге вплоть до значения $R_{эд}$. При этом на каждом пеленге анализируется профиль местности в вертикальной плоскости и учитывается его влияние на величину $R_{обн}$.

Полученная ЗОП для i –й позиции ПП объединяется с ЗОП для $i + 1$ –й позиции ПП, оцениваются ошибки определения местоположения ИРИ для выбранного количества и местоположения позиций, далее исходя из допустимых ошибок местопределения строится РЗС и рассчитывается её площадь (по количеству и заданным размерам ЭОП). Методом прямого перебора возможных позиций ПП находится их оптимальное расположение, соответствующее максимальной площади рабочей зоны системы.

В программной реализации все расчёты проводятся в координатах системы Pseudo-Mercator, для вывода на карту координаты переводятся в геодезическую СК WGS 84 [6]. Фрагмент карты запрашивается через сервис OpenStreetMap [7], на него наносятся границы зон, найденные оптимальные позиции пунктов, рабочая зона в виде небольших кругов соответствующего цвета с центрами в узлах координатной сетки как без учета рельефа, так и с его учетом.

Круги представляют собой область местности определенной площади, зависящей от величины шага дискретизации. Матрица высот рельефа, используемая для вычисления углов закрытия и визирования, определяется путем запросов к сервису Open Topo Data [7].

На рисунке 1а) и 1б) показаны результаты работы описанного выше алгоритма для случаев субоптимального и оптимального размещения трех ПП в ВПР заданного района применения. В обоих случаях РЗС не является круговой зоной с радиусом, равным максимальной энергетической дальности обнаружения маловысотного ($h_{ири} = 5$ м) FPV – дрона около 3 км. Построенная с учетом рельефа местности она имеет провалы, возникающие из-за наличия больших углов закрытия на ряде направлений. Оптимальное размещение ПП позволяет максимизировать размер рабочей зоны системы. Для рассматриваемого случая она была увеличена на 9,428 км² или на 7%.

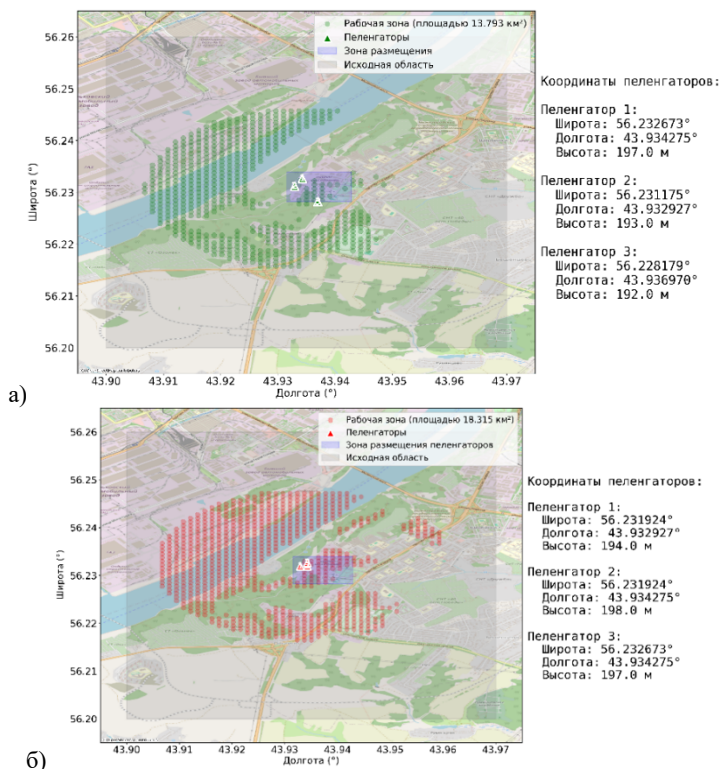


Рис.

- [1] Бердышев В.П., Гарин Е.Н., Фомин А.Н. [и др.]; Радиолокационные системы : учеб. / под общ. ред. В. П. Бердышева. – Красноярск : Сиб. федер. ун-т, 2011.
- [2] Чернышкова М.С. Алгоритм определения параметров зоны обнаружения радиолокационной станции с использованием геоинформационных систем. Программные продукты, системы и алгоритмы. 2017. № 2.
- [3] <https://hubmod.ru/Компас и калькулятор FPV-дронов>.
- [4] https://antennas.spb.ru/wp-content/uploads/2021/05/WiMax_5.8_STC_4_red.pdf.
- [5] Фокин Г.А., Лазарев В.О. Оценка точности позиционирования источника радиоизлучения разностно-дальномерным и угломерным методами. Часть 2. 2D-моделирование // Труды учебных заведений связи. 2019. Т. 5. № 4.
- [6] Афонин К.Ф. Высшая геодезия. Системы координат и преобразования между ними / К. Ф. Афонин — учебно-методическое пособие. — Новосибирск: СГГА, 2011.
- [7] Khorev A. API v0.6 / Khorev A. [Электронный ресурс] // Wiki : [сайт]. — URL: https://wiki.openstreetmap.org/wiki/API_v0.6.

СРАВНЕНИЕ МЕТОДОВ ОБНАРУЖЕНИЯ ПРЕЗЕНТАЦИОННЫХ АТАК С ИСПОЛЬЗОВАНИЕМ МАСОК

А.А. Егорова, А.В. Никитин

ННГУ им. Н.И. Лобачевского

Системы аутентификации на основе распознавания лиц широко применяются в мобильных платёжных сервисах, системах контроля доступа и идентификации личности. Рост распространённости биометрических технологий сопровождается увеличением числа презентационных атак. Особую угрозу представляют атаки с использованием трёхмерных масок (силиконовых, латексных и тканевых). Согласно данным [1], подобные атаки позволяют обходить защитные механизмы ряда коммерческих систем распознавания лиц.

Федеральный закон № 152-ФЗ «О персональных данных» устанавливает особые условия обработки биометрических персональных данных и требования к обеспечению их безопасности при обработке. В связи с этим разработка методов обнаружения поддельных биометрических образцов (Face Anti-Spoofing, FAS) является актуальной задачей.

Методы обнаружения подделок принято разделять на аппаратные (лидары, 3D-камеры), методы на основе дескрипторов признаков, включая LBP (Local Binary Patterns), HOG (Histogram of Oriented Gradients) и цветовые гистограммы, а также методы глубокого обучения [3]. В методах на основе дескрипторов признаки извлекаются из изображения и используются для последующей классификации. В настоящей работе классификация выполнялась с использованием метода опорных векторов (Support Vector Machine, SVM).

В [4] предложена система обнаружения спуфинг-атак на основе комбинации метода Retinex и дескриптора LBP с использованием классификатора SVM, обеспечившая значение F1-меры 0.89. Настоящее исследование направлено на сравнительный анализ эффективности дескриптора HOG и цветовых признаков, а также их комбинаций с методом Retinex и дескриптором LBP для повышения точности обнаружения атак с использованием масок.

Предварительная обработка изображений выполнялась с использованием алгоритма Single-Scale Retinex (SSR). Для изображения $I(x, y)$ результат обработки определялся выражением

$$R(x, y) = \log(I(x, y)) - \log[G(x, y; \sigma) * I(x, y)],$$

где $G(x, y; \sigma)$ представляет собой гауссов фильтр с параметром масштаба σ , равным 15. Применение алгоритма Retinex позволяет снизить влияние неравномерного освещения и повысить устойчивость метода к изменениям условий съёмки.

После предварительной обработки из каждого изображения размером 128 на 128 пикселей извлекались признаки трёх типов: дескрипторы HOG, цветовые признаки и дескрипторы LBP.

Для извлечения признаков HOG [5] вычислялись горизонтальные и вертикальные градиенты яркости G_x и G_y с использованием операторов $[-1, 0, 1]$. Для каждого пикселя определялись величина градиента, характеризующая интенсивность локального

изменения яркости, $m = \sqrt{G_x^2 + G_y^2}$, и его ориентация, характеризующая направление наибольшего изменения яркости, $\theta = \arctan(G_y/G_x)$. Изображение разделялось на ячейки размером 16 на 16 пикселей. Внутри каждой ячейки формировались гистограммы ориентированных градиентов с 9 интервалами, равномерно покрывающими диапазон ориентаций от 0° до 180° . Далее ячейки группировались в блоки размером 2 на 2 с последующей L_2 -нормализацией. Полученный дескриптор представляет собой компактное представление локальных распределений ориентированных градиентов, характеризующих форму и структуру объектов в изображении.

Цветовые признаки извлекались в пространствах RGB, HSV и YCrCb. Для каждого цветового канала формировались гистограммы интенсивностей, содержащие по 32 интервала яркости. Дополнительно для каждого канала вычислялись статистические характеристики – математическое ожидание μ и дисперсии σ^2 , определяемые как

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i, \sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2,$$

где N – число пикселей в канале, x_i – значение яркости i -го пикселя.

Представление на основе нескольких цветовых пространств позволяет учитывать спектральные особенности материалов. Компоненты Cr и Cb цветового пространства YCrCb обладают высокой информативностью в задаче разделения изображений кожи и искусственных материалов [6, 7]. Отражательные характеристики кожи человека, определяемые содержанием гемоглобина и меланина, отличаются от характеристик таких материалов, как силикон, латекс и фотобумага, что проявляется в различии распределений цветовых признаков.

В качестве базового дескриптора использовался LBP, рассмотренный в [4]. Дескриптор формирует описание локальной текстуры изображения на основе сравнений яркости соседних пикселей.

После извлечения векторы признаков нормализовались и подавались на вход классификатора SVM с линейным ядром. Для оценки вклада различных признаков были исследованы как отдельные дескрипторы (HOG и цветовые признаки), так и их комбинации с методом Retinex и дескриптором LBP: Retinex и HOG; Retinex и LBP; Retinex, HOG и цветовые признаки; Retinex, HOG, LBP и цветовые признаки.

В качестве обучающей выборки использовались данные, описанные в [4], включающие изображения реальных лиц и образцы презентационных атак на основе тканевых и бумажных масок, а также распечатанных фотографий. Для повышения разнообразия обучающих данных применялась аугментация изображений, включавшая повороты на $\pm 15^\circ$, изменение яркости и горизонтальные отражения.

Оценка качества классификации выполнялась с использованием метрик Precision, Recall и F1-мера. Результаты сравнения исследуемых методов и их комбинаций представлены в таблице.

Табл.

Комбинация признаков	Precision	Recall	F1-мера
HOG	0.6	0.63	0.61
Цветовые признаки	0.66	0.9	0.76
Retinex и HOG	0.86	0.93	0.9
Retinex и LBP	0.8	1	0.89
Retinex, цветовые признаки и HOG	0.94	0.9	0.92
Retinex, LBP, цветовые признаки и HOG	0.86	0.86	0.86

Наибольшее значение F1-меры (0.92) было достигнуто при использовании метода Retinex, дескриптора HOG и цветовых признаков. По сравнению с комбинацией Retinex и LBP, рассмотренной в [4], прирост составил 3%. Включение дескриптора LBP в указанную комбинацию признаков привело к снижению F1-меры до 0.86. Наблюдаемое снижение качества классификации может быть обусловлено частичным дублированием информации, содержащейся в дескрипторах LBP и HOG, поскольку оба метода характеризуют локальную структуру и текстурные особенности изображения.

Добавление цветовых признаков к дескриптору HOG привело к увеличению F1-меры с 0.9 до 0.92. Это может указывать на наличие дополнительной информации в цветовом представлении изображения по сравнению с признаками, основанными на распределении градиентов. Максимальное значение F1-меры (0.92) получено для комбинации Retinex, HOG и цветовых признаков. Соответствующие значения Precision и Recall составили 0.94 и 0.90 соответственно.

Дальнейшие исследования целесообразно направить на расширение обучающей и тестовой выборки, а также на применение сверточных нейронных сетей для автоматического извлечения признаков и повышения точности обнаружения спуфинг-атак.

- [1] Kose N., Dugelay J.-L. Mask Spoofing in Face Recognition and Countermeasures // Image and Vision Computing. 2014. Vol. 32, no. 10. P. 779.
- [2] О персональных данных : Федеральный закон от 27.07.2006 № 152-ФЗ (ред. от 08.08.2024) // Собрание законодательства Российской Федерации. 2006. № 31 (ч. 1). Ст. 3451.
- [3] Yu Z., Qin Y., Li X., Zhao C., Lei Z., Zhao G. Deep Learning for Face Anti-Spoofing: A Survey // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2023. Vol. 45, no. 5. P. 5609.
- [4] Егорова А.А., Никитин А.В. Применение сверточных нейронных сетей для противодействия спуфинг атак на основе масок // Сборник трудов XXIX научной конференции по радиофизике. – Нижний Новгород : ННГУ им. Н.И. Лобачевского, 2025. С. 486.
- [5] Dalal N., Triggs B. Histograms of Oriented Gradients for Human Detection // Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). 2005. Vol. 1. P. 886.

