ISSN 2225-5435

Вестник УрФО.

БЕЗОПАСНОСТЬ В ИНФОРМАЦИОННОЙ СФЕРЕ

№ 3(57) / 2025

Подписной индекс 73852 в каталоге «Почта России»

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

Свидетельство ПИ № ФС77-65765 от 20.05.2016

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

16+

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

Journal of the Ural Federal District.

ISSN 2225-5435

INFORMATION SECURITY

№ 3(57) / 2025

Subscription index 73852 in the «Russian Post» catalog

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

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

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

16+

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

РАДИОТЕХНИКА, В ТОМ ЧИСЛЕ СИСТЕМЫ И УСТРОЙСТВА ТЕЛЕВИДЕНИЯ

СИСТЕМНЫЙ АНАЛИЗ, УПРАВЛЕНИЕ И ОБРАБОТКА ИНФОРМАЦИИ

МЕТОДЫ И СИСТЕМЫ ЗАЩИТЫ ИНФОРМАЦИИ, ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ

IN THIS ISSUE

RADIO ENGINEERING, INCLUDING TELEVISION SYSTEMS AND DEVICES

SYSTEM ANALYSIS, MANAGEMENT AND INFORMATION PROCESSING

METHODS AND SYSTEMS OF INFORMATION PROTECTION, INFORMATION SECURITY

УДК 621.396.9 Вестник УрФО № 3(57) / 2025, с. 5–16

Бердюгин-Малиновский Д. А., Шабунин С. Н. 10.14529/secur250301

ВИДЫ НАВИГАЦИОННЫХ ИМИТИРУЮЩИХ ПОМЕХ И МЕТОДЫ БОРЬБЫ С НИМИ

Проведен анализ влияния преднамеренных помех на сигналы глобальных навигационных спутниковых систем (ГНСС), принимаемых навигационной аппаратурой потребителя (НАП). Кратко рассмотрено воздействие маскирующих помех и способы повышения помехоустойчивости НАП. Дается описание типов имитирующих помех и приводится расширенное описание технологий их постановки. Выполнена классификация спуфинг-атак. Рассмотрены методы обнаружения имитирующих помех на основе временного и пространственного анализа принимаемого сигнала. Описаны методы противодействия имитирующим помехам и выполнено сравнение их эффективности в борьбе с спуфинг-атаками.

Ключевые слова: *глобальная навигационная спутниковая система, маскирующие помехи, имитирующие помехи, спуфинг.*

Berdiugin-Malinovskiy D. A., Shabunin S. N.

TYPES OF NAVIGATION SIMULATION INTERFERENCE AND METHODS OF DEALING WITH THEM

The analysis of the impact of intentional interference on the signals from global navigation satellite systems as received by consumer navigation equipment is conducted. The effects of masking interference and methods to enhance noise immunity are briefly discussed. A description of the types of simulated interference is provided, along with an extended overview of the technologies used to create them. Classification of spoofing attacks has also been performed. Techniques for detecting simulated interference based on the temporal and spatial characteristics of the received signal are explored. Methods to counter imitating interference are outlined, and their efficacy in combating spoofing assaults is compared.

Keywords: *Global Navigation Satellite Systems, jamming, deception jamming, spoofing.*

Введение

Свое развитие ГНСС начали с Глобальной системы позиционирования (GPS), созданной в США, и Глобальной навигационной спутниковой системы (ГЛОНАСС), разработанной в Советском Союзе. Современная ГНСС так же включает в себя такие системы, как Galileo – Европейский союз, BeiDou – Китайская народная республика, а также региональные навигационные спутниковые системы: индийская региональная навигационная спутниковая система (IRNSS) и квазизенитная спутниковая система (QZSS) в Японии [1]. НАП ГНСС широко востребована по всему миру, предоставляя точные данные о местоположении, скорости и времени.

Услуги предоставления ГНСС открытых высокоточных синхронизированных данных о времени нашли применение и стали незаменимы в отраслях, не связанных с навигацией, таких как энергетика, банковское дело, коммуникации. Это делает актуальным анализ уязвимостей ГНСС, связанных с помехоустойчивостью из-за крайне низкого уровня сигнала, получаемого со спутников, расположенных на высоте приблизительно в 20 000 километров, открытой структуры общедоступных гражданских навигационных сигналов (НС), используемых в навигации, которые легко подделать ложными с помощью технологии, известной как спуфинг (spoofing) [2].

При изучении уязвимости ГНСС, кроме возможных естественных помех, исследователи выделяют два основных типа преднамеренных помех НАП: маскирующие помехи (jamming) [3] и имитирующие помехи (спуфинг) [4]. Отношение мощности НС на входе приемного устройства к уровню шума обычно не превышает 30 дБ в полосе пропускания приемника. Так для ГНСС ГЛОНАСС в диапазоне L1 уровень сигнала на входе приемника может составлять – 161 дБВт, а для ГНСС GPS – 166 дБВт [1]. Для внесения помех и потери работоспособности НАП GPS достаточно передатчика с мощностью 0,25 Вт на удалении 4 км [5]. Постановщики помех (ПП) используют передатчики с широкополосным гауссовским шумом [6] или другие формы сигналов [7] на частотах работы ГНСС для снижения отношения сигнал/шум на входе приемника или полного подавления работы НАП.

Для повышения помехоустойчивости от маскирующих помех используются методы пространственной режекции сигналов помехи. Адаптивное формирование диаграммы направленности с помощью комбинации нескольких антенн [8] или антенной решеткой (АР) осуществляется за счет подстройки фаз сигналов [9], получаемых с антенных элементов, для достижения максимального отношения сигнал / (помеха + шум) на входе приемника НС [10]. Применение цифровых антенных решеток позволяет подавить уровень множественных помех на 90–100 дБ [5]. Таким образом, методы борьбы с разными видами маскирующих помех НС ГНСС достаточно глубоко исследованы [11] и далее не рассматриваются. Основной целью работы является раскрытие и описание наиболее значимой проблемы – имитирующих помех НС.

Имитирующие помехи (спуфинг) НС ГНСС

Спуфинг НС ГНСС классифицируется как преднамеренная передача ложных НС с задачей дезинформировать НАП с целью получения ею неверных данных о положении, скорости и времени [12]. В отличие от помехи глушения цель атаки тайно заставить приемник НС отслеживать ложные сигналы ГНСС для предоставления неверного навигационного решения и получения контроля над ней. Воздействие имитирующей помехи на НАП намного опаснее маскирующей [13].

Термин «спуфинг» (имитирующая помеха) был заимствован из области информационной безопасности [12]. В области информационной безопасности целью атаки спуфинга является получение разведданных с помощью человека или программы путем маскировки под видом другого человека или программы с помощью фальсификации данных [14]. Имитирующие помехи НС ГНСС имеют сходство с информационными спуфинг-помехами, однако воздействие помехи на НАП не является главной целью. ПП рассчитывает получить полный контроль над объектом с помощью атаки. Цель ошибочно думает, что это настоящий сигнал и использует его. Чтобы добиться своего результата ПП используют различные стратегии и технологии.

Технологии имитирующих помех

При приеме навигационных сигналов со значительного удаления, возникают искажения, связанные с прохождением атмосферы, влияющие на прием [15]. Принимаемый НАП уровень НС настолько низкий, что уязвим для любого вида преднамеренных и непреднамеренных радиочастотных помех [16]. Чтобы на-

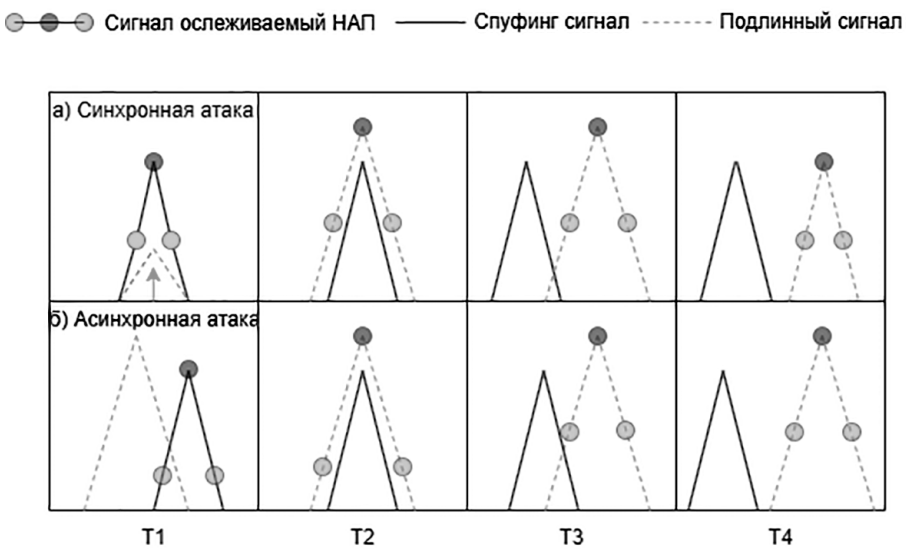

Рис. 1. Синхронная атака спуфинга (а), асинхронная атака (б) с этапами:
Т1 – Выравнивание фазы спуфинг-сигнала с реальным сигналом,
Т2 – Захват контроля над НАП с повышением мощности, Т3 – Изменение параметров сигнала (фазы),
Т4 – Контроль ПП над НАП

чать атаку, ПП сначала передает синхронизированные по фазе ложные НС, но более низкой мощности по сравнению с подлинными. На следующем этапе мощность спуфинг-сигналов увеличивается до уровня, незначительно превышающего подлинные [17]. С увеличением мощности НАП начинает отслеживать ложный сигнал вместо подлинного, получая таким образом контроль над НАП. ПП медленно изменяет параметры сигнала уводя их от исходных. Как только ложные сигналы перемещают приемник на 2 микросекунды во времени или на 600 метров в сторону от подлинного навигационного решения, можно констатировать, что НАП подчинена ПП [18]. Описанная технология постановки имитирующих помех называется синхронной атакой (или мягкий захват), в противном случае это будет асинхронная атака [19] (или грубый захват) как показано на рис. 1.

Для синхронной атаки требуется точная синхронизация помеховых сигналов по фазе, доплеровскому сдвигу, положению и времени. Сигналы должны содержать соответствующие данные альманаха и эфемерид. Для обеспечения точности синхронизации постановщику помех необходим приемник ГНСС.

Несинхронная атака является одной из самых простых. Для нее не требуется синхронизация. Для успешной атаки уровень сигнала помехи должен превышать подлинный сигнал не менее, чем на 30 дБ. Во время атаки происходит срыв слежения за НС, после чего НАП переходит в режим поиска и захватывает более мощный спуфинг-сигнал. В случае, когда НАП находится в режиме поиска НС, для проведения атаки помеховому сигналу достаточно небольшой разницы в мощности [20].

Постановка помехи на основе генерации НС

Такие сигналы генерируются самим оборудованием ПП для передачи на НАП ложных данных [21]. Сгенерированные сигналы имеют хорошую гибкость в плане изменения навигационных параметров о местоположении и времени. В основном такие атаки выполняются после предварительного глушения сигналов ГНСС, чтобы заставить НАП захватить ложный сигнал, превышающий мощность подлинных сигналов.

Постановка помехи на основе ретрансляции НС

Источник помех на основе ретрансляции НС состоит из ретранслятора и приемника НС. Перед ретрансляцией сигналы синхронизируются с подлинными, после этого генерируются спуфинг-сигналы, усиленные по мощности относительно подлинных сигналов и с небольшой задержкой во времени, направленные на приемную антенну цели. Атаки на небольшом расстоянии от целевой НАП наиболее синхронизированы с подлинными НС [22].

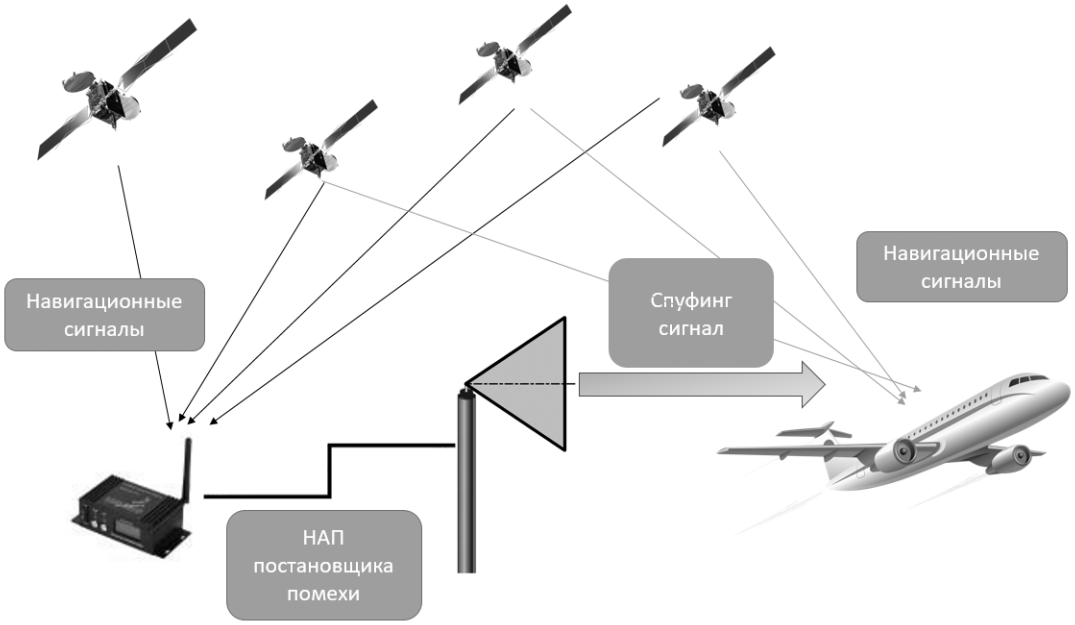

Рис. 2. Схема синхронной спуфинг-атаки на основе ретрансляции НС

На рис. 2 показана базовая схема синхронной атаки на основе ретрансляции НС.

Постановка помехи с автосинхронизацией

Как наиболее продвинутая исследователями рассматривается технология скрытной постановки имитирующей помехи с постепенной синхронизацией к подлинным НС, подстраиваемой под контур слежения НАП [23]. После обработки подлинных НС, ПП осуществляет задержку по дальности и доплеровскому сдвигу в соответствии с динамикой объекта атаки, с целью скрытного контроля задержки приема НС [24]. Особенность технологии автосинхронизации помехи заключается в эффективной скрытной подмене подлинного НС. Для общедоступных гражданских и военных НС технология постановки помех отличается [25].

Постановка помехи с использованием нескольких передатчиков

Атака с несколькими передатчиками считается наиболее сложной помехой с технологической точки зрения. На рис. 3 изображена схема атаки на основе ретрансляции НС с несколькими источниками помех. В этой технологии используется несколько передатчиков, что значительно усложняет пространственную оценку прихода сигналов для подавления таких помех [26]. Как при генерации, так и при ретрансляции НС, работа источников помех должна быть синхронизирована, что делает её сложно реализуемой [27].

Выбор технологии постановки имитирующих помех зависит от целей ПП [28], их можно разделить на несколько категорий:

- подмена информации в определении местоположения цели [29];
- подмена информации текущего времени цели [30];
- получение контроля над НАП ГНСС [31].

По отличительным признакам описанных методов и технологий, имитирующие помехи можно классифицировать в виде отдельных групп (рис. 4).

Методы детектирования имитирующих помех

Большая работа по детектированию имитирующих помех была проведена еще в 1995 году [32]. Были проанализированы технологии постановки помех, а также сформулированы методы их обнаружения:

- по амплитуде сигнала;
- по углу прихода сигнала;
- по времени прибытия сигнала;
- проверка согласованности с другим навигационным оборудованием;
- аутентификация шифрования сигнала;

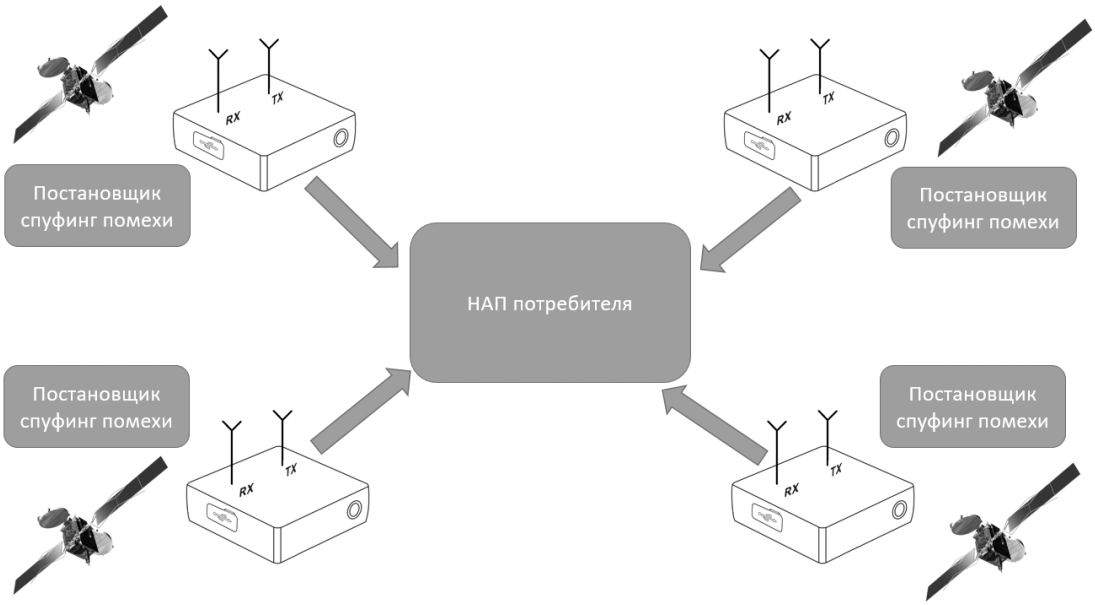

Рис. 3. Схема синхронизированной спуфинг-атаки на основе ретрансляции с несколькими источниками помех

- определение направления поляризации сигнала;
- обнаружение векторных петель слежения.

Анализ амплитуды принимаемого НС

Отношение мощности полезного сигнала к мощности шума является одной из характеристик навигационного сигнала. В нормальных условиях работы НАП это отношение находится в определенных пределах и не подвержено значительным изменениям [33]. Когда с помощью атаки ПП пытается воздействовать на НАП, это отношение может резко измениться. Контролируя данный параметр НС, можно определить воздействие спуфинга. Аналогичным способом можно оценивать уровень мощности самого принимаемого НС. Еще одним методом, основанном на анализе уровня принимаемых НС, служит оценка разности мощности принимаемых сигналов разных диапазонов частот с одного спутника. Изменение разности или потеря одного из сигналов, может свидетельствовать об атаке [34].

Пространственная оценка НС

Подлинные сигналы передаются с разных спутников и направлений. Если спуфинг-атака совершается с одного передатчика, то такие сигналы приходят с одной антенны. Тем

Тип помехового сигнала		
сгенерированный сигнал	ретранслированный сигнал с задержками	автосинхронизированный помеховый сигнал
Стратегия атаки		
несинхронная атака	синхронная атака	
Цель атаки		
ложная информация о местоположении цели	ложная информация о времени	контроль над целью
Структура атаки		
одиночный постановщик помех	несколько помеховых передатчиков	

Рис. 4. Классификация спуфинг-атак

самым, с помощью пространственной обработки сигналов можно выявлять направление источника помехи для управления ДН антенных систем НАП.

Способы пространственной обработки могут быть разными. За счет оценки разности фаз принимаемых сигналов [35], оценки изменения угла пересечения НС между двумя направлениями прибытия на одной антенне [36], разности фаз от двух разнесенных антенн [37], использования многоэлементной антенной решетки (АР), сравнения фазовых измерений коррелятора НАП от разных спутников, определения угла прибытия сигналов и исключения тех, которые были получены из одной пространственной области [38].

Методы, основанные на оценке параметров времени задержки прибытия НС

Для случаев, когда имитирующие помехи используют в своей основе навигационный приемник, необходимо время для получения НС, его обработки и генерации на его основе помехи. Методы по оценке времени задержки могут отличаться. К примеру, алгоритм обнаружения имитирующих помех, основанный на анализе разницы псевдодальности и данных эфемерид [39], метод оценки разницы во времени прибытия (TDOA), измеряемой как дифференциальное отношение псевдодальности к несущей частоте между сигналами имитирующих помех и подлинными НС с помощью корреляции данных с нескольких навигационных приемников [40].

Распознавание спуфинга с использованием инерциальных навигационных систем (ИНС)

Корреляция данных НАП и датчиков ИНС позволяет в различной степени и на разных этапах настраивать и применять методы обнаружения имитирующих помех, даже для медленных атак спуфинга, когда ПП пытается подстраиваться под подлинные НС [41].

Мониторинг качества сигнала

Метод основан на мониторинге качества пика корреляции в условиях многолучевости. НАП на выходе коррелятора подвергается аналогичному воздействию от атак спуфинга, поэтому данный метод может применяться для детектирования имитирующих помех [42]. Эффективность метода снижается в условиях многолучевости.

Методы противодействия имитирующим помехам

Еще в 1994 году проведена оценка использования инерциальной навигационной системы на основе высокоточного кольцевого лазерного гироскопа и акселерометра, позволяющей корректировать курс при отказе или подмене данных со спутников [43]. Интеграция навигационных данных, инерциальной системы и одометрии на основе разработанного алгоритма фильтра Калмана позволяет отфильтровать резкие изменения местоположения и осуществлять кратковременную навигацию без НС ГНСС под воздействием на НАП имитирующих помех. Здесь дрейф инерциальной системы, накапливающийся из-за ошибок в измерениях, ограничивается калибровкой при запуске, и непрерывной оценкой погрешности изменяющихся данных с помощью фильтра Калмана [44]. Использование услуги высокоточного позиционирования для ГНСС GPS с использованием широкополосного сигнала, модулированного дальномерным P(Y) кодом, делает его значительно менее уязвимым к имитирующим помехам по сравнению с открытым кодом C/A (Coarse/Acquisition – грубый прием), но его применение ограничено санкционированными пользователями [45]. Тем не менее, использование даже полубескодовых приемников (semi-codeless receiver), использующих косвенные методы обработки НС с P(Y) в сочетании с ИНС, снижает влияние имитирующих помех по сравнению с C/A [46]. Применение метода двумерного поиска по задержке кода НС и доплеровского сдвига с использованием асимметричного быстрого преобразования Фурье позволяет быстрее определять сдвиг и задержку в условиях слабого сигнала и помех и с меньшими вычислительными затратами, тем самым снижая уязвимость к примитивным имитирующим помехам, направленным на срыв синхронизации [47]. В борьбе с несложными некогерентными спуфинг-атаками при подмене НС хорошо зарекомендовали себя методы использования референсных станций ДНС, вычисляющие поправки и передающие их другим НАП по каналам связи, защищенным от спуфинг-сигналов [48]. Более высокоточной, чем дифференциальная, является технология RTK (Real Time Kinematic). Метод использует гибридную систему, состоящую из НС улучшенной целостности с технологией RTK. Данные устройств ИНС и относитель-

ной навигации на основе ультразвуковых и оптических датчиков позволяют проводить перекрестную проверку навигационных данных для выявления и возможной фильтрации спуфинг-сигналов [49]. Для гражданской НАП, не имеющей доступа к широкополосным защищённым военным сигналам, надежными методами борьбы с имитирующими помехами являются методы пространственной обработки сигнала, использующие пространственную фильтрацию для адаптивного формирования нулей диаграммы направленности АР в направлении источников помех [50]. Уровень помехоустойчивости при применении многодиапазонных гибридных антенн сопоставим с применением более громоздких многоэлементных АР [33]. Применение АР эффективно при обработке сложной навигационной помеховой обстановки.

В контексте обзора методов противодействия имитирующим помехам нужно рассмотреть перспективные системы навигации для воздушных судов по физическим полям: магнитного и гравитационного поля Земли [51]. Существуют методы использования оптической навигации, позволяющие корректировать навигационное решение или осуществлять независимую навигацию по видео изображению [52]. Для навигации морских судов предлагается метод корреляции измерений глубины рельефа бортовыми эхолотами с цифровыми батиметрическими картами в сочетании с ИНС [53].

Инженерный университет в городе Харбин в 2022 году провел большое исследование на тему обзора технологий имитирующих помех и методов борьбы с ними [4]. В работе были объединены труды 121 научного исследования и на их основе сформированы оценочные выводы об эффективности рассмотренных методов. Наивысшие оценки получили следующие методы:

- методы оценки мощности и векторных контуров слежения принимаемых НС;
- оценка отклонений от нормы параметров в принимаемых НС в сочетании с использованием инструментов криптозащиты и шифрования;
- анализ дрейфа параметров НС, где наивысшую оценку получили методы, сочетающие в себе корреляцию оцениваемых параметров с инерциальными датчиками ориентации или оптическими датчиками;
- определение местоположения источников излучения принимаемых сигналов;
- использование нейронных сетей и других методов машинного обучения, позволяющих комбинировать различные методы защиты в зависимости от стратегии атаки.

Заключение

Несмотря на развитие технологий и методов борьбы с имитирующими помехами, технологии спуфинга НС ГНСС на основе генерации НС или использующие подлинные НС все меньше отличаются от реальных НС. Исходя из этого, приоритеты развития методов обнаружения и борьбы с имитирующими помехами будут смещаться к комбинации интегрированных НАП, ИНС и методов машинного обучения в различных сочетаниях.

Для гражданской НАП эффективным и недорогим методом борьбы с помехами является применение адаптивных антенных систем или АР. При разработке адаптивных навигационных антенных систем для борьбы с помехами, в дополнение к методам пространственной оценки источников сигналов, повысить эффективность поможет оценка дрейфа параметров НС в сочетании с инерциальными датчиками или полноценной ИНС.

Работа выполнена при финансовой поддержке Министерства науки и высшего образования РФ (тема № FEUZ-2023-0015)

Литература

1. Карутин С.Н., Власов И.Б., Дворкин В.В. Дифференциальная коррекция и мониторинг глобальных навигационных спутниковых систем. 2014. 463 с.

2. Warner J.S., Johnston R.G., and Alamos C.L. A Simple Demonstration that the Global Positioning System (GPS) is Vulnerable to Spoofing. The Journal of Security Administration, 2002, vol. 25, no. 10, pp. 19–28.

3. Перов А.И. ГЛОНАСС. Модернизация и перспективы развития. Монография. М: Радиотехника, 2020. 1072 с.

4. Meng L. et al. A Survey of GNSS Spoofing and Anti-Spoofing Technology. Remote Sensing, 2022, vol. 14, no. 19, p. 4826.

5. Слюсар В. Цифровые антенные решетки – будущее радиолокации. Электроника: Наука, технология, бизнес, 2001, № 3(33), с. 42–47.

6. Lubbers B. GNSS Accuracy Under White Gaussian Noise Jamming. Engineering Proceedings, 2025, vol. 88, no. 1, p. 26.

7. Elghamrawy H. et al. Experimental Evaluation of the Impact of Different Types of Jamming Signals on Commercial GNSS Receivers. Applied Sciences, 2020, vol. 10, no. 12, p. 4240.

8. Ruegamer A. et al. Jamming and Spoofing of GNSS Signals – An Underestimated Risk?! Proc. Wisdom Ages Challenges Modern World, 2015, vol. 3, pp. 17–21.

9. Кашин В.А. Методы фазового синтеза антенных решеток. Зарубежная радиоэлектроника. Успехи современной радиоэлектроники, 1997, № 1, с. 47–60.

10. Монзинго Р.А., Миллер Т.У. Адаптивные антенные решетки: введение в теорию; пер. с англ. В.Г. Челпанова, В.А. Лексаченко; под общ. ред. В.А. Лексаченко. М.: Радио и связь, 1986, 448 с.

11. Sharifi-Tehrani O., Ghasemi M.H. A Review on GNSS-Threat Detection and Mitigation Techniques. Cloud Computing and Data Science, 2023, pp. 161–185.

12. Broumandan A., Jafarnia-Jahromi A., Lachapelle G. Spoofing Detection, Classification and Cancelation (SDCC) Receiver Architecture for a Moving GNSS Receiver. GPS Solutions, 2015, vol. 19, pp. 475–487.

13. Tippenhauer N. O. et al. On the Requirements for Successful GPS Spoofing Attacks. Proceedings of the 18th ACM Conference on Computer and Communications Security, 2011, pp. 75–86.

14. Psiaki M. L. et al. GNSS lies, GNSS truth: Spoofing Detection with Two-Antenna Differential Carrier Phase. GPS World, 2014, vol. 25, no. 11, pp. 36–44.

15. Glomsvoll O. Jamming of GPS & GLONASS signals. Department of Civil Engineering, Nottingham Geospatial Institute, 2014, pp. 25–32.

16. Borio D. et al. Impact and Detection of GNSS Jammers on Consumer Grade Satellite Navigation Receivers. Proceedings of the IEEE, 2016, vol. 104, no. 6, pp. 1233–1245.

17. Jafarnia-Jahromi A. et al. Detection and Mitigation of Spoofing Attacks on a Vector Based Tracking GPS Receiver. Proceedings of the 2012 International Technical Meeting of the Institute of Navigation, 2012, pp. 790–800.

18. Kyle W., Daniel S., Todd H. Straight Talk on Anti-Spoofing Securing the Future of PNT. GPS World, 2012, vol. 23, pp. 32–43.

19. Ouyang X. et al. Analysis and Evaluation of Spoofing Effect on GNSS Receiver. 2015 IEEE 12th Intl Conf. on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf. on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf. on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom). IEEE, 2015, pp. 1388–1392.

20. Aissou G. et al. Instance-Based Supervised Machine Learning Models for Detecting GPS Spoofing Attacks on UAS. 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC). IEEE, 2022, pp. 0208–0214.

21. Caparra G., Wullems C., Ioannides R.T. An Autonomous Gnss Anti-Spoofing Technique. 2016 8th ESA Workshop on Satellite Navigation Technologies and European Workshop on GNSS Signals and Signal Processing (NAVITEC). IEEE, 2016, pp. 1–8.

22. Ledvina B.M. et al. An In-Line Anti-Spoofing Device for Legacy Civil GPS Receivers. Proceedings of the 2010 International Technical Meeting of the Institute of Navigation, 2010, pp. 698–712.

23. Wesson K., Shepard D., Humphreys T. Straight Talk on Anti-Spoofing. GPS World, 2012, vol. 23, no. 1, pp. 32–39.

24. Humphreys T.E. Detection Strategy for Cryptographic GNSS Anti-Spoofing. IEEE Transactions on Aerospace and Electronic Systems, 2013, vol. 49, no. 2, pp. 1073–1090.

25. Peng C. et al. Research of Intermediate Spoofing without Precise Target Information. China Satellite Navigation Conference (CSNC) 2019 Proceedings: Volume II. Springer Singapore, 2019. pp. 615–624.

26. Manfredini E.G. Signal Processing Techniques for GNSS Anti-Spoofing Algorithms. Polytechnic University of Turin, Italy, 2017.

27. Psiaki M.L., Humphreys T.E. GNSS Spoofing and Detection. Proceedings of the IEEE, 2016, vol. 104, no. 6, pp. 1258–1270.

28. Неровный В.В. и др. Характеристики уязвимости аппаратуры потребителей глобальных навигационных спутниковых систем к спуфинг-атакам. Труды учебных заведений связи, 2023, т. 9, № 6, с. 95–100.

29. Zen K. C. et al. A Practical GPS Location Spoofing Attack in Road Navigation Scenario. Proceedings of the 18th International Workshop on Mobile Computing Systems and Applications, 2017, pp. 85–90.

30. Khan I., Centeno V. Undetectable GPS-Spoofing Attack on Time Series Phasor Measurement Unit Data, arXiv preprint arXiv:2206.12440, 2022.

31. Zaragoza S., Zumalt E. Spoofing a Superyacht at Sea. Cockrell School of Engineering. Univ. of Texas/ Austin, 2013, available at http://www. utexas. edu/know/2013/07/30/spoofing-asuperyacht-at-sea.

32. Key E.L. Techniques to Counter GPS Spoofing. Internal Memorandum. MITRE Corporation, Bedord, 1995.

33. Криницкий Г.В. Повышение помехоустойчивости навигационной аппаратуры потребителей с использованием различных конфигураций антенных систем. Радиотехника, 2022, т. 86, № 9, с. 31–39.

34. Jafarnia-Jahromi A. et al. GPS Vulnerability to Spoofing Threats and a Review of Antispoofing Techniques. International Journal of Navigation and Observation, 2012, vol. 2012, no. 1, pp. 127072.

35. Montgomery P.Y., Humphreys T.E., Ledvina B.M. Receiver-Autonomous Spoofing Detection: Experimental Results of a Multi-Antenna Receiver Defense against a Portable Civil GPS Spoofer. Proceedings of the 2009 International Technical Meeting of the Institute of Navigation, 2009, pp. 124–130.

36. Chen S. et al. GNSS Spoofing Detection via the Intersection Angle between Two Directions of Arrival in a Single Rotating Antenna. Sensors, 2024, vol. 24, no. 4, p. 1116.

37. Dinh T.N. et al. A Method to Build Feature Descriptor for GNSS Spoofing Detection by Carrier Phase Double Difference Measurement. Measurement Science and Technology, 2024, vol. 35, no. 6, p. 066307.

38. McDowell C. E. GPS Spoofer and Repeater Mitigation System Using Digital Spatial Nulling: pat. № 7250903, USA, 2007.

39. Liu K. et al. Spoofing Detection Algorithm Based on Pseudorange Differences. Sensors, 2018, vol. 18, no. 10, p. 3197.

40. Zhang Z., Zhan X. GNSS Spoofing Network Monitoring Based on Differential Pseudorange. Sensors, 2016, vol. 16, no. 10, p. 1771.

41. Gao Y., Li G. A Slowly Varying Spoofing Algorithm on Loosely Coupled GNSS/IMU Avoiding Multiple Anti-Spoofing Techniques. Sensors, 2022, vol. 22, no. 12, pp. 4503.

42. Pini M. et al. Signal Quality Monitoring Applied to Spoofing Detection. Proceedings of the 24th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS 2011), 2011, pp. 1888–1896.

43. Gilster G. High Accuracy Performance Capabilities of the Military Standard Ring Laser Gyro Inertial Navigation Unit. Proceedings of 1994 IEEE Position, Location and Navigation Symposium-PLANS'94. IEEE, 1994, pp. 464–473.

44. Beckmann H., Niedermeier H., Eissfeller B. Improving GNSS Road Navigation Integrity Using MEMS INS and Odometry. 2012 6th ESA Workshop on Satellite Navigation Technologies (Navitec 2012) & European Workshop on GNSS Signals and Signal Processing. IEEE, 2012, pp. 1–7.

45. Wolfert R. et al. Direct P (Y)-Code Acquisition Under a Jamming Environment. IEEE 1998 Position Location and Navigation Symposium (Cat. No. 98CH36153). IEEE, 1996, pp. 228–235.

46. Tang X., Yuan M., Wang F. Analysis of GPS P(Y) Signal Power Enhancement Based on the Observations with a Semi-Codeless Receiver. Satellite Navigation, 2022, vol. 3, no. 1, p. 26.

47. Lu H. et al. Two Dimension Direct GPS Signal Acquisition Based on Unsymmetrical FFT. 2009 4th IEEE Conference on Industrial Electronics and Applications. IEEE, 2009, pp. 1769–1772.

48. Jin M. H. et al. GPS Spoofing Signal Detection and Compensation Method in DGPS Reference Station. 2011 11th International Conference on Control, Automation and Systems. IEEE, 2011, pp. 1616–1619.

49. Brewer E. et al. A Low SWaP Implementation of High Integrity Relative Navigation for Small UAS. 2014 IEEE/ION Position, Location and Navigation Symposium-PLANS 2014. IEEE, 2014, pp. 1183–1187.

50. Magiera J., Katulski R. Applicability of Null-Steering for Spoofing Mitigation in Civilian GPS. 2014 IEEE 79th Vehicular Technology Conference (VTC Spring). IEEE, 2014, pp. 1–5.

51. Каршаков Е.В. и др. Перспективные системы навигации летательных аппаратов с использованием измерений потенциальных физических полей. Гироскопия и навигация, 2021, Т. 29, №. 1, с. 112.

52. Веселов Г.Е. и др. Разработка и испытания навигационной системы на основе оптического потока для беспилотных транспортных средств. 2024 IEEE 3-я Международная конференция по проблемам информатики, электроники и радиотехники (PIERE). IEEE, 2024, с. 90–95.

53. Hagen O.K., Ånonsen K.B., Skaugen A. Robust Surface Vessel Navigation Using Terrain Navigation. 2013 MTS/IEEE OCEANS-Bergen. IEEE, 2013, pp. 1–7.

References

1. Karutin S.N., Vlasov I.B., Dvorkin V.V. Differential Correction and Monitoring of Global Navigation Satellite Systems [Differencial'naja korrekcija i monitoring global'nyh navigacionnyh sputnikovyh system], 2014, 463 p.

2. Warner J.S., Johnston R.G., and Alamos C.L. A Simple Demonstration that the Global Positioning System (GPS) is Vulnerable to Spoofing. The Journal of Security Administration, 2002, vol. 25, no. 10, pp. 19–28.

3. Perov A.I. GLONASS. Modernization and Development Prospects Monograph. [Modernizacija i perspektivy razvitija. Monografija]. Moscow, Radio Engineering, 2020, 1072 p.

4. Meng L. et al. A Survey of GNSS Spoofing and Anti-Spoofing Technology. Remote Sensing, 2022, vol. 14, no. 19, p. 4826.

5. Slyusar V. Digital Antenna Arrays – the Future of Radar [Cifrovye antennye reshetki – budushhee radiolokacii]. Electronics: Science, Technology, Business, 2001, no. 3, pp. 42–47.

6. Lubbers B. GNSS Accuracy Under White Gaussian Noise Jamming. Engineering Proceedings, 2025, vol. 88, no. 1, p. 26.

7. Elghamrawy H. et al. Experimental Evaluation of the Impact of Different Types of Jamming Signals on Commercial GNSS Receivers. Applied Sciences, 2020, vol. 10, no. 12, p. 4240.

8. Ruegamer A. et al. Jamming and Spoofing of GNSS Signals – An Underestimated Risk?! Proc. Wisdom Ages Challenges Modern World. 2015, vol. 3, pp. 17-21.

9. Kashin V.A. Methods of phase synthesis of antenna arrays [Metody fazovogo sinteza antennyh reshetok]. Foreign radioelectronics. Successes Modern Radioelectronics [Zarubezhnaja radiojelektronika. Uspehi sovremennoj radiojelektroniki], 1997, vol. 1, pp. 47–60.

10. Monzingo R.A., Miller T.W. Adaptive Antenna Arrays: Introduction to the Theory [Adaptivnye antennye reshetki: vvedenie v teoriju]; translated from English by V. G. Chelpanova, V. A. Leksachenko; edited by V. A. Leksachenko. Moscow: Radio i svyaz, 1986, 448 p.

11. Sharifi-Tehrani O., Ghasemi M.H. A Review on GNSS-Threat Detection and Mitigation Techniques. Cloud Computing and Data Science, 2023, pp. 161–185.

12. Broumandan A., Jafarnia-Jahromi A., Lachapelle G. Spoofing Detection, Classification and Cancelation (SDCC) Receiver Architecture for a Moving GNSS Receiver. GPS Solutions, 2015, vol. 19, pp. 475–487.

13. Tippenhauer N. O. et al. On the Requirements for Successful GPS Spoofing Attacks. Proceedings of the 18th ACM Conference on Computer and Communications Security, 2011, pp. 75–86.

14. Psiaki M. L. et al. GNSS Lies, GNSS Truth: Spoofing Detection with Two-Antenna Differential Carrier Phase. GPS World, 2014, vol. 25, no. 11, pp. 36–44.

15. Glomsvoll O. Jamming of GPS & GLONASS signals. Department of Civil Engineering, Nottingham Geospatial Institute, 2014, pp. 25–32.

16. Borio D. et al. Impact and Detection of GNSS Jammers on Consumer Grade Satellite Navigation Receivers. Proceedings of the IEEE, 2016, vol. 104, no. 6, pp. 1233–1245.

17. Jafarnia-Jahromi A. et al. Detection and Mitigation of Spoofing Attacks on a Vector Based Tracking GPS Receiver. Proceedings of the 2012 International Technical Meeting of the Institute of Navigation, 2012, pp. 790–800.

18. Kyle W., Daniel S., Todd H. Straight Talk on Anti-Spoofing Securing the Future of PNT. GPS World, 2012, vol. 23, pp. 32–43.

19. Ouyang X. et al. Analysis and Evaluation of Spoofing Effect on GNSS Receiver. 2015 IEEE 12th Intl Conf. on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf. on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf. on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom). IEEE, 2015, pp. 1388–1392.

20. Aissou G. et al. Instance-Based Supervised Machine Learning Models for Detecting GPS Spoofing Attacks on UAS. 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC). IEEE, 2022, pp. 0208–0214.

21. Caparra G., Wullems C., Ioannides R.T. An Autonomous Gnss Anti-Spoofing Technique, 2016 8th ESA Workshop on Satellite Navigation Technologies and European Workshop on GNSS Signals and Signal Processing (NAVITEC). IEEE, 2016, pp. 1–8.

22. Ledvina B.M. et al. An In-Line Anti-Spoofing Device for Legacy Civil GPS Receivers. Proceedings of the 2010 International Technical Meeting of the Institute of Navigation, 2010, pp. 698–712.

23. Wesson K., Shepard D., Humphreys T. Straight Talk on Anti-Spoofing. GPS World, 2012, vol. 23, no. 1, pp. 32–39.

24. Humphreys T.E. Detection Strategy for Cryptographic GNSS Anti-Spoofing. IEEE Transactions on Aerospace and Electronic Systems, 2013, vol. 49, no. 2, pp. 1073–1090.

25. Peng C. et al. Research of Intermediate Spoofing without Precise Target Information. China Satellite Navigation Conference (CSNC) 2019 Proceedings: Volume II. Springer Singapore, 2019. pp. 615–624.

26. Manfredini E.G. Signal Processing Techniques for GNSS Anti-Spoofing Algorithms. Polytechnic University of Turin, Italy, 2017.

27. Psiaki M.L., Humphreys T.E. GNSS Spoofing and Detection. Proceedings of the IEEE, 2016, vol. 104, no. 6, pp. 1258–1270.

28. Nerovny V. V. et al. Characteristics of vulnerability of consumer equipment of global navigation satellite systems to spoofing attacks [Harakteristiki ujazvimosti apparatury potrebitelej global'nyh navigacionnyh sputnikovyh sistem k spufing-atakam]. Proceedings of educational institutions of communication [Trudy uchebnyh zavedenij svjazi], 2023, vol. 9, no. 6. pp. 95–100.

29. Zen K.C. et al. A Practical GPS Location Spoofing Attack in Road Navigation Scenario. Proceedings of the 18th International Workshop on Mobile Computing Systems and Applications, 2017, pp. 85–90.

30. Khan I., Centeno V. Undetectable GPS-Spoofing Attack on Time Series Phasor Measurement Unit Data, arXiv preprint arXiv:2206.12440, 2022.

31. Zaragoza S., Zumalt E. Spoofing a Superyacht at Sea. Cockrell School of Engineering. Univ. of Texas/ Austin, 2013, available at http://www. utexas. edu/know/2013/07/30/spoofing-asuperyacht-at-sea.

32. Key E.L. Techniques to Counter GPS Spoofing. Internal Memorandum. MITRE Corporation, Bedord, 1995.

33. Krinitsky, G.V. Increasing the Noise Immunity of Consumer Navigation Equipment Using Various Configurations of Antenna Systems [Povyshenie pomehoustojchivosti navigacionnoj apparatury potrebitelej s ispol'zovaniem razlichnyh konfiguracij antennyh system]. Radio Engineering, 2022, vol. 86, no. 9, pp.31–39.

34. Jafarnia-Jahromi A. et al. GPS Vulnerability to Spoofing Threats and a Review of Antispoofing Techniques. International Journal of Navigation and Observation, 2012, vol. 2012, no. 1, pp. 127072.

35. Montgomery P.Y., Humphreys T.E., Ledvina B.M. Receiver-Autonomous Spoofing Detection: Experimental Results of a Multi-Antenna Receiver Defense against a Portable Civil GPS Spoofer. Proceedings of the 2009 International Technical Meeting of the Institute of Navigation, 2009, pp. 124–130.

36. Chen S. et al. GNSS Spoofing Detection via the Intersection Angle between Two Directions of Arrival in a Single Rotating Antenna. Sensors, 2024, vol. 24, no. 4, p. 1116.

37. Dinh T.N. et al. A Method to Build Feature Descriptor for GNSS Spoofing Detection by Carrier Phase Double Difference Measurement. Measurement Science and Technology, 2024, vol. 35, no. 6, p. 066307.

38. McDowell C. E. GPS Spoofer and Repeater Mitigation System Using Digital Spatial Nulling: pat. № 7250903, USA, 2007.

39. Liu K. et al. Spoofing Detection Algorithm Based on Pseudorange Differences. Sensors, 2018, vol. 18, no. 10, p. 3197.

40. Zhang Z., Zhan X. GNSS Spoofing Network Monitoring Based on Differential Pseudorange. Sensors, 2016, vol. 16, no. 10, p. 1771.

41. Gao Y., Li G. A Slowly Varying Spoofing Algorithm on Loosely Coupled GNSS/IMU Avoiding Multiple Anti-Spoofing Techniques. Sensors, 2022, vol. 22, no. 12, pp. 4503.

42. Pini M. et al. Signal Quality Monitoring Applied to Spoofing Detection. Proceedings of the 24th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS 2011), 2011, pp. 1888–1896.

43. Gilster G. High Accuracy Performance Capabilities of the Military Standard Ring Laser Gyro Inertial Navigation Unit. Proceedings of 1994 IEEE Position, Location and Navigation Symposium-PLANS'94. IEEE, 1994, pp. 464–473.

44. Beckmann H., Niedermeier H., Eissfeller B. Improving GNSS Road Navigation Integrity Using MEMS INS and Odometry. 2012 6th ESA Workshop on Satellite Navigation Technologies (Navitec 2012) & European Workshop on GNSS Signals and Signal Processing. IEEE, 2012, pp. 1–7.

45. Wolfert R. et al. Direct P (Y)-Code Acquisition Under a Jamming Environment. IEEE 1998 Position Location and Navigation Symposium (Cat. No. 98CH36153). IEEE, 1996, pp. 228–235.

46. Tang X., Yuan M., Wang F. Analysis of GPS P(Y) Signal Power Enhancement Based on the Observations with a Semi-Codeless Receiver. Satellite Navigation, 2022, vol. 3, no. 1, p. 26.

47. Lu H. et al. Two Dimension Direct GPS Signal Acquisition Based on Unsymmetrical FFT. 2009 4th IEEE Conference on Industrial Electronics and Applications. IEEE, 2009, pp. 1769–1772.

48. Jin M. H. et al. GPS Spoofing Signal Detection and Compensation Method in DGPS Reference Station. 2011 11th International Conference on Control, Automation and Systems. IEEE, 2011, pp. 1616–1619.

49. Brewer E. et al. A Low SWaP Implementation of High Integrity Relative Navigation for Small UAS. 2014 IEEE/ION Position, Location and Navigation Symposium-PLANS 2014. IEEE, 2014, pp. 1183–1187.

50. Magiera J., Katulski R. Applicability of Null-Steering for Spoofing Mitigation in Civilian GPS. 2014 IEEE 79th Vehicular Technology Conference (VTC Spring). IEEE, 2014, pp. 1–5.

51. Karshakov E.V. et al. Promising aircraft navigation systems using measurements of potential physical fields [Perspektivnye sistemy navigacii letatel'nyh apparatov s ispol'zovaniem izmerenij potencial'nyh fizicheskih polej]. Gyroscopy and navigation, 2021, vol. 29, no. 1, pp. 112.

52. Veselov G.E. et al. Development and testing of an optical flow-based navigation system for unmanned vehicles [Razrabotka i ispytanija navigacionnoj sistemy na osnove opticheskogo potoka dlja bespilotnyh transportnyh sredstv]. 2024 IEEE 3rd International Conference on Problems of Informatics, Electronics, and Radio Engineering (PIERE) [3-ja Mezhdunarodnaja konferencija po problemam informatiki, jelektroniki i radiotehniki]. IEEE, 2024, pp. 90–95.

53. Hagen O.K., Ånonsen K.B., Skaugen A. Robust Surface Vessel Navigation Using Terrain Navigation. 2013 MTS/IEEE OCEANS-Bergen. IEEE, 2013, pp. 1–7.

БЕРДЮГИН-МАЛИНОВСКИЙ Денис Александрович, студент магистратуры ФГАОУ ВО «Уральский федеральный университет им. первого Президента России Б.Н. Ельцина». 620002, г. Екатеринбург, ул. Мира, 32. E-mail: Berdiugin-Malinovs@urfu.me

ШАБУНИН Сергей Николаевич, доктор технических наук, профессор, заведующий кафедрой радиоэлектроники и телекоммуникаций ФГАОУ ВО «Уральский федеральный университет им. первого Президента России Б.Н. Ельцина». 620002, г. Екатеринбург, ул. Мира, 32. E-mail: s.n.shabunin@urfu.ru

BERDYUGIN-MALINOVSKIY Denis Alexandrovich, master's student of the Federal State Autonomous Educational Institution of Higher Education «Ural Federal University named after the first President of Russia B.N. Yeltsin». 620002, Ekaterinburg, Mira st., 32. E-mail: Berdiugin-Malinovs@urfu.me

SHABUNIN Sergey Nikolaevich, Doctor of Technical Sciences, Professor, Head of the Department of Radioelectronics and Telecommunications, Ural Federal University named after the first President of Russia B.N. Yeltsin. 620002, Ekaterinburg, Mira st., 32. E-mail: s.n.shabunin@urfu.ru

УДК 004.891.2 Вестник УрФО № 3(57) / 2025, с. 17–23

Кушнир Н. В., Власенко А. В., Колесников Д. С., Яцкевич Е. С.

10.14529/secur250302

ОБНАРУЖЕНИЕ АНОМАЛИЙ В ФИНАНСОВЫХ ТРАНЗАКЦИЯХ С ПОМОЩЬЮ АВТОЭНКОДЕРОВ

В данной работе рассматривается задача выявления подозрительных финансовых транзакций с использованием нейросетевых автоэнкодеров. Подобные транзакции зачастую могут быть индикатором мошеннической активности, что делает проблему их своевременного обнаружения критически важной для систем финансовой безопасности.

Вначале проведён обзор существующих подходов к детектированию аномалий, включая классические статистические методы и современные алгоритмы машинного обучения. Отмечается, что большинство традиционных моделей сталкиваются с трудностями при работе с данными, в которых наблюдается существенный дисбаланс классов, характерный для финансовых операций (мошеннические транзакции составляют крайне малую долю от общего объёма).

На этом фоне использование автоэнкодеров представляется целесообразным решением. Данный тип нейронных сетей позволяет строить компактное внутреннее представление нормальных данных, а затем выявлять аномалии на основе величины ошибки восстановления. В исследовании применён полносвязный симметричный автоэнкодер, обученный на открытом наборе данных транзакций.

Результаты экспериментов показывают, что предложенная модель способна эффективно выделять отклонения от типичного поведения. Чтобы подтвердить качество разработанного метода, был проведён сравнительный анализ с признанными подходами, такими как метод изолирующего леса и логистической регрессии. Анализ показал, что использование нейронных сетей позволяет значительно повысить точность выявления подозрительных транзакций, что свидетельствует о перспективности внедрения такого подхода на практике. Исследование подчёркивает важность интеграции моделей автоэнкодинга и глубокого обучения для эффективного решения проблем финансового мониторинга. Практическое применение полученных результатов открывает возможности для совершенствования существующих механизмов предотвращения мошенничества, наделяя их способностью оперативно реагировать на появление новых форм противоправных действий.

Ключевые слова: *обнаружение аномалий, автоэнкодер, нейросети, финансовые транзакции, выявление мошенничества, ошибка восстановления, несбалансированные данные, глубокое обучение.*

DETECTING ANOMALIES IN FINANCIAL TRANSACTIONS USING AUTOENCODERS

This paper considers the problem of detecting suspicious financial transactions using neural network autoencoders. Such transactions can often be an indicator of fraudulent activity, which makes the problem of their timely detection critically important for financial security systems.

First, an overview of existing approaches to anomaly detection is provided, including classical statistical methods and modern machine learning algorithms. It is noted that most traditional models face difficulties when working with data in which there is a significant class imbalance characteristic of financial transactions (fraudulent transactions account for an extremely small proportion of the total volume).

Against this background, the use of autoencoders seems to be an appropriate solution. This type of neural network allows you to build a compact internal representation of normal data, and then identify anomalies based on the magnitude of the recovery error. The study uses a fully connected symmetric autoencoder trained on an open set of transaction data.

The experimental results show that the proposed model is able to effectively identify deviations from typical behavior. To assess its quality, a comparison was made with alternative methods, including Isolation Forest and logistic regression. The analysis showed that the neural network architecture demonstrates a higher sensitivity in detecting fraudulent transactions, which confirms its prospects for practical application.

Thus, the conducted research highlights the importance of using autoencoders and neural network technologies in general when solving financial analytics problems. The results obtained can serve as a basis for the development of more reliable protection systems capable of adapting to new types of fraudulent schemes.

Keywords: *anomaly detection, autoencoder, neural networks, financial transactions, fraud detection, reconstruction error, imbalanced data, deep learning.*

Введение

Современная финансовая индустрия находится в условиях интенсивной цифровизации. Это явление значительно упрощает проведение транзакций и расширяет возможности пользователей, однако одновременно повышает уязвимость системы перед различными угрозами безопасности. Одной из наиболее острых проблем является рост числа мошеннических операций, связанных с использованием украденных персональных данных и созданием поддельных аккаунтов. Несмотря на то что подавляющее большинство транзакций является легитимным, даже относительно небольшая доля мошеннических действий способна привести к серьёзным финансовым потерям и подорвать доверие клиентов к системе.

В этой связи задача своевременного и надёжного обнаружения аномальных транзакций приобретает особую значимость. Традиционные методы, основанные на фиксированных правилах или простых статистических моделях, показывают ограниченную эффективность в условиях постоянно меняющихся мошеннических схем. Злоумышленники непрерывно адаптируют свои стратегии, что снижает результативность жёстко детерминированных алгоритмов и требует внедрения более гибких и интеллектуальных подходов [1].

Одним из перспективных направлений в этой области является использование методов машинного обучения, а именно нейросетевых архитектур, способных выявлять скры-

тые зависимости и обучаться на больших массивах данных. Среди таких моделей особое внимание привлекают автоэнкодеры. Их ключевая особенность заключается в способности восстанавливать входные данные после их сжатия, формируя компактное внутреннее представление. Аномальные транзакции, как правило, плохо соответствуют выученным закономерностям и, следовательно, воспроизводятся сетью с высокой ошибкой. Этот эффект позволяет использовать автоэнкодеры в качестве эффективного инструмента для обнаружения отклонений.

Применение автоэнкодеров имеет ряд преимуществ перед традиционными методами. Во-первых, их работа не требует большого количества размеченных данных, что особенно важно при выраженном дисбалансе классов: количество мошеннических операций значительно уступает числу нормальных транзакций. Во-вторых, архитектура модели позволяет классифицировать операции на нормальные и подозрительные, опираясь исключительно на величину ошибки реконструкции. Таким образом, подход не только обеспечивает адаптивность к новым угрозам, но и повышает надёжность функционирования финансовых систем.

Целью данной работы является исследование эффективности применения автоэнкодеров в задаче выявления аномальных транзакций, а также анализ направлений возможного совершенствования данной модели.

Автоэнкодер представляет собой разновидность искусственной нейронной сети, предназначенную для построения сжатого представления исходных данных при сохранении ключевых признаков и их последующего восстановления. Обучение, как правило, проводится в безучительном режиме, когда целевая переменная совпадает со входным вектором. Архитектура автоэнкодера включает три основных блока: энкодер, выполняющий преобразование входных данных в скрытое представление; бутылочное горлышко (или латентное пространство), содержащее упрощённое описание исходных данных; и декодер, восстанавливающий входную информацию на основе скрытого представления. Архитектура автоэнкодера приведена на рисунке 1. Такая структура обеспечивает возможность выявления несоответствий и делает автоэнкодеры ценным инструментом анализа данных в условиях высокой изменчивости угроз.