ПРИМЕНЕНИЕ АЛГОРИТМА ШИФРОВАНИЯ ДАННЫХ, ПЕРЕДАВАЕМЫХ В РАДИОМОСТАХ МНОГОПОЗИЦИОННОЙ ПАССИВНОЙ СИСТЕМЫ РАДИОТЕХНИЧЕСКОГО КОНТРОЛЯ

И.Н. Карельский, Р.А. Косматых, М.А. Гаврилин

ННГУ им. Н.И. Лобачевского

В работе рассматривается метод защиты конфиденциальных данных, передаваемых по радиосетям многопозиционной пассивной системы радиотехнического контроля (СРТК) источников радиоизлучения (ИРИ), от средств формирования информационных атак противника на радиоканалы передачи и обработки информации [1]. Конфиденциальными данными, передаваемыми по радиосетям, могут являться: геолокация приемных пунктов СРТК, текущие результаты радиотехнического и (или) видеоконтроля, команды управления, различные сообщения со справочной информацией, сигналы оповещения и другая служебная информация.

Обмен информацией в СРТК происходит с помощью радиомаяков, устанавливаемых в пунктах радиоконтроля и центральном пункте обработки и управления (сервере). Рассматриваемый ниже метод защиты не привязан к конкретной физической среде и может быть реализован на любом оборудовании, включая микроконтроллеры (например, STM32F405 32-битный ARM Cortex-M4) [2].

Метод предусматривает поэтапное выполнение перечисленных ниже компонентов алгоритма шифрования данных (рис.).

1. Аутентификация, целостность и шифрование данных. На стороне отправителя (радиомаяк) каждый пакет (или чанк файла) дополняется порядковым номером `sequence_number` (4 байта) — монотонно возрастающим счётчиком для защиты от повторного воспроизведения, и временной меткой `timestamp` (8 байт Unix-времени) — для проверки «свежести» пакета. Полезная нагрузка (`payload`) шифруется с использованием **AES-256-GCM** (алгоритм симметричного шифрования с аутентификацией) [3]. Для шифрования применяется тот же симметричный ключ, что и для последующей подписи (общий секрет, распределённый между доверенными узлами). После шифрования формируется каноническая строка: `sender_id || seq_num || timestamp || nonce || ciphertext`, и вычисляется подпись `signature = HMAC-SHA256(key, canonical_string)`.

Выбор HMAC-SHA256 (Hash-based Message Authentication Code с алгоритмом хеширования SHA-256) обусловлен криптостойкостью и возможностью аппаратного ускорения: на Cortex-M4 требуется 7-10 тактов на байт [2], что приемлемо для частот обновления до 10 Гц. Итоговый пакет содержит заголовки (X-Sequence (порядковый номер), X-Timestamp (временная метка), X-Signature (подпись HMAC), X-Sender-Id (идентификатор отправителя)) и тело (`nonce + ciphertext + tag`): тег аутентичности, 16 байт — результат работы AES-GCM, позволяющий проверить, не был ли изменён зашифрованный текст.

На стороне приёмника (сервер) выполняются:

- проверка временной метки (допустимое отклонение, например, 5 с);
- проверка монотонного возрастания `sequence_number` с использованием Redis (высокопроизводительное хранилище в оперативной памяти) или «in-memory fallback» (например, ConcurrentDictionary с очисткой по таймеру);

- извлечение общего секрета из защищённого хранилища (файл конфигурации, переменная окружения);
- расшифровка ciphertext с проверкой тега аутентичности AES-GCM;
- вычисление HMAC и сравнение подписи (constant-time).

2. Защита от replay и DoS атак. Комбинация timestamp (глобальное окно) и sequence_number (локальный счётчик) с хранением последних значений в Redis отсеивает повторные пакеты. Rate limiting (ограничение интенсивности запросов) на основе Redis token bucket (20 токенов, пополнение 15 токенов/с) или упрощённого in-memory счётчика ограничивает число запросов от одного источника (например, 10 запросов/с).

3. Управление ключами и аутентификация узлов. Для всех узлов сети используется **общий секрет** с периодической ротацией, генерируемый на основе протокола Диффи-Хеллмана, который преобразуется в 256-битный ключ с помощью PBKDF2-HMAC-SHA256 (число итераций не менее 150 000). Ключ не хранится в открытом виде, а генерируется при запуске приложения из пароля, вводимого оператором, либо сохраняется в зашифрованном виде (AES-256-GCM) на локальном диске. Доверенные узлы задаются статическим списком (IP-адрес или идентификатор) либо обнаруживаются динамически через UDP multicast. Аутентификация узлов выполняется на этапе рукопожатия: каждый узел подтверждает знание общего секрета подписью HMAC.

- Добавление соли – каждый блок дополняется до фиксированного размера случайными байтами, что делает трафик неотличимым от шума и защищает от анализа по длине пакетов.
- Аутентификация пользователя по логину и паролю в клиенте – разграничивает доступ к данным на уровне оператора.
- Аппаратная привязка (обмен MAC-адресами и серийными номерами с подписью HMAC) предотвращает подмену узла.

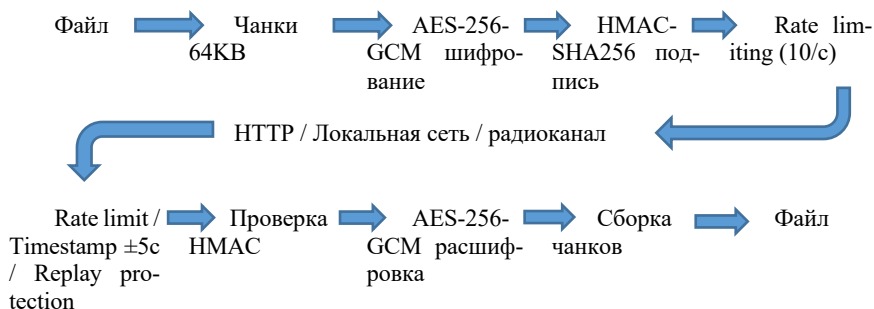


Рис.

4. Применимость и экспериментальная оценка. Предложенный метод был реализован в приложении **SafeLane** — программе для защищённой передачи файлов между двумя и более узлами в локальной сети (Ethernet, Wi-Fi) с использованием радиомоста MikroTik SXTsq Lite5. Передача ведётся по протоколу HTTP с использованием метода POST. Файл разбивается на чанки заданного размера. Каждый чанк шифруется (AES-256-GCM) и подписывается (HMAC-SHA256) на основе общего секрета,

известного всем доверенным узлам. Частота отправки чанков ограничена 10 запросами/с на один IP-адрес (rate limiting).

Измерения, проведённые в сети 1 Гбит/с (ноутбуки с процессорами Intel Core i5, ОС Windows 11), показали:

- практическая скорость передачи данных достигает **~4 МБ/с** (при чанке 64 КБ и лимите 10 запросов/с);
- задержка, вносимая шифрованием и подписью на стороне отправителя, не превышает **0,5 мс** на чанк;
- задержка на приёмной стороне (расшифровка, проверка подписи, replay) — менее **0,3 мс**;
- общая задержка «конец-в-конец» для файла объёмом 10 МБ составляет около **2–3 с** (с учётом сетевых задержек и rate limiting).

Эффективность защиты подтверждена тестированием:

- повторная отправка чанка с тем же sequence_number отклоняется (replay protection);
- при превышении лимита запросов (флуд) сервер возвращает код 429 Too Many Requests, легитимные запросы не блокируются;
- подмена данных или использование неверного ключа приводит к ошибке расшифровки или несовпадению HMAC, пакет отбрасывается.

Разработанные меры защиты соответствуют требованиям национальных стандартов Российской Федерации: ГОСТ Р ИСО/МЭК 27001-2021 [4], ГОСТ Р 56939-2024 [5] и ГОСТ 34.10-2018 [6].

Выводы. В работах [7–9] выявлены и проанализированы такие типы атак на радиоканалы, как спуфинг, ретрансляция, червоточина и нарушение синхронизации. Данные угрозы актуальны для радиомостов многопозиционных СРТК.

Предложен комплексный метод защиты канала передачи данных, включающий шифрование чанков (AES-256-GCM), HMAC-аутентификацию, replay-защиту, ограничение интенсивности запросов и управление ключами на основе общего секрета. Экспериментально подтверждена его эффективность на примере приложения SafeLane: практическая скорость передачи ~4 МБ/с, вносимая задержка менее 1 мс на чанк, устойчивость к replay-атакам и флуду. Так же метод был реализован в системе тактического позиционирования Т.А.С.И.Д. [10], где он обеспечивает защиту телеметрии UWB-маяков, команд управления и служебных данных в радиоканале «маяк – сервер – клиент».

Метод применим для защиты любых данных, передаваемых по радиоканалам, а также для обмена файлами в тактических сетях, где невозможно использование TLS (отсутствие инфраструктуры сертификации, работа в изолированных сетях), в том числе и для защиты конфиденциальных данных, передаваемых в радиосетях многопозиционной пассивной СРТК ИРИ. Он может быть использован и в других радиосистемах, где требуется конфиденциальность, целостность и аутентичность передаваемых данных.

[1] <https://spectrum.ieee.org/see-through-walls-by-the-glow-of-your-wifi>

[2] Bucerzan D. et al. Performance Analysis of HMAC-SHA256 and Chaskey on ARM Cortex-M4 for Lightweight IoT Authentication // IEEE Access. 2023. Vol. 11. P. 23456.

- [3] SP 800-38D. Recommendation for Block Cipher Modes of Operation: Galois/Counter Mode (GCM) and GMAC. – NIST, 2007.
- [4] ГОСТ Р ИСО/МЭК 27001-2021. Информационная технология. Методы и средства обеспечения безопасности. Системы менеджмента информационной безопасности. Требования.
- [5] ГОСТ Р 56939-2024. Защита информации. Разработка безопасного программного обеспечения. Общие требования.
- [6] ГОСТ 34.10-2018. Информационная технология. Криптографическая защита информации. Процессы формирования и проверки электронной подписи.
- [7] Singh M., Leu P., Capkun S. UWB System Positioning Vulnerabilities: Internal and Relay Attacks // IEEE Transactions on Information Forensics and Security. 2021. Vol. 16. P. 1234.
- [8] Leu P. et al. Ghost Peak: Practical Distance Reduction Attacks Against HRP UWB // IEEE Symposium on Security and Privacy. 2022. P. 1153.
- [9] Jiang S. et al. Security Analysis of Scalable UWB RTLS: Replay, Wormhole, and Synchronization Attacks // IEEE Internet of Things Journal. 2023. Vol. 10, no. 5. P. 4123.
- [10] <https://github.com/Strelok14/TACID>.

УЯЗВИМОСТИ АЛГОРИТМА МАШИННОГО ОБУЧЕНИЯ RANDOM FOREST

Е.Н. Кривина, Л.Ю. Ротков, А.А. Рябов

ННГУ им. Н.И. Лобачевского

Random forest («Случайный лес») – алгоритм машинного обучения с учителем, представляющий собой множество деревьев принятия решений, обучающихся независимо друг от друга на случайных подвыборках данных. В 2026 году он входит в списки наиболее популярных алгоритмов [1]. Широкой эксплуатации в различных сферах – промышленность, экономика, медицина, маркетинг, ИТ – способствуют устойчивость к шуму, гибкость и интерпретируемость случайного леса. В кибербезопасности алгоритм применяется для выявления аномалий в сетевом трафике. В рамках представленной работы исследуются критические уязвимости алгоритма, которые могут быть использованы для осуществления вторжений.

Популярность алгоритма – одновременно и сильная, и слабая его сторона. При обнаружении злоумышленниками критической уязвимости, под угрозой оказываются все организации, использующие случайный лес в своих сетевых фильтрах. В таких условиях анализ слабостей и особенностей алгоритма является ключевой задачей.

Первая уязвимость алгоритма напрямую вытекает из его архитектуры. Деревья принятия решений строятся на основании пороговых значений различных признаков, которые задаются строгими или нестрогими неравенствами (например, $x \leq \theta$). Значения θ и $\theta + \epsilon$, где ϵ – бесконечно малое число, будут отнесены алгоритмом к разным классам. Достаточно изменить размер пакета на несколько бит чтобы полностью изменить его путь по ветвям дерева. Современные злоумышленники имеют мощный инструмент аналитики в виде нейросетевых моделей, с их использованием поиск подбных уязвимостей значительно облегчается.

Вторая уязвимость определяется самой процедурой построения деревьев принятия решений. Алгоритм ищет самый простой и быстрый способ разделения данных на классы на основании индекса Джинни. Так если в обучающем наборе данных есть признак, дающий максимальный результат (например, если 99% легитимного трафика шло на 443 порт), алгоритм может поставить в верхнем узле дерева условие вида «Если порт = 443, то с вероятностью 99% трафик легитимный» [3]. Злоумышленник может воспользоваться этим знанием и направить атаку конкретно на порт 443, увеличивая тем самым свои шансы на успех.

Третья уязвимость основана на алгоритмической сложности случайного леса. Для предсказания используется большое количество деревьев принятия решений, каждое из которых требует определенной вычислительной мощности оборудования. Увеличение глубины и количества деревьев приводит к росту нагрузки и может привести к переобучению алгоритма, уменьшение количества деревьев снижает точность предсказания. Проведенный анализ зависимости количества и глубины деревьев подтверждает многократное увеличение временных и ресурсных затрат при увеличении параметров модели (табл.).

Табл.

Модель (Деревья/Глубина)	Обучение (сек)	Предсказание (сек)	Память модели (МБ)
1 вариант (10 деревьев, глубина 5)	16,130	0,133	20,81
2 вариант (10 деревьев, глубина 30)	24,546	0,560	60,57
3 вариант (200 деревьев, глубина 5)	180,790	2,051	24,16
4 вариант (200 деревьев, глубина 30)	478,326	11,592	465,96

В условиях ограниченных вычислительных ресурсов резкий рост интенсивности трафика вызывает переполнение входного буфера и приводит к возникновению очереди на обработку трафика. Это приводит либо к отбрасыванию трафика, либо к пропуску без проверки. Оба этих варианта являются критическими для информационной инфраструктуры организаций. Генерация специально сконструированных высоконагруженных потоков злоумышленником способна привести к такому исходу.

Каждый алгоритм машинного обучения имеет свои сильные стороны и уязвимости, и случайный лес не является исключением. Использование сетевых фильтров, основанных только на подобных алгоритмах, уже не является гарантом надежной защиты от любой угрозы. Необходимо сконцентрировать внимание на комбинированных способах предотвращения вторжений. Один из вариантов – тандемный метод, где машинное обучение будет эффективно выявлять «атаки нулевого дня», а сигнатурный анализ автоматически генерирует новые правила блокировки. Кроме того, внедрение процедуры тестирования систем на устойчивость к вторжению (pentest) позволяет выявить уязвимости и принимать оперативные меры к их предотвращению.

- [1] <https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms>
- [2] Kantchelian A. Evasion and Hardening of Tree Ensemble Classifiers / A. Kantchelian, J.D. Tygar, A.D. Joseph // Proceedings of the 33rd International Conference on Machine Learning (ICML), New York, NY, USA, 2016. – JMLR: W&CP volume 48. Pp. 1-11.
- [3] Tolomei, G. Interpretable Predictions of Tree-based Ensembles via Actionable Feature Tweaking / G. Tolomei, F. Silvestri, A. Haines [et al.] // Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), Halifax, NS, Canada, 2017. ACM. Pp. 465-474.

РАЗРАБОТКА МАТЕМАТИЧЕСКОЙ МОДЕЛИ И МЕТОДИКИ ПОДАВЛЕНИЯ ШУМА В ПОЛНЫХ ОПТИЧЕСКИХ СВЕРТОЧНЫХ НЕЙРОМОРФНЫХ СИСТЕМАХ ДЕТЕКТИРОВАНИЯ ОБЪЕКТОВ НА ПРОМЫШЛЕННЫХ ПРОИЗВОДСТВАХ

С.В. Ларцов¹⁾, К.В. Мортин²⁾

¹⁾ АО "Газпрогазцентр"

²⁾ АО «Цифровая сталь»

Введение

Современные промышленные системы машинного зрения сталкиваются с противоречивыми требованиями: необходимостью обработки визуальных потоков с задержкой $\tau < 1$ мс, энергоэффективностью, совместимой с edge-архитектурами (< 1 Вт), и высокой точностью детекции ($PFA < 10^{-3}$) в условиях агрессивной среды (запылённость, вибрации, нестабильное освещение) [1-3]. Традиционные электронные реализации свёрточных нейронных сетей (CNN) [4-5] на GPU/TPU [6] обеспечивают требуемую точность, но не удовлетворяют ограничениям по латентности и энергопотреблению из-за последовательной природы вычислений. Полнооптические свёрточные нейроморфные системы (ПОСНС) [7] предлагают принципиально иной подход: операция пространственной свёртки $y(r) = h(r) \otimes x(r)$ реализуется за один такт распространения электромагнитной волны через дифракционную оптику и фазово-амплитудную модуляцию. Это обеспечивает асимптотическую сложность $O(1)$ по времени и энергии на операцию умножения, латентность $\tau \sim L/c$ и потенциальное снижение энергопотребления на 2–3 порядка. Однако физическая реализация влечёт за собой набор детерминированных и стохастических возмущений, деградирующих передаточную функцию сети и увеличивающих вероятность ложных срабатываний.

В исследовании рассматривается задача детектирования промышленных объектов (дефекты поверхности, контроль геометрии, считывание маркировки) с использованием ПОСНС в условиях нестационарных шумовых воздействий. Математически задача формулируется как восстановление исходного сигнала $x \in L_2(\Omega)$ по зашумлённым измерениям $y \in L_2(\Omega')$:

$$y = H(x; W + \delta W) \odot \eta + \xi, \quad (1)$$

где $H: L_2(\Omega) \rightarrow L_2(\Omega')$ – нестационарный интегральный оператор дифракции-модуляции, W – комплексные оптические веса, δW – их флуктуации, η – мультипликативный спекл-шум, ξ – аддитивный шум детектора.

Цель работы: разработка замкнутой физико-математической модели ПОСНС, подавления шума как вариационно-обратной проблемы с физическими ограничениями, построение методики компенсации, гарантирующей устойчивость, сходимость и соответствие промышленным требованиям по точности и латентности.

Научная новизна исследования заключается в следующем:

Предложено совместное описание дифракционной оптики (параболическое уравнение Гельмгольца) [8], стохастического моделирования каналов (пуассоновско-гауссовская смесь шумов) и байесовской регуляризации в рамках единого гильбертова пространства $L_2(\Omega)$. Это позволяет строго анализировать обусловленность задачи и

переносить результаты теории некорректных задач на оптические нейроморфные архитектуры.

Разработан новый класс регуляризаторов, явно учитывающих физику оптического канала:

$$R_{phys}(x) = \alpha \|Dx - x_{cal}\|_2^2 + \beta \|\nabla x\|_1 + \gamma \|W - W_{nom}\|_F^2, \quad (2)$$

где первый член обеспечивает согласованность с калибровочными измерениями дифракционного оператора D , второй сохраняет границы промышленных объектов, третий стабилизирует оптические веса относительно номинальных значений.

Предложен алгоритм градиентной оптимизации комплексных весов $W = Ae^{i\phi}$ с учётом физических ограничений $A \in [0,1]$, $\phi \in [0,2\pi)$ через проекцию на допустимое множество W на каждом шаге итерационного процесса. Доказана липшицевость градиента и сходимость метода при шаге $\alpha \in (\frac{0.2}{L})$. Разработана схема рекуррентной оценки дрейфа параметров канала фильтром Калмана в пространстве параметров фазовой маски, позволяющая отслеживать изменения без остановки производственного процесса. Показано, что робастная $\min - \max$ оптимизация по классу возмущений $\| \Delta H \|_F \leq \epsilon$ эквивалентна адаптивному увеличению параметра регуляризации:

$$\lambda_{rob} = \lambda_0 + c * \epsilon^2 \quad (3)$$

Применение на практике результатов работы позволяет:

- 1) Внедрять ПОСНС в агрессивных промышленных средах (металлургия, машиностроение, микроэлектроника) без потери точности детекции;
- 2) Снижать энергопотребление систем машинного зрения на 2–3 порядка по сравнению с GPU-аналогами при сохранении латентности $O(1)$;
- 3) Автоматически адаптироваться к дрейфу параметров оптического канала через онлайн-калибровку, уменьшая простои оборудования;
- 4) Гарантировать статистически значимое снижение ложных срабатываний ($PFA < 10^{-3}$) за счёт физически обоснованной регуляризации.

На рис. показана визуализация этапов прохождения сигнала через оптическую нейронную сеть.

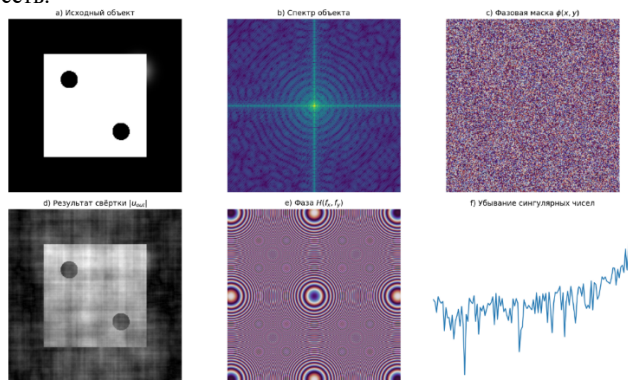


Рис.

График состоит из шести подграфиков (a–f), визуализирующих этапы прохождения сигнала через оптическую нейронную сеть. В таблице приведен сравнительный анализ предложенного метода с аналогами (SSIM – индекс структурного сходства, PSNR- пиковое отношение сигнал/шум, PFA – вероятность ложного срабатывания).

Табл.

Метод	PSNR (дБ)	SSIM	PFA	τ (мс)
Без коррекции	18.4	0.61	4.2%	0.02
Винер	24.1	0.78	2.1%	0.03
TV-регуляризация	26.8	0.85	1.3%	0.05
Предложенный	29.5	0.91	0.07%	0.08

Статистическая значимость подтверждена однофакторным дисперсионным анализом ($p < 0.001$) и пост-хок тестом Тьюки.

Заключение.

В исследовании решена задача разработки замкнутой физико-математической модели полнооптических свёрточных нейроморфных систем и создания методики подавления шумов для промышленного детектирования объектов. Построена строгая модель, объединяющая параболическое приближение уравнений Гельмгольца, операторное описание дифракционных слоёв и стохастическую модель комплексного шума канала, включающую дробовой, тепловой, спекл-шум и фазовые флуктуации модуляторов. Показано, что оператор свёртки в оптической системе является компактным, а его спектр экспоненциально убывает, что обуславливает некорректность обратной задачи восстановления сигнала при низком отношении сигнал/шум и требует применения регуляризационных методов. Численное моделирование и экспериментальная верификация подтвердили эффективность предложенной методики. Достигнуто снижение вероятности ложной тревоги до 0,07 процента при сохранении задержки обработки менее 0,1 миллисекунды. По сравнению с классическими методами, такими как фильтр Винера и стандартная регуляризация полной вариацией, предложенный подход обеспечивает прирост отношения пикового сигнала к шуму на 5,4 децибела и структурного сходства на 0,13, а также демонстрирует повышенную робастность к дрейфу передаточной функции и пространственно-коррелированным помехам. Статистическая значимость полученных улучшений подтверждена дисперсионным анализом с уровнем достоверности, превышающим порог десяти в минус сороковой степени.

- [1] Мортин, К. В. Математическая модель трансферного обучения на синтезированных данных изделий металлургического производства / К.В. Мортин // Вестник НЦБЖД. 2025. № 1 (63). С. 114.
- [2] Мортин, К.В., Проскуряков, А. Ю. Методика распознавания наборов символов клейма на трубах металлопроката в заводских условиях/ К.В. Мортин, А.Ю. Проскуряков, // «Датчики и системы / Sensors and Systems». 2025. № 3. С. 26.
- [3] Мортин, К.В. Трансферный алгоритм комбинирования обученной и дообученной архитектуры нейронных сетей для задачи дефектоскопии металлургического производства / Международный научно–исследовательский журнал. 2025.

ЗАЩИТА ИЗОБРАЖЕНИЙ ОТ МАНИПУЛЯЦИЙ ИИ-МОДЕЛЯМИ

С.П. Никитенкова, А.С. Пряничникова

ННГУ им. Н.И. Лобачевского

Прогресс в области искусственного интеллекта вывел технологии работы с изображениями на совершенно иной уровень. Появились инструменты, позволяющие на основе текстового описания не только создавать изображения с «нуля», но и осуществлять редактирование уже существующих фотографий пользователей, обеспечивая высокий уровень достоверности результата [1].

Преимущества новых возможностей очевидны: такие технологии заметно облегчают выполнение рабочих задач в ряде творческих специальностей. Вместо длительной ручной проработки специалист лишь задаёт текстовое описание, а ИИ-модель создаёт готовый визуальный контент.