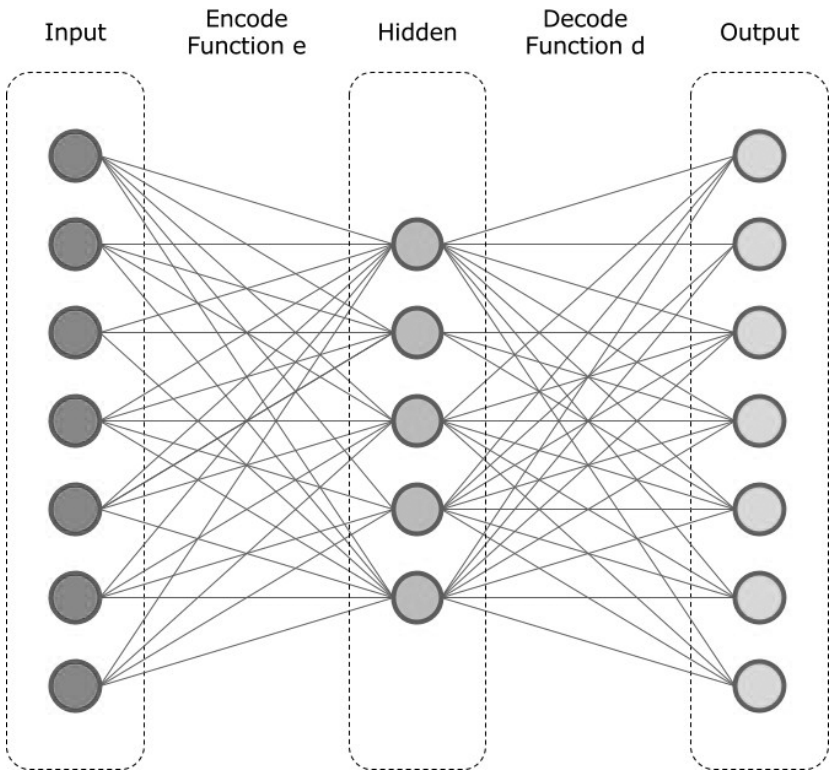

Рис.1. Схема архитектуры энкодера

Обнаружение аномалий в финансовых транзакциях с помощью автоэнкодеров

В задачах обнаружения аномалий автоэнкодеры обучаются преимущественно на примерах, относящихся к «нормальному» поведению системы. При этом транзакции, не соответствующие общей закономерности, воспроизводятся сетью с заметными искажениями. Это приводит к увеличению ошибки восстановления, которая и используется как основной критерий для классификации операции как аномальной. Таким образом, величина ошибки играет ключевую роль: при превышении заранее установленного порогового значения транзакция рассматривается как подозрительная. Следует отметить, что подбор оптимального порога является эмпирическим процессом и во многом определяет баланс между чувствительностью модели и её специфичностью.

Автоэнкодеры могут иметь различную архитектуру и уровень сложности: от простейших полносвязных структур до более продвинутых модификаций на основе сверточных и рекуррентных слоёв. Для анализа финансовых транзакций чаще всего применяются полносвязные автоэнкодеры, так как данные в подобных задачах представлены числовыми векторами признаков. Эффективность такой модели напрямую зависит как от качества исходных данных, так и от объема обучающей выборки. В этом контексте автоэнкодеры выступают как мощный инструмент безнадзорного выявления аномалий, особенно в ситуациях, когда число реальных примеров мошенничества крайне ограничено. Подобный подход находит применение не только в финансовой сфере, но и в более широком контексте кибербезопасности.

Для проведения эксперимента в настоящем исследовании был выбран открытый датасет **Credit Card Fraud Detection**, характеризующийся ярко выраженным дисбалансом классов: доля мошеннических транзакций составляет менее 1% от общего числа записей. Предобработка данных включала нормализацию числовых признаков, а также разделение выборки на тренировочную, тестовую и валидационную части.

Для задачи обнаружения аномалий использовалась архитектура автоэнкодера, включающая энкодер и декодер. Энкодер сжимал входные данные в латентное представление размерности 10, после чего декодер выполнял реконструкцию исходного вектора признаков. Обучение модели проводилось с использованием оптимизатора **Adam**, при этом применялась стратегия ранней остановки, позволяющая предотвратить переобучение и повысить обобщающую способность сети [3].

Оценка качества модели осуществлялась с применением таких метрик, как **MSE (среднеквадратичная ошибка), Precision, Recall, F1-score и AUC.** Результаты тестирования приведены на рисунке 2. Исследование продемонстрировало высокую эффективность разработанной модели: автоэнкодер продемонстрировал более высокие показатели по сравнению с такими традиционными методами, как логистическая регрессия и Isolation Forest.

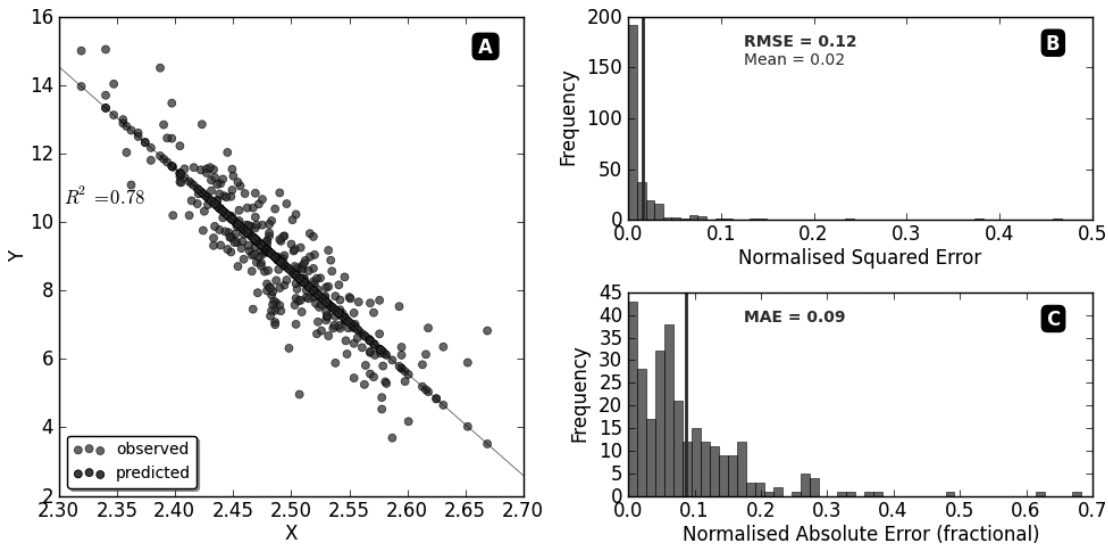

Рис. 2. График ошибок восстановления (MSE) для нормальных и аномальных транзакций

Проведенный анализ показал, что предложенная модель автоэнкодера работает лучше, чем традиционные методы (логистическая регрессия, алгоритм изоляции леса). Автоэнкодер точнее определяет необычные финансовые операции. Это говорит о том, что нейросети-автоэнкодеры очень хорошо выявляют подозрительные транзакции и могут быть полезны в финансах.

Главным критерием проверки качества стало среднее квадратичное отклонение (MSE), которое чётко отображало разницу между исходными и реконструированными финансовыми записями. Малое значение ошибки означало отсутствие подозрений относительно конкретной транзакции, а высокий показатель ошибки мог указывать на возможное нарушение нормального режима поведения. Экспериментально доказано, что автоэнкодер эффективно восстанавливает обычные финансовые события, чему соответствует минимальное значение среднего квадрата ошибок. Оценка успешности распознавания злоумышленников проводилась с использованием таких показателей, как точность (Precision), полнота (Recall) и коэффициент сбалансированности (F1-score), критически важных в ситуации существенного перекоса распределения типов событий. Предложенная модель показала отличные способности по обнаружению редких нарушений. Показатель F1-score обеспечил оптимальное соотношение между уровнем чувствительности и специфичности, сведя количество ошибочных сигналов к минимуму. Ещё одним убедительным аргументом стало изучение кривых ROC и вычисленной площади под ними (AUC). Методы дали стабильное разделение обычных и нестандартных записей. Обзор методов также зафиксировал превосходство предлагаемого подхода перед классическими моделями, включая логистическую регрессию и Isolation Forest, практически по всем важным критериям оценки, таким как точность, полнота и площадь под ROC-кривой. Немаловажным аспектом оказалось определение порога принятия решений, основанного на величине ошибки реконструкции. Точная настройка этого показателя оказала решающее влияние на равновесие между количеством найденных преступных операций и числом случайных предупреждений. Следует подчеркнуть важное свойство автоэнкодеров — независимость от заранее помеченных примеров отклонений и способность самостоятельно раскрывать скрытую структуру сложных многомерных данных.

Итоги эксперимента, которые приведены на рисунке 3, ещё раз подтверждают значимость в задаче нахождения сомнительных финансовых манипуляций. В перспективе качество выявления мошенничества может быть улучшено за счёт применения более сложных модификаций, таких как вариационные автоэнкодеры (VAE), которые учитывают вероятностную природу данных и способны обеспечить более точное разделение транзакций [5].

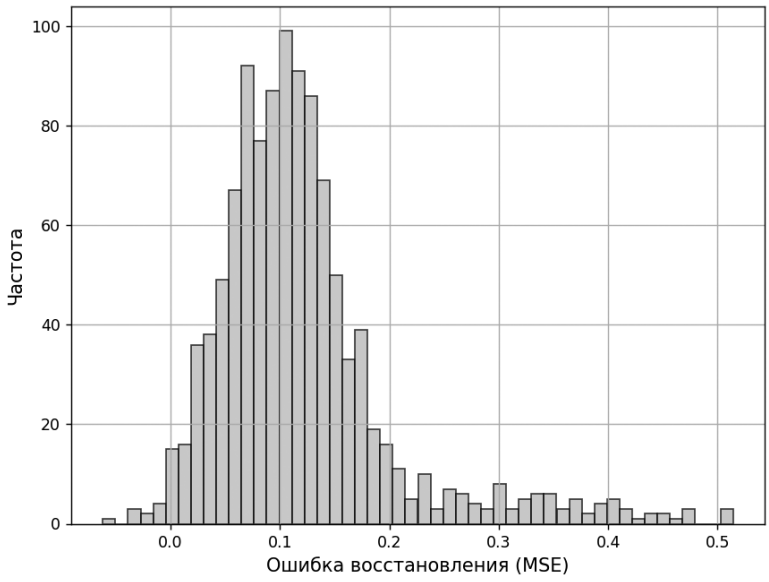

Рис. 3. Гистограмма распределения транзакций по ошибке восстановления (MSE)

Проведённое исследование наглядно подтвердило, что автоэнкодеры являются крайне действенным инструментом для выявления аномалий в финансовых транзакциях. Модель не только успешно воссоздаёт нормальные операции, глубже понимая их типичные черты, но и уверенно обнаруживает редкие, подозрительные отклонения, к которым относятся и мошеннические действия.

Как показало сравнение с другими методами, автоэнкодеры работают результативнее классических подходов — логистической регрессии и Isolation Forest — особенно когда данные сильно несбалансированны. Эти выводы свидетельствуют о том, что такие технологии легко адаптируются к реальным условиям эксплуатации, где выявление отдельных эпизодов злоупотреблений имеет ключевое значение для обеспечения защищённости и поддержания уверенности пользователей. Применение автоэнкодеров способствует созданию инновационных подходов к контролю финансовых потоков. Благодаря снижению числа неверных сигналов тревоги и улучшению выявления рисков, повышается устойчивость всей банковской инфраструктуры. Поскольку каждая допущенная погрешность способна привести к серьёзным финансовым потерям и удару по имиджу организации, подобные инструменты приобретают особую важность. Однако возможности дальнейшего роста сохраняются. Перспективным подходом является создание сложных моделей, например, вариационных автоэнкодеров. Эти модели умеют учитывать случайные факторы, что делает их более гибкими и эффективными. Такое сочетание технологий может значительно повысить точность обнаружения мошенничества и устойчивость к новым, более сложным схемам обмана.

В целом, автоэнкодеры показали свою эффективность в выявлении подозрительных операций в финансовой сфере. Их применение в системах безопасности помогает существенно уменьшить риски, связанные с мошенничеством. Дальнейшее исследование и улучшение этих методов является важной задачей, как для научных исследований, так и для практического применения.

Литература

1. Развитие алгоритмов и программ для защиты информации в сфере финансовых технологий. URL: https://www.elibrary.ru/item.asp?id=58908413 (дата обращения: 07.04.2025).

2. Финансовое мошенничество в современном мире. URL: https://www.elibrary.ru/item.asp?id=54279278 (дата обращения: 07.04.2025).

3. Повышение точности кластеризации QRS-комплексов с помощью вариационного автоэнкодера. URL: https://www.elibrary.ru/item.asp?id=68018062 (дата обращения: 07.04.2025).

4. Исследование влияния batch-size на качество обучения нейронных сетей. URL: https://elibrary.ru/item.asp?id=65663440 (дата обращения: 07.04.2025).

5. Анализ эффективности алгоритмов обучения на основе градиентного спуска в задаче дообучения предварительно обученной нейросети. URL: https://elibrary.ru/item.asp?id=48115024 (дата обращения: 07.04.2025).

References

1. Razvitiye algoritmov i programm dlya zashchity informatsii v sfere finansovykh tekhnologiy. URL: https://www.elibrary.ru/item.asp?id=58908413 (data obrashcheniya: 07.04.2025).

2. Finansovoye moshennichestvo v sovremennom mire. URL: https://www.elibrary.ru/item.asp?id=54279278 (data obrashcheniya: 07.04.2025).

3. Povysheniye tochnosti klasterizatsii QRS-kompleksov s pomoshch'yu variatsionnogo avtoenkodera. URL: https://www.elibrary.ru/item.asp?id=68018062 (data obrashcheniya: 07.04.2025).

4. Issledovaniye vliyaniya batch-size na kachestvo obucheniya neyronnykh setey. URL: https://elibrary.ru/item.asp?id=65663440 (data obrashcheniya: 07.04.2025).

5. Analiz effektivnosti algoritmov obucheniya na osnove gradiyentnogo spuska v zadache doobucheniya predvaritel'no obuchennoy neyroseti. URL: https://elibrary.ru/item.asp?id=48115024 (data obrashcheniya: 07.04.2025).

КУШНИР Надежда Владимировна, старший преподаватель кафедры информационных систем и программирования ФГБОУ ВО «Кубанский государственный технологический университет». 350072, г. Краснодар, ул. Московская, д. 2. E-mail: kushnir.06@mail.ru

ВЛАСЕНКО Александра Владимировна, кандидат технических наук, доцент, Врио заведующего кафедрой информационной безопасности, профессор кафедры ФГКОУ ВО «Краснодарский университет МВД России». 350005, г. Краснодар, ул. Ярославская 128. E- mail: alex_vlasenko@list.ru

КОЛЕСНИКОВ Даниил Сергеевич, студент 4 ФГБОУ ВО «Кубанский государственный технологический университет». 350072, г. Краснодар, ул. Московская, д. 2. E-mail: dan-kol@bk.ru

ЯЦКЕВИЧ Евгений Сергеевич, магистрант 2 курса ФГБОУ ВО «Кубанский государственный технологический университет». 350072, г. Краснодар, ул. Московская, д. 2. E-mail: es_yatskevich@internet.ru

KUSHNIR Nadezhda Vladimirovna, Senior Lecturer at the Department of Information Systems and Programming of the Federal State Budgetary Educational Institution of Higher Education «Kuban State Technological University». 2 Moskovskaya St., Krasnodar, 350072. E-mail: kushnir.06@mail.ru

VLASENKO Alexandra Vladimirovna, Candidate of Technical Sciences, Associate Professor, Acting Head of the Department of Information Security Professor of the Department, of the Krasnodar University of the Ministry of Internal Affairs of Russia, 128 Yaroslavskaya str., Krasnodar, 350005. E-mail: alex_vlasenko@list.ru

KOLESNIKOV Daniil Sergeevich, 4th year student of the Federal State Budgetary Educational institution of Higher Education "Kuban State Technological University". 2 Moskovskaya St., Krasnodar, 350072. E-mail: dan-kol@bk.ru

YATSKEVICH Evgeny Sergeevich, 2nd year Master's student at the Federal State Budgetary Educational Institution of Higher Education «Kuban State Technological University». 2 Moskovskaya St., Krasnodar, 350072. E-mail: es_yatskevich@internet.ru

УДК 004.42 Вестник УрФО № 3(57) / 2025, с. 24–29

Неклюдов Д. Н., Лебедь А. С., Кузьмина У. В. DOI: 10.14529/secur250303

ПРОГРАММНО-АППАРАТНЫЙ КОМПЛЕКС МОНИТОРИНГА WI-FI СЕТЕЙ ДЛЯ ОБНАРУЖЕНИЯ И ПРОТИВОДЕЙСТВИЯ БПЛА

В данной статье рассматривается разработка программно-аппаратного комплекса для мониторинга Wi-Fi сетей с целью обнаружения и противодействия беспилотным летательным аппаратам (БПЛА). Предложен вероятностный подход к идентификации сетей БПЛА на основе анализа ключевых параметров беспроводного трафика. Описаны методы обработки зашумленных данных. Представлены способы определения местоположения БПЛА и оператора. В качестве меры активного противодействия рассмотрена атака деаутентификации, обеспечивающая разрыв связи между БПЛА и пультом управления.

Ключевые слова: *информационная безопасность, беспилотные летательные аппараты, мониторинг, обнаружение дронов, анализ трафика, трилатерация, атака деаутентификации.*

Neklyudov D. N., Lebed A. S., Kuzmina U. V.

SOFTWARE-HARDWARE SYSTEM FOR WI-FI NETWORK MONITORING TO DETECT AND COUNTERACT UAVS

This paper presents the development of a software-hardware system for Wi-Fi network monitoring designed to detect and counteract unmanned aerial vehicles (UAVs). We propose a probabilistic method for identifying UAV networks through analysis of key wireless traffic parameters. The study describes effective techniques for processing noise-contaminated signal data and demonstrates reliable methods for locating both UAVs and their operators. As an active countermeasure, we examine the implementation of deauthentication attacks to sever communication between UAVs and their control devices.

Keywords: *information security, unmanned aerial vehicles, network monitor-ing, drone detection, traffic analysis, signal trilateration, deauthentication attack.*

Современные беспилотные летательные аппараты находят широкое применение в различных сферах, включая доставку грузов, мониторинг террито-рий, видеосъемку и даже в военных и разведывательных операциях. Однако их активное использование создает серьезные угрозы безопасности, такие как несанкционированное наблюдение, вмешательство в работу критической инфраструктуры и потенциальные кибератаки. В связи с этим, компании и организации вынуждены внедрять меры противодействия, включая глушение радиоканалов, перехват управления и обнаружение дронов с помощью радиолокационных систем.

Множество существующих традиционных методов подавления БПЛА, такие как радиоэлектронное подавление каналов связи и радионавигации, обладают существенным недостатком: они могут нарушать работу собственных беспроводных сетей организации, включая сети Wi-Fi, Bluetooth, GSM и другие протоколы связи [1]. Это происходит из-за того, что подобные системы работают неизбирательно, создавая мощные помехи во всем рабочем диапазоне частот и подавляя не только сигналы БПЛА, но и легитимные устройства. Такие методы особенно критичны для предприятий, где стабильная работа сетевой инфраструктуры является одним из условий функционирования бизнес-процессов.

В связи с этим актуальной задачей становится разработка специализиро-ванных систем мониторинга, способных обнаруживать и идентифицировать БПЛА по их взаимодействию с Wi-Fi сетями без негативного влияния на существующую инфраструктуру связи. Подобные решения должны точно детектировать подозрительные сети и интегрироваться в комплексные системы защиты.

В качестве одного из способов идентификации подобных Wi-Fi сетей возможно применить анализ обмена кадров при общении БПЛА с пультом управления. Поскольку точное определение принадлежности Wi-Fi сети к БПЛА не всегда возможно, предлагается использовать вероятностный подход, основанный на анализе совокупности параметров беспроводного трафика [2].

К основным параметрам для оценки, перехваченных с помощью сниффера пакетам относятся:

- Анализ OUI (Organizationally Unique Identifier) MAC-адреса
- Количество устройств в сети
- SSID (имя сети)
- Анализ силы сигнала (RSSI) и динамики его изменения
- История и время жизни сети

В качестве базовой вычислительной платформы системы мониторинга была выбрана одноплатная микрокомпьютерная система Raspberry Pi 4 Model B с установленной операционной системой Raspberry Pi OS (ранее известной как Raspbian). Данный выбор обусловлен сочетанием вычислительной мощности, энергоэффективности, необходимых при анализе большого количества пакетов из множества сетей, а также компактных размеров устройства [3].

Хотя плата Raspberry Pi 4 оснащена встроенным беспроводным сетевым адаптером Broadcom BCM4345 с поддержкой стандартов IEEE 802.11 b/g/n/ac (2,4/5 ГГц), способным работать в режиме мониторинга (monitor mode), для задач данного исследования его возможностей оказалось недостаточно. Для преодоления ограничений, таких как, недостаточная мощность и ограниченная чувствительность передатчика, был дополнительно подключен внешний беспроводной адаптер ALFA AWUS036ACM с чипсетом MediaTek MT7612U, а также резервные адаптеры [4].

Для захвата и анализа беспроводных пакетов в разработанной системе используется библиотека Scapy для языка Python, которая предоставляет воз-можности для низкоуровневой работы с сетевыми пакетами [5]. С помощью данного инструментария осуществляется перехват радиокадров, из которых извлекаются ключевые параметры для последующего анализа. Особое значение имеет получение показателя мощности принимаемого сигнала (RSSI), которое реализуется через обращение к полю dBm_AntSignal в RadioTap-заголовке пакета (*pkt[RadioTap].dBm_AntSignal*) [6].

При практическом применении системы в условиях плотной городской застройки или на промышленных объектах возникает существенная проблема, связанная с одновременным присутствием большого количества беспроводных сетей и подключенных к ним устройств. Массовый параллельный анализ всех обнаруженных сетей приводит к значительной нагрузке на вычислительные ресурсы системы и снижению оперативности обработки.

Для оптимизации работы системы был разработан многоуровневый подход к анализу сетей, основанный на принципе приоритезации. Новые, только что обнаруженные сети подвергаются наиболее тщательному и детальному анализу по всем доступным параметрам. По мере накопления статистики о сети и подтверждения ее стабильности и легитимности, интенсивность проверок постепенно снижается. Сети, длительное время присутствующие в зоне мониторинга и демонстрирующие предсказуемое поведение, переходят в категорию "известных" и проверяются по упрощенной схеме. Приоритезация позволяет снизить вычислительную нагрузку на систему, сохраняя при этом эффективность обнаружения подозрительной активности.

В условиях работы с сильно зашумленными беспроводными сигналами возникает необходимость в специальных алгоритмах обработки данных, позволяющих одновременно решать две задачи: снижение влияния случайных помех и сохранение актуальной информации о динамике изменения сигнала. Для этих целей в разработанной системе применяется метод взвешенного скользящего среднего (Weighted Moving Average, WMA).

$$WMA_i = \frac{\sum_{i=0}^{k-1} \omega_i \cdot x_{t-i}}{\sum_{i=0}^{k-1} \omega_i},$$

где: x_t – последовательность измеренных значений RSSI

ω_i – весовые коэффициенты

k – размер окна усреднения

Практические испытания показали, что применение WMA для обработки RSSI-сигналов в условиях городской застройки снижает среднеквадратичную ошибку позиционирования на 60–70% в сильнозашумленных сетях и на 13–20% в стабильных условиях по сравнению с использованием необработанных данных. На рисунке 1 визуализирована работа фильтра взвешенного, скользящего, среднего, примененного к искусственно зашумленному сигналу имитирующему динамику приближающегося и удаляющегося источника беспроводного сигнала.

Исходные данные содержат значительный уровень случайных помех, характерный для реальных условий работы в городской среде. Фильтр сглаживает высокочастотных шумовых составляющих, при этом сохраняя тенденции изменения уровня сигнала.

В дополнение к описанным ранее методам идентификации беспилотных летательных аппаратов по характеристикам Wi-Fi сетей, в разработанной сис-теме реализован альтернативный подход, основанный на перехвате и анализе кадров протокола MAVLink [7]. Данный протокол используется в большинстве коммерческих и любительских БПЛА.

Перехваченные кадры MAVLink имеют четко определенную структуру, включающую несколько обязательных полей: начальный маркер пакета (STX), длина полезной нагрузки (LEN), последовательность пакетов (SEQ), идентификаторы системы (SYS) и компонента (COMP), тип сообщения (MSG), полезные данные (PAYLOAD), контрольная сумма (CKA, CKB) (рисунок 2).

Анализ перехваченных кадров MAVLink позволяет получить текущие координаты БПЛА. В полезной нагрузке кадров часто содержатся данные от бортового GPS-приемника, включающие точные координаты аппарата.

Для вычисления расстояния до беспроводного источника (в частности, БПЛА) в разработанной системе применяется метод трилатерации, основанный на анализе уровня принимаемого сигнала. Трилатерация позволяет оценить местоположение излучающего устройства без необходимости использования специализированного оборудования [8].

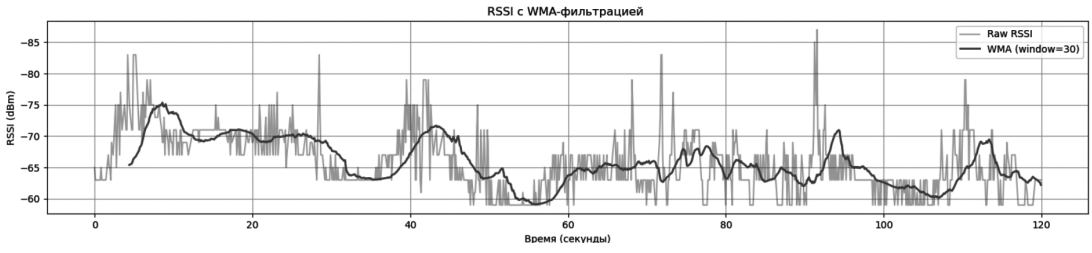

Рис. 1. Динамика RSSI-сигнала до и после применения WMA-фильтра

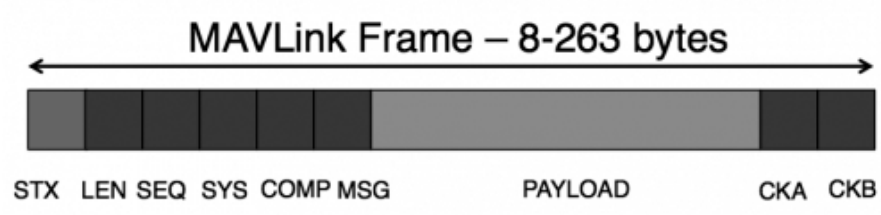

Рис. 2. Структура кадра протокола MAVLink

Метод базируется на известной модели затухания радиосигнала в свободном пространстве, которая математически выражается соотношением:

$$RSSI = P_{tx} - 10 \cdot n \cdot \lg(d) + X_{\sigma},$$

где: P_{tx} – мощность передатчика источника сигнала (дБм)

n – коэффициент затухания, зависящий от характеристик среды распро-странения (значения 2 - 4)

d – расстояние от приемника до источника (метры)

X_{σ} – случайная составляющая, учитывающая эффекты многолучевого распространения и шумы

Подход, несмотря на определенные ограничения по точности (погреш-ность в условиях городской застройки может составлять 5-15 метров), обеспе-чивает достаточную эффективность для задач обнаружения и примерного позиционирования БПЛА, особенно при использовании в сочетании с другими методами анализа.

Для определения местоположения оператора (пульта дистанционного управления) также используется метод трилатерации на основе показателей RSSI, полученных в трех различных точках пространства [9].

Математически задача решается с помощью системы уравнений:

$$\begin{cases} E0 = \sqrt{(x_o - x_e)^2 + (y_o - y_e)^2 + (z_o - z_e)^2}, \\ B0 = \sqrt{(x_o - x_b)^2 + (y_o - y_b)^2 + (z_o - z_b)^2}, \\ C0 = \sqrt{(x_o - x_c)^2 + (y_o - y_c)^2 + (z_o - z_c)^2}, \end{cases}$$

где: (x_o, y_o, z_o) – искомые координаты оператора,

$(x_e, y_e, z_e), (x_b, y_b, z_b), (x_c, y_c, z_c)$ – координаты БПЛА в разные моменты времени.

Основным преимуществом метода является его эффективность при работе с большинством коммерческих БПЛА, использующих протокол MAVLink.

После успешного обнаружения БПЛА с использованием вероятностного подхода и подтверждения его принадлежности к потенциально опасным объектам, система инициирует механизм активного противодействия [10]. В качестве эффективного метода нейтрализации применяется атака деаутентификации, направленная на разрыв установленного соединения между БПЛА и пультом управления оператора. Данный подход основан на отправке специально сформированных управляющих кадров, которые принудительно завершают активную сессию связи. По результатам атаки оператор теряет возможность управления дроном и получать видеопоток с его бортовой камеры, что вынуждает аппарат либо автоматически возвращаться в точку взлета (при наличии соответствующей функции), либо совершать аварийную посадку.

В результате проведённого исследования был разработан программно-аппаратный комплекс мониторинга Wi-Fi сетей для противодействия БПЛА, сочетающий в себе несколько взаимодополняющих методов обнаружения и нейтрализации беспилотников. Комплекс реализует многоуровневый вероятностный подход к идентификации сетей БПЛА на основе анализа совокупности параметров беспроводного трафика. В качестве активного средства противодействия используется селективная атака деаутентификации, обеспечивающая разрыв соединения между БПЛА и пультом управления без нарушения работы легитимных беспроводных устройств.

Несмотря на достигнутые результаты, следует отметить некоторые огра-ничения системы. Точность определения местоположения в условиях плотной городской застройки требует дальнейшего совершенствования, а эффек-

тивность противодействия может снижаться при работе с новыми моделями БПЛА. Кроме того, система требует оптимизации вычислительных ресурсов для обработки трафика в режиме реального времени при большом количестве одновременно активных сетей и подключенных к ним устройств.