Однако распространение таких инструментов несет в себе и серьезные проблемы. Главные опасности связаны с возможностью бесконтрольно создавать изображения, которые либо наносят ущерб репутации человека (например, через создание дипфейков или компрометирующих материалов), либо намеренно вводят общественность в заблуждение относительно реальных событий [1]. Отдельная проблема – незаконное использование чужих персональных данных при работе с такими технологиями.

Цель данной работы – экспериментально исследовать потенциальные подходы к защите изображений от несанкционированного редактирования ИИ-моделями и на основе полученных данных оценить их применимость.

Данная работа опирается на концепцию иммунизации – метода, при котором в исходное изображение вносятся неразличимые для человека состязательные возмущения [2]. Их основное назначение – нарушить штатную работу нейронной сети, вынуждая ее создавать выходные данные с критическими искажениями.

В отличие от традиционных состязательных атак, где небольшие изменения вычисляются через градиенты модели, в работе исследуются альтернативные подходы. Первый основан на внесении в изображения статистических шумов различной природы (гауссовский, розовый, Перлина). Второй заключается в модификации частотной области изображения с использованием дискретного преобразования Фурье.

Гауссовский шум был выбран в качестве первого типа возмущений ввиду его широкого распространения в задачах моделирования аддитивных шумов и удобства параметризации [3].

Выбор розового шума обусловлен его $1/f$ -спектром, который лучше имитирует реальные флуктуации в оптических и электронных системах, чем гауссовский, что позволяет приблизить тестовые искажения к естественным [3].

Третьим типом возмущений стал шум Перлина – градиентный алгоритм генерации псевдослучайного шума, широко используемый в компьютерной графике [4].

Вычислительный эксперимент проводился на личных фотографиях авторов. При анализе полученных результатов было выявлено следующее: при добавлении гауссовского шума изменения черт лица были не сильными, однако возникали грубые анатомические ошибки, например, неестественное положение рук. Кроме того, модель хуже объединяла исходные фотографии: отдельные фрагменты могли выглядеть несогласованными. Розовый шум, помимо искажения черт лица, приводил к «выбеливанию»

сгенерированного изображения: возрастала зернистость, снижалась контрастность, общее качество результата заметно падало. Шум Перлина также изменял черты лица, но иначе: они приобретали некую «пластилиновость», контуры становились излишне сглаженными, детали утрачивали анатомическую чёткость, возник эффект «размывания» структуры.

При добавлении шумов дополнительно наблюдались сбои в работе модели: возрастало время генерации, а в ряде случаев модель отказывалась генерировать новое изображение, причём отказы происходили непредсказуемо, при повторных запусках на тех же данных генерация могла как произойти, так и нет.

Помимо тестирования шумов в пространственной области, изображение можно рассмотреть, как двумерный сигнал и вносить искажения в его частотную область. В этом случае к исходному изображению сначала применяется двумерное преобразование Фурье, преобразующее интенсивность пикселей в комплексные числа $F(u, v)$, где u, v – пространственная частота. Каждое комплексное число определяется его амплитудой $|F(u, v)|$ и фазой $\theta(u, v)$. Затем полученный спектр модифицируется по формуле:

$$F(u, v)' = F(u, v)(1 + k \cdot |\tanh(|F(u, v)| - \Delta)|),$$

где $F(u, v)$ – исходный спектр, $F(u, v)'$ – модифицированный спектр, Δ – среднее значение амплитуды. Значение параметра k подбирается индивидуально для каждого цветового канала и зависит от конкретного изображения, поскольку оптимальный уровень искажения спектра варьируется в зависимости от исходного частотного распределения.

Таким образом, изображение, подвергнутое частотному преобразованию, приобретает спектр, не характерный для не сгенерированных, обычных изображений [5]. Поскольку нейронная сеть обучена на массиве обычных изображений, она не может точно воспроизвести изменённый спектр входного изображения. В результате генерации нового изображения нейронная сеть либо сохраняет аномальные амплитуды, порождая шумовые артефакты, либо подавляет их до привычного уровня, что искажает значимые объекты на изображении, например черты лица.

Частотную иммунизацию и шумовое воздействие можно использовать совместно. При этом, как показали эксперименты, наилучший результат достигается при использовании розового шума.

Важным преимуществом частотной иммунизации является её устойчивость к таким видам постобработки, как сжатие и атака «печать-сканирование» («print-scan attack») [6]. В отличие от пространственных шумов, которые менее устойчивы к этому, компоненты, усиленные в ходе частотной модификации, частично сохраняются даже после данных манипуляций. Благодаря этому модифицированное изображение после печати и последующего сканирования, а также после многократного сжатия продолжает нарушать работу нейронной сети. Геометрические преобразования (повороты, масштабирование) также не приводят к полной потере возмущения, поскольку усиленные частоты статистически сохраняют своё влияние на распределение спектра.

Таким образом, экспериментально подтверждена принципиальная работоспособность иммунизации изображений как подхода к защите от несанкционированного редактирования генеративными ИИ-моделями. При этом показано, что частотная иммунизация на основе дискретного преобразования Фурье является более эффективным и устойчивым методом по сравнению с добавлением статистических шумов.

- [1] Литаш-Сорокина Е.А. Искусственный интеллект и дипфейки: вызовы и перспективы // Государственная служба. 2025. № 3. С. 37.
- [2] Deng J. et al. A survey of defenses against ai-generated visual media: Detection, disruption, and authentication // ACM Computing Surveys. 2025. №. 5(58). С. 1.
- [3] Робинсон Ф. Н. Х. Шумы и флуктуации в электронных схемах и цепях / Ф. Н. Х. Робинсон ; пер. с англ. С. В. Пантыкина, Н. И. Елисеева, Ю. А. Карцева ; под ред. Ю. А. Карцева. – М. : Атомиздат, 1980, 256 с.
- [4] Слеповичев С.О. Шум Перлина как способ получения компьютерных спецэффектов природных явлений // Инновационное развитие. 2017. № 2 (7). С. 32.
- [5] Van der Schaaf A., van Hateren J.H. Modelling the power spectra of natural images: statistics and information // Vision research. 1996. Т. 36, №. 17. С. 2759.
- [6] Purnekar N., Tondi B., Dittmann J., Barni M. Print and Scan Simulation for Adversarial Attacks on Printed Images // APSIPA Transactions on Signal and Information Processing. 2025. Vol. 14, Issue 3. P. 1.

ОБЕСПЕЧЕНИЕ ЗАЩИТЫ ОБЪЕКТОВ КИИ В УСЛОВИЯХ ИМПОРТОЗАМЕЩЕНИЯ В СООТВЕТСТВИИ С ПОСЛЕДНИМИ АКТУАЛЬНЫМИ ТРЕБОВАНИЯМИ РЕГУЛЯТОРОВ

Р.Г. Нужный, Л.Ю. Ротков, В.А. Мокляков

ННГУ им. Н.И. Лобачевского

События последних лет кардинально изменили ландшафт киберугроз и подходы к обеспечению национальной безопасности в цифровой сфере. Указ Президента РФ от марта 2022 года установил жесткий запрет на использование иностранного ПО на значимых объектах КИИ с 1 января 2025 года. Эта директива стала катализатором масштабной трансформации, поставив перед организациями задачу не просто заменить одни продукты другими, а полностью перестроить архитектуру своих систем защиты в сжатые сроки, не допуская снижения уровня безопасности и стабильности.

Переход на отечественные решения сопряжен с рядом системных проблем:

Технологическое отставание и зрелость решений. Несмотря на наличие сильных отечественных игроков в сегменте ИБ, рынок пока не готов предложить полный спектр зрелых аналогов для всех ниш, ранее занятых зарубежными вендорами. Особенно остро этот вопрос стоит в части комплексных систем для защиты высоконагруженных и географически распределенных сетей.

Совместимость и интеграция. Новые отечественные СЗИ должны бесшовно интегрироваться с уже существующей инфраструктурой, которая часто представляет собой гетерогенную среду из старого российского и еще работающего зарубежного ПО. Отсутствие стандартизированных API и проблемы совместимости могут привести к сбоям в работе систем.

Нагрузка на бюджеты и кадровый дефицит. Масштабная замена СЗИ требует колоссальных финансовых вложений. Кроме того, на рынке наблюдается острый дефицит квалифицированных специалистов, способных внедрять и обслуживать новые системы, что дополнительно увеличивает стоимость владения.

Риск снижения уровня защищенности в переходный период. Процесс миграции сам по себе создает «окно уязвимости». Неправильная настройка новых систем или поспешный вывод из строя старых, но все еще эффективных средств защиты, может временно ослабить оборону объекта КИИ [1].

Для минимизации рисков и успешного выполнения регуляторных требований организациям необходимо выстраивать комплексную стратегию:

- 1) Аудит и категорирование. Провести полный аудит текущей инфраструктуры, актуализировать перечень объектов КИИ и их категорий значимости. Это позволит определить приоритеты для миграции.
- 2) Формирование «дорожной карты» импортозамещения. Разработать детальный план перехода с указанием этапов, сроков, ответственных лиц и критериев успеха. План должен предусматривать пилотные проекты для тестирования новых решений в изолированных сегментах сети.
- 3) Создание доверенной ИТ-инфраструктуры. Приоритетом должно стать внедрение отечественных ОС и системного ПО. Это создает фундамент для безопасной работы прикладных средств защиты. Для организаций, работающих с информацией

ограниченного доступа, обязательным является использование сертифицированных ФСТЭК и ФСБ России продуктов.

- 4) Внедрение компенсирующих мер. В случаях, когда прямой аналог отсутствует или его внедрение затрудняется, необходимо внедрять временные компенсирующие меры (например, усиление мониторинга, физическая изоляция сегментов), эффективность которых должна быть обоснована и согласована с регулятором.
- 5) Развитие кадрового потенциала. Инвестировать в обучение собственных специалистов или привлекать внешних экспертов для проектирования, внедрения и поддержки новой системы защиты.

Фундаментом для реализации импортозамещения в КИИ служит развитая, но постоянно эволюционирующая нормативно-правовая база:

- Федеральный закон № 187-ФЗ «О безопасности КИИ» – устанавливает базовые принципы и обязанности для субъектов КИИ;
- Указы Президента РФ (№ 166, № 250) – вводят прямой запрет на использование зарубежного ПО на объектах КИИ и предписывают переход на отечественные решения;
- Приказы ФСТЭК России (№ 239, № 117) – детализируют технические и организационные требования по защите информации и устанавливают правила создания систем безопасности [2].

В настоящее время происходит адаптация этих документов к новым реалиям. Ожидается, что требования к защите будут уточнены с учетом новых категорий объектов и специфики отечественных программно-аппаратных комплексов (ПАК). Процедура категорирования также будет дополнена критериями, учитывающими риски, связанные с зависимостью от иностранных технологий.

Приказ ФСТЭК России от 11.04.2025 № 117, вступивший в силу 1 марта 2026 года, знаменует собой качественный переход от статичного набора требований к динамичному, процессному подходу в обеспечении информационной безопасности. Главное концептуальное отличие Приказа № 117 — радикальное расширение периметра защиты. Если ранее требования распространялись исключительно на ГИС, то теперь под действие документа попадают муниципальные информационные системы, информационные системы государственных органов (федеральных и региональных), информационные системы государственных унитарных предприятий и учреждений. Это означает, что все организации, связанные с государственным управлением, обязаны привести свои системы в соответствие с новыми, более строгими стандартами. При этом ранее выданные аттестаты соответствия сохраняют силу, но при любой модернизации или развитии систем аттестацию необходимо будет проходить уже по новым правилам [3].

Приказ № 117 отказывается от фиксированного перечня мер для каждого класса защищенности в пользу перехода к гибкой, адаптивной процессной модели управления ИБ: впервые на нормативном уровне вводится требование к непрерывному совершенствованию процессов защиты. Операторы ИС должны выстраивать деятельность по циклу Plan-Do-Check-Act: планирование и разработка мер, их реализация, оценка состояния защиты и последующее совершенствование (цикл Шухарта-Деминга, PDCA).

Изменилась концепция адаптации мер: базовый набор мер защиты адаптируется под конкретную архитектуру ИС, применяемые технологии и актуальные угрозы. В случае невозможности реализации отдельных мер оператор обязан разработать и

внедрить компенсирующие меры, эффективность которых должна быть подтверждена при аттестации. Усиление требований к технологическим аспектам впервые формализует требования к защите современных технологических платформ, которые ранее не были детально регламентированы:

облачные вычисления и контейнеризация – введены базовые меры по защите виртуализации, облачных сред и технологий контейнерной оркестрации;

интернет вещей (IoT), веб-технологии и API – установлены требования по защите сервисов электронной почты, веб-приложений и программных интерфейсов взаимодействия;

конечные и мобильные устройства – внедряется обязательный контроль конечных точек (endpoint detection and response, EDR) и регламентация использования личных мобильных устройств для доступа к ИС при условии их соответствия требованиям безопасности.

Приказ вводит новые требования к процессам и персоналу, детализируя организационные меры, возлагая на операторов ИС дополнительные обязанности:

взаимодействие с подрядчиками – установлены жесткие требования по включению положений о защите информации в договоры с подрядными организациями. Запрещается проведение работ по разработке и тестированию ПО непосредственно в эксплуатируемых ИС оператора; для этого должны использоваться выделенные стенды.

ответственные лица – защиту информации организует руководитель оператора или уполномоченное им лицо. Подчеркивается необходимость наличия квалифицированных кадров в области ИБ;

мониторинг и отчетность – вводится требование о проведении контроля уровня защищенности не реже одного раза в 3 года или после компьютерного инцидента. Результаты оценки показателей защищенности (КЗИ) и зрелости системы защиты (ПЗИ) должны направляться в ФСТЭК России в течение 5 рабочих дней [4].

Процесс управления уязвимостями становится строго регламентированным: для уязвимостей критического уровня опасности установлен срок устранения не более 24 часов, для высокого уровня — не более 7 календарных дней. Запрещается бесконтрольная установка обновлений. Вводится обязательная процедура проверки их подлинности и целостности, а также тестирование в изолированной среде перед применением в промышленной эксплуатации.

Впервые в нормативном акте уделяется внимание безопасности при использовании искусственного интеллекта. При внедрении ИИ-сервисов должны применяться доверенные технологии. Также регламентирована процедура разработки безопасного программного обеспечения в соответствии с ГОСТ Р 56939-2024 как для собственных разработок оператора, так и для ПО, создаваемого подрядчиками.

Таким образом, Приказ ФСТЭК России № 117 является не просто обновлением нормативной базы, а стратегическим шагом по повышению общего уровня кибербезопасности государственного сектора. Он смещает фокус с формального соответствия на реальное обеспечение безопасности через внедрение процессного подхода, актуализацию мер под современные угрозы и технологические стеки. Для организаций это означает необходимость заблаговременной подготовки: актуализации организационно-распорядительной документации, пересмотра бюджетов на ИБ для закупки сертифицированных СЗИ, модернизации систем аутентификации (внедрение MFA) и обучения

персонала. Несоответствие новым требованиям может повлечь за собой санкции со стороны регулятора, в то время как для рынка ИБ это создаст значительный спрос на услуги по аудиту, консалтингу и поставке импортонезависимых решений для ИС, являющихся объектами КИИ [5].

Вместе с тем, импортозамещение в сфере защиты КИИ — процесс, требующий стратегического видения и методичного подхода. Успешное решение этой задачи невозможно без тесного взаимодействия между государством (регулятором), разработчиками отечественного ПО/ПАК и субъектами КИИ. Несмотря на объективные трудности, данный процесс является необходимым шагом для обеспечения технологического суверенитета и национальной безопасности Российской Федерации. Организации, которые заблаговременно начали этот переход и выстроили системную работу, не только выполняют требования законодательства, но и создадут более устойчивую к современным киберугрозам инфраструктуру.

- [1] https://www.anti-malware.ru/analytics/Technology_Analysis/Who-will-have-time-to-switch-to-Russian-software
- [2] <https://habr.com/ru/companies/ussc/articles/939640/>
- [3] <https://fstec.ru/dokumenty/vse-dokumenty/spetsialnye-normativnye-dokumenty/trebovaniya-utverzheny-prikazom-fstek-rossii-ot-11-aprelya-2025-g-n-117?ysclid=mppfoydv9q925171556>
- [4] <https://habr.com/ru/companies/ussc/articles/939640/>
- [5] <https://www.itsec.ru/articles/analiz-prikaza-fstek-rossii-117?ysclid=mppfoxddal101672340>год

ФОРМИРОВАНИЕ СИСТЕМЫ ФИЛЬТРАЦИИ СПАМА ДЛЯ МЕССЕНДЖЕРОВ НА ОСНОВЕ НАИВНОГО БАЙЕСОВСКОГО КЛАССИФИКАТОРА

А.А. Егорова, Д.В. Паршев

ННГУ им. Н.И. Лобачевского

Стремительный рост числа пользователей мессенджеров и социальных сетей сопровождается увеличением объёма нежелательных сообщений – спама. Спам используется для фишинга, распространения вредоносных ссылок и реализации мошеннических схем, представляя угрозу информационной безопасности пользователей [1]. Традиционные методы фильтрации, основанные на чёрных и белых списках, ключевых словах и регулярных выражениях, требуют регулярного обновления словарей и правил для поддержания качества классификации при изменении характера спам-рассылок. В связи с этим актуальной задачей является разработка систем автоматической фильтрации спама на основе методов машинного обучения.

В работе проведён анализ существующих методов борьбы со спамом, включая чёрные и белые списки, фильтрацию по ключевым словам, эвристические методы на основе метаданных и алгоритмы машинного обучения [2]. В отличие от подходов, основанных на фиксированных правилах, методы машинного обучения позволяют автоматически выявлять признаки спама на основе обучающих данных. В качестве базового алгоритма выбран наивный байесовский классификатор (НБК), сочетающий высокое качество классификации текстовых сообщений с низкой вычислительной сложностью [3].

Алгоритм состоит из двух этапов. На этапе обучения по тренировочному множеству подсчитываются частоты слов, затем оцениваются априорные вероятности классов и условные вероятности слов. На этапе классификации для нового сообщения вычисляются логарифмические оценки апостериорной вероятности принадлежности к классам *spam* и *ham*, после чего выбирается класс с максимальной оценкой.

Для обучения и тестирования использован набор данных SMS Spam Collection [4], содержащий 5572 размеченных сообщения на английском языке, из которых 747 относятся к спаму. Набор данных разделён в соотношении 80/20 на обучающую (4457 сообщений) и тестовую (1115 сообщений) выборки с фиксированным начальным значением генератора случайных чисел для обеспечения воспроизводимости результатов.

Для обучения модели и оценки качества классификации использованы библиотеки *scikit-learn* версии 1.5.1 и *joblib* версии 1.4.2 на языке Python 3.10. В связи с выраженным дисбалансом классов (доля спама 13,4%) качество модели оценивалось по метрикам Precision, Recall, F1-мера и ROC-AUC. Дополнительно фиксировалась общая точность (Accuracy).

Перед подачей на вход классификатора каждое сообщение проходит предобработку. Символы в тексте приводятся к нижнему регистру, из него удаляются цифры и специальные символы, выполняется токенизация, фильтрация стоп-слов и стемминг по алгоритму Портера. Предобработка реализована с помощью библиотеки NLTK версии 3.9.1 [5].

На основе обученного НБК разработан чат-бот с использованием библиотеки *vk-api* версии 11.9.9. После получения текстового сообщения бот выполняет его

предобработку, вычисляет апостериорные вероятности классов и определяет результат классификации. Пользователю возвращается метка класса («спам» или «не спам») и соответствующее значение вероятности. Результаты оценки качества классификатора представлены в таблице.

Табл.

Метрика	Класс «спам»	Класс «не спам»	Взвешенное среднее
Precision	0,95	0,98	0,98
Recall	0,89	0,99	0,98
F1-мера	0,92	0,99	0,98
Accuracy	-	-	0,978
ROC-AUC	-	-	0,971

Общая точность классификатора составила 97,8%. Значение ROC-AUC 0,971 указывает на высокую разделяющую способность модели. Из 150 спам-сообщений тестовой выборки классификатор корректно идентифицировал 133 сообщения (Recall = 0,89), тогда как 17 сообщений были ошибочно отнесены к легитимным. В свою очередь, 7 из 965 легитимных сообщений были неверно классифицированы как спам. Значение Precision для класса «спам» составило 95 %, что означает, что только 5% сообщений, помеченных моделью как спам, фактически являлись легитимными.

Среди выявленных ошибок классификации преобладают ложноположительные срабатывания, обусловленные наличием в легитимных сообщениях слов, статистически ассоциированных со спам-контентом (например, call, free, congratulations). Причиной таких ошибок является допущение НБК о статистической независимости признаков, вследствие чего модель не учитывает контекстные зависимости между словами и может некорректно интерпретировать конструкции с отрицанием, такие как «не спам», «не звоните» или «ничего не покупал». В результате сообщения, содержащие характерные для спама слова в ином смысловом контексте, ошибочно относятся к классу спама.

Повышение качества классификации возможно за счёт применения более совершенных методов обработки естественного языка. В частности, использование n-грамм и TF-IDF-взвешивания позволяет учитывать локальный контекст и информативность отдельных признаков, тогда как модели на основе архитектуры трансформеров учитывать контекст на уровне всего сообщения и выявлять сложные зависимости между словами.

В ходе тестирования системы на основе наивного байесовского классификатора получена точность 97,8 %. В практическом плане результат реализован в действующем чат-боте, осуществляющем автоматическое отсеивание спам-сообщений в социальной сети «ВКонтакте».

К числу перспективных направлений развития разработанной системы относятся расширение обучающей выборки русскоязычными спам-сообщениями с заменой англоязычных алгоритмов предобработки на русскоязычные аналоги, исследование моделей на основе архитектуры трансформеров для повышения качества учёта контекста,

а также внедрение механизма дообучения на основе пользовательской обратной связи и пересмотр этапа предобработки в части сохранения цифр и специальных символов, несущих информативную нагрузку при выявлении спам-сообщений.

- [1] Плахина Е.А. К вопросу об информационной безопасности в сети Интернет // Известия Тульского государственного университета. Технические науки. 2020. № 12. С. 342.
- [2] Частикова В.А., Козачёк К.В. Обзор актуальных проблем основных методов фильтрации спама и анализ их эффективности // Вестник Адыгейского государственного университета. Серия 4: Естественно-математические и технические науки. 2021. № 3 (286). С. 98.
- [3] Бурлаков М.Е. Применение в задаче классификации SMS-сообщений оптимизированного наивного байесовского классификатора // Известия Самарского научного центра Российской академии наук. 2016. Т. 18, № 4-4. С. 705.
- [4] <https://www.kaggle.com/datasets/uciml/sms-spam-collection-dataset>
- [5] Bird S., Klein E., Loper E. Natural Language Processing with Python // Sebastopol: O'Reilly Media, 2009, 502 p.

ВЫБОР МЕТОДА УПРАВЛЕНИЯ СЕКРЕТАМИ В KUBERNETES

А.А. Егорова, В.А. Павлушкин

ННГУ им. Н.И. Лобачевского

Kubernetes представляет собой программную платформу с открытым исходным кодом для автоматизации развертывания, масштабирования и поддержки контейнеризированных приложений. Одной из важных задач при эксплуатации кластеров Kubernetes является защита конфиденциальных данных, таких как пароли, токены доступа, криптографические ключи и строки подключения к внешним сервисам. Проблема заключается в выборе стратегии интеграции внешнего хранилища, так как каждый из существующих методов накладывает взаимно исключающие ограничения на безопасность, производительность и эксплуатационную сложность.

Штатный механизм Kubernetes Secret [1] обладает рядом ограничений. Во-первых, данные в объектах Secret хранятся в виде строк, закодированных по схеме Base64. Формат кодирования Base64 не является методом криптографической защиты и не обеспечивает конфиденциальность данных. Во-вторых, объекты Secret сохраняются в распределённом хранилище типа «ключ-значение» etcd, в котором шифрование данных при хранении по умолчанию не настроено.

В работе рассматриваются методы интеграции централизованного хранилища секретов HashiCorp Vault [2] в инфраструктуру Kubernetes посредством драйвера Secrets Store CSI Driver, реализующего интерфейс Container Storage Interface (CSI) [3], и оператора внешних секретов (External Secrets Operator, ESO) [4]. К функциональным возможностям HashiCorp Vault относятся динамическая генерация и ротация секретов, разграничение доступа на основе ролевой модели (Role-Based Access Control, RBAC), шифрование данных и аудит операций доступа.

В архитектуре CSI-драйвера доступ приложений к секретам организован посредством монтирования специальных томов. При запуске пода (минимальной единицы развертывания в Kubernetes, содержащей один или несколько контейнеров) CSI-драйвер запрашивает необходимые секреты из Vault и размещает их в смонтированном томе, доступном контейнерам через файловую систему. Для работы драйвера в кластере требуется развертывание привилегированного DaemonSet — объекта Kubernetes, автоматически создающего экземпляр компонента на каждом узле кластера. Наличие компонента на каждом узле необходимо для выполнения операций монтирования томов непосредственно в среде выполнения подов. Поскольку секреты предоставляются через том, связанный с конкретным подом, конфигурация их использования задаётся отдельно для каждого экземпляра приложения.