Применимые в исследовании методы в рамках разработки комплекса сохраняют практическую значимость, поскольку основаны на фундаментальных характеристиках беспроводной связи, не зависящих от конкретных моделей БПЛА или используют общепринятые протоколы взаимодействия. Эти способы в совокупности образуют сбалансированное решение, сочетающее достаточную эффективность с относительно низкими требованиями к вычислительным ресурсам, что делает комплекс практичным для широкого внедрения.

Литература

1. Способы подавления радиозакладных устройств / И. И. Баранкова, У. В. Кузьмина, А. Р. Федорова, Ю. А. Кульевич // Безопасность информационного пространства: Сборник трудов XXII Всероссийской научно-практической конференции студентов, аспирантов и молодых учёных, Челябинск, 30 ноября 2023 года. – Челябинск: Издательский центр ЮУрГУ, 2024. – С. 77-83. – EDN KVEVUI.

2. Неклюдов, Д. Н. Разработка алгоритма обнаружения Wi-Fi-сетей БПЛА / Д. Н. Неклюдов, А. С. Лебедь, У. В. Кузьмина // Актуальные проблемы современной науки, техники и образования. – 2024. – Т. 15, №1. – С. 29-30.

3. Баранова, Ж. М. Основные требования к функционалу сетевого анализатора трафика на основе Raspberry Pi / Ж. М. Баранова, Д. С. Большаков, Д. В. Ковтунович // Энергетика, информатика, инновации - 2022 (электроэнергетика, электротехника и теплоэнергетика, математическое моделирование и информационные технологии в производстве) : сборник трудов XII международной научно-технической конференции, Смоленск, 23 ноября 2022 года / филиал ФГБОУ ВО «НИУ «МЭИ» в г. Смоленске. Том 1. – Смоленск: Б. и., 2022. – С. 149-150. – EDN MFSEUM.

4. Фаткуллин, А. Р. современные средства радиоразведки / А. Р. Фаткуллин, У. В. Михайлова // Актуальные проблемы современной науки, техники и образования: Тезисы докладов 79-й международной научно-технической конференции, Магнитогорск, 19–23 апреля 2021 года. Том 1. – Магнитогорск: Магнитогорский государственный технический университет им. Г.И. Носова, 2021. – С. 402. – EDN NXUIZG.

5. Федорова, А. Р. Анализ защищенности скрытых беспроводных сетей / А. Р. Федорова, В. А. Шпак // Этнополитический и религиозный экстремизм в России: социально-культурные истоки, угрозы распространения в информационной среде, методы противодействия: Сборник материалов Всероссийской молодежной научной школы-конференции, Уфа, 04–05 декабря 2020 года. – Уфа: ООО "Издательство "Диалог", 2020. – С. 184-189. – EDN WJJJUC.

6. Пахомов С. Анатомия беспроводных сетей // КомпьютерПресс. 2018. №1. С. 155-158.

7. Киричек, Р. В. Метод обнаружения беспилотных летательных аппаратов на базе анализа трафика / Р. В. Киричек, А. А. Кулешов, А. Е. Кучерявый // Труды учебных заведений связи. – 2016. – Т. 2, № 1. – С. 77-82.

8. Мельничук, А. И. К проблеме синтеза многопозиционной радиолокационной станции обнаружения беспилотных летательных аппаратов / А. И. Мельничук, Н. В. Горячев, Н. К. Юрков // Надежность и качество сложных систем. – 2022. – № 3(39). – С. 33-41. – DOI 10.21685/2307-4205-2022-3-4.

9. Kosmanos D., Pappas A., Tzes A., Cantos-Romanos D. RSSI-Based Indoor Localization with the Internet of Things. J. of Sensor and Actuator Networks, 2018, vol. 7, no.4, available at: https://www.researchgate.net/publication/325561044_RSSI-Based_Indoor_Localization_with_the_Internet_of_Things (accessed [1.04.2025]).

10. Гриценко Н.С., Лукьянов Г.И., Михайлова У.В. Безопасность сетей беспроводной передачи данных / Безопасность информационного пространства. Сборник трудов XVII Всероссийской научно-практической конференции студентов, аспирантов и молодых ученых: в 2 томах. 2018. С. 64-69.

References

1. Sposoby podavlenija radiozakladnyh ustrojstv / I. I. Barankova, U. V. Kuz'mina, A. R. Fedorova, Ju. A. Kul'evich // Bezopasnost' informacionnogo prostranstva : Sbornik trudov XXII Vserossijskoj nauchno-prakticheskoj konferencii studentov, aspirantov i molodyh uchjonyh, Cheljabinsk, 30 nojabrja 2023 goda. – Cheljabinsk: Izdatel'skij centr JuUrGU, 2024. – S. 77-83. – EDN KVEVUI.

2. Nekljudov, D. N. Razrabotka algoritma obnaruzhenija Wi-Fi-setej bpla / D. N. Nekljudov, A. S. Lebed', U. V. Kuz'mina // Aktual'nye problemy sovre-mennoj nauki, tehniki i obrazovanija. – 2024. – T. 15, №1. – S. 29-30.

3. Baranova, Zh. M. Osnovnye trebovanija k funkcionalu setevogo analizatora trafika na osnove Raspberry Pi / Zh. M. Baranova, D. S. Bol'shakov, D. V. Kovtunovich // Jenergetika, informatika, innovacii - 2022 (jelektrojenergetika, jelektrotehnika i teplojenergetika, matematicheskoe modelirovanie i informacionnye tehnologii v proizvodstve): sbornik trudov XII mezhdunarodnoj nauchno-tehnicheskoj konferencii, Smolensk, 23 nojabrja 2022 goda / filial FGBOU VO «NIU «MJel» v g. Smolenske. Tom 1. – Smolensk: B. i., 2022. – S. 149-150.

4. Fatkullin, A. R. sovremennye sredstva radiorazvedki / A. R. Fat-kullin, U. V. Mihajlova // Aktual'nye problemy sovremennoj nauki, tehniki i obrazovanija : Tezisy dokladov 79-j mezhdunarod-noj nauchno-tehnicheskoj konferencii, Magnitogorsk, 19–23 aprelja 2021 goda. Tom 1. – Magnitogorsk: Magnitogorskij gosudarstven-nyj tehnicheskij universitet im. G.I. Nosova, 2021. – S. 402.

5. Fedorova, A. R. Analiz zashhishhennosti skrytyh besprovodnyh setej / A. R. Fedorova, V. A. Shpak // Jetnopoliticheskij i religioznyj jekstremizm v Rossii: social'no-kul'turnye istoki, ugrozy raspro-stranenija v informacionnoj srede, metody protivodejstvija: Sbornik materialov Vserossijskoj molodezhnoj nauchnoj shkoly-konferencii, Ufa, 04–05 dekabrja 2020 goda. – Ufa: OOO "Izda-tel'stvo "Dialog", 2020. – S. 184-189. – EDN WJJJUC.

6. Pahomov S. Anatomija besprovodnyh setej // Komp'juterPress. 2018. №1. S. 155-158.

7. Kirichek, R. V. Metod obnaruzhenija bespilotnyh letatel'nyh appa-ratov na baze analiza trafika / R. V. Kirichek, A. A. Kuleshov, A. E. Kucherjavyj // Trudy uchebnyh zavedenij svjazi. – 2016. – T. 2, № 1. – S. 77-82. – EDN XCGQHF.

8. Mel'nichuk, A. I. K probleme sinteza mnogopozicionnoj radiolo-kacionnoj stancii obnaruzhenija bespilotnyh letatel'nyh appara-tov / A. I. Mel'nichuk, N. V. Gorjachev, N. K. Jurkov // Nadezhnost' i kachestvo slozhnyh sistem. – 2022. – № 3(39). – S. 33-41. – DOI 10.21685/2307-4205-2022-3-4. – EDN GPQOJN.

9. Kosmanos D., Pappas A., Tzes A., Cantos-Romanos D. RSSI-Based Indoor Localization with the Internet of Things. J. of Sensor and Actuator Networks, 2018, vol. 7, no.4, available at: https://www.researchgate.net/publication/325561044_RSSI-Based_Indoor_Localization_with_the_Internet_of_Things (accessed [1.04.2025]).

10. Gricenko N.S., Luk'janov G.I., Mihajlova U.V. Bezopasnost' setej besprovodnoj peredachi dannyh / Bezopasnost' informacionnogo prostran-stva. Sbornik trudov XVII Vserossijskoj nauchno-prakticheskoj konferencii studentov, aspirantov i molodyh uchenyh: v 2 tomah. 2018. S. 64-69.

НЕКЛЮДОВ Данил Николаевич, студент 5 курса по специальности «Информационная безопасность автоматизированных систем» кафедры информатики и информационной безопасности, ФГБОУ ВО «Магнитогорский государственный технический университет им Г.И. Носова». 455000, г. Магнитогорск, пр. Ленина, 38. E-mail: dejeksdan@gmail.com

ЛЕБЕДЬ Александр Сергеевич, студент 5 курса по специальности «Информационная безопасность автоматизированных систем» кафедры информатики и информационной безопасности, ФГБОУ ВО «Магнитогорский государственный технический университет им Г.И. Носова». 455000, г. Магнитогорск, пр. Ленина, 38. E-mail: mail23032019@mail.ru

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

NEKLYUDOV Danil Nikolaevich, 5th year student majoring in "Information Security of Automated Systems" department of Computer Science and Information security, Nosov Magnitogorsk State Technical University, 455000, Magnitogorsk, Lenin Ave., 38. E-mail: dejeksdan@gmail.com

LEBED Alexander Sergeevich, 5th year student majoring in "Information Security of Automated Systems" department of Computer Science and Information security, Nosov Magnitogorsk State Technical University, 455000, Magnitogorsk, Lenin Ave., 38. E-mail: mail23032019@mail.ru

KUZMINA Ulyana Vladimirovna, candidate of Technical Sciences, asso-ciate professor of the department of Computer Science and Information security, Nosov Magnitogorsk State Technical University, 455000, Magnitogorsk, Lenin Ave., 38. E-mail: Ylianapost@gmail.com

УДК 004.056 Вестник УрФО № 3(57) / 2025, с. 30–39

Мельников Д. Ю.

DOI: 10.14529/secur250304

МОДЕЛЬ ХРАНЕНИЯ ДАННЫХ В NOSQL-СУБД «APACHE CASSANDRA» В ЗАДАЧАХ ОБНАРУЖЕНИЯ КОМПЛЕКСНЫХ КОМПЬЮТЕРНЫХ АТАК

В статье предлагается модель представления данных в NoSQL-СУБД «Apache Cassandra» для системы обнаружения комплексных компьютерных атак на основе анализа действий пользователей, сетевых взаимодействий и сопоставления с базой знаний MITRE ATT&CK. Сформулированы шесть целевых требований к системе хранения данных, такие как: высокая пропускная способность приема событий, предсказуемые ключ-ориентированные чтения, гибкость схемы для полуструктурированных данных, управление жизненным циклом, масштабируемость и совместимость с ML-модулями. На их основе разработана практическая модель с десятью ключевыми таблицами, предложены правила группировки по временным окнам, стратегия оптимизации хранения для временных рядов TWCS и механизмы автоматического удаления устаревших событий по времени жизни TTL для оперативного окна хранения. Приведено теоретическое обоснование проектных решений, выполнены расчеты объемов хранения и размеров партиций, а также построена аддитивная модель end-to-end задержки. Выполнено аналитическое моделирование трех сценариев нагрузки, показаны ориентировочные требования к ресурсам и p50/p95/p99 латентности. Обозначены ограничения моделирования.

Ключевые слова: *Apache Cassandra, NoSQL, UEBA, MITRE ATT&CK, комплексная компьютерная атака.*

DATA STORAGE MODEL IN THE NOSQL DATABASE SYSTEM «APACHE CASSANDRA» FOR DETECTING COMPLEX COMPUTER ATTACKS

The article proposes a model for representing data in the Apache Cassandra NoSQL database management system for a system that detects complex computer attacks based on the analysis of user actions, network interactions, and comparison with the MITRE ATT&CK knowledge base. Six target requirements for the data storage system are formulated, such as high event reception throughput, predictable key-oriented reads, schema flexibility for semi-structured data, lifecycle management, scalability, and compatibility with ML modules. Based on these requirements, a practical model with ten key tables has been developed, rules for grouping by time windows have been proposed, a storage optimization strategy for TWCS time series has been proposed, and mechanisms for automatic deletion of outdated events based on TTL lifetime for the operational storage window have been proposed. A theoretical justification of the design solutions is provided, calculations of storage volumes and partition sizes are performed, and an additive end-to-end delay model is constructed. Analytical modeling of three load scenarios has been performed, and approximate resource requirements and p50/p95/p99 latency have been shown. The limitations of the modeling are indicated.

Keywords: *Apache Cassandra, NoSQL, UEBA, MITRE ATT&CK, complex computer attack.*

Введение

Современные информационные системы все чаще становятся объектами комплексных компьютерных атак [1]. В предыдущей работе [2] автором была предложена модель обнаружения комплексных компьютерных атак (далее – ККА), основанная на построении и динамическом обновлении профилей действий пользователей (далее – UEBA) [3], обнаружении аномалий с помощью метода машинного обучения - автокодировщика и графовых сверточных сетей (далее – GCN), а также сопоставлении последовательностей событий с тактиками и техниками базы MITRE ATT&CK. Модель позволяет выявлять не только единичные аномальные действия, но и объединять их в логически связанные группы, соответствующие этапам ККА. В основе подхода — представление потенциальной комплексной компьютерной атаки как последовательности связанных групп событий, каждая из которых может соответствовать определенной тактике/технике базы знаний MITRE [4]. Данное представление позволяет переводить поток телеметрии в структурированные цепочки и оценивать их согласованность во времени и по контексту. В статье подробно описаны каналы сбора и нормализации событий, логика построения контекстных связей и механизм сопоставления групп событий с базой знаний MITRE ATT&CK

Реализация предлагаемой модели требует надежной и масштабируемой инфраструктуры хранения данных, способной обрабатывать большие объемы разнородной информации в реальном времени. Реляционные системы управления базами данных (далее – СУБД) являются недостаточно гибкими и не справляются с требованиями к скорости записи, масштабируемости и структурной неоднородности данных [5]. Альтернативный подход реализован в нереляционных СУБД

(далее – NoSQL-СУБД), которые обеспечивают гибкость схемы хранения, горизонтальную масштабируемость и устойчивость к нагрузкам.

1. Требования к хранилищу данных

В условиях постоянного роста объемов телеметрии и разнообразия источников, подход к организации хранения данных напрямую влияет на эффективность системы обнаружения. Предлагаемый подход обнаружения ККА, требует одновременной работы с различными типами данных, такими как журналы аутентификации, сетевыми соединениями, данные от средств защиты. Вышеперечисленные данные могут поступать в асинхронном режиме, отличаться по структуре и длительности хранения. Следовательно, хранилище должно не просто выполнять функцию долговременного хранения, но и обеспечивать эффективный доступ, агрегацию и подготовку данных для последующего анализа. Формализуем целевой набор требований к хранилищу данных исходя из предложенной модели обнаружения ККА:

R1 — высокая пропускная способность записи. Ключевым критерием яв-ляется способность хранилища поддерживать устойчивую запись при высоко-интенсивном и неравномерном потоке телеметрии с минимальной задержкой операций записи и отсутствием блокировок, связанных с транзакционными ог-раничениями. Данное свойство критично для предотвращения потери событий и обеспечения непрерывности.

R2 — предсказуемые и быстрые ключ-ориентированные чтения. Для ана-лиза действий пользователей и корреляции аномальных событий необходимы быстрые запросы типа «последние N событий по ключу». Поэтому важным критерием является низкая латентность при доступе к данным по ключевым атрибутам. Это обеспечивает оперативность аналитики и принятия решений в реальном времени.

R3 — гибкость схемы хранения и поддержка полуструктурированных данных. Структура событий может быть изменчива, могут появляться новые поля или меняться формат записи. Поэтому возможность хранения полуструктурированных данных и добавления новых атрибутов без полной миграции схемы является необходимым критерием.

R4 — управление жизненным циклом данных. Для снижения объема хра-нилища и устранения шума необходимо наличие встроенных механизмов автоматического удаления устаревших событий по времени жизни (Time To Live, далее — TTL) [6]. Критерием является поддержка политики выборочного хранения, а также возможность архивации и изоляции событий.

R5 — горизонтальная масштабируемость и отказоустойчивость. Хранилище должно линейно масштабироваться при росте объема данных и обеспечивать круглосуточную доступность. Возможность расширения кластера за счет добавления узлов без остановки сервиса, наличие встроенной репликации и балансировки нагрузки, а также устойчивость к сбоям отдельных узлов являются важными факторами. Это гарантирует стабильность системы даже при увеличении нагрузки и в условиях отказов оборудования.

R6 — совместимость с внешними аналитическими и модулями машинного обучения (далее – ML модули). Способность хранилища предоставлять данные для аналитики и машинного обучения. Важными свойствами являются: эффективная выборка признаков по ключам, хранение векторных представлений объектов (далее – эмбеддинги), формируемых ML-модулями, совместно с исходными событиями, поддержка агрегированных профилей пользователей и экспорт данных в форматы, удобные для обучения моделей.

Таким образом, каждое из требований к хранилищу (R1–R6) соответствует функциональным и архитектурным особенностей предложенной модели и определяет необходимость использования специализированного подхода к проектированию хранилища данных для задач обнаружения ККА. Все перечисленные характеристики в наибольшей степени реализуются в NoSQL-СУБД [4, 5], что и определяет выбор типа СУБД в качестве технологической основы.

2. Выбор NoSQL - СУБД и обоснование

Для выбора NoSQL-СУБД был проведен сравнительный анализ распро-страненных решений, на основе научно-технической литературы [5,7,8,9] и критериев, сформулированных ранее. Оценка производилась по пятибалльной шкале, однако в ходе анализа значения 1 и 2 не использовались, поскольку все рассматриваемые NoSQL-СУБД демонстрируют удовлетворительный уровень соот-

Таблица 1

Сравнительный анализ NoSQL-СУБД

СУБД	R1 (0.20)	R2 (0.15)	R3 (0.15)	R4 (0.15)	R5 (0.20)	R6 (0.15)	Итоговая оценка
Cassandra	5	5	4	5	5	4	4,7
HBase	4	4	3	4	4	3	3,7
Elasticsearch	3	3	4	4	3	5	3,8
MongoDB	3	4	5	3	3	4	3,8

ветствия требованиям. Итоговый балл рассчитан как средневзвешенное значение с учетом весов критериев. Весовые коэффициенты определялись автором на основе приоритета каждого критерия для системы хранения данных.

По итогам сравнительного анализа, можно сделать вывод о том, что NoSQL-СУБД «Apache Cassandra» демонстрирует наиболее высокую степень соответствия целевым требованиям, особенно по приоритетным критериям R1 R2, что позволяет рассматривать ее в качестве приоритетной платформы для построения системы обнаружения ККА.

3. Описание архитектуры модели хранения данных

Предлагаемая архитектура модели хранения данных (Рисунок 1) разрабо-тана с учетом обозначенных требований и реализует принцип запрос-ориентированного проектирования (query-driven design), при котором схема хранения определяется набором типовых запросов системы.

Проектирование модели хранения данных основано на ключевых принципах: денормализации и дублировании данных в соответствии со сценариями доступа для мини-

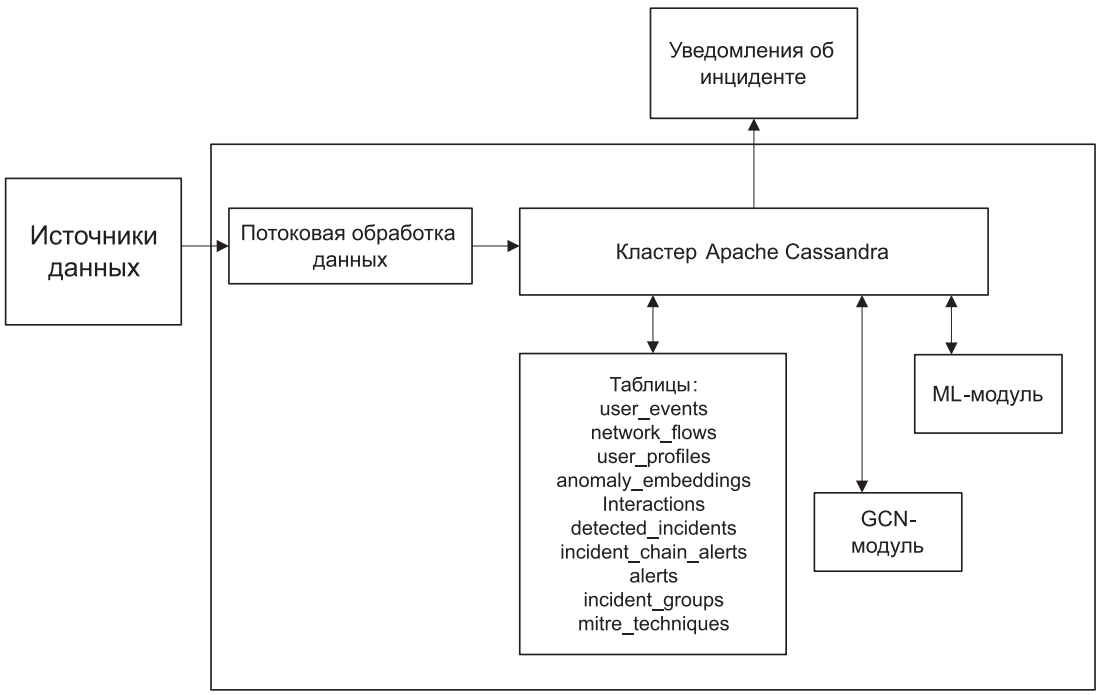

Рис. 1. Архитектура модели хранения данных в NoSQL-СУБД «Apache Cassandra»

мизации задержек выборки, применении группировки по временным окнам (далее – временная сегментация) для предотвращения неравномерной нагрузки и учете жизненного цикла данных через TTL и стратегии оптимизации хранения для временных рядов (Time Window Compaction Strategy, далее - TWCS) [10,11].

Предложенная архитектура охватывает полный цикл анализа событий от приема необработанных данных до генерации итоговых уведомлений. Ключевые сущности системы, представленные в Таблице 2, спроектированы на осно-ве характерных шаблонов доступа и требований аналитических запросов.

Поток данных в архитектуре реализован следующим образом (Рисунок 2):

1. Данные поступают и сохраняются в таблицах «user_events» и «network_flows».

2. На их основе формируются агрегаты и профили и записываются в «user_profiles», а также вспомогательные представления для расследования.

3. ML-модуль извлекает признаки из «user_events»/ «user_profiles» и вы-числяет эмбеддинги. Эмбеддинги сохраняются в «anomaly_embeddings».

4. Информация о последовательностях взаимодействий и сессиях строится в «interactions», это позволяет восстановить цепочки действий.

5. ML-модуль на основе выявленных аномалий формирует предваритель-ные оповещения и записывает их в таблицу «alerts».

6. Оповещения логически связываются в упорядоченные цепочки в «incident_chain_ alerts».

7. На основе цепочек оповещений формируются окончательные записи инцидентов в «detected_incidents».

8. Инциденты, обладающие сходными признаками и относящиеся к одной ККА, объединяются в таблице «incident_groups».

9. На каждом этапе события и инциденты сопоставляются с тактиками и техниками из базы знаний MITRE ATT&CK, обеспечивая классификацию и добавление атрибутов для анализа.

Таблица 2

Сводная таблица ключевых сущностей

Таблица	Описание	Ключи (Partition Key, Clustering Key)
user_events	Информация о действиях пользователей	PK: (event_type, event_date); CK: event_time
network_flows	Сетевые соединения	PK: (user_id, event_date) CK: event_time
user_profiles	Профили действий пользователей	PK: user_id
anomaly_embeddings	Эмбеддинги событий, полученные с помощью моделей машинного обучения	PK: (user_id, event_date); CK: event_time
interactions	Взаимодействия между субъектами и объектами	PK: (user_id, interaction_date); CK: time
detected_incidents	Обнаруженные инциденты (аномалии)	PK: (user_id, detected_date); CK: time
incident_chain_alerts	Агрегированные инциденты Цепочки инцидентов	PK: (alert_date); CK: (time, alert_id)
alerts	Финальные оповещения	PK: (created_date); CK: (time, alert_id)
incident_groups	Объединяет инциденты в логические группы с агрегированной оценкой	PK: (group_id) CK:group_date
mitre_techniques	Справочная таблица с описаниями техник и тактик базы знаний MITRE ATT&CK	PK: technique_id

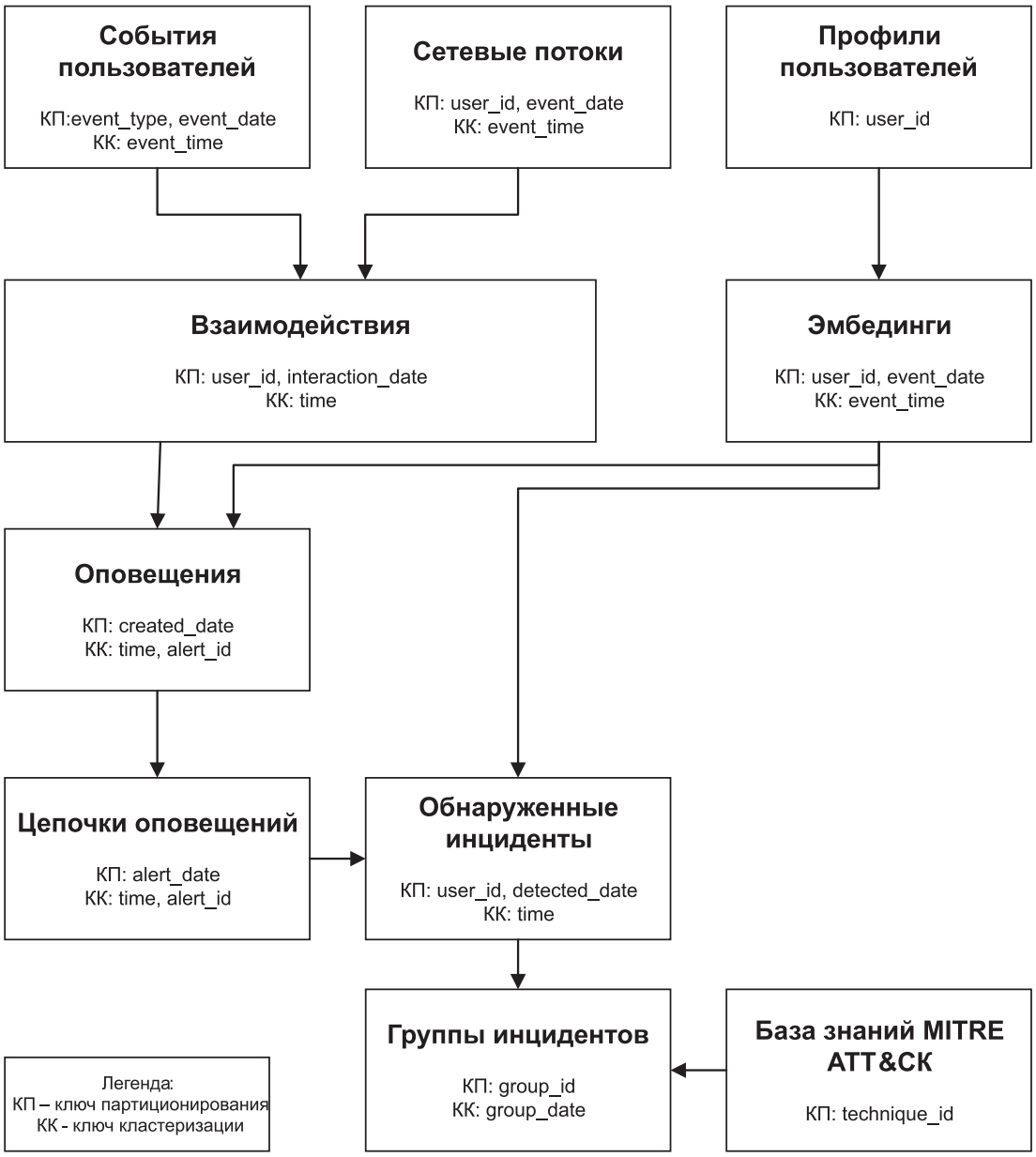

Рис. 2. Модель представления данных в NoSQL-СУБД «Apache Cassandra»

Использование временной сегментации, встроенных TTL, а также ориен-тированной на запросы схемы позволяет обеспечить стабильную работу систе-мы при высоких нагрузках. Вся архитектура организована таким образом, чтобы ускорить работу аналитических модулей и обеспечить устойчивость к сбоям.