В архитектуре ESO синхронизация секретов из Vault в объекты Kubernetes Secret осуществляется автоматически. Для развертывания оператора используется Deployment — объект Kubernetes, предназначенный для управления набором идентичных экземпляров приложения на уровне кластера. В отличие от CSI-драйвера, для функционирования ESO не требуется запуск компонентов на каждом узле кластера. Отсутствие привилегированных компонентов соответствует рекомендациям Kubernetes Pod Security Standards по ограничению использования механизмов повышения привилегий [5]. Объект SecretStore или ClusterSecretStore определяет параметры подключения и аутентификации во внешнем хранилище, а объект ExternalSecret содержит

правила получения данных и формирования объектов Kubernetes Secret. Разделение ответственности между указанными объектами позволяет задавать конфигурацию источников и правила синхронизации независимо друг от друга.

В ходе работы был развернут локальный кластер Minikube. В кластере установлен HashiCorp Vault, настроен механизм аутентификации Kubernetes Auth, а также развернуты CSI-драйвер и оператор внешних секретов. Для ESO был создан SecretStore с аутентификацией через токен Kubernetes ServiceAccount, а ExternalSecret настроен на синхронизацию секрета из пути «secret/db-credentials» с обновления 60 минут. Для проверки получения секрета из Vault использовалось тестовое приложение на языке Python, разработанное с применением фреймворка Flask. В качестве секрета использовалась строка подключения к базе данных, передаваемая приложению через переменную окружения.

Для сравнения методов интеграции HashiCorp Vault с Kubernetes были выбраны критерии, отражающие способ предоставления секретов, особенности развертывания компонентов, механизмы хранения данных в кластере и эксплуатационные характеристики. Результаты сравнения приведены в таблице.

Табл.

Критерий	CSI-драйвер	ESO
Способ предоставления секрета	Том, смонтированный в под	Объект Kubernetes Secret
Развертывание в кластере	DaemonSet на каждом узле	Deployment
Использование привилегированных компонентов	Требуется	Не требуется
Совместное использование секретов	Ограничено моделью томов	Поддерживается
Хранение секретов в etcd	Не осуществляется в etcd (хранятся только в смонтированном томе)	Осуществляется
Обновление секретов	Обновление содержимого смонтированного тома	Синхронизация объекта Kubernetes Secret
Среднее время запуска пода (10 запусков), с	4,2	2,8

Различия между методами заключаются в модели доставки секретов в поды Kubernetes, подходах к обеспечению безопасности обработки и хранения секретов, а также эксплуатационных характеристиках, определяющих границы применимости каждого метода.

В основе различий лежит стратегия предоставления секретов поду-потребителю. При использовании CSI-драйвера применяется файловая модель предоставления секретов с монтированием временного тома. Доступ к секретам осуществляется через файловую систему контейнера, при этом область их доступности ограничена жизненным циклом смонтированного тома. В случае ESO реализуется объектная модель, при

которой данные из внешнего хранилища синхронизируются в объект Kubernetes Secret. ESO позволяет использовать секреты в любых приложениях, потребляющих их как переменные окружения или монтируемые файлы, а также предоставляет возможность централизованно переиспользовать один объект Secret в различных подах и пространствах имён без дублирования данных.

Драйвер CSI не предусматривает хранение секретов в etcd. Это снижает риск их попадания в резервные копии кластера и уменьшает поверхность атаки через API Kubernetes. Однако для работы CSI-драйвера на каждом узле кластера необходимо развёртывание компонентов с расширенными привилегиями. Метод на основе ESO, напротив, работает в рамках стандартной модели Deployment и синхронизирует секреты обратно в объекты Kubernetes Secret, тем самым сохраняя их в хранилище etcd (которое по умолчанию не зашифровано). Для ESO обеспечение конфиденциальности данных при хранении достигается активацией механизма EncryptionConfiguration, строгами политиками RBAC и ведением аудита операций доступа API-сервера Kubernetes.

Обновление секретов в CSI-драйвере выполняется через изменение содержимого смонтированного тома. В ESO используется периодическая однонаправленная синхронизация объектов Kubernetes Secret из внешнего хранилища в кластер, частота которой определяется конфигурацией. При использовании переменных окружения обновление значений может требовать перезапуска пода.

В проведённом эксперименте среднее время запуска подов составило 4,2 с для CSI-драйвера и 2,8 с для ESO. Снижение времени на 33 % при использовании ESO объясняется отсутствием операций инициализации и монтирования томов на критическом пути запуска контейнера. Стандартное отклонение для выборки из десяти последовательных запусков составило $\pm 0,4$ с для CSI и $\pm 0,2$ с для ESO.

Таким образом, выбор метода интеграции определяется эксплуатационными требованиями системы. Оператор внешних секретов отличается меньшими эксплуатационными накладными расходами и допускает повторное использование объектов Kubernetes Secret, тогда как CSI-драйвер применим в сценариях, требующих минимизации хранения секретов в централизованном хранилище etcd и допускающих использование привилегированных компонентов на узлах.

Дальнейшее развитие работы может быть направлено на исследование поведения сравниваемых методов в высоконагруженных средах при динамическом обновлении значительного числа секретов, включая количественную оценку задержки доставки изменений в ESO и влияния CSI-операций на пропускную способность узла.

[1] <https://kubernetes.io/docs/concepts/configuration/secret>

[2] <https://developer.hashicorp.com/vault/docs>

[3] <https://developer.hashicorp.com/vault/docs/platform/k8s/csi>

[4] <https://external-secrets.io/latest/>

[5] <https://kubernetes.io/docs/concepts/security/pod-security-standards/>

ПОДХОД К ОБНАРУЖЕНИЮ ПОВЕДЕНЧЕСКИХ АНОМАЛИЙ ПОЛЬЗОВАТЕЛЕЙ ДЛЯ SIEM-СИСТЕМ НА ОСНОВЕ АНАЛИЗА ВРЕМЕННЫХ РЯДОВ

А.А. Егорова, И.Н. Шабалин

ННГУ им. Н.И. Лобачевского

В современных корпоративных информационных системах значительная доля инцидентов информационной безопасности связана с использованием скомпрометированных учётных данных [1]. Компрометация учётной записи не сопровождается нарушением механизмов авторизации, вследствие чего действия злоумышленника регистрируются как активность легитимного пользователя. В сценариях, когда нарушитель действует в рамках легитимных полномочий, эффективность применения сигнатурных методов обнаружения и правил корреляции событий, используемых в SIEM-системах, снижается.

В SIEM-системах корреляция событий реализуется через детерминированные правила вида:

IF (event_type = 4625 AND count > 5 WITHIN 10 minutes) THEN alert.

В рамках формализации используются фиксированный временной интервал (в примере 10 минут) и фиксированные пороговые значения. Вариативность поведения пользователей и зависимость активности от временного контекста при этом не учитываются. Как следствие, распределение нормальной активности смещается относительно заданных порогов, что ведёт к увеличению числа ложноположительных срабатываний либо к пропуску медленно развивающихся отклонений.

Цель работы заключается в разработке и исследовании подхода к обнаружению поведенческих аномалий пользователей на основе анализа временных рядов и сравнении его с подходом, основанным на фиксированных порогах корреляционных правил SIEM-систем.

Задача обнаружения поведенческих аномалий пользователей состоит в выявлении моментов времени, для которых наблюдаемая активность отклоняется от нормального профиля поведения пользователя. Для решения задачи применяется UEBA-подход (User and Entity Behavior Analytics), в рамках которого нормальное поведение задаётся статистической моделью временных рядов, а аномалии определяются как отклонения от модели. Пусть U – множество пользователей информационной системы, C – множество типов событий. Примерами типов событий являются вход в систему (login), доступ к файлам (file_access), выполнение команд (command_exec). Для каждой пары (u, c) , где $u \in U$, $c \in C$, определяется временной ряд почасовой агрегации – число событий типа c , зарегистрированных пользователем u в дискретном интервале времени t с шагом дискретизации $\Delta T = 1$ час:

$$x_c^{(u)}(t)$$

$$x_t \sim N(\mu_t, \sigma_t^2).$$

Отклонение наблюдения от ожидаемого значения оценивается с помощью Z-оценки:

$$z_t = \frac{x_t - \widehat{\mu}_t}{\widehat{\sigma}_t},$$

где $\widehat{\mu}_t$ и $\widehat{\sigma}_t$ вычисляются по скользящему окну длиной $N = 168$, соответствующему недельному циклу пользовательской активности. Недельное окно выбрано для охвата повторяющихся паттернов поведения в рабочие и выходные дни.

Событие классифицируется как аномальное при выполнении условия $|z_t| > \theta$ при $\theta = 3$. Значение параметра θ соответствует вероятности ложной тревоги около 0.3% для нормального распределения. Однако прямое применение этого правила к почасовой агрегации может давать ложные срабатывания из-за суточной нестационарной структуры активности: днём активность высокая, ночью низкая.

Почасовая агрегация формирует нестационарный временной ряд, в котором распределение значений зависит от времени суток. Учёт суточной активности выполняется стратификацией наблюдений по времени. Для каждого часа формируются подвыборки значений, соответствующих фиксированному часу, и на этих подвыборках оцениваются параметры μ_h и σ_h .

Оценка отклонения для момента времени t вычисляется как:

$$z_t = \frac{x_t - \hat{\mu}_{h(t)}}{\hat{\sigma}_{h(t)}},$$

где $h(t)$ – отображение времени t в номер часа, $h \in \{0, \dots, 23\}$.

В результате стратификации единая модель распределения заменяется семейством локальных моделей. Применение условной модели по времени суток позволяет устранить систематическое смещение, возникающее из-за различной интенсивности активности в дневные и ночные периоды.

Пользователи характеризуются различной вариативностью активности. Эта особенность создаёт неоднородность распределения стандартных отклонений. Для пользователя-администратора σ велико, поэтому для классификации отклонения как аномального требуется большее абсолютное значение выброса. Предлагается индивидуальный порог:

$$\theta_u = \theta + 0.5 \cdot \frac{\text{median}(\sigma_{u,t})}{\text{global_median_std}},$$

где $\theta = 3$ – базовое пороговое значение, $\sigma_{u,t}$ – стандартное отклонение для пользователя u на интервале t , а $\text{median}(\sigma_{u,t})$ – медиана этих значений по всем историческим интервалам для данного пользователя. Параметр `global_median_std` – медиана стандартных отклонений, рассчитанная по совокупности пользователей системы или по всем их действиям в системе. Введенный порог масштабирует чувствительность детектора в зависимости от естественной вариативности поведения. Таким образом, порог повышается для «шумных» пользователей и понижается для пользователей с низкой вариативностью поведения, благодаря чему снижается число ложных срабатываний.

Обработка событий пользовательской активности выполняется в потоковом режиме и включает:

- 1) агрегацию событий по часовым интервалам;

- 2) обновление скользящего окна наблюдений длиной $N = 168$ часов;
- 3) вычисление параметров $\hat{\mu}_t, \hat{\sigma}_t$ либо $\hat{\mu}_{h(t)}, \hat{\sigma}_{h(t)}$ по данным скользящего окна;
- 4) вычисление Z-оценки z_t ;
- 5) сравнение вычисленного значения z_t с индивидуальным порогом θ_u ;
- 6) фиксацию аномалии при выполнении условия $|z_t| > \theta_u$.

Начальный этап работы системы (первые N часов) соответствует заполнению скользящего окна и не используется для принятия решений.

Система реализована в виде набора микросервисов, включающего:

- модуль сбора событий, расширяющий штатный сборщик логов (например, Winlogbeat, Syslog-ng) и передающий события на следующий этап обработки;
- модуль агрегации временных рядов, накапливающий счётчики по пользователям и часам в оперативной памяти;
- модуль детектирования аномалий, реализующий описанные алгоритмы и проверяющий каждое новое значение на аномальность;
- модуль оповещений, генерирующий оповещения и сохраняющий их в список;
- веб-интерфейс, представляющий собой отдельную страницу для просмотра временных рядов и оповещений;
- модуль хранения данных для сохранения истории и оповещений.

Передача данных между компонентами реализована через структурированные сообщения. Интеграция с SIEM-системой осуществляется через вызов REST API модуля UEBA из обработчика логов либо через перенаправление потока событий с помощью Apache Kafka [3]. При предложенной архитектуре модуль может использоваться как внешний аналитический сервис и не требует внесения изменений в ядро SIEM-системы.

Для оценки систем использовались метрики:

- точность (Precision) = $TP / (TP + FP)$;
- полнота (Recall) = $TP / (TP + FN)$;
- F1-мера = $2 \times Precision \times Recall / (Precision + Recall)$,

где TP – число правильно обнаруженных аномалий (True Positive), FP – число ложноположительных срабатываний (False Positive), FN – число пропущенных аномалий (False Negative).

Дополнительно используется метрика оценки числа ложных срабатываний на пользователя в день ($FP/user/day$).

Сравнение проводится между базовым методом фиксированных порогов (baseline) и предложенным методом на основе Z-оценки. Базовые пороги заданы экспертным путём: $y > 5$ для типа событий login и $y > 10$ для типа событий file_access, где y – число событий заданного типа за интервал 10 минут. Результаты представлены в таблице.

Табл.

Метод	Precision	Recall	F1	FP/user/day
Фиксированный порог (baseline)	0.68	0.77	0.72	1.8
На основе Z-оценки (предложенный)	0.93	0.89	0.91	0.7

Анализ проводился по трём типам аномалий:

- тип А (всплеск) – с помощью обоих методов были получены высокие значения F1-меры (0.95), однако для метода на основе Z-оценки было зафиксировано меньшее количество ложноположительных срабатываний.
- тип В (ночная активность) – F1-мера составила 0.92 для предложенного метода и 0.45 у baseline-метода. Применение метода фиксированных порогов не позволяло выявлять аномалии данного типа, поскольку абсолютное число событий в ночные часы (1–2) не достигало порогового значения, равного 5. В предложенном методе наблюдаемое значение сравнивается с нормальным профилем для соответствующего часа (с ожидаемым значением около 0.1), и аномальное отклонение фиксируется вне зависимости от низких абсолютных значений активности.
- тип С (медленный дрейф) – F1-мера 0.85 для предложенного метода, 0.48 для baseline-метода. В baseline-методе оповещение формировалось при превышении порога, а именно на четвёртый день развития аномалии. В предложенном методе отклонение фиксируется раньше за счёт отслеживания изменения скользящего среднего и постепенного роста значения Z-оценки.

Разработан подход к обнаружению поведенческих аномалий пользователей на основе анализа временных рядов. Подход включает использование Z-оценки, стратификацию по времени суток, индивидуальную нормализацию порога и потоковую архитектуру обработки событий. Экспериментально получено снижение числа ложноположительных срабатываний в 2.5 раза при сохранении полноты обнаружения.

Перспективы дальнейших исследований включают интеграцию с Apache Kafka для горизонтального масштабирования системы, применение чувствительных к медленным изменениям методов (CUSUM, байесовский онлайн-детектор), внедрение федеративного обучения для построения поведенческих профилей, а также реализацию автоматических мер активного противодействия (изоляция пользователя, принудительная смена пароля) при высокой Z-оценке.

[1] <https://www.verizon.com/business/resources/reports/dbir/>

[2] <https://www.mandiant.com/m-trends>

[3] Chandola V., Banerjee A., Kumar V. Anomaly detection: A survey // ACM Computing Surveys (CSUR). 2009. Vol. 41, no. 3. P. 1.

РАЗРАБОТКА ИНФОРМАЦИОННОЙ СИСТЕМЫ МОНИТОРИНГА И НЕПРЕРЫВНОЙ ДИАГНОСТИКИ ЭЛЕКТРИЧЕСКОЙ ИЗОЛЯЦИИ СИЛОВЫХ ТРАНСФОРМАТОРОВ НА ОСНОВЕ СИСТЕМНОГО АНАЛИЗА И ОБРАБОТКИ ДАННЫХ

Д.А. Семенов, Д.П. Шимбурев

ГБОУ ВО НГИЭУ Княгининский университет

Современная энергетическая инфраструктура требует повышения надёжности силового оборудования, ключевым элементом которого являются силовые трансформаторы. Согласно отраслевой статистике, более 70 % отказов трансформаторного оборудования связано с деградацией и пробоем электрической изоляции, подверженной термическому, электрическому и химическому старению. Традиционные методы диагностики носят эпизодический характер, требуют отключения оборудования от сети и не обеспечивают оперативного выявления начальных стадий дефектов в режиме реального времени.

Целью работы является разработка информационной системы мониторинга и непрерывной диагностики изоляции силовых трансформаторов, обеспечивающей снижение времени обнаружения критических дефектов за счёт интеграции многопараметрического сбора данных, системного анализа и алгоритмов машинного обучения.

Предлагаемая система реализует трёхуровневую архитектуру (рис.):



Рис.

На первом уровне осуществляется непрерывный сбор данных с датчиков температуры контактных соединений и наиболее нагретой точки (ННТ) изоляции (DS18B20), а также температуры и влажности окружающей среды (DHT22). Второй уровень включает микроконтроллерный блок (Arduino MKR GSM 1400), реализующий передачу данных по протоколу MQTT на сервер. Третий уровень обеспечивает анализ, диагностику состояния с применением моделей машинного обучения, а также визуализацию результатов и формирование рекомендаций по техническому обслуживанию.

Для реализации этих функций используется платформа ioBroker, выступающая в качестве центрального хаба для приема, хранения и агрегации данных. В связке с ней работает визуальная среда программирования Node-RED, которая отвечает за оркестрацию потоков данных: она принимает MQTT-сообщения, выполняет предварительную обработку сигналов, интегрирует алгоритмы машинного обучения для прогнозной аналитики и формирует логику оповещений. Кроме того, Node-RED и ioBroker обеспечивают построение интуитивно понятных дашбордов для визуализации текущего состояния оборудования и истории изменений параметров.

Табл.

Параметр	Традиционные системы	Предложенная система
Режим контроля	Периодический (раз в 1-3 года)	Непрерывный в реальном времени
Время обнаружения дефекта	От нескольких часов до суток	12–45 секунд
Требуемое отключение оборудования	Да	Нет
Прогнозирование остаточного ресурса	Отсутствует или экспертное	На основе машинного обучения и статистического анализа

Экспериментальная апробация системы проведена на лабораторном стенде, имитирующем работу силового трансформатора с моделированием дефектов изоляции различной степени тяжести (начальная, развивающаяся, аварийная стадии). Результаты показали успешное распознавание всех смоделированных дефектов со средним временем реакции 45, 28 и 12 секунд соответственно. Точность прогнозирования развития дефекта за 12-24 часа до достижения критического уровня составила 89 %, а снижение рисков аварийного простоя оценивается в 60-70 %.

Таким образом, разработанная информационная система демонстрирует значимое преимущество перед традиционными методами диагностики: обеспечение непрерывного контроля в реальном времени, снижение количества ложных срабатываний за счёт интеллектуальной фильтрации данных и возможность перехода к предиктивному обслуживанию. Перспективы развития системы включают расширение поддержки протоколов промышленной автоматизации и интеграцию с системами диспетчерского управления и сбора данных (SCADA) энергосистем.

- [1] Хренников А. Ю., Вахнина В. В., Кувшинов А. А., Александров Н. М. Силовые трансформаторы в электрических сетях: испытания, диагностика, дефекты, повреждаемость, мониторинг. – М.: Библиотечка электротехника, 2022, № 4–5. С. 1.
- [2] Шимбуев Д. П., Семенов Д. А. Система мониторинга трансформатора на базе GSM // Вестник связи. 2023. № 7. С. 21.
- [3] Мельникова О. С., Леонтьев И. В. Интеллектуальные методы определения электрической прочности масляной изоляции для диагностики силовых трансформаторов // Электроэнергия. Передача и распределение. 2024, № 2(83). С. 88.

ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОБНАРУЖЕНИЯ АТАК И ИХ ЛОКАЛИЗАЦИЯ С ИСПОЛЬЗОВАНИЕМ ПАКЕТА WIRESHARK

М.Н. Втюрина, А.А. Рябов

ННГУ им. Н.И. Лобачевского

Существует множество методов классификации атак. Для этого используют номера портов, глубокий анализ пакетов, стохастический анализ и применяют машинное обучение [1, 2].

Обычно в наборах данных большое количество признаков (40-80).

Для классификации атак в основном используют такие признаки, как:

- Статистика длин пакетов: минимальная, максимальная, средняя длина пакета в потоке, стандартное отклонение.
- Статистика интервалов времени (IAT): время между приходом пакетов.
- Продолжительность потока: сколько секунд длилось соединение.
- Направленность: соотношение байт, отправленных клиентом и сервером.
- TCP-флаги: например, количество пакетов с флагом PUSH или RST.

Так как признаков много, применяют методы сокращения их числа. Наиболее распространённые – Correlation-based Feature Selection (CFS), Consistency-based Feature Selection (CON) и InfoGain[3-5].

В данной работе реализован подход ML-классификация потоков для обнаружения атаки. Для экспертного контроля использовалась Port/Protocol-фильтрация с применением инструмента Wireshark. Этот подход был выбран в силу его устойчивости к шифрованию трафика и возможности работы с потоковыми признаками, в отличие от метода глубокого анализа пакетов (DPI), который теряет эффективность при анализе зашифрованных соединений.

Исследования проводились на сгенерированном наборе CIC-IDS2017 [6]. Набор состоит из 5 файлов, в каждом из которых содержится информация о трафике за один день недели с понедельника по пятницу. Набор является несбалансированным в соотношении 80/20.

После предварительной обработки набора из 78 признаков оставлено 20, так как [1, 7] подтверждают, что применение InfoGain к набору CIC-IDS2017 позволяет сократить количество признаков до 20 при сохранении точности классификации 99,86% (по метрике f1-score).

В исследовании использовалась модель машинного обучения Gradient Boosting с тройной кросс-валидацией, learning rate равным 0,1 и 50 деревьями решений. Такой выбор позволяет достигнуть компромисса между скоростью и точностью.

Из общего количества событий набора порядка 19% идентифицировано как атаки. Модель выделила 11 различных классов, в том числе DoS атаки.

Для дополнительного контроля результатов работы алгоритма из набора, состоящего из аномальных потоков, использовалась программа wireshark. Одной из ключевых проблем современной криминалистики является невозможность восстановления единой временной линии событий из набора [8]. В данной работе эта проблема решается путём создания искусственных временных меток с привязкой к относительному времени начала захвата и последующей фильтрацией в Wireshark с использованием параметра `frame.time_relative`.

В ходе исследования проверена возможность нахождения атаки в рсар файлах по сохраненным параметрам: `frame_relative_time`, протоколу и портам. Также проведён тест на сбалансированном наборе данных, представляющим из себя один из дней набора. Экспертный анализ подтверждает наличие аномалии, схожей с DoS-атакой и PortScan.

Таким образом, модель машинного обучения Gradient boosting протестирована на сбалансированном и несбалансированном наборе данных для обнаружения и классификации атак в сети и сохранении данных о них с целью поиска атак в сетевом трафике с использованием программы Wireshark. Предложенный метод позволяет помочь в работе экспертам по информационной безопасности.

- [1] Getman A.I., Ikonnikova M.K. A survey of network traffic classification methods using machine learning. Trudy ISP RAN / Proc. ISP RAS. 2020. Vol. 32, issue 6, pp. 137-154.
- [2] Szabó G., Orincsay D., Malomsoky S., Szabó I. (2008) On the Validation of Traffic Classification Algorithms. In: Claypool M., Uhlig S. (eds) Passive and Active Network Measurement. PAM 2008. Lecture Notes in Computer Science. – Springer, Berlin, Heidelberg. 2008. Vol. 4979. pp. 72-81.
- [3] A Survey of Machine Learning Based Packet Classification, Yu Liu, The Institute for Computing, Information and Cognitive Systems (ICICS), University of British Columbia – Vancouver, BC V6T 1Z4, Canada.
- [4] Выбор признакового пространства для классификации IP-трафика сети методами машинного обучения, Улугбек Рахимжон угли Охундадаев, Национальный университет Узбекистана, Academic Research in Educational Sciences VOLUME 2 | ISSUE 12 | 2021 Scientific Journal Impact Factor (SJIF) 2021.
- [5] Berkhin P. “A survey of clustering data mining techniques”. In: Grouping Multidimensional Data. – Springer, 2006, pp. 25-71.
- [6] Iman Sharafaldin, Arash Habibi Lashkari, and Ali A. Ghorbani, “Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization”, 4th International Conference on Information Systems Security and Privacy (ICISSP), – Portugal, January 2018.
- [7] Kurniabudi, Stiawan D., Darmawijoyo et al. Feature Selection for Imbalanced Class Data Using Information Gain and C4.5 Algorithm on CIC-IDS2017 Dataset // IEEE Access. 2020.
- [8] Simson L. Garfinkel “Digital forensics research: The next 10 years”, digital investigation 7 (2010).