Предложенная архитектура модели хранения данных в NoSQL-СУБД «Apache Cassandra» обеспечивает выполнение целевых требований R1–R6. За-пись телеметрии (R1) реализуется за счет оптимизированных таблиц «user_events» и «network_flows» с применением TTL и стратегий слияния дан-ных, оптимизированных для временных рядов [12]. Предсказуемые и быстрые ключ-запросы (R2) поддерживаются через специализированные таблицы «user_profiles», «interactions» и «alerts». Гибкость в работе с полуструктурированными событиями (R3) достигается использованием денормализации и хранения вложенных атрибутов. Управление жизненным циклом данных (R4) реализуется посредством TTL и дифференцированных сроков хранения для событий, уведомлений и инцидентов. Горизонтальная масштабируемость и

Таблица 3

Основные параметры моделирования

Параметр	Значение
Число узлов в кластере (N)	3
Число реплик фрагмента данных (RF)	3
Средний размер события (s)	500 байт
Сценарии нагрузки (ε)	5 000 / 20 000 / 100 000 событий/в секунду
Максимально допустимый размер партиции	100 MB
TTL для events	7 дней
Compaction strategy	TWCS
Уровень согласованности при записи	ONE
Уровень согласованности при оповещении	QUORUM

отказоустойчивость (R5) обеспечиваются встроенными средствами NoSQL-СУБД «Apache Cassandra», а именно распределенным партиционированием, репликацией и балансировкой нагрузки. Интеграция с аналитическими модулями (R6) достигается через хранение агрегатов в «user_profiles», векторных представлений в «anomaly_ embeddings» и связи с базой знаний MITRE ATT&CK в «mitre_techniques».

4. Экспериментальная оценка и моделирование

Для оценки производительности предложенной модели хранения данных при различных нагрузках необходимо:

• произвести аналитическое моделирование объема хранимых данных и нагрузку на вычислительные ресурсы в зависимости от сценариев нагрузки;

• рассчитать задержку записи и чтения (далее – латентность) в условиях нагрузки (p50/p95/p99);

• оценить совокупную задержку поступления события до генерации соответствующего оповещения (далее - end-to-end задержка).

В качестве методики использованы аналитические расчеты и приближен-ное моделирование на основе характерных параметров NoSQL-СУБД «Apache Cassandra». При моделировании использовались опубликованные результаты тестов производительности и официальная документация, применявшиеся для ориентировочной калибровки параметров модели и оценки компонент латент-ности [10,13,14].

Для оценки эффективности предложенной модели хранения данных была смоделирована работа кластера «Apache Cassandra» с заданными характеристиками. В таблице 3 представлены ключевые параметры, которые использовались в ходе экспериментов.

5. Моделирование объема хранения и потребления ресурсов

Из приведенных данных видно, что объем требуемого хранилища линейно зависит от интенсивности входного потока событий. При росте нагрузки от 5 000 до 100 000 событий/ в секунду суточный объем данных увеличивается почти на порядок. Эти оценки позволяют заранее определить минимально необходимый объем дискового пространства и требования к TTL.

Таблица 4

Результаты моделирования объема хранения данных

Нагрузка (событий/с)	Запись на узел (MB/s)	Суточный объем на узел (ГБ)	Объем (7 дн) на узел (ГБ)
5 000	2,5	216	1 512
20 000	10	864	6 048
100 000	50	4 320	30 240

Таблица 5

Результаты моделирования латентности

Сценарий нагрузки (событий/с)	Запись p50/p95/p99 (мс)	Чтение p50/p95/p99 (мс)
5 000	03.10.1930	04.08.2015
20 000	5 / 25 / 80	6 / 20 / 30
100 000	12 / 120 / 400	15 / 150 / 300

6. Моделирование латентности

Для оценки латентности запросов применен упрощенный метод, основан-ный на опубликованных экспериментальных результатах и их линейной экст-раполяции на более высокие нагрузки. Зависимость времени отклика от интен-сивности входного потока для каждой функциональной компоненты аппроксимировалась посредством масштабирования базового значения с учетом эмпирических коэффициентов чувствительности к росту нагрузки.

Результаты моделирования задержек показывают предсказуемый рост ла-тентности при увеличении интенсивности записи. Для сценариев средней на-грузки система сохраняет p95-латентность в пределах десятков миллисекунд, что подтверждает возможность работы в режиме реального времени. При экс-тремальной нагрузке наблюдается рост задержек, однако система продолжает обеспечивать приемлемую доступность и предсказуемость откликов.

7. Моделирование end-to-end задержки

Дополнительно к оценкам объема хранения и базовых задержек операций была рассчитана end-to-end задержка (таблица 6). В отличие от отдельных метрик записи и чтения, данная характеристика учитывает полный путь события в системе: сохранение в хранилище, формирование агрегированных профилей, обработку методами анализа аномалий и генерацию уведомлений.

Как видно из таблицы, при низкой и средней нагрузке медианная задержка (p50) остается в пределах десятков миллисекунд, а 95 % событий обрабатываются быстрее 350 мс. При высокой нагрузке наблюдается рост задержек, однако даже в этом случае система обеспечивает end-to-end обработку в пределах секунды. Таким образом, предложенная модель хранения данных демонстрирует устойчивость и предсказуемость временных характеристик в условиях различной интенсивности телеметрии.

Аналитическое моделирование показывает, что предложенная модель хранения данных обеспечивает ожидаемую способность выдерживать высокие нагрузки записи и поддерживать низкие p95-латентности чтений при корректном подборе параметров временной сегментации и TTL. Приведенные оценки являются репрезентативными и требуют практического подтверждения. Использование временной сегментации позволяет ограничивать размер каждой партиции и снижает риск образования неравномерно нагруженных партиций, что обеспечивает стабильное время доступа к недавним событиям.

Таблица 6

Результаты моделирования end-to-end задержки

Сценарий нагрузки (ε, событий/с)	E2E p50 (мс)	E2E p95 (мс)
5 000	~35	~120
20 000	~60	~350
100 000	~200	~800+

Заключение

В статье предложена и обоснована модель представления данных для NoSQL-СУБД «Apache Cassandra», ориентированная на задачу обнаружения комплексных компьютерных атак с учетом анализа действий пользователей, сетевых взаимодействий и сопоставления с базой знаний MITRE ATT&CK. В основе работы лежит формализованный набор требований (R1–R6), определяющих ключевые характеристики системы в условиях высоких нагрузок. На основе этих требований была разработана детализированная модель данных с обоснованием выбора схемы временной сегментации и ключей кластеризации, а также стратегий денормализации.

Проведенный анализ демонстрирует, что предложенная архитектура соответствует сформулированным требованиям. Временная сегментация, контроль размеров партиций, политика TTL и стратегия TWCS обеспечивают приемлемые значения объема хранения и латентности при заданных сценариях. Модель демонстрирует устойчивость к росту объема входных данных за счет распределенного хранения и репликации; при этом сохранение прогнозируемых p50/p95-значений достигается путем ограничения размера партиции, подбора числа реплик фрагмента данных и выбора уровней согласованности. Оценки, полученные в ходе моделирования объема хранения и латентности, подтверждают практическую реализуемость подхода при оговоренных допущениях.

Вместе с тем, необходимо отметить ограничения исследования. Численные результаты получены методом аналитического моделирования и синтетической симуляции. Оценка качества ML-модулей требует отдельного исследования на реальных данных, так как текущие ML-результаты носят иллюстративный характер и служат для демонстрации влияния дизайна хранилища на ML-производительность.

Литература

1. IBM. «What is Security Information and Event Management (SIEM)?» // IBM Think, 23 June 2023. – URL: https://www.ibm.com/topics/siem (дата обращения: 20.08.2025).

2. D. Melnikov, «A Model for Detecting Complex Computer Attacks Based on the Analysis of User Actions, Network Interactions and Comparison with the MITRE ATT&CK Knowledge Base» 2025 IEEE Ural-Siberian Conference on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT), Yekaterinburg, Russian Federation, 2025, pp. 221-224, doi: 10.1109/USBEREIT65494.2025.11054096.

3. Fortinet. «What is User and Entity Behavior Analytics (UEBA)?» // Fortinet CyberGlossary. – 2024. – URL: https://www.fortinet.com/resources/cyberglossary/what-is-ueba (дата обращения: 20.08.2025).

4. MITRE. «ATT&CK® Design and Philosophy». – March 2020. – URL: https://attack.mitre.org/docs/ATTACK_Design_and_Philosophy_March_2020.pdf (дата обращения: 20.08.2025).

5. Sadalage P. J., Fowler M. NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence. – Addison-Wesley Professional, 2012. – 192 с. – ISBN 978-0321826626.

6. DataStax Docs. «Expiring data with time-to-live (TTL)». – 2022. – URL: https://docs.datastax.com/en/cql-oss/3.3/cql/cql_using/useExpire.html (дата обращения: 20.08.2025).

7. Chebotko A., Kashlev A., Lu S. «A Big Data Modeling Methodology for Apache Cassandra» // 2015 IEEE International Congress on Big Data (BigData Congress). – 2015. – С. 238-245. – DOI: 10.1109/BigDataCongress.2015.41.

8. Apache Cassandra Documentation. «Evaluating and Refining Data Models». – 2024. – URL: https://cassandra.apache.org/doc/latest/cassandra/data_modeling/evaluating.html (да-та обращения: 20.08.2025).

9. Lakshman A., Malik P. Cassandra: a decentralized structured storage system // ACM SIGOPS Operating Systems Review. – 2010. – Т. 44, № 2. – С. 35-40. – DOI: 10.1145/1773912.1773922.

10. Apache Cassandra Documentation. «Time Window Compaction Strategy (TWCS)». – 2024. – URL: https://cassandra.apache.org/doc/latest/cassandra/operating/compaction/twcs.html (дата обращения: 20.08.2025).

11. DataStax Tutorial. «Getting Started with Time Series Data Modeling». – 2016. – URL: https://docs.datastax.com/en/tutorials/Time_Series.pdf (дата обраще-ния: 20.08.2025).

12. Ellis J. «Apache Cassandra Performance Benchmarking (4.0 brings new collectors) » // DataStax Blog, Apr. 2021. – URL: https://www.datastax.com/blog/apache-cassandra-benchmarking-40-brings-heat-new-garbage-collectors-zgc-and-shenandoah (дата обращения: 20.08.2025).

13. Instaclustr. «Apache Cassandra Data Partitioning». – 2022. – URL: https://www.instaclustr.com/blog/cassandra-data-partitioning/ (дата обращения: 20.08.2025).

14. O'Neil P., Cheng E., Gawlick D., O'Neil E. «The Log-Structured Merge-Tree (LSM-Tree) » // Acta Informatica. – 1996. – Т. 33, № 4. – С. 351-385. – DOI: 10.1007/s002360050048.

References

1. IBM. «What is Security Information and Event Management (SIEM)?» // IBM Think, 23 June 2023. URL: https://www.ibm.com/topics/siem (дата обращения: 20.08.2025).

2. D. Melnikov, «A Model for Detecting Complex Computer Attacks Based on the Analysis of User Actions, Network Interactions and Comparison with the MITRE ATT&CK Knowledge Base» 2025 IEEE Ural-Siberian Conference on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT), Yekaterinburg, Russian Federation, 2025, pp. 221-224, doi: 10.1109/USBEREIT65494.2025.11054096.

3. Fortinet. «What is User and Entity Behavior Analytics (UEBA)?» // Fortinet CyberGlossary, 2024. URL: https://www.fortinet.com/resources/cyberglossary/what-is-ueba (accessed: 20.08.2025).

4. MITRE. «ATT&CK® Design and Philosophy», March 2020. URL: https://attack.mitre.org/docs/ATTACK_Design_and_Philosophy_March_2020.pdf (accessed: 20.08.2025).

5. Sadalage P. J., Fowler M. NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence. – Addison-Wesley Professional, 2012. – 192 с. – ISBN 978-0321826626.

6. DataStax Docs. «Expiring data with time-to-live (TTL)», 2022. URL: https://docs.datastax.com/en/cql-oss/3.3/cql/cql_using/useExpire.html (accessed: 20.08.2025).

7. Chebotko A., Kashlev A., Lu S. «A Big Data Modeling Methodology for Apache Cassandra» // 2015 IEEE International Congress on Big Data (BigData Congress). – 2015. – С. 238-245. – DOI: 10.1109/BigDataCongress.2015.41.

8. Apache Cassandra Documentation. «Evaluating and Refining Data Models», 2024. URL: https://cassandra.apache.org/doc/latest/cassandra/data_modeling/evaluating.html (ac-cessed: 20.08.2025).

9. Lakshman A., Malik P. Cassandra: a decentralized structured storage system // ACM SIGOPS Operating Systems Review. –2010. – Т. 44, № 2. – С. 35-40. – DOI: 10.1145/1773912.1773922.

10. Apache Cassandra Documentation. «Time Window Compaction Strategy (TWCS)», 2024. URL: https://cassandra.apache.org/doc/latest/cassandra/operating/compaction/twcs.html (accessed: 20.08.2025).

11. DataStax Tutorial. «Getting Started with Time Series Data Modeling», 2016. URL: https://docs.datastax.com/en/tutorials/Time_Series.pdf (accessed: 20.08.2025).

12. Ellis J. «Apache Cassandra Performance Benchmarking (4.0 brings new collectors) » // DataStax Blog, Apr. 2021. URL: https://www.datastax.com/blog/apache-cassandra-benchmarking-40-brings-heat-new-garbage-collectors-zgc-and-shenandoah (accessed: 20.08.2025).

13. Instaclustr. «Apache Cassandra Data Partitioning», 2022. URL: https://www.instaclustr.com/blog/cassandra-data-partitioning/ (accessed: 20.08.2025).

14. O'Neil P., Cheng E., Gawlick D., O'Neil E. «The Log-Structured Merge-Tree (LSM-Tree) » // Acta Informatica. – 1996. – Т. 33, № 4. – С. 351-385. – DOI: 10.1007/s002360050048.

МЕЛЬНИКОВ Дмитрий Юрьевич, аспирант Учебно-научного центра «Информационная безопасность» федерального государственного автономного образовательного учреждения высшего образования «Уральский федеральный университет им. первого Президента России Б.Н. Ельцина». 620002, г. Екатеринбург, ул. Мира, 32. E-mail: melnikovdima517@yandex.ru

MELNIKOV Dmitriy Yuryevich, graduate student at the Educational and Scientific Center «Information Security» of the Federal State Autonomous Educa-tional Institution of Higher Education «Ural Federal University named after the first President of Russia B.N. Yeltsin». 620002, Yekaterinburg, st. Mira, 32. E-mail: mel-nikovdima517@yandex.ru

УДК 004.056 Вестник УрФО № 3(57) / 2025, с. 40–50

Нагибин Д. В.

DOI: 10.14529/secur250305

МОДЕЛЬ ОЦЕНКИ ЗРЕЛОСТИ MLOPS С ТОЧКИ ЗРЕНИЯ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ И ОБЪЯСНИМОСТИ

В статье рассматривается модель оценки зрелости MLOps с точки зрения информационной безопасности и объяснимости (XAI). Предлагается структурированный подход к оценке зрелости MLOps, учитывающий специфику XAI, и демонстрируется его потенциальная эффективность в повышении устойчивости моделей к угрозам информационной безопасности. Модель включает критерии оценки, основанные на интеграции XAI в жизненный цикл MLOps, автоматизации анализа угроз, балансе объяснимости и производительности, а также скорости реагирования на угрозы. Представлены уровни зрелости, аналогичные DSOMM OWASP, с примерами применения.

Ключевые слова: *MLOps, информационная безопасность, объяснимый ИИ (XAI), уровни зрелости, безопасность машинного обучения, объяснимость, угрозы информационной безопасности.*

Nagibin D. V.

MLOPS MATURITY MODEL FROM THE PERSPECTIVE OF INFORMATION SECURITY AND EXPLAINABILITY

This article presents a maturity model for MLOps from the perspective of infor-mation security and explainability (XAI). It proposes a structured approach to assessing MLOps maturity, considering the specifics of XAI, and demonstrates its potential effec-tiveness in enhancing the resilience of models against information security threats. The model includes assessment criteria based on the integration of XAI into the MLOps lifecycle, automation of threat analysis, balance of explainability and performance, and speed of response to threats. Maturity levels, similar to DSOMM OWASP, are presented with examples of application.

Keywords: *MLOps, information security, explainable AI (XAI), maturity levels, machine learning security, explainability, information security threats.*

Введение

В современном мире, где системы машинного обучения (МО) все глубже интегрируются в критически важные процессы: от здравоохранения и финансов до автономного транспорта и национальной безопасности – вопросы информационной безопасности (ИБ) приобретают первостепенное значение и требуют комплексного подхода. Однако наряду с потенциальными преимуществами, возрастают и риски информационной безопасности, связанные с уязвимостями, возникающими вследствие использования сложных, зачастую непрозрачных, моделей машинного обучения. Уязвимости, специфичные для систем машинного обучения, такие как отравление данных, атаки на конфиденциальность и эксплуатация уязвимостей в развертывании, представляют серьезную угрозу для надежности и безопасности таких систем [1; 2].

В связи с этим, концепция MLOps становится ключевым элементом обеспечения стабильной и безопасной работы моделей МО на протяжении всего жизненного цикла – от разработки и обучения до развертывания и мониторинга. Однако для достижения максимальной эффективности MLOps, необходимо интегрировать принципы объяснимости искусственного интеллекта (ИИ). Объяснимость ИИ (от англ. explainable artificial intelligence, или XAI) позволяет не только повысить доверие к моделям, но и предоставляет инструменты для выявления и смягчения угроз информационной безопасности, делая системы более устойчивыми [3].

На данный момент уже существует ряд моделей оценки зрелости MLOps (от англ. machine learning operations), таких, как: Microsoft AI Maturity Model, Gartner AI Maturity Model, MITRE AI Framework и OWASP AI Maturity Assessment (AIMA). Но в основном данные модели рассматривают XAI как отдельный инструмент для повышения прозрачности, а не как неотъемлемую часть комплексной модели оценки зрелости. В то же время, интеграция XAI в процессы оценки зрелости MLOps позволяет выявлять и смягчать угрозы, которые не обнаруживаются традиционными методами, а также повышает доверие к моделям и упрощает аудит, что подчеркивает важность интеграции XAI на всех этапах жизненного цикла модели машинного обучения [4].

Таким образом, представляется актуальной разработка комплексной модели оценки зрелости MLOps, интегрированной с принципами объяснимого искусственного интеллекта (XAI), которая позволит систематически оценивать и повышать уровень безопасности и надежности систем машинного обучения.

Данная статья предлагает модель оценки зрелости MLOps с интеграцией принципов объяснимого ИИ (XAI), разработанную для повышения безопасности и интерпретируемости систем машинного обучения. Цель данной работы – предложить структурированный подход к оценке зрелости MLOps, учитывающий специфику XAI, и продемонстрировать его потенциальную эффективность в повышении устойчивости моделей к угрозам информационной безопасности.

Угрозы информационной безопасности в жизненном цикле MLOps

Все более широкое применение систем, использующих машинное обучение, требует разработки новых стратегий обеспечения их безопасности и надежности. MLOps становится ключевым элементом в этом процессе, представляя собой комплексный подход к управлению жизненным циклом моделей машинного обучения.

MLOps включает в себя ряд ключевых аспектов: управление данными, управление версиями моделей, автоматизированное развертывание, непрерывный мониторинг и эффективную интеграцию данных, инфраструктуры и процессов [5]. Однако для достижения максимальной эффективности MLOps, необходимо учитывать вопросы информационной безопасности на каждом этапе жизненного цикла модели МО.

Важными этапами, требующими особого внимания с точки зрения информационной безопасности, являются: управление данными, обучение модели, ее развертывание, мониторинг и обновление. Базовая схема этапов жизненного цикла моделей представлена на рисунке 1.

Применение систем машинного обучения сопряжено с уникальными вызовами к обеспечению информационной безопасности. Например, фаза обучения подвержена риску отравления данных (data poisoning), когда злоумышленник внедряет вредоносные данные для манипулирования результатами обучения. Этап развертывания требует защиты от несанкционированного доступа и

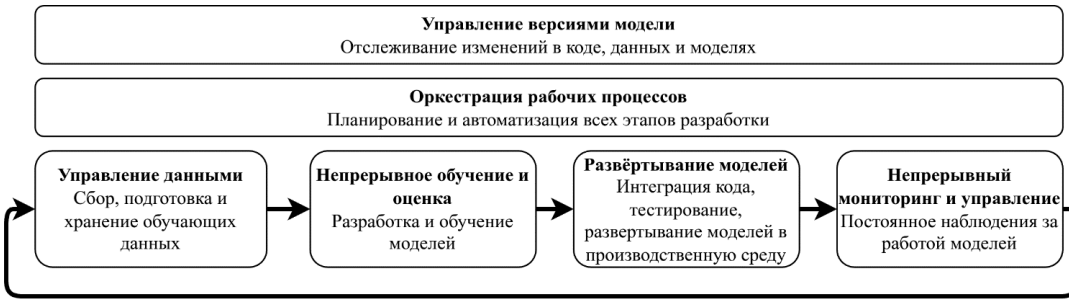

Рис. 1. Базовая схема этапов жизненного цикла MLOps

модификации модели, а также от атак, направленных на эксплуатацию уязвимостей в инфраструктуре. Мониторинг модели должен включать механизмы обнаружения атак, направленных на обход системы защиты, таких как атаки уклонения (evasion attacks). Обновление модели, необходимое для поддержания ее актуальности, должно осуществляться с соблюдением строгих протоколов безопасности, исключающих возможность внедрения вредоносного кода [5].

Также следует отметить атаки, направленные на извлечение конфиденциальной информации из обученных моделей (model extraction), а также манипулирование моделью в процессе ее эксплуатации (adversarial attacks) [2; 6]. Эти атаки могут приводить к непредсказуемым последствиям, включая утечку конфиденциальных данных, нарушение целостности системы и потерю доверия к модели. Особую опасность представляют атаки, эксплуатирующие непрозрачность моделей [7].

Эффективная защита систем машинного обучения требует понимания этих угроз и применения соответствующих мер безопасности на каждом этапе жизненного цикла MLOps, а также активной интеграции методов объяснимого ИИ (XAI) для повышения прозрачности, надежности и верифицируемости моделей.

Объяснимость ИИ как фактор информационной безопасности в MLOps

Объяснимости ИИ (XAI) важна не только для понимания работы моделей, но и для повышения их устойчивости к кибератакам, поскольку она позволяет выявлять потенциальные уязвимости и вредоносные паттерны, например, в «черных ящиках» – моделях, решения которых трудно понять и интерпретировать. В частности, интеграция XAI позволяет выявлять предвзятости в данных и моделях, а также прогнозировать поведение модели в различных сценариях, что критически важно для обеспечения безопасности и надежности систем МО.

Прямое влияние объяснимости на защиту от кибератак заключается в возможности более глубокого анализа моделей. Модели, решения которых можно понять и интерпретировать, легче анализировать на предмет наличия скрытых уязвимостей и предвзятостей. Объяснимость позволяют выявлять потенциальные точки отказа и предсказывать поведение модели в различных сценариях, что существенно повышает устойчивость модели к различным типам атак. Например, понимание того, какие признаки оказывают наибольшее влияние на принятие решения, позволяет выявлять признаки, которые могут быть использованы злоумышленниками для манипулирования моделью [3]. Такой анализ позволяет не только выявлять уязвимости, но и разрабатывать контрмеры для их предотвращения.

Значение объяснимости ИИ трудно переоценить. Например, в медицине объяснимость позволяет врачам понимать, почему модель поставила определенный диагноз или рекомендовала конкретное лечение, что повышает доверие к системе и позволяет принимать более взвешенные решения. В финансах объяснимость может быть полезна для обеспечения прозрачности при принятии решений о кредитовании, страховании и инвестициях [8].

Для эффективного применения объяснимости ИИ в целях защиты моделей, важно понимать различные категории методов и их специфические возможности. Методы объяснимости можно разделить на три основные категории: встроенные (intrinsic), постхок (post-hoc) и априорные (prior) [9].

Априорные (prior) методы разрабатываются с учетом объяснимости изначально, ча-

сто используя простые, линейные модели, которые по своей природе легко интерпретируются. Такие модели могут быть менее точными, но прозрачность делает их идеальными для понимания процесса принятия решений. Например, использование линейной регрессии вместо глубокой нейронной сети для прогнозирования кредитного риска является примером априорного подхода.

Встроенные (intrinsic) методы интегрируются непосредственно в архитектуру модели для обеспечения объяснимость на каждом этапе обучения. Например, механизм внимания (attention mechanism) в нейронных сетях позволяет понять, какие части входных данных наиболее важны для принятия решения [10].

Постхок (post-hoc) методы применяются к уже обученной модели для получения объяснений ее поведения. Они не требуют изменений в архитектуре модели и могут быть применены к любой модели, независимо от ее сложности. Постхок (post-hoc) методы являются наиболее распространенными и гибкими, но могут быть менее точными, чем встроенные (intrinsic) методы [5].

Важно отметить, что объяснимость может быть локальной или глобальной. Локальная объяснимость позволяет понять, почему модель приняла конкретное решение для конкретного экземпляра данных, в то время как глобальная объяснимость позволяет понять общие закономерности, которые использует модель для принятия решений на всем наборе данных [8].

Существует широкий спектр методов объяснимости ИИ, каждый из которых имеет свои особенности и области применения. Рассмотрим некоторые из них.

SHAP (SHapley Additive exPlanations) позволяет достигать как локальной, так и глобальной объяснимости, вычисляя вклад каждого признака в принятие решения [3]. LIME (Local Interpretable Model-agnostic Explanations), в свою очередь, создает локальные, интерпретируемые модели вокруг конкретного предсказания, позволяя понять, какие признаки наиболее важны для данного конкретного случая [8]. What-If Tool позволяет интерактивно исследовать поведение модели и визуализировать влияние различных факторов на результат [11]. ELI5 (Explain Like I'm 5) предоставляет простые и понятные объяснения работы моделей, особенно полезные для неспециалистов [12].

Объяснимость помогает выявлять уязвимости моделей, которые могут быть использованы для атак. Выбор метода объяснимости должен основываться на конкретных требованиях задачи, доступных ресурсах и необходимом уровне детализации. Например, SHAP может использоваться для обнаружения предвзятости, а LIME – для объяснения конкретных предсказаний [3; 8].

Интеграция XAI в MLOps-инфраструктуру

Интеграция методов объяснимого искусственного интеллекта (XAI) в инфраструктуру MLOps требует комплексной работы на каждом этапе жизненного цикла MLOps. Для достижения этой цели необходимо автоматизировать процессы безопасности и использовать специализированные инструменты, позволяющие отслеживать объяснимость моделей, валидировать их поведение и автоматизировать рутинные задачи.

Интеграция объяснимости (XAI) в процессы MLOps находит опору в ряде нормативных документов и инициатив. Так в стандарте NIST IR 8312 выделяются четыре принципа объяснимого ИИ: объяснимость, точность, ограниченность и согласованность [13]. Стандарты ISO/IEC 27001 и сопутствующие из серии 27000, а также ГОСТ Р ИСО/МЭК 27001, устанавливают более общие требования к системам управления информационной безопасностью, акцентируя внимание на объяснимости (XAI).

Ключевым элементом успешной интеграции объяснимости (XAI) является её внедрение в конвейеры непрерывной интеграции и доставки (CI/CD). CI/CD позволяет автоматизировать проверку моделей на предмет объяснимости и безопасности перед развертыванием. Например, при помощи SHAP можно оценивать важность признаков и выявлять потенциальные предвзятости, останавливать развертывание при выходе результатов анализа за допустимые границы, а также выявлять признаки, непропорционально влияющие на предсказания для определенных групп пользователей, что позволяет своевременно устранять предвзятости и повышать доверие к моделям ИИ [5].

Кроме прочего, важно обеспечить постоянный мониторинг моделей в процессе эксплуатации. Изменение статистических свойств входных данных, используемых для обучения модели, – явление, известное как

дрейф данных (data drift), может привести к снижению производительности и увеличению вероятности ошибок [11]. Например, LIME может использоваться для анализа локального поведения модели и выявлять признаки, которые внесли наибольший вклад в предсказание для конкретного экземпляра данных. Мониторинг LIME-анализа может быть полезен для своевременного выявления изменений в важности признаков и сигнализировать о потенциальном дрейфе данных, требующем переобучения модели или корректировки входных данных [14].

Для эффективной интеграции объяснимого искусственного интеллекта (XAI) в процессы MLOps необходим специализированный инструментарий. Платформы, такие как MLflow, позволяют управлять жизненным циклом моделей машинного обучения, включая отслеживание и анализ объяснимости моделей при интеграции с инструментами XAI, например, SHAP. Kubeflow Pipelines, в свою очередь, позволяют автоматизировать проверку моделей на предмет объяснимости и безопасности [11]. Сочетание этих инструментов позволяет не только отслеживать, но и активно управлять объяснимостью моделей в рамках MLOps.

Одним из ключевых вызовов при внедрении XAI в MLOps является поиск оптимального баланса между объяснимостью модели и её точностью. В некоторых случаях, использование более простых, но интерпретируемых моделей, может оказаться предпочтительнее, чем сложные нейронные сети. Данный выбор должен основываться на детальном анализе специфики решаемой задачи и требований к интерпретируемости результатов [15]. Кроме того, необходимо учитывать экономические аспекты внедрения XAI, включая потребность в дополнительных вычислительных ресурсах и временных затратах, связанных с обучением персонала.

Для успешной интеграции XAI рекомендуется придерживаться поэтапного подхода, начиная с оценки текущего уровня зрелости MLOps и определения приоритетных областей для внедрения XAI. Использование гибридных подходов, таких как применение локальной объяснимости для «черных ящиков», может помочь сбалансировать требования к объяснимости и производительности, обеспечивая при этом возможность анализа сложных моделей.

Предлагаемая модель оценки зрелости MLOps с точки зрения информационной безопасности и объяснимости

Интеграция XAI в MLOps может быть формализована как система управления рисками, где каждый уровень зрелости соответствует определенному этапу снижения рисков, характеризующемуся переходом от первого уровня (реактивного), ориентированного на устранение последствий угроз, к проактивному уровню, направленному на предотвращение угроз.

Кроме того, перспективным подходом является рассмотрение MLOps как киберфизической системы. Киберфизические системы (от англ. cyber-physical systems, или CPS) представляют собой интеграцию вычислительных процессов и физических процессов, обеспечивая взаимодействие между программным обеспечением (ПО) и оборудованием. Рассматривая MLOps в рамках CPS, XAI позволяет лучше понимать, как работают модели машинного обучения и почему они принимают те или иные решения, что повышает доверие к моделям и облегчает их использование в критически важных приложениях. Автоматизация безопасности, в свою очередь, снижает вероятность ошибок, которые могут быть вызваны человеческим фактором, например, неправильной настройкой параметров или несоблюдением процедур безопасности [16].

Оценка зрелости MLOps с точки зрения ИБ и объяснимости основывается на четырех ключевых критериях: степень интеграции XAI в этапы жизненного цикла модели (обучение, развертывание, мониторинг), уровень автоматизации анализа угроз (переход от реактивного к предиктивному подходу), баланс между объяснимостью и производительностью (выбор методов, соответствующих требованиям задачи) и скорость реагирования на угрозы с использованием инструментов XAI.

Предлагаемая модель зрелости основана на принципах OWASP's DSOMM [17], адаптированных под специфику MLOps. Цель создания модели – повышение уровня безопасности и надежности систем машинного обучения посредством интеграции XAI в процессы MLOps.

Модель фокусируется на интеграции XAI на всех этапах жизненного цикла моделей в рамках процессов MLOps с обеспечением баланса между объяснимостью и производительностью, а также автоматизации анализа

Таблица 1

Уровни зрелости предлагаемой модели с точки зрения ИБ и XAI

Уровень зрелости	Описание	Интеграция XAI	Автоматизация анализа угроз	Скорость реагирования на угрозы
1. Initial/Ad-hoc (Начальный)	Отсутствие формализованных процедур и структурированного подхода	Отсутствует или минимальная	Реактивный подход (реагирование на инциденты после их возникновения)	Низкая
2. Repeatable (Повторяемый)	Начало структурирования процессов, документирование базовых процедур, внедрение простых инструментов автоматизации	XAI рассматривается как дополнительная функция	Базовый мониторинг	Низкая
3. Defined (Определенный)	Формализованные процессы, внедрение стандартизированных процедур, интеграция XAI в CI/CD-конвейеры	Начальная интеграция объяснимости на отдельных этапах жизненного цикла	Начальный переход к автоматизации	Средняя
4. Managed (Управляемый)	Проактивный подход к управлению рисками, автоматизированный анализ угроз, баланс объяснимости и производительности	Базовая интеграция в процессы разработки и эксплуатации моделей	Предиктивный подход (прогнозирование атак)	Высокая
5. Optimized (Оптимизированный)	Постоянное улучшение процессов, использование AI для выявления и предотвращения угроз, непрерывный мониторинг и адаптация	Интегрирована во все этапы жизненного цикла моделей	Проактивный подход (автоматическое исправление уязвимостей и предотвращение атак)	Очень высокая

угроз с использованием инструментов XAI. Модель включает пять уровней, отражающих прогресс в интеграции XAI и повышении уровня безопасности – таблица 1.

Уровень 1 (initial/ad-hoc – начальный) представляет собой начальный этап развития MLOps, характеризующийся ручным управлением, отсутствием формализованных процедур и минимальной автоматизацией. Обеспечение безопасности строится на основе реактивных мер, а объяснимость не является приоритетом.

Уровень 2 (repeatable – повторяемый) – это этап, на котором процессы MLOps начинают автоматизироваться, появляются базовые процедуры мониторинга и управления версиями моделей. Безопасность обеспечивается на основе стандартных процедур, но отсутствует активное обнаружение атак. Объяснимость рассматривается как дополнительная функция.

Уровень 3 (defined – определенный) представляет собой этап стандартизации и документирования процессов MLOps, охватывающий основные этапы жизненного цикла моделей. Внедряются базовые механизмы мониторинга и обнаружения атак, начинается интеграция методов объяснимости для анализа и интерпретации моделей.

Уровень 4 (managed – управляемый) – это этап, на котором процессы MLOps подвергаются количественной оценке и оптимизации. Внедряются продвинутые механизмы мониторинга и обнаружения атак, и методы

объяснимости интегрированы в процессы разработки и эксплуатации моделей, обеспечивая прозрачность и подотчетность.

Уровень 5 (optimized – оптимизированный) – это высший уровень развития MLOps, характеризующийся непрерывным улучшением процессов на основе данных и обратной связи. Внедряются проактивные меры безопасности и автоматическое исправление уязвимостей, а методы объяснимости интегрированы во все этапы жизненного цикла моделей, обеспечивая максимальную прозрачность и доверие.

В качестве примера можно привести организацию, которая планирует внедрить практики MLOps для автоматизации процессов разработки и развертывания моделей машинного обучения.

Организация начинает свой путь с уровня 1 (initial/ad-hoc – начальный). На данном этапе MLOps существует в виде ряда ручных процессов, связанных с обучением и развертыванием моделей, но без формализованной политики безопасности или объяснимости. Этапы MLOps включают: ручную подготовку данных, обучение моделей с использованием локальных вычислительных ресурсов, ручное тестирование моделей и развертывание моделей в производственную среду. Отсутствует мониторинг объяснимости предсказаний модели.

По мере развития, организация переходит на уровень 2 (repeatable – повторяемый). Разработчики начинают проводить ручное тестирование моделей на предмет предвзятости, используя простые метрики. Начинается сбор данных о предсказаниях модели, но без автоматизированного анализа. Появляются первые попытки автоматизации процессов MLOps. Начинают использоваться инструменты визуализации важности признаков, но анализ объяснимости проводится вручную.

На уровне 3 (defined – определенный) организация разрабатывает и внедряет политику безопасности для моделей MLOps, определяющую требования к тестированию и мониторингу. Процессы MLOps документируются, что позволяет обеспечить воспроизводимость и отслеживаемость. Внедряются базовые инструменты для автоматизации сборки и развертывания моделей, например, использование CI/CD-конвейера для автоматизации развертывания моделей. Инструменты для визуализации важности признаков начинают использоваться систематически, но анализ объяснимости пока в основном проводится вручную.

Переходя на уровень 4 (managed – управляемый), организация интегрирует SHAP в CI/CD-конвейер для автоматической проверки объяснимости моделей перед развертыванием. Используется LIME для мониторинга дрейфа данных и выявления потенциальных угроз безопасности. Разработаны автоматизированные сценарии отката модели в случае обнаружения аномалий. Активно применяются инструменты MLOps.

Наконец, организация достигает уровня 5 (optimized – оптимизированный), что позволяет перейти к проактивным методам защиты от потенциальных угроз безопасности, например на основе анализа исторических данных. Постоянно оптимизируется процессы обеспечения безопасности и объяснимости. Инструменты MLOps позволяют автоматизировать все этапы жизненного цикла модели, включая сбор данных, обучение, развертывание, мониторинг и переобучение. Внедряется автоматизированный анализ причин возникновения аномалий в работе моделей на основе объяснимости.

Для количественной оценки уровня развития практик информационной безопасности и объяснимости в жизненном цикле машинного обучения, можно использовать индекс зрелости MLOps (Maturity Index, MI), вычисляемый например так (1):

$$MI = (A \cdot \alpha) + (B \cdot \beta) + (C \cdot \gamma) + (D \cdot \delta), \quad (1)$$

где A – степень интеграции XAI в этапы жизненного цикла модели;

B – уровень автоматизации анализа угроз;

C – баланс между объяснимостью и производительностью;

D – скорость реагирования на угрозы с использованием инструментов XAI;

коэффициенты α, β, γ и δ отражают относительную важность каждого критерия в конкретном контексте (α для A, β для B, γ для C и δ для D), позволяя адаптировать оценку зрелости к специфическим требованиям и приоритетам. Коэффициенты между собой связаны и влияют на друг друга, поэтому их сумма должна равняться единице, обеспечивая получение осмысленного значения индекса зрелости.

Степень интеграции XAI в этапы жизненного цикла модели может быть определена как:

$$A = \frac{\text{Количество этапов } MLOps \text{ с } XAI}{\text{Общее количество этапов } MLOps},$$

уровень автоматизации анализа угроз:

$$B = \frac{\text{Число автоматизированных процессов}}{\text{Общее число процессов}},$$

баланс между объяснимостью и производительностью:

$$C = \frac{\text{Точность}}{\text{Объяснимость}},$$

скорость реагирования на угрозы с использованием инструментов XAI:

$$D = \frac{1}{T_{\text{реакции}}}.$$

Приведём примерный расчёт индекса зрелости MLOps для организации, описанной в примере выше, демонстрирующий эволюцию её практик информационной безопасности и объяснимости – таблица 2.

Представленная таблица иллюстрирует, как индекс зрелости (MI) увеличивается с каждым уровнем, отражая прогресс организации в области MLOps. В результате применения предложенной модели, организация получает возможность не только объективно оценить текущий уровень информационной безопасности и объяснимости своих моделей, но и спланировать дальнейшие шаги по оптимизации процессов, снижению рисков и повышению общей производительности MLOps-инфраструктуры.

Важно отметить, что определение коэффициентов и оценок должно проводиться экспертами, обладающими знаниями в области MLOps, XAI и информационной безопасности, а формулы и расчёты по ним представлены исключительно для примера, демонстрирующего потенциальное применение предложенной модели. Реальные значения будут зависеть от специфики организации, её приоритетов и используемых технологий.

Заключение

Проведенное исследование продемонстрировало, что интеграция принципов MLOps, подкрепленная активным использованием методов объяснимого искусственного интеллекта (XAI), может оказаться полезной в процессе создания надежных, безопасных и прозрачных моделей ИИ.

Предложенная модель уровней зрелости MLOps предоставляет структурированный подход к оценке и улучшению процессов разработки и эксплуатации моделей машинного обучения. Переход от начального уровня, характеризующегося ручным управлением и отсутствием формализованных процедур, к оптимизированному уровню, обеспечивающему проактивные меры безопасности и автоматическое исправление уязвимостей, требует целенаправленных усилий и инвестиций.

С практической точки зрения, модель может найти применение в различных отраслях, включая финансовый сектор, здравоохранение и государственное управление, где безопасность и надежность моделей ИИ критически важны.

С научной точки зрения, модель вносит вклад в развитие методологии оценки зрелости процессов разработки и эксплуатации ИИ, предлагая интегрированный подход, учитывающий как технические аспекты безопасности, так и принципы объяснимости.

Таблица 2

Примерный расчёт индекса зрелости по формуле (1)

Уровень Зрелости	A	B	C	D	α	β	γ	δ	MI
1. Initial/Ad-hoc – начальный	0	0	0	0	0,3	0,25	0,25	0,2	0
2. Repeatable – повторяемый	0,25	0,25	0,25	0	0,3	0,25	0,25	0,2	0,2
3. Defined – определенный	0,5	0,5	0,5	0,25	0,3	0,25	0,25	0,2	0,45
4. Managed – управляемый	0,75	0,75	0,75	0,5	0,3	0,25	0,25	0,2	0,7
5. Optimized – оптимизированный	1	1	1	1	0,3	0,25	0,25	0,2	1

Предложенная модель способствует решению задач, связанных с повышением доверия к моделям ИИ, снижением рисков кибератак и манипуляций, а также обеспечением соответствия нормативным требованиям. При этом, важно помнить о необходимости достижения оптимального баланса между объяснимостью и сложностью моделей, выбирая стратегии, позволяющие максимизировать точность и прозрачность без существенной потери производительности.

Дальнейшие исследования могут быть направлены на разработку новых методов объяснимости, адаптированных к специфическим требованиям различных областей применения, а также на создание автоматизированных инструментов для оценки и улучшения зрелости MLOps. Предлагаемая модель может оказаться полезной и эффективной при оценке и повышении уровня безопасности и надежности систем МО, однако ее практическая ценность требует дальнейшей верификации и адаптации к конкретным условиям и требованиям.

Литература

1. Намиот Д.Е. О работе AI Red Team / Д.Е. Намиот, Е.В. Зубарева // International Journal of Open Information Technologies. – 2023. – Т. 11. – № 10. – С. 130-139.

2. Намиот Д.Е. Схемы атак на модели машинного обучения / Д.Е. Намиот // International Journal of Open Information Technologies. – 2023. – Vol. 11. – № 5. – P. 68-86.

3. Explainable artificial intelligence for cybersecurity: a literature survey / F. Charmet [et al.] // Annals of Telecommunications. – 2022. – Vol. 77. – Explainable artificial intelligence for cybersecurity. – № 11-12. – P. 789-812.

4. A MLOps Architecture for XAI in Industrial Applications / L. Faubel [et al.] // 2024 IEEE 29th International Conference on Emerging Technologies and Factory Au-tomation (ETFA). – 2024. – P. 1-4.

5. Kreuzberger D. Machine Learning Operations (MLOps): Overview, Definition, and Architecture / D. Kreuzberger, N. Kühl, S. Hirschl // IEEE Access. – 2023. – Vol. 11. – Machine Learning Operations (MLOps). – P. 31866-31879.

6. Намиот Д.Е. Атаки на системы машинного обучения - общие проблемы и методы / Д.Е. Намиот, Е.А. Ильюшин, И.В. Чижов // International Journal of Open Information Technologies. – 2022. – Vol. 10. – № 3. – P. 17-22.

7. Saputra A. The Robustness of Machine Learning Models Using MLSecOps: A Case Study On Delivery Service Forecasting / A. Saputra, E. Suryani, N.A. Rakhmawati // 2023 14th International Conference on Information & Communication Technology and System (ICTS). – 2023. – The Robustness of Machine Learning Models Using MLSecOps. – P. 265-270.

8. Шевская Н.В. Объяснимый искусственный интеллект и методы интерпретации результатов / Н.В. Шевская // Моделирование, оптимизация и информационные технологии. – 2021. – Т. 9. – № 2(33). – С. 24-25.

9. Explainable Artificial Intelligence (XAI): What we know and what is left to at-tain Trustworthy Artificial Intelligence / S. Ali [et al.] // Information Fusion. – 2023. – Vol. 99. – Explainable Artificial Intelligence (XAI). – P. 101805.

10. Soydaner D. Attention mechanism in neural networks: where it comes and where it goes / D. Soydaner // Neural Computing and Applications. – 2022. – Vol. 34. – Attention mechanism in neural networks. – № 16. – P. 13371-13385.

11. Mona M. Google Cloud Certified Professional Machine Learning Engineer Study Guide / M. Mona, P. Ramamurthy. – Indianapolis: John Wiley and Sons, 2023.

12. Мишра П. Объяснимые модели искусственного интеллекта на Python. Модель искусственного интеллекта. Объяснения с использованием библиотек, расширений и фреймворков на основе языка Python. Объяснимые модели искусственного интеллекта на Python / П. Мишра. – ДМК Пресс. – Москва: Apress, 2022. – 344 с.

13. Four principles of explainable artificial intelligence / P.J. Phillips [et al.]. – Gaithersburg, MD: National Institute of Standards and Technology (U.S.), 2021.

14. Digital Image Security: techniques and applications. Digital Image Security / eds. A.K. Singh [et al.]. – S.l.: CRC PRESS, 2024. – 351 p.

15. Investigating the Duality of Interpretability and Explainability in Machine Learning / M. Garouani [и др.] // 2024 IEEE 36th International Conference on Tools with Artificial Intelligence (ICTAI). – 2024. – С. 861-867.

16. MLOps for Cyberphysical Production Systems: Challenges and Solutions / L. Faubel [и др.] // IEEE Software. – 2025. – Т. 42. – MLOps for Cyberphysical Production Systems. – № 1. – С. 65-73.

17. OWASP DevSecOps Maturity Model [Электронный ресурс]. – URL: https://owasp.org/www-project-devsecops-maturity-model/ (дата обращения: 31.03.2025).

References

1. Namiot D.Ye. O rabote AI Red Team / D.Ye. Namiot, Ye.V. Zubareva // In-ternational Journal of Open Information Technologies. – 2023. – T. 11. – № 10. – S. 130-139.

2. Namiot D.Ye. Skhemy atak na modeli mashinnogo obucheniya / D.Ye. Namiot // International Journal of Open Information Technologies. – 2023. – Vol. 11. – № 5. – P. 68-86.

3. Explainable artificial intelligence for cybersecurity: a literature survey / F. Charmet [et al.] // Annals of Telecommunications. – 2022. – Vol. 77. – Explainable artificial intelligence for cybersecurity. – № 11-12. – P. 789-812.

4. A MLOps Architecture for XAI in Industrial Applications / L. Faubel [et al.] // 2024 IEEE 29th International Conference on Emerging Technologies and Factory Au-tomation (ETFA). – 2024. – P. 1-4.

5. Kreuzberger D. Machine Learning Operations (MLOps): Overview, Definition, and Architecture / D. Kreuzberger, N. Kühl, S. Hirschl // IEEE Access. – 2023. – Vol. 11. – Machine Learning Operations (MLOps). – P. 31866-31879.

6. Namiot D.Ye. Ataki na sistemy mashinnogo obucheniya - obshchiye problemy i metody / D.Ye. Namiot, Ye.A. Il'yushin, I.V. Chizhov // International Journal of Open Information Technologies. – 2022. – Vol. 10. – № 3. – P. 17-22.

7. Saputra A. The Robustness of Machine Learning Models Using MLSecOps: A Case Study On Delivery Service Forecasting / A. Saputra, E. Suryani, N.A. Rakhmawati // 2023 14th International Conference on Information & Communication Technology and System (ICTS). – 2023. – The Robustness of Machine Learning Models Using MLSecOps. – P. 265-270.

8. Shevskaya N.V. Ob"yasnimyy iskusstvennyy intellekt i metody inter-pretatsii rezul'tatov / N.V. Shevskaya // Modelirovaniye, optimizatsiya i infor-matsionnyye tekhnologii. – 2021. – T. 9. – № 2(33). – S. 24-25.

9. Explainable Artificial Intelligence (XAI): What we know and what is left to at-tain Trustworthy Artificial Intelligence / S. Ali [et al.] // Information Fusion. – 2023. – Vol. 99. – Explainable Artificial Intelligence (XAI). – P. 101805.

10. Soydaner D. Attention mechanism in neural networks: where it comes and where it goes / D. Soydaner // Neural Computing and Applications. – 2022. – Vol. 34. – Attention mechanism in neural networks. – № 16. – P. 13371-13385.

11. Mona M. Google Cloud Certified Professional Machine Learning Engineer Study Guide / M. Mona, P. Ramamurthy. – Indianapolis: John Wiley and Sons, 2023.

12. Mishra P. Ob"yasnimyye modeli iskusstvennogo intellekta na Python. Model' iskusstvennogo intellekta. Ob"yasneniya s ispol'zovaniyem bibliotek, rasshireniy i freymvorkov na osnove yazyka Python. Ob"yasnimyye modeli is-kusstvennogo intellekta na Python / P. Mishra. – DMK Press. – Moskva: Apress, 2022. – 344 s.

13. Four principles of explainable artificial intelligence / P.J. Phillips [et al.]. – Gaithersburg, MD: National Institute of Standards and Technology (U.S.), 2021.

14. Digital Image Security: techniques and applications. Digital Image Security / eds. A.K. Singh [et al.]. – S.l.: CRC PRESS, 2024. – 351 p.

15. Investigating the Duality of Interpretability and Explainability in Machine Learning / M. Garouani [и др.] // 2024 IEEE 36th International Conference on Tools with Artificial Intelligence (ICTAI). – 2024. – С. 861-867.

16. MLOps for Cyberphysical Production Systems: Challenges and Solutions / L. Faubel [и др.] // IEEE Software. – 2025. – Т. 42. – MLOps for Cyberphysical Production Systems. – № 1. – С. 65-73.

17. OWASP DevSecOps Maturity Model [Электронный ресурс]. – URL: https://owasp.org/www-project-devsecops-maturity-model/ (дата обращения: 31.03.2025).

НАГИБИН Дмитрий Викторович, аспирант кафедры «Автоматика и телемеханика» федерального государственного автономного образовательного учреждения высшего образования «Пермский национальный исследовательский политехнический университет». 614990, г. Пермь, Комсомольский проспект, д. 29. E-mail: dvnagibin@pstu.ru

NAGIBIN Dmitriy Viktorovich, post-graduate student of the Department of «Automation and Telemechanics» of the Federal Autonomous Educational Institution of Higher Education «Perm National Research Polytechnic University». 614990, Perm, Komsomolsky prospekt, 29. E-mail: dvnagibin@pstu.ru

УДК 004.89+004.056 Вестник УрФО № 3(57) / 2025, с. 51–60

Грибачёв А. С., Ручай А. Н. DOI: 10.14529/secur250306

ОПТИМИЗАЦИЯ МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ЭФФЕКТИВНОГО ПРЕДОТВРАЩЕНИЯ КИБЕРАТАК В СИСТЕМАХ ОБНАРУЖЕНИЯ СЕТЕВЫХ ВТОРЖЕНИЙ[1]

В данной статье рассматривается оптимизация моделей машинного обучения (ML), предназначенных для улучшения предотвращения кибератак в системах обнаружения сетевых вторжений. В исследовании делается упор на повышение производительности и надежности этих моделей путем точной настройки их гиперпараметров с использованием передовых методов оптимизации. Используя широко известный набор данных CIC-IDS-2018, проведенисчерпывающий сравнительный анализ различных алгоритмов классификации, включая деревья принятия решений, методы опорных векторов, логистическую регрессию и продвинутые групповые подходы, такие как случайные леса и градиентный бустинг. Эмпирические данные свидетельствуют о существенном повышении производительности благодаря оптимизированным конфигурациям, особенно при выявлении сложных киберугроз. Важнейшие показатели оценки, такие как показатель F1, правильность, полнота, и точность, подтверждают превосходство оптимизированных моделей ML в защите жизненно важных инфраструктур от текущих киберугроз.

Ключевые слова: *кибербезопасность, информационная безопасность, машинное обучение, алгоритмы классификации.*

[1] Исследование поддержано грантом Российского научного фонда (проект No 22-71-10095-П).

OPTIMIZING MACHINE LEARNING MODELS FOR EFFICIENT PREVENTION OF CYBER ATTACKS IN NETWORK INTRUSION DETECTION SYSTEMS

This paper explores the optimization of machine learning (ML) models designed to improve cyber-attack prevention in network intrusion detection systems. The study emphasizes improving the performance and reliability of these models by fine-tuning their hyperparameters using advanced optimization techniques. Using the widely recognized CIC-IDS-2018 dataset, we conducted an exhaustive comparative analysis of various classification algorithms, including decision trees, support vector machines, logistic regression, and advanced ensemble approaches such as random forests and gradient boosting. Empirical evidence demonstrates substantial performance improvements from optimized configurations, especially in identifying complex and emerging cyber threats. Critical evaluation metrics, such as F1-score, accuracy, recall, and precision, confirm the superiority of optimized ML models in protecting vital infrastructures against ongoing cyber risks.

Keywords: *cybersecurity, information security, machine learning, classification algorithms.*

Introduction

In today's digital age, cyberthreats pose unprecedented challenges to organizations. Attackers are employing increasingly sophisticated strategies that render conventional protection measures inadequate [1]. Consequently, there is a growing need for intelligent systems capable of processing large volumes of network traffic to detect anomalous behavior and predict potential breaches [2-3].

Machine learning is emerging as a powerful tool for improving cyberdefenses. Advanced algorithms can uncover hidden patterns in user activity and network events, identifying deviations that indicate possible attacks. By leveraging state-of-the-art methodologies, organizations can proactively respond to new forms of cyberassaults, strengthening their defenses [4, 5].

This study aims to evaluate the efficacy of different machine learning models in detecting network intrusions. Using the CIC-IDS-2018 dataset, we compare common classification techniques and introduce an automated procedure for periodically retrainingthe model's knowledge base with new data. Through rigorous testing and systematic analysis, we will determine the most effective solution for cyber-attack prevention [6].

Related works

Numerous studies have examined machine learning-driven approaches to improve cybersecurity [7-9]. Earlier studies primarily focused on basic algorithms,such as linear analysis and artificial neural networks [10]. However, current trends favor more advanced techniques such as random forests, gradient boosting, and deep learning architectures. These methods provide greater precision and robustness, offering improved insights into complex attack patterns [11-12].

Data preprocessing is pivotal to the success of any machine learning endeavor. Researchers emphasize the importance of cleansing, normalizing, and removing anomalies prior to modeling. Studies show that proper preprocessing substantially improves the predictive power of classification models [13].

Supervised learning dominates the field of cyberattack detection since labeled datasets al-

low for the precise identification of known attack types. Cross-validation is frequently employed to reliably assess model performance. For instance, kk-fold cross-validation minimizes bias by repeatedly dividing the dataset into training and testing folds, offering a comprehensive evaluation framework [14].

In article [15], the authors propose an approach thatcombines cross-validation with t-SNE visualization and applies it to four well-known datasets: UNSW-NB15, CIC-CSE-IDS2018, BoT-IoT, and ToN-IoT. The results highlight that high accuracy achieved on one dataset often fails to transfer effectively to the others due to varying class distributions. Additionally, generalization capability differs between benign and malicious traffic classes. Therefore, it is essential to carefully select training data that match dataset classes with real-world traffic profiles. While merging datasets could increase diversity, thorough validation is required prior to practical implementation.

Work [16] evaluated 15 machine learning algorithms for both binary and multi-class classification tasks focused on IoT traffic conditions. Performance was assessed based on temporal and spatial dimensions. The conclusion recommends using nonlinear algorithms, particularly tree-based ones, to train or retrain models aimed at detecting malicious network traffic in IoT intrusion detection systems.

Finally, maintaining models becomes vital in dynamic environments where attack signatures evolve rapidly. Techniques such as "on-demand retraining" allow for continuous updates, ensuring that the defensive systems are updated with the latest threats [17-18].

Methods and data

CIC-IDS-2018 Dataset

The CIC-IDS-2018 Dataset (Canadian Institute for Cybersecurit y Intrusion Detection System Dataset) is a widely recognized benchmark dataset specifically designed to evaluate intrusion detection systems (IDSs) and advanced network security mechanisms. Developed by researchers at the University of New Brunswick's Centre for Cyber Security Research, this dataset contains a comprehensive collection of both legitimate and malicious network traffic traces [19].

It contains detailed logs of network activities captured during controlled experiments that simulate various types of cyberattacks. These include brute force attacks, denial-of-service (DoS/DDoS), botnet activity, SQL injection attempts, cross-site scripting (XSS), port scans, and others. In addition to these anomalous behaviors, the dataset also captures regular user behavior (benign traffic), making it well-balanced for building and validating machine-learning-driven intrusion detection systems [20].

Each record in the dataset represents an individual flow between network endpoints and includes numerous features extracted from each flow, such as flow duration, inter-packet arrival times, packet count, bytes transferred, TCP/IP flags, source and destination IP addresses, ports, protocols, etc. All entries are structured in tabular form, facilitating easy integration with popular tools and libraries such as Pandas and Scikit-Learn [21].

Additionally, it can be employed for feature engineering tasks aimed at extracting discriminatory characteristics indicative of potential attacks.By offering a balanced mix of benign and malicious samples, alongside granularity in feature representation, the CIC-IDS-2018 dataset remains indispensable for anyone aiming to understand, simulate, or defend against sophisticated cyber assaults.

Data Preprocessing

The original dataset, obtained from the CIC-IDS-2018 repository, contained incomplete and noisy entries. Several steps were taken to prepare the data for modeling:

First, duplicates were removed to eliminate redundancy. Next, timestamps were converted into numerical representations, to facilitate time-series analyses. Missing values were filled using column medians, mitigating the impact of gaps in the data. Finally, feature scaling normalized the dataset, reducing disparities caused by varying attribute magnitudes.

After preprocessing, the cleaned dataset (base_preprocessed.csv) was ready for subsequent modeling tasks.