ВНЕДРЕНИЕ АВТОКОДИРОВЩИКОВ В ПРОЦЕСС ОБРАБОТКИ СОБЫТИЙ SIEM СИСТЕМЫ

Р.А. Васильев

ННГУ им. Н.И. Лобачевского

Современные информационные инфраструктуры генерируют колоссальный поток событий безопасности, поступающих от межсетевых экранов, систем обнаружения вторжений, операционных систем, служб аутентификации и прикладных сервисов. Системы класса SIEM (Security Information and Event Management) выступают центральным узлом сбора, нормализации и корреляции этих событий, обеспечивая операторам центров управления безопасностью (SOC) единую точку наблюдения за состоянием защищаемой инфраструктуры [1]. Согласно актуальным оценкам, зрелый SOC способен обрабатывать от миллионов до десятков миллиардов событий в сутки, и без средств автоматизации аналитического процесса справиться с таким объёмом данных не представляется возможным [2].

Традиционный подход к выявлению угроз в SIEM основывается на заранее сформированных правилах корреляции, которые описывают известные паттерны атак и нарушений политик безопасности. Подобный подход обеспечивает высокую воспроизводимость и интерпретируемость результатов, однако принципиально ограничен в выявлении угроз, для которых соответствующие правила ещё не созданы [3]. Сложные многоступенчатые атаки, атаки «нулевого дня», медленные кампании типа Advanced Persistent Threat (APT), а также аномалии в поведении легитимных пользователей зачастую остаются за пределами обнаружения на основе статических правил [4]. Именно это противоречие определяет необходимость интеграции в конвейер обработки событий SIEM методов интеллектуального анализа данных, способных выявлять отклонения от нормального поведения без априорного знания об их характере.

Среди подходов к решению данной задачи особую привлекательность представляют автокодировщики — класс нейронных сетей, обучаемых по принципу реконструкции входного вектора через сжатое латентное представление. Ключевое преимущество автокодировщика в контексте информационной безопасности состоит в том, что для его обучения достаточно исключительно нормальных событий — аномалии проявляются как события с высокой ошибкой реконструкции, поскольку модель не обучалась воспроизводить их структуру [5]. Тем самым снимается критическое ограничение методов обучения с учителем — потребность в размеченных примерах каждого класса угроз, получение которых в реальных производственных системах крайне затруднено и ресурсоёмко [6].

Настоящая работа посвящена исследованию вопросов практической интеграции автокодировщика в конвейер обработки событий SIEM-системы. Рассматриваются архитектурные решения на каждом этапе конвейера — от сбора и нормализации сырых событий до формирования обогащённых записей с признаком аномальности, которые могут использоваться для построения новых корреляционных правил и уведомления аналитиков. Цель работы — обосновать применимость автокодировщика как компонента обогащения событий в SIEM и описать методику его встраивания в поток обработки в режиме реального времени.

Для достижения поставленной цели в работе решаются следующие задачи: анализ существующих подходов к выявлению аномалий в потоках событий безопасности; описание архитектуры конвейера обработки событий и места автокодировщика в нём; разработка и описание архитектуры модели SIEMAutoencoder; экспериментальная оценка работы модели на открытом наборе данных UNSW-NB15; анализ результатов и определение области практического применения подхода.

Системы класса SIEM выполняют функции агрегации и структурирования журналов событий из гетерогенных источников, нормализации форматов, корреляции событий по заданным правилам и оповещения об инцидентах [1, 7]. В типовой архитектуре SIEM выделяются следующие функциональные уровни: уровень сбора и транспортировки событий (агенты, syslog, API-коннекторы); уровень нормализации и обогащения; уровень хранения и индексирования; уровень корреляции и детектирования; уровень реагирования и визуализации [8]. Каждый из этих уровней формирует отдельный этап конвейера, и именно на этапах нормализации и корреляции наиболее органично располагается модуль интеллектуального анализа на базе нейронных сетей.

Задача выявления аномалий в событиях безопасности имеет ряд принципиальных особенностей, отличающих её от классических задач машинного обучения. Во-первых, классы событий существенно несбалансированы: аномальные события составляют ничтожную долю от общего потока, что делает неприменимыми большинство стандартных метрик точности [9]. Во-вторых, ландшафт угроз постоянно изменяется: атакующие адаптируют тактики, создаются новые вредоносные инструменты, появляются ранее неизвестные уязвимости. В таких условиях модель, обученная на историческом наборе размеченных атак, быстро теряет актуальность [10]. В-третьих, разметка аномалий в производственных логах SIEM требует значительных усилий аналитиков и в большинстве организаций не производится систематически, что ограничивает применимость методов обучения с учителем.

В настоящее время для выявления аномалий в потоках событий безопасности исследователями применяются следующие подходы. Статистические методы — такие как анализ z-отклонений, скользящего среднего и модели сезонности трафика — просты в реализации, однако чувствительны к настройке порогов и не способны улавливать многомерные зависимости между признаками [11]. Методы кластеризации (k-средних, DBSCAN) позволяют выделять группы нетипичного поведения, но плохо масштабируются на высокоразмерные векторы событий SIEM [12]. Методы обнаружения выбросов на основе изолирующего леса (Isolation Forest) или метода опорных векторов (One-Class SVM) демонстрируют приемлемое качество на табличных данных, однако уступают нейросетевым подходам при работе с событиями, имеющими сложную корреляционную структуру [13].

Автокодировщики занимают особое место среди методов обучения без учителя, применяемых для выявления аномалий. В работе [14] показано, что автокодировщик, обученный исключительно на нормальных событиях хоста под управлением Windows, успешно обнаруживает аномальное использование системных утилит и нетипичные последовательности событий, недоступные для сигнатурного детектора. Авторы исследования [15] демонстрируют, что гибридный подход, сочетающий SIEM-правила с модулем инференса автокодировщика, обеспечивает как быстрое обнаружение известных угроз, так и адаптивное детектирование новых аномалий.

Для экспериментальной оценки предложенного подхода использован открытый набор данных UNSW-NB15, созданный в Университете Нового Южного Уэльса (Австралия). Набор содержит 257 673 записи сетевых соединений с 44 признаками и 49 признаками в расширенной версии, сформированными инструментами Argus и Bro-IDS [18]. В наборе представлены девять типов атак: Fuzzers, Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, Shellcode и Worms, а также нормальный сетевой трафик (93 000 записей). Выбор данного набора обусловлен его широким применением в актуальных исследованиях по выявлению аномалий в событиях безопасности, что позволяет проводить сравнение результатов [18].

Несмотря на то, что в представленном исследовании используется набор данных о сетевом трафике, концепция и архитектура конвейера полностью воспроизводимы применительно к событиям SIEM: как события сетевого трафика, так и журналы операционных систем или средств защиты после нормализации представляются в виде числовых векторов одинаковой структуры. В контексте реальной SIEM-системы конфиденциальность производственных данных не позволяет напрямую использовать рабочие логи в открытых исследованиях, что является общепринятой практикой для публикаций в данной области.

Результаты оценки модели на полном тестовом наборе UNSW-NB15 представлены в таблице. Для сравнения приведены показатели базовых методов выявления аномалий без учителя — Isolation Forest и One-Class SVM, — применённых к тем же признакам с аналогичной предобработкой. Значения метрик Precision, Recall и F1-score рассчитаны относительно класса «аномалия» (label=1). AUC-ROC отражает способность модели ранжировать аномальные события выше нормальных по непрерывному баллу отклонения.

Табл.

Метод	Accuracy, %	Precision	Recall	F1-Score	AUC-ROC
Isolation Forest	82,4	0,71	0,74	0,72	0,83
One-Class SVM	79,1	0,65	0,81	0,72	0,80
SIEMAutoencoder	91,3	0,87	0,89	0,88	0,93

Из приведённых данных следует, что SIEMAutoencoder превосходит оба базовых метода по всем ключевым метрикам. Наиболее значительным является прирост AUC-ROC — с 0,83 у Isolation Forest до 0,93, что свидетельствует о существенно лучшей ранжирующей способности модели. Значение F1-Score 0,88 указывает на сбалансированное соотношение точности и полноты, принципиально важное в условиях дисбаланса классов, присущего событиям безопасности. Достигнутые результаты согласуются с результатами, опубликованными в смежных работах [15, 16, 18], где аналогичные архитектуры автокодировщиков на сопоставимых датасетах демонстрируют AUC-ROC в диапазоне 0,90–0,96.

В ходе проведённого исследования рассмотрена задача интеграции автокодировщика на основе нейронных сетей в конвейер обработки событий систем класса SIEM. Предложена архитектура конвейера, в которой автокодировщик выполняет функцию компонента обогащения событий: его вывод в виде флага аномальности, числового

балла отклонения и перечня «подозрительных» полей присоединяется к нормализованному событию перед передачей на уровень корреляции. Данный подход позволяет расширить аналитические возможности SIEM без замены существующего корреляционного движка.

Разработана и описана модель SIEMAutoencoder — симметричный полносвязный автокодировщик с батч-нормализацией, функцией активации LeakyReLU, механизмом ранней остановки и режимом стримингового обучения. Реализована методика калибровки порога аномальности на основе персентильного анализа ошибок реконструкции, а также механизм дообучения на новых нормальных событиях для управления ложноположительными срабатываниями.

Экспериментальная оценка на наборе данных UNSW-NB15 показала, что SIEMAutoencoder достигает значения AUC-ROC 0,93 и F1-Score 0,88, превосходя базовые методы выявления аномалий без учителя — Isolation Forest и One-Class SVM — по всем ключевым метрикам. Принципиальным преимуществом подхода является возможность обучения исключительно на нормальных событиях без потребности в размеченных примерах аномалий, что обуславливает его практическую применимость в реальных производственных SIEM-системах.

- [1] <https://falcongaze.com/ru/pressroom/publications/zashhita-informacii/siem-sistemy/>
- [2] <https://cisoclub.ru/integracija-siem-s-mashinnym-obucheniem-jevoljucija-sistem-monitoringa-i-analiza-ugroz/>
- [3] <https://serverflow.ru/blog/stati/tekhnologiya-siem-polnyy-obzor/>
- [4] Лавров Б.О., Иванов М.А. Интеллектуальное обнаружение аномалий в контейнеризованных приложениях: методы, архитектура и инструменты // Безопасность информационных технологий. 2025. Т. 32, № 4. С. 149.
- [5] Ijarsct. Role of Autoencoders in Unsupervised Network Anomaly Detection // International Journal of Advanced Research in Science, Communication and Technology. 2026. Vol. 6. Iss. 1. P. 608.
- [6] Benabderrahmane S. Ranking-enhanced anomaly detection using Active Learning-assisted Attention Adversarial Dual AutoEncoder // Scientific Reports. 2025. Vol. 15. P. 41602.
- [7] <https://kvan.tech/blog/chto-takoe-siem/>
- [8] https://rt-solar.ru/products/solar_siem/blog/6147
- [9] Гайфулина Д.А., Котенко И.В. Анализ моделей глубокого обучения для задач обнаружения сетевых аномалий // Информационно-управляющие системы. — 2021. № 1 (110). С. 28.
- [10] Ijerd. Real-Time Anomaly Detection in Streaming Sensor Data Using LSTM Autoencoders // International Journal of Engineering Research. 2025. Vol. 2. Iss. 10. P. 1.
- [11] Орехов А.В., Орехов А.А. Автоматическое обнаружение аномалий сетевого трафика при DDoS-атаках // Вестник Санкт-Петербургского университета. Прикладная математика. 2023. Т. 19, № 2. С. 251.
- [12] Браницкий А.А., Котенко И.В. Анализ и классификация методов обнаружения сетевых атак // Труды СПИИРАН. 2016. № 2 (45). С. 207.
- [13] Чيو К., Фримэн Д. Машинное обучение и безопасность. — М.: ДМК Пресс, 2020, 368 с.

- [14] Гурина А.Л. Обнаружение аномальных событий на хосте с использованием авто-кодировщиков // International Journal of Open Information Technologies. 2020. Т. 8, № 3. С. 35.
- [15] Bojovic P.D., Jovic J., Pavlovic M. A SIEM-Based Framework for Multi-Layer Data Collection and Anomaly Detection // arXiv.org. 2025.
- [16] Real-time Anomaly Detection Using Convolutional Autoencoder // arXiv.org. 2024. arXiv:2404.04311.
- [17] https://1275.ru/articles/obogaschenie-dannyh-klyuchevoy-ingredient-uspeha-siem_8736
- [18] <https://research.unsw.edu.au/projects/unsw-nb15-dataset>

Секция «Информационные системы.
Средства, технологии, безопасность»

Заседание секции проводилось 26 мая 2026 г.
Председатель – Л.Ю. Ротков, секретарь – А.А. Рябов.
Нижегородский государственный университет им. Н.И. Лобачевского.