Data Partitioning

Proper data separation is crucial for successful machine learning because it affects the quality of model evaluation and its ability to generalize knowledge to new data. This study used a standard approach to divide the dataset into

three parts:80% of the initial dataset constituted the training data, which was essential for directly training the model and adjusting its internal parameters. The remaining 20% was distributed as follows: 10% formed the validation set, which was used for training hyperparameters and intermediate evaluation of model quality, thereby avoiding overfitting.Another 10% composed the test set, deliberately isolated from the training process and reserved exclusively for the final check of the model's operational readiness in real-life conditions. Additionally, target labels were separately extracted and recorded for each of the three subsets, stored in the files y_train_val.csv, y_valid.csv, and y_test.csv, ensuring a clear distinction between features and target variables. This division allowed for a reliable assessment of the model's quality and verified its ability to make accurate predictions on previously unseen data.

Model Evaluation and Selection

Multiple classification algorithms were tested systematically:

Logistic Regression is a fundamental classification technique that employs a linear combination of features passed through a sigmoid function to estimate the probability of an observation belonging to one of two classes (e.g., attack or no attack). A hypothesis is defined as follows:

$$z = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_3 + \cdots, \quad \#(1)$$

the resulting probability is calculated using the sigmoid function:

$$p(y = 1|x) = \frac{1}{1 + e^{-z}}, \quad \#(2)$$

transforming the linear combination into a probability score ranging from 0 to 1. Objects are assigned to class 1 if their probability exceeds 0.5. Despite its simplicity and computational efficiency, logistic regression struggles with non-linear problems and performs poorly on high-dimensional datasets. Nevertheless, it serves as a baseline for comparison and demonstrates interpretability in its coefficients, allowing easy understanding of feature influence on classification outcomes [22-23].

Support Vector Machines (SVM) is a powerful classification method aimed at finding an optimal hyperplane that separates two classes with maximum margin—the largest possible gap between the closest points (support vectors) of opposing classes. The boundary is described by the equation as follows:

$$w^T x + b = 0, \quad \#(3)$$

where w denotes the weight vector perpendicular to the hyperplane, and b is the offset along the axis. The distance from the boundary to the nearest point (a support vector) is calculated as follows:

$$\frac{|w^T x + b|}{||w||}. \quad \#(4)$$

For non-linear problems, SVM projects data into a higher-dimensional space using special functions called kernels (common choices include linear, radial basis function (RBF), and polynomial kernels). Although SVM delivers excellent separation quality even with small numbers of features and possesses strong generalization abilities, it requires careful selection of the kernel type and regularization parameters, and its computational demand increases with larger datasets [24].

Naive Bayes is a quick and computationally efficient classification algorithm based on Bayes' theorem and the simplifying assumption that features are statistically independent. The posterior probability of assigning an object to a particular class C_k is calculated as follows:

$$P(C_K|X) = \frac{P(X|C_K) \cdot P(C_K)}{P(X)}, \quad \#(5)$$

where $(X|C_K)$, the likelihood of observing features X given class C_K, is approximated as the product of individual feature probabilities under the naive independence assumption. This approach greatly reduces computational complexity, making Naïve Bayes particularly suitable for tasks involving many features. Its advantages include speed and low memory usage, but it suffers from limitations when features exhibit dependencies or correlations, potentially degrading its performance [25].

Decision Trees are simple and intuitively understandable classification models that construct hierarchical branching structures to guide decision-making processes. They recursively split data by selecting the best features for parti-

tioning, creating nodes representing logical checks and leaf nodes symbolizing final class assignments. One major advantage lies in their transparent nature, as the reasoning behind predictions is easily traceable through the tree structure. Decision trees effectively handle both categorical and numerical data, aiming to minimize impurity measures such as entropy or Gini index during splits. However, they tend to suffer from overfitting when grown too deeply, meaning they may memorize training data rather than generalizing well to new observations. Despite this limitation, decision trees remain invaluable in domains where interpretability and visualization play a crucial role, such as diagnostic systems or customer profiling [26].

Ensemble Methods (Random Forest, Gradient Boosting) arepowerful predictors combining weak learners into stronger ones, albeit requiring greater computing capacity [27].Each algorithm was tuned individually to maximize its predictive performance.Ensemble methods significantly enhance classification accuracy and robustness by aggregating multiple weak learners into a collective model. Unlike standalone decision trees, they are less prone to overfitting, yielding reliable performance even on complex, noisy, or heterogeneous datasets. However, these methods come with increased computational costs and reduced interpretability compared to single-tree models. Nonetheless, they are widely preferred in classification and regression tasks where high accuracy and reliability are paramount [18-30].

Random Forest builds an ensemble consisting of numerous decision trees, each trained on randomly sampled subsets of data and features. Final predictions are made either by majority voting (for classification) or averaging (for regression). Suppose we haveN decision trees$T_1, T_2, \ldots, T_N$, each trained on a random subsample of data and features. The prediction of the random forest is formed as follows:

• For regression:

$$\hat{Y}_{RF} = \frac{1}{N}\sum_{i=1}^{N} T_i(x). \qquad \#(6)$$

• For classification, choose the class with the highest number of votes:

$$\hat{C}(x) = arg_{C\in\{C_1,\ldots,C_K\}} \max\left\{\sum_{i=1}^{N} I(T_i(x) = c)\right\}, \#(7)$$

where$I(\cdot)$ is an indicator function equal to 1 if the condition holds, otherwise 0.

Gradient Boosting constructs a series of trees sequentially, where each subsequent tree tries to improve upon the errors of the previous ones. The next element in the ensemble learns from residuals (errors) produced by the preceding tree, gradually refining the overall prediction towards the correct output.Let us assume we have a sequence of trees$T_1, T_2, \ldots, T_M$. The prediction of gradient boosting is expressed as:

$$\hat{\boldsymbol{Y}}_{\boldsymbol{GB}} = \sum_{\boldsymbol{m=1}}^{\boldsymbol{M}} \boldsymbol{\alpha_m T_m(x)}, \qquad \#(7)$$

where α_m controls the contribution of each tree to the final outcome.

Explanation of Metrics

Metrics used here follow standard definitions in binary classification:

Accuracy represents the proportion of correctly classified instances out of the total number of predictions.

$$Accuracy = \frac{True\ Positives + True\ Negatives}{Total\ Predictions}, \#(8)$$

where *True Positives* are objects that truly belong to class 1 and were correctly classified as such; *True Negatives* are objects that truly belong to class 0 and were correctly classified as such; *False Positives* areobjects that were incorrectly classified as class 1 (false alarm); *False Negatives* are objects that were incorrectly classified as class 0 (missed detections).This is a measure of the overall correctness rate but doesnot reveal details about error distribution.

Recall measures the fraction of actual positive examples correctly identified by the model.

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}. \#(9)$$

It refers tohow well the model captured the positive examples. Having a high recall is important, when finding all relevant positive cases is crucial, like disease diagnosis,.

Precision calculates what portion of positively predicted objects actually belong to the positive class.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}, \quad \#(10)$$

Of all the objects predicted as positive, it may not be clear how many are really positive. When misclassifying something as positive is costly (such as false diagnoses), achieving high precision is desirable.

F1-Score combines precision and recall into a single metric, taking their harmonic meaning. Thus, it balances both factors equally.

$$F1 - Score = 2 \cdot \frac{Recall \times Precision}{Recall + Precision}. \quad \#(11)$$

Values closer to 1 indicate better classification performance. It is useful when you want to balance both precision and recall, especially in situations where neither should dominate (e.g., early-stage medical diagnostics).

These four metrics provide a comprehensive evaluation of the classification model's performance. Combining their outputs provides insight into strengths and weaknesses, helping to select the best model depending on task-specific needs [31]. Given these trade-offs, ensemble methods are the preferred choice for real-world intrusion detection tasks.

Results

Several classification algorithms were thoroughly tested, including Logistic Regression, Support Vector Machines (SVM), Naïve Bayes, Decision Trees, and ensemble methods such as Random Forest and Gradient Boosting. Logistic Regression,for example, yielded training accuracies ranging from 0.902672 to 0.9998309005344181, whereas Decision Trees scored perfectly (1.0) every time. Notably, ensemble methods demonstrated remarkable stability and precision, achieving almost flawless results across epochs. Similarly, SVM exhibited nearly identical performance, highlighting its effectiveness in binary classification tasks.

Experiments confirmed the superiority of ensemble methods over other classification techniques. Specifically, Random forest and Gradient boosting delivered near-perfect accuracy rates (close to 1.0), surpassing alternatives such as Logistic regression and Decision trees. Furthermore, incorporating cross-validation into the evaluation process improved reliability. With cross-validation, the model maintained stable performance across multiple iterations, minimizing the risk of overfitting.Table 1 summarizes the results comparing performance metrics.

After evaluating various machine learning algorithms according to predefined criteria, the results revealed that Random Forest and Gradient Boosting models outperformedother models. This outcome aligns with existing literature that has previously discussed the strengths and weaknesses of each algorithm. Specifically, Logistic Regression and Linear Support Vector Machines tend to perform better in binary classification settings, however, intrusion detection in-

Table 1

Summarized results comparing performance metrics.

Algorithm	Train Accuracy	Test Accuracy	Recall	Precision	F1-Score
Logistic Regression	0.99961	0.99959	0.9996	0.9996	0.9996
SVM	0.99991	0.99991	0.9999	0.9999	0.9999
Native Bayes	0.99997	0.99997	0.9999	0.9999	0.9999
Decision Tree	0.99999	0.99999	0.9999	0.9999	0.9999
Random Forest	0.99999	1.0	1.0	1.0	1.0
Gradient Boosting	1.0	0.99999	1.0	1.0	1.0
AdaBoost	0.99962	0.99965	0.9996	0.9996	0.9996

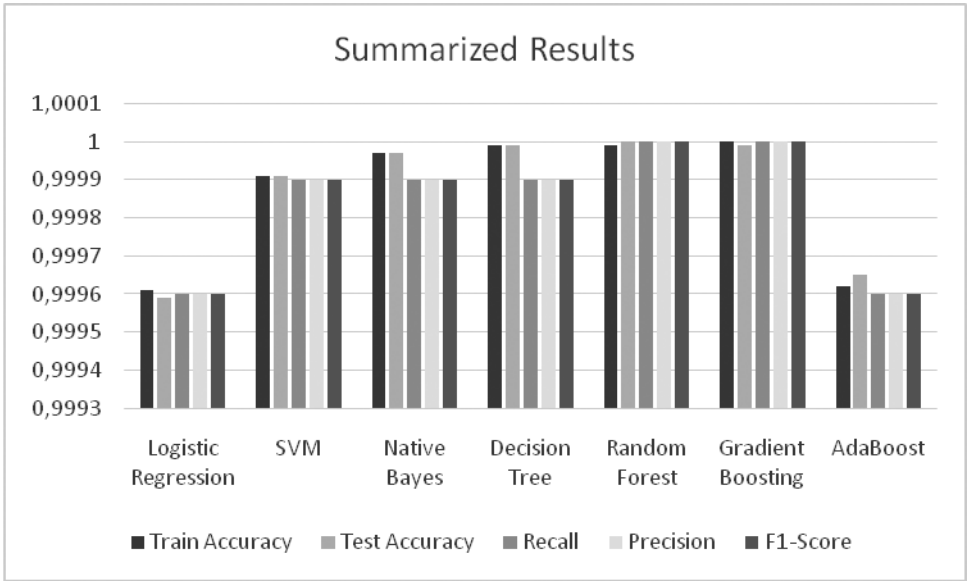

Figure 1. Summary of the results

herently involves multi-class classification. Additionally, theoretical expectations suggest that Gaussian Naïve Bayes would underperform in real-world scenarios, a hypothesis that our findings validate.

Our study confirms that Random Forest and Gradient Boosting, when coupled with training and grouping techniques, substantially improve the handling of imbalanced datasets and the prediction of cyberattacks. Across key metrics, such as precision, recall, F1-score, and accuracy, these models consistently outperformed alternative algorithms. Figure 1 illustrates a clear disparity between Random Forest, Gradient Boosting, and the remaining four models (including Logistic Regression and AdaBoost). Despite relatively high accuracy figures for certain models, this metric alone cannot fully capture their overall effectiveness.

Discussion

The authors [32] propose a NIDS system that uses machine learning to detect modern attack types with a high detection rate. Several ML algorithms were implemented and evaluated using a cutting-edge dataset containing modern attacks. The results demonstrate that the Random Forest model achieves superior performance, with a detection rate of 97%. These results suggest that machine learning has significant potential for creating highly effective NIDS systems.

Researchers in [33] introduce IWHO-XGBoost—a novel machine learning-based intrusion detection system optimized with Intelligent Wild Horse Algorithm. Trained on diverse simulated network intrusion datasets, it combines Z-score normalization and Kernel PCA for preprocessing and feature extraction. Hyperparameter optimization enhances XGBoost's intrusion identification abilities. Evaluation metrics—F1-score (98.42%), accuracy (98.45%), recall (98.4%), and precision (98.38%)—demonstrate superior detection performance compared to conventional methods.

Our results confirm that ensemble methods excel in cyber-intrusion detection. Their high accuracy makes them ideal candidates for deploying scalable, resilient cyber-defense systems. In contrast, simpler models struggle to match their performance, underscoring the value of ensembling weaker learners into robust classifiers. Performance measures such as F1-score (100.0%), accuracy (99.99%), recall (100.0%), and precision (100.0%) clearly illustrate that the proposed method significantly outperforms traditional approaches in terms of intrusion detection efficiency.

Cross-validation helped stabilize the evaluation process, ensuring consistency across tests. The automated training mechanism also played a critical role in maintaining model relevance despite shifting attack profiles.

However, caution must be exercised regarding ensemble sizes and configurations. Overly complex ensembles may lead to overfitting, necessitating vigilance in choosing appropriate parameters.

Conclusion

The present study has demonstrated the significant advantages of using advanced machine learning algorithms, such as Random Forest and Gradient Boosting,to identify cyber threats. These methods offer superior performance compared to traditional approaches by effectively handling complex data distributions and high-dimensional feature spaces commonly encountered in cybersecurity datasets.

A key innovation of this research is an automated training mechanism that enables the model to dynamically adapt to emerging attack vectors without requiring manual intervention or reconfiguration. This ensures sustained effectiveness over extended periods despite evolving adversarial strategies.

Future work may focus on exploring hybrid architecturesthat combine multiple algorithmic paradigms to enhance detection capabilities further. Additionally, more comprehensive studies on feature engineering techniques are warranted to refine input representations and maximize prediction accuracy. Investigating the process of retrofitting the model with smallportions of newdatainrealtime is also worthwhile.These advancements will contribute significantly to the development of robust and scalable solutions that can effectively mitigate modern cyber risks.

References

1. Goswami, Maloy Jyoti. (2024). AI-Based Anomaly Detection for Real-Time Cybersecurity. 3006-1075.

2. Banketov D.A., Fedorov M.A., Palmov S.V. (2024). The role of artificial intelligence in cybersecurity. Industrial Economics. 143-148.

3. GribachevA.S, Kal'shchikovV.V., RuchayA.N..Metody, algoritmyibazydannykhobnaruzheniyakomp'y uternykhincidentov. Vestnik URFU. Bezopasnost' v informatsionnoy sfere.1(51) 2024, 42-52.

4. Ejjami, Rachid. (2024). Enhancing Cybersecurity through Artificial Intelligence: Techniques, Applications, and Future Perspectives. Journal of Next-Generation Research 5 0. 1. 10.70792/jngr5.0.v1i1.5.

5. Vlasenko A.V. Obzorinstrumentovmashinnogoobucheniya i ikhprimeneniya v oblastikiberbezopasnosti / A.V. Vlasenko, P.I. Dzyoban, R.V. Zhuk // Prikaspiyskiyzhurnal: upravlenie i vysokietekhnologii. – 2020. – № 1(49). – pp. 144-155. – DOI 10.21672/2074-1707.2020.49.4.144-155. – EDN IPJDMM.

6. A.S. Gribachev, V.V. Kalschikov, A.N. Ruchay. Metody, algoritmyibazydannykhobnaruzheniyakomp'y uternykhincidentov / // VestnikUrFO. Bezopasnost' v informatsionnoysfere.– 2024. – № 1(51). – pp. 45-52. – DOI 10.14529/secur240106. – EDN JDVEUY.

7. Nagaraju, Bharath. (2025). Machine Learning For Cybersecurity: Enhancing Intrusion Detection Systems And Threat Mitigation.

8. Oyetoro, Amos & Mart, Joseph & Okoroafor, Nneoma & Amah, Joshua. (2022). Using Machine learning Techniques Random Forest and Neural Network to Detect Cyber Attacks. 10.13140/RG.2.2.27484.05763/1.

9. Y. Wang, Q. Li, and Z. Huang, "A Random Forest Approach for Detecting Botnets," in Proceedings of the International Conference on Advanced Cloud and Big Data, pp. 68-74, 2020

10. Ravi, Vinayakumar& Hb, Barathi Ganesh & Poornachandran, Prabaharan &Madasamy, Anand Kumar &Kp, Soman. (2018). Deep-Net: Deep Neural Network for Cyber Security Use Cases. 10.13140/RG.2.2.13593.06243.

11. Alharthi, Ayesha &Alaryani, Meera & Kaddoura, Sanaa. (2025). A Comparative Study of Machine Learning and Deep Learning Models in Binary and Multiclass Classification for Intrusion Detection Systems. Array. 26. 100406. 10.1016/j.array.2025.100406.

12. Jeamaon A, Khemapatapan C. Development cyber risk assessment for intrusion detection using enhanced random forest. ECTI Transactions on Computer and Information Technology (ECTI-CIT) 2024; 18:429–42.

13. Dhongade, Gauri &Chandrakar, Dr &Khande, Rajeshree. (2024). Enhancing Cyber Security: A Study of Data Preprocessing Techniques for Cyber Security Datasets. International Journal of Scientific Research in Science and Technology. 11. 71-75. 10.32628/IJSRST2411427.

14. Lumumba, Victor & Sang, Dennis &Mpaine, Mary & Makena, Njoka & Musyimi, Daniel &Kavita. (2024). Comparative Analysis of Cross-Validation Techniques: LOOCV, K-folds Cross-Validation, and Repeated K-folds Cross-Validation in Machine Learning Models. American Journal of Theoretical and Applied Statistics. 13. 127-137. 10.11648/j.ajtas.20241305.13.

15. Iwanowski, M.; Olszewski, D.; Graniszewski, W.; Krupski, J.; Pelc, F. The Choice of Training Data and the Generalizability of Machine Learning Models for Network Intrusion Detection Systems. Appl. Sci. 2025, 15, 8466. https://doi.org/10.3390/app15158466

16. Li, Shuren & Lu, Yifei & Li, Jingqi. (2022). Can We Half the Work with Double Results: Rethinking Machine Learning Algorithms for Network Intrusion Detection System. 242-247. 10.1109/CBD54617.2021.00049.

17. Ruchay A.N., Tokarev I.V., Gribachev A.S. Metody mashinnogoobucheniyaiiskusstvennogointellekta v sfereinformatsionnoybezopasnosti: analizsovremennogososto yaniyaiperspektivyrazvitiya. Vestnik URFO. Bezopasnost' v informatsionnoysfere. 4(46) 2022, 76-87.

18. Gribachev A.S., Kalschikov V.V. Issledovaniemetodov i algoritmovobnaruzheniyakomp'yuternykhincidentov. Sborniktrudov XXII Vserossiyskoynauchno-prakticheskoykonferentsiistudentov, aspirantovimolodykhuchenykh. Chelyabinsk, 2024. БЕЗОПАСНОСТЬ ИНФОРМАЦИОННОГО ПРОСТРАНСТВА, 290-294

19. Liu, Lan & Wang, Pengcheng& Lin, Jun & Liu, Langzhou. (2020). Intrusion Detection of Imbalanced Network Traffic Based on Machine Learning and Deep Learning. IEEE Access. PP. 1-1. 10.1109/ACCESS.2020.3048198.

20. Shandilya, Shishir K & Ganguli, Chirag &Izonin, Ivan & Nagar, Atulya. (2022). Cyber Attack Evaluation Dataset for Deep Packet Inspection and Analysis. Data inBrief. 46. 108771. 10.1016/j.dib.2022.108771.

21. Luay, Majed &Layeghy, Siamak &Noorbin, Faezeh & Sarhan, Mohanad & Moustafa, Nour & Portmann, Marius. (2025). Temporal Analysis of NetFlow Datasets for Network Intrusion Detection Systems. 10.48550/arXiv.2503.04404.

22. Hristov, Alexander & Trifonov, Roumen & Hristov, Valentin. (2023). Cyberattacks Detector with Logistic Regression Module. 1-4. 10.1109/COMSCI59259.2023.10315802.

23. Sperandei S. Understanding logistic regression analysis // Bio-chemia Medica. 2014. No. 24 (1), pp. 12-8.

24. Ghanem, Kinan & Aparicio-Navarro, Francisco & Kyriakopoulos, Konstantinos & Lambotharan, S. & Chambers, Jonathon. (2017). Support Vector Machine for Network Intrusion and Cyber-Attack Detection. 1-5. 10.1109/SSPD.2017.8233268.

25. Wickramasinghe, I., Kalutarage, H. Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation. Soft Comput 25, 2277–2293 (2021). https://doi.org/10.1007/s00500-020-05297-6

26. Lukiewicz, K. et al. (2025). Decision Trees and Machine Learning for Cybersecurity: How Model Settings Affect Attack Detection. In: Paszynski, M., Barnard, A.S., Zhang, Y.J. (eds) Computational Science – ICCS 2025 Workshops. ICCS 2025. Lecture Notes in Computer Science, vol 15907. Springer, Cham. https://doi.org/10.1007/978-3-031-97554-7_4

27. Shrivastav, Lokesh & Kumar, Ravinder. (2022). An Ensemble of Random Forest Gradient Boosting Machine and Deep Learning Methods for Stock Price Prediction. Journal of Information Technology Research. 15. 1-19. 10.4018/JITR.2022010102.

28. G. R. Bagadi and Y. B. Adimulam, "Random Forest and Gradient Boosting for Superior Intrusion Detection and Anomaly Classification," 2025 International Conference on Machine Learning and Autonomous Systems (ICMLAS), Prawet, Thailand, 2025, pp. 1618-1625, doi: 10.1109/ICMLAS64557.2025.10968997.

29. Gulhane, Kimsy& Saxena, Surabhi &Deogaonkar, Anant & Kumar, Vinod &Vichoray, Chandan & Goyal, Shweta. (2024). Overview of Machine Learning Techniques in Cybersecurity Data Science using Gradient Boosting and Random Forest Algorithm. 19-24. 10.1109/ICoICI62503.2024.10696688.

30. Kumar, Chaitanya & Sabeena, Jasmine. (2023). Cybersecurity Attack Detection using Gradient Boosting Classifier. 10.21203/rs.3.rs-3711213/v1.

31. Soto, Jean. (2023). Binary classification of malware by analyzing its behavior in the network using machine learning. Innovare: Revista de ciencia y tecnología. 12. 30-36. 10.5377/innovare.v12i1.15956.

32. Ali, Md & Thakur, Kutub& Schmeelk, Suzanna &Debello, Joan & Dragos, Denise. (2025). Deep Learning vs. Machine Learning for Intrusion Detection in Computer Networks: A Comparative Study. Applied Sciences. 15. 1903. 10.3390/app15041903.

33. Manivannan, R. & Selvaraj, Senthilkumar. (2025). Intrusion Detection System for Network Security Using Novel Adaptive Recurrent Neural Network-Based Fox Optimizer Concept. International Journal of Computational Intelligence Systems. 18. 10.1007/s44196-025-00767-x.

ГРИБАЧЁВ Антон Сергеевич, аспирант (соискатель) математического факультета, Челябинский государственный университет. 454001, г. Челябинск, ул. Братьев Кашириных, 129. E-mail: a.gribachev@yandex.ru.

РУЧАЙ Алексей Николаевич, доктор технических наук, доцент, заведующий кафедрой компьютерной безопасности и прикладной алгебры, Челябинский государственный университет. 454001, г. Челябинск, ул. Братьев Кашириных, 129.; профессор кафедры защиты информации, Южно-Уральский государственный университет (национальный исследовательский университет), 454080, г. Челябинск, пр. Ленина, 76. E-mail: ran@csu.ru.

GRIBACHEV Anton Сергеевич, PhD candidate of the Faculty of Mathematics, Chelyabinsk State University. 454001, Chelyabinsk, st. Brothers Kashirin, 129. E-mail: a.gribachev@yandex.ru.

RUCHAY Alexey Nikolaevich, Doctor ofTechnical Sciences, Associate Professor, Head of the Department of Computer Security and Applied Algebra, Chelyabinsk State University. 454001, Chelyabinsk, st. Brothers Kashirin, 129.; Professor, Department of Information Security, South Ural State University (National Research University). 454080, Chelyabinsk, Lenina avenue, 76. E-mail: ran@csu.ru.

УДК: 004.93'1 Вестник УрФО № 3(57) / 2025, с. 61–72

Иниватов Д. П.

DOI: 10.14529/secur250307

АУТЕНТИФИКАЦИЯ ПО ГОЛОСОВОМУ ПАРОЛЮ С УЧЁТОМ ФУНКЦИОНАЛЬНОГО СОСТОЯНИЯ ПОЛЬЗОВАТЕЛЯ[1]

В статье представлена методика биометрической аутентификации по голосу, учитывающая функциональное состояние (ФС) пользователя, такое как нормальное состояние, различные стадии алкогольной интоксикации и сонливость. Предложенный подход объединяет нейросетевой преобразователь «биометрия-код» (НПБК) с моделью классификации ФС на основе ансамблевого алгоритма CatBoost, обеспечивая высокую точность (макро-F1 0.954) в определении состояния пользователя. Методика позволяет не только подтвердить личность, но и ограничить доступ в системе в случае выявления неприемлемого ФС при наиболее важных действиях.

Ключевые слова: *голосовая аутентификация, функциональное состояние, защита данных, НПБК, CatBoost, биометрическая аутентификация, алкогольная интоксикация.*

Inivatov D. P.

VOICE PASSWORD AUTHENTICATION WITH CONSIDERATION OF USER'S FUNCTIONAL STATE

The article presents a voice-based biometric authentication method that accounts for the user's functional state, including normal state, various stages of alcohol intoxication, and drowsiness. The proposed approach integrates a neural fuzzy extractor with a functional state classification model based on the CatBoost ensemble algorithm, achieving high accuracy (macro-F1 0.954) in state detection. The method not only verifies user identity but also dynamically restricts access based on functional state, minimizing risks of errors in critical systems.

Keywords: *voice authentication, functional state, neural fuzzy extractor, CatBoost, biometric authentication, alcohol intoxication, data security.*

Работа выполнена ОмГТУ в рамках государственного задания Минобрнауки России на 2023-2025 годы № FSGF-2023-0004

Введение

Голосовая биометрия используется повсеместно: в системах аутентификации банков, в кол-центрах, умных устройствах. Современные биометрические системы, основанные на голосе, уже достигают значений точности в 98-99% [1] для задач аутентификации пользователя, предполагая, что речевой образ остаётся стабильным. Данный прогресс позволил переориентировать исследования с повышения базовой точности на решение задач, связанных с вариативностью биометрических данных, среди которых функциональное состояние (ФС) пользователя является одним из основных факторов.

ФС, включая сонливость, стресс, утомление и алкогольную интоксикацию, вызывают значительные изменения в амплитудно-частотных характеристиках голоса [2], что не только снижает точность аутентификации, но и создает побочные угрозы. Выдача доступа пользователю в неприемлемом ФС, таком как сильное опьянение или крайняя усталость, может привести к компрометации данных. В банковской системе сотрудник в подобном состоянии может инициировать ошибочные транзакции, повредив финансовые записи, или стать мишенью для фишинговой атаки, открыв доступ к защищаемой информации. Не все отклонения от нормального ФС представляют угрозу, но состояния, влияющие на когнитивные и моторные функции, требуют строгого контроля.

Для минимизации таких рисков традиционная аутентификация, подтверждающая только легальность пользователя, недостаточна. Необходимо внедрять методы, оценивающие ФС. Подобные решения позволят не только верифицировать личность, но и предотвратить угрозу целостности данных от действий в нежелательных состояниях.

Обзор существующих методов определения изменё нного функционального состояния пользователя

Методы машинного обучения (МО) находят широкое применение в системах голосовой аутентификации, позволяя анализировать такие характеристики речевых сигналов, как тембр, высота тона, ритмические особенности, паузы и др. Такие подходы как машина опорных векторов (SVM), нейронные сети (НС), модели Гауссовой смеси и ансамблевые методы, способны определять схожие признаки и проводить классификацию пользователей по соответствующим категориям [3].

Для анализа голосовых данных в задачах классификации речевой сигнал сначала проходит этап предобработки, направленный на устранение шумов и нормализацию, после чего из него выделяются признаки, такие как мел-кепстральные коэффициенты или спектральные характеристики, которые затем обрабатываются алгоритмами МО [4]. Задача оценки ФС может быть сформулирована как бинарная классификация с двумя категориями – нормальным и изменённым функциональным состоянием, либо как многоклассовая, в которой каждый класс отражает своё состояние. Для оценки эффективности предлагаемых решений исследователи активно применяют метрики UAR (невзвешенная средняя точность), которая учитывает сбалансированность классов, и accuracy (доля пра работку системы проверки состояния водителей для предупреждения неприемлемых ФС за рулём. В исследовании используется сразу 2 базы данных: ALC для классификации состояния алкогольной интоксикации и BESD для распознавания эмоций (нейтральное, гнев, радость, грусть, страх, отвращение). Признаки извлекаются с помощью мелчастотных кепстральных коэффициентов (MFCC), линейных предсказательных коэффициентов и частоты пересечения нуля для ALC, а также высоты тона, энергии и просодических характеристик для определения эмоций. Полученные признаки объединяются, их обработка осуществляется с помощью алгоритмов МО: SVM, метод k-ближайших соседей, метод случайного леса, градиентный бустинг и метод экстремально случайных деревьев, с применением анализа главных компонент и селекции признаков.

Предложенное авторами решение позволило достичь точности в 80% для выявления состояния алкогольного опьянения, 94.7% в подтверждении его отсутствия, 75% для эмоции «радость», 66,7% для «нейтрального» состояния, 72.41% для «гнева» и 57.77% «грусти». Объединение признаков улучшило точность по сравнению с использованием отдельных наборов данных, что делает метод перспективным для определения ФС.

В исследовании [10] представлен алгоритм ADLAIA, использующий глубокое обучение для определения алкогольного опьянения по 12-секундным речевым записям. Метод основан на анализе акустических призна-

ков речи, таких как тембр и ритм, с использованием свёрточных нейронных сетей для классификации. Тестирование на корпусе ALC показало значение по метрике UAR = 68.09% для выявления изменённого ФС при концентрации алкоголя в крови 0.5‰ и 75.7% для 0.12‰, что является лучшим результатом на ALC среди аналогичных методов по метрике UAR.

Исследование [11] посвящено методам определения изменённого ФС на примере корпуса ALC. Базовая система использует набор ComParE (6373 акустических признака, таких как основная частота F0, спектральные параметры и др.) и SVM, достигнув значения UAR 0.616 без нормализации. Нормализация по испытуемым повысила результат до 0.669, а кластерная нормализация до 0.665. Модель ResNet18 (18-слойная сеть с остаточными связями, где свёрточные слои выделяют признаки, слой глобального усреднения GAP формирует фиксированное представление, а два полносвязных слоя классифицируют данные) без нормализации достигла UAR 0.633. Z-Нормализация позволила повысить значение метрики до 0.677 (на 4.4 %), а кластерная нормализация до 0.671.

Можно сделать вывод, что с 2011 года, после публикации проекта ALC в открытом доступе, в научной среде произошло заметное развитие исследований по распознаванию ФС по голосу, что стало возможным благодаря обширному набору речевых данных. Этот ресурс послужил основой для создания и проверки множества методов, способствуя прогрессу в данной области.

Набор данных для экспериментальных исследований

В рамках одного из предыдущих исследований [2] был сформирован свой архив голосовых записей AIC-spkr-130, ориентированный на задачу аутентификации по голосовому паролю с защитой биометрического шаблона. Предлагаемый архив состоит из записей 130 испытуемых возраста от 18 до 50 лет в формате wav и частотой дискретизации 8 кГц. Для сбора архива голосовых данных было отобрано 130 участников, не страдающих серьёзными заболеваниями или неврологическими расстройствами. Запись проводилась в утренние часы после достаточного отдыха испытуемых, чтобы минимизировать влияние усталости. Участники по-

Таблица 1

Сравнение эффективности методов определения изменённого ФС

Описание архитектуры	Точность, %	UAR, %	База данных
Архитектура ResNeXt50 и мел-спектрограммы [5]		59,2	ALC
Метод адаптивной обратной фильтрации и модель Лильенкранца-Фанта [7]	69,3 – 77,0		Собственная
Метод случайного леса (10 деревьев) [8]	95,3		Собственная
Метод опорных векторов и метод случайного леса [9]	57,7 – 94,7		BESD, ALC
Алгоритм глубокого обучения на основе аудио для выявления алкогольного опьянения (ADLAIA) [10]		68,1 – 75,7	ALC
ResNet-18 + глобальный средний уровень пула + z-нормализация [11]		67,1	ALC
Система супервекторов модели Гауссовой смеси, объединенная с линейной машиной опорных векторов для классификации и дополнительными методами нормализации [12]		70,54	ALC
Модели Гауссовой смеси + линейная машина опорных векторов [13]	68,6	68,5	ALC
Двунаправленная рекуррентная нейронная сеть со стробируемыми рекуррентными единицами [14]	75,9	69,2	ALC

следовательно вводились в разные функциональные состояния, в каждом из которых повторяли >60 голосовые пароли, представляющие из себя фразы из 2 слов на русском языке.

Собранная база голосовых образов содержит записи испытуемых в 5 различных функциональных состояниях:

1. Нормальное состояние. В этом состоянии испытуемый не подвергался каким-либо внешним воздействиям.

2. Градации стадий алкогольной интоксикации. В соответствии с классификацией стадий алкогольного опьянения, приведённой в методических рекомендациях Минздрава СССР [15] выделяются 3 стадии, каждая из которых характеризует определённое функциональное состояние:

2.1. Состояние незначительной алкогольной интоксикации (интоксикация 1), при котором концентрация алкоголя в крови находится в диапазоне от 0.3‰ до 0.5%. Она характеризуется отсутствием выраженных клинических проявлений воздействия алкоголя.

2.2. Лёгкое опьянение (интоксикация 2), при котором концентрация алкоголя находится в диапазоне от 0.5 до 1.5‰. Данное содержание токсина воздействует на когнитивные и моторные функции, способность к принятию решений, замедляет реакцию. Также в данном ФС наблюдается ощущение тепла, комфортного состояния, мышечное расслабление, понижение сдержанности. В рамках проводимого эксперимента у испытуемых данное значение составляло около 0,8‰.

2.3. Третья стадия (интоксикация 3). Концентрация алкоголя находится в диапазоне от 1.5 до 2.5‰. Характеризуется нарушением моторных функций, способности к рассуждению, нечеткой речью, реакцией зрачка на свет и эмоциональной неустойчивостью. В рамках эксперимента уровень ограничивался значением порядка 1,6‰.

3. Сонное. Это состояние характеризуется снижением общей активности и когнитивной продуктивности. Для его воспроизведения участники использовали натуральные растительные седативные средства.

К формированию базы голосовых образов привлечены русскоговорящие добровольцы без неврологических или речевых нарушений.

Архитектура нейросетевой модели для распознавания ФС пользователя

Эксперимент проводился AIC-spkr-130. Голосовые данные представляли собой сырые аудиосигналы с приведением частоты дискретизации к 16000 Гц. Предобработка включала удаление самых зашумлённых и нечётких записей, а также нормализацию амплитуды в диапазоне [-1, 1]. Извлечение признаков выполнялось комбинированным подходом: MFCC через библиотеку openSMILE захватывали спектральные характеристики речи, такие как формантные частоты, а высокоуровневые представления извлекались тремя предобученными моделями: Wav2Vec2 XLS-R, HuBERT-large и ECAPA-TDNN.

Традиционные мел-кепстральные коэффициенты обеспечивали базовое описание амплитудно-частотных свойств речи. Признаки, извлечённые моделями Wav2Vec2 XLS-R и HuBERT, формировались путём агрегации выходов последнего скрытого слоя с использованием усреднения и стандартного отклонения. Модели XLS-R и HuBERT формировали 2048-мерные вектора признаков, а ECAPA-TDNN - 192-мерные вектора. Итоговый набор получался путём объединения этих 3 наборов и содержал 4376 признаков.

Для классификации ФС были сформированы и обучены три модели. Первой был многослойный персептрон (MLP) [16], состоящий из входного слоя на 4376 признаков, трёх полносвязных слоёв с 512, 256 и 128 нейронами для постепенного уменьшения размерности и выделения ключевых признаков. В качестве функции активации применялась ReLu после каждого слоя. Оптимизация проводилась с помощью Adam с начальной скоростью обучения 0.0001. Обучение продолжалось 72 эпохи при размере батча 256. Параметры нейронной сети приведены в таблице 2.

В качестве ядра SVM использовались радиальные базисные функции (RBF). Для обучения и тестирования применялись только мел-кепстральные коэффициенты. Параметр регуляризации, отвечающий за величину штрафа за ошибки в классификации, был настроен на 10. Автоматически адаптируемый параметр ширины ядра определялся как функция дисперсии данных. Основные параметры SVM приведены в таблице 3.

Для обучения и тестирования модели CatBoost [17] (метод МО, использующий дере-

Таблица 2

Параметры модели MLP

Параметр	Значение
Входной слой	4376 признаков
Скрытые слои	512, 256, 128 нейронов
Dropout	0.3, 0.3, 0.2 соответственно
Функция активации	ReLU
Оптимизатор	Adam
Число эпох в обучении	72
Размер батча	256

Таблица 3

Параметры модели SVM

Параметр	Значение
Тип ядра	RBF
Признаки	MFCC
Регуляризация	10
Параметр ширины ядра	Адаптивный (по дисперсии данных)

Таблица 4

Параметры ансамблевой модели MO CatBoost

Параметр	Значение
Число признаков	4376
Функция потерь	Многоклассовая классификация
Глубина дерева	8
Количество итераций	3000
Скорость обучения	0.1
Бутстрэппинг	Байесовский (регуляризация 0.7)
Балансировка весов классов	Автоматическая
Начальное значение	Фиксированное

вья решений для задачи классификации), уже показавшей свою эффективность в задаче распознавания возраста и пола по голосу, использовался весь набор из 4376 признаков, включая мел-кепстральные коэффициенты и высокоуровневые представления. В качестве целевой функции при обучении использовалась многоклассовая функция потерь. Глубина дерева 8 и Байесовский бутстрэппинг (bootstrapping) с параметром регуляризации 0.7 были подобраны опытным путём для избегания переобучения. А количество итераций 3000 и скорость обучения 0.1 позволило модели в достаточной степени обучиться. Автоматическая балансировка весов классов увеличивала вес наименьшего из классов (интоксикация 2). Значения параметров указаны в таблице 4.

Таблица 5

Результаты 5-блочной кросс-валидации и средние значения точности, сбалансированной точности и F1

№ кросс-валидации	MLP accuracy	MLP balanced acc	MLP macro F1	SVM accuracy	SVM balanced acc	SVM macro F1	CatBoost accuracy	CatBoost balanced acc	CatBoost macro F1
1	0.834	0.810	0.810	0.723	0.702	0.704	0.957	0.953	0.950
2	0.844	0.837	0.837	0.714	0.701	0.698	0.954	0.953	0.951
3	0.801	0.781	0.778	0.714	0.694	0.694	0.961	0.957	0.957
4	0.844	0.822	0.824	0.714	0.702	0.698	0.956	0.952	0.954
5	0.835	0.820	0.821	0.717	0.706	0.706	0.951	0.947	0.947
Среднее	0.832	0.814	0.814	0.716	0.700	0.700	**0.956**	**0.952**	**0.954**

Таблица 6

Средние метрики по классам для MLP и CatBoost

Класс	MLP precision	MLP recall	MLP F1	CatBoost precision	CatBoost recall	CatBoost F1
Интоксикация 1	0.76	0.81	0.79	0.96	0.94	0.95
Интоксикация 2	0.59	0.68	0.63	0.93	0.92	0.93
Интоксикация 3	0.91	0.75	0.82	0.95	0.96	0.95
Нормальное	0.96	0.89	0.93	0.97	0.97	0.97
Сонное	0.87	0.93	0.90	0.96	0.97	0.97

Эксперимент проводился с использованием 5-блочной кросс-валидации, где данные разделялись на обучающую и тестовую выборки в соотношении 80/20 с сохранением пропорций классов (нормальное, интоксикация 1–3, сонное). Кросс-валидация обеспечивала оценку устойчивости моделей путём разделения очищенного набора данных на пять подмножеств, где каждое подмножество использовалось как тестовая выборка один раз в то время, как оставшиеся четыре – для обучения. Результаты всех 5 кросс-валидационных экспериментов и их усреднённые значения представлены в таблице 5. Ансамблевый алгоритм МО CatBoost оказался наиболее эффективным в задаче распознавания ФС: средняя точность 0,956, сбалансированная точность 0,952, макро-F1 0,954. MLP уступала с точностью 0.826, сбалансированной 0.818 и макро-F1 0,814. SVM, ограниченная только MFCC признаками, имела самые низкие показатели: точность 0,716, сбалансированная 0,7, макро-F1 0,7.

Детальный анализ по классам (табл. 6) показал, что CatBoost стабильно превосходит другие модели, особенно для интоксикации 1–3 (F1 0.92–0.96) и нормального состояния (F1 0.97). MLP хорошо справлялась с нормальным и сонным состояниями (F1 0.89–0.95), но слабее для интоксикации 2 (F1 0.56–0.73). SVM показала умеренные результаты (F1 0.54–0.84), с наибольшими трудностями для интоксикации 2.

Для визуализации представлена гистограмма данных результатов (Рис. 2), где CatBoost демонстрирует стабильное превосходство.

Матрицы ошибок (Рис. 3–5) иллюстрируют распределение предсказаний: для CatBoost (Рис. 5) ошибки классификации минимальны, с высокой точностью для всех классов, тогда как MLP (Рис. 3) показывает заметное число ошибок между интоксикацией 1 и 2, а SVM (Рис. 4) – между всеми стадиями опьянения. По ошибочным решениям всех 3 классификаторов можно выявить общую закономерность. Она заключается в том, что состояние «интоксикация 2» может быть ошибочно классифицировано как «интоксикация 1» или «интоксикация 3». А состояние «интоксикация 3» может быть ошибочно классифицировано с наибольшей вероятностью как «интоксикация 1» или «интоксикация 2». Также можно выявить, что ансамблевый алго-

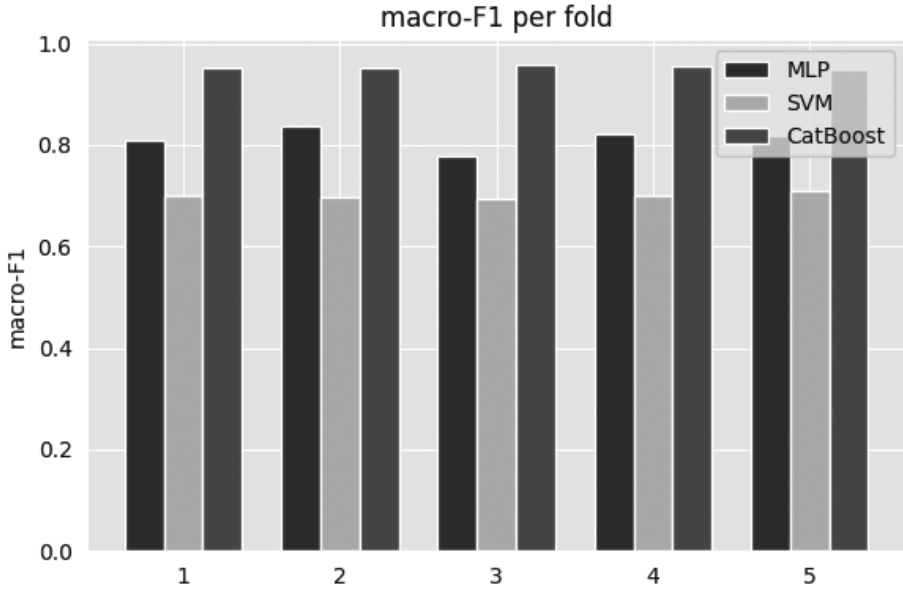

Рис. 2. Гистограмма метрики макро-F1 по 3 моделям 5 кросс-валидационных экспериментов

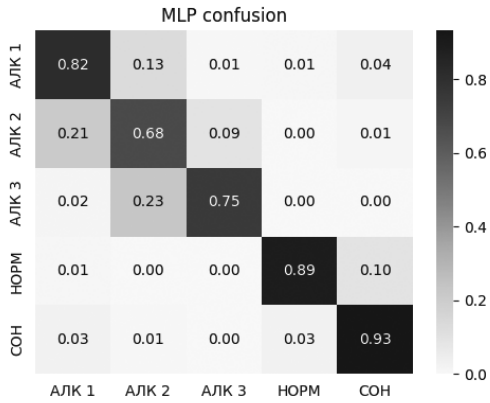

Рис. 3. Матрица ошибок модели MLP при классификации ФС

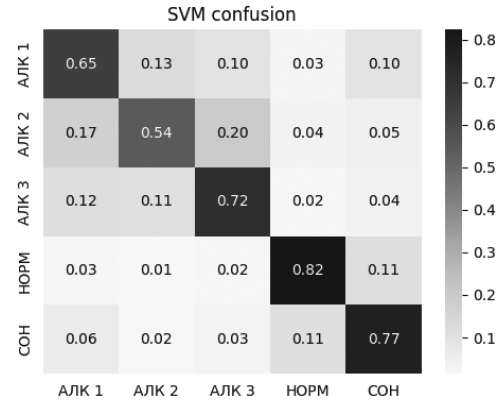

Рис. 4. Матрица ошибок модели SVM при классификации ФС

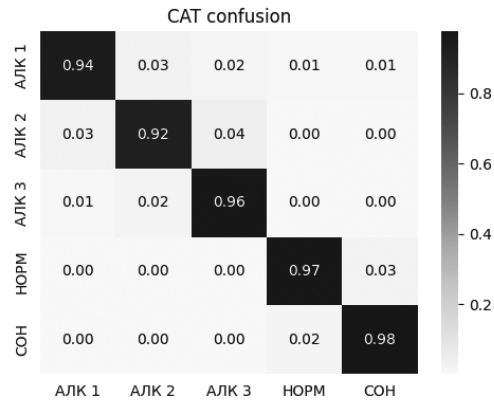

Рис. 5. Матрица ошибок модели CatBoost при классификации ФС

ритм МО CatBoost не допустил ошибок в решениях классификатора между такими состояниями как нормальное и интоксикация 2/3.

Исходя из полученных результатов, можно сделать вывод, что наиболее эффективным оказался подход с использованием объединённого набора признаков (мелкепстральные коэффициенты и высокоуровневые представления) и модели CatBoost, который позволил добиться результата 0.954 по метрике макро-F1.

Для анализа эффективности разработанной модели проведём сравнение с лучшими работами из уже проведённого литературного обзора (таблица 1). В работе [8] на собственной базе данных (голоса мужчин 22–34 лет) с методом случайного леса достигнута точность 95,3% для бинарной классификации (нормальное состояния или алкогольное опьянения) без градаций опьянения или сонного состояния. В [9] на архиве ALC с опорными векторами и случайным лесом получена точность 80%, также для бинарной задачи. Несмотря на различия в используемых базах голосовых образов (собственная в [8], ALC в [9], в настоящей работе – AIC-spkr-130), предложенная в данной работе модель обеспечивает точность 95,6% (сбалансированная точность 95,2%), но с учётом трёх стадий опьянения и сонного состояния, что повышает её возможности и способы применения.

Методика аутентификации пользователя по голосу с учётом его функционального состояния

Нейросетевой преобразователь «биометрия-код» (НПБК) представляет собой технологию, которая преобразует биометрический образ пользователя в бинарный код, используемый для аутентификации или генерации криптографических ключей, обеспечивая при этом защиту исходных биометрических данных от компрометации. В настоящей работе в качестве классификатора, осуществляющего подтверждение личности пользователя, предлагается применить уже апробированный НПБК, предложенный в [9], где используются два типа тригонометрических корреляционных нейронов для обработки голосовых образов, что позволяет генерировать стабильные ключи длиной 1024 бита с низким уровнем ошибок (EER = 2,1%).

НПБК рассматривается как технология, способная превращать биометрический образ в бинарный код. Этот код не является простым решением о доступе (True или False), а представляет собой последовательность бит, которая может служить либо паролем для аутентификации, либо основой для генерации закрытого ключа электронной подписи. Объединяя НПБК с моделью классификации ФС, которая классифицирует состояние пользователя по пяти категориям, мы создаём систему, в которой доступ предоставляется с адаптацией к текущему состоянию пользователя. Предлагаемая методика аутентификации по голосу, основанная на ансамблевых методах МО, позволяет проводить не только верификацию личности, но и выполнять классификацию ФС пользователя, позволяя учитывать его при авторизации. Основная особенность данной методики заключается в возможности ограничения прав доступа пользователя при выявлении у него изменённого функционального состояния для минимизации рисков при работе с наиболее важными данными.

Основная идея комплексирования заключается в последовательном алгоритме: сначала голосовой сигнал поступает в НПБК для генерации бинарного кода и проверки легальности пользователя. Если пользователь признан легальным, сигнал обрабатывается моделью ФС для определения текущего состояния. Если выявленное состояние считается приемлемым (например, нормальное или интоксикация 1), код используется для аутентификации или служит для генерации ключевой пары закрытого ключа. В случае критического состояния, такого как интоксикация 3, система может ограничить или отложить генерацию кода, предлагая альтернативные меры, включая повторную аутентификацию через определённое время или уведомление администратора.

Подобное объединение может быть полезно в системах с дискреционным или мандатным доступом, где права пользователя динамически корректируются. Например, в корпоративной сети при нормальном состоянии пользователь получает полный доступ: чтение, запись, подпись документов. При интоксикации 2 или 3 доступ ограничивается чтением и просмотром, но запись или подпись блокируются, чтобы избежать ошибок под влиянием интоксикации. В сонном состоянии система может предоставить базовый доступ (ограничивающий действия с персональными данными, конфиденциальной информацией и выполнение финансовых или

Рис. 6. Схема алгоритма комплексирования решения НПБК и модели классификации ФС, обученных для генерации бинарного кода для аутентификации

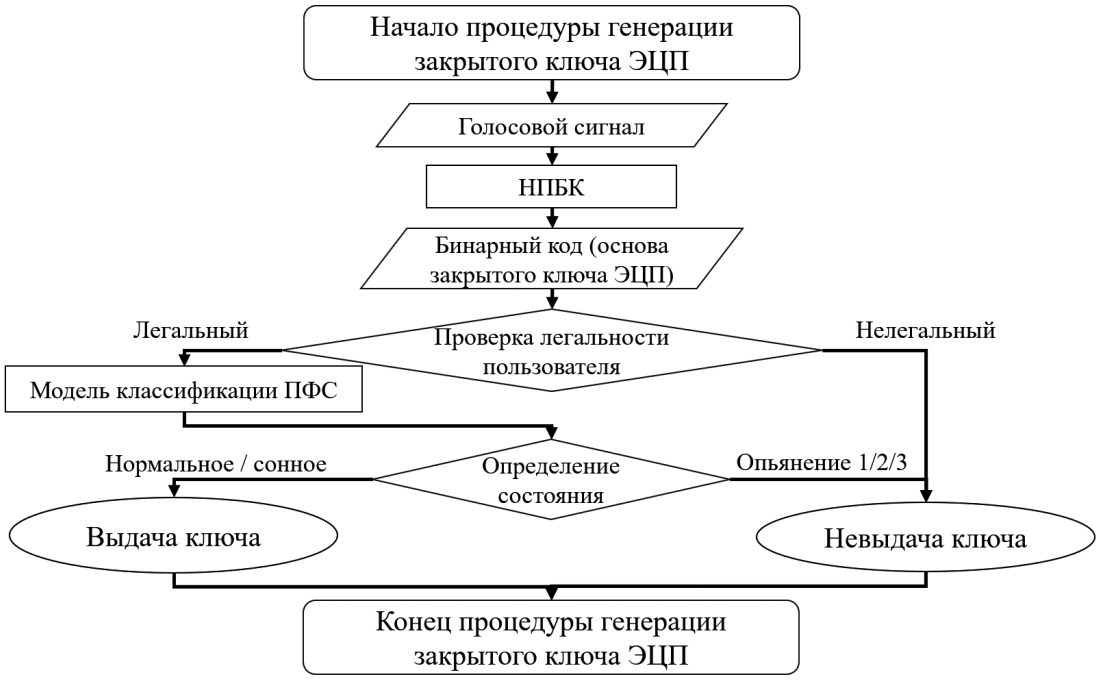

Рис. 7. Схема алгоритма комплексирования решения НПБК и модели классификации ФС, обученных для генерации закрытого ключа электронной подписи

административных операций), но полный только через какое-то число минут (время на "просыпание"), с автоматическим повторным запросом голосового образца. В банковских приложениях, где используется электронная подпись, голос генерирует закрытый ключ только в нормальном состоянии. Подобное ограничение позволит не совершать действий с повышенной ответственностью, например, взятие кредита на крупную сумму удалённо. Похожий способ применения распознавания ФС может быть полезен для приложений по заказу такси. Водитель, перед "выходом на линию" проходит верификацию личности и состояния машины через фотоподтверждение. Прохождение данной верификации с помощью голосового образа позволит не только убедиться в том, что за рулём сам водитель, но и определить его ФС, что очень важно. Такое применение биометрической системы добавляет дополнительный слой безопасности, делая биометрию не только идентификатором личности, но и состояния. На рисунках 6 и 7 предложены алгоритмы регулирования уровня доступа в зависимости от состояния пользователя и функционального назначения НПБК (генерация бинарного кода для прохождения аутентификации или закрытого ключа электронной подписи).

В данном работе приведены только основные способы объединения данных классификаторов. Варианты их совместного использования могут зависеть от специфики применения и прав, предоставляемых пользователю. Так, в информационной системе для образовательных целей, направленной на получение знаний и освоение навыков, может и вовсе не требоваться проверка на ФС, а в оборонных системах, с их мандатным контролем, оно потребует дополнительных механизмов, таких как временные блокировки или логирование попыток в изменённом ФС.

Заключение

В результате настоящего исследования разработана методика аутентификации пользователя по голосу, учитывающая его ФС при выдаче прав доступа, что существенно повышает надёжность и безопасность систем аутентификации по голосу. Предлагаемая методика сочетает нейросетевой преобразователь «биометрия-код» с моделью классификации ФС, что позволяет не только подтвердить личность пользователя, но и адаптировать его права в зависимости от выявленного состояния для предотвращения пользовательских ошибок в критически важных модулях. Разработанная модель классификации ФС, основанная на ансамблевом алгоритме МО CatBoost, достигла высокой точности (метрика макро-F1 0.954), что позволяет различать сразу несколько уровней алкогольной интоксикации, а также нормальное и сонное состояния. Интеграция НПБК с моделью классификации ФС даёт возможность генерировать бинарный код для аутентификации или криптографического ключа, с динамическим ограничением прав в зависимости от состояния.

Литература

1. Иниватов, Д. П. Аналитическое исследование проблемы биометрической идентификации и аутентификации субъектов по голосу / Д. П. Иниватов // Вопросы защиты информации. – 2023. – № 3(142). – С. 28-38. – DOI 10.52190/2073-2600_2023_3_28.

2. Влияние психофизиологического состояния диктора на параметры его голоса и результаты биометрической аутентификации по речевому паролю / А.Е. Сулавко, А.В. Еременко, Р.В. Борисов, Д.П. [и др.] // Международное издание «Компьютерные инструменты в образовании», г. Санкт-Петербург, 2017 г., с. 29-47.

3. Siddiqui N., Pryor L., Dave R. User authentication schemes using machine learning methods—a review //Proceedings of International Conference on Communication and Computational Technologies: ICCCT 2021. – Springer Singapore, 2021. – pp. 703-723.

4. Сулавко А.Е., Иниватов Д.П., Стадников Д.Г., Чобан А.Г. Преобразователь образов голосовых паролей дикторов в криптографический ключ на основе комитета предварительно обученных сверточных нейронных сетей // Вопросы защиты информации. – 2021. - №4. – С. 23-33.

5. Rezvani S. Intoxication detection from audio using deep learning //Bachelor Thesis. – 2019.

6. Schiel F., Heinrich C., Barfüsser S. Alcohol language corpus: the first public corpus of alcoholized German speech //Language resources and evaluation. – 2012. – Т. 46. – №. 3. – С. 503-521.

7. Sigmund M., Zelinka P. Analysis of voiced speech excitation due to alcohol intoxication //Information Technology and Control. – 2011. – vol. 40. – no. 2. – pp. 143-150.

8. Terlapu P. V. Intelligent Novel Approach for Identification of Alcohol Consumers using Incremental Hidden Layer Neurons ANN (IHLN-ANN)-Based Model on Vowelized Voice Dataset. – 2023.

9. Shenoi V. V., Kuchibhotla S., Kotturu P. An efficient state detection of a person by fusion of acoustic and alcoholic features using various classification algorithms //International Journal of Speech Technology. – 2020. – Т. 23. – pp. 625-632.

10. Bonela A. A. et al. Audio-based Deep Learning Algorithm to Identify Alcohol Inebriation (ADLAIA) //Alcohol. – 2023. – Т. 109. – С. 49-54.

11. Wang W., Wu H., Li M. Deep neural networks with batch speaker normalization for intoxicated speech detection //2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). – IEEE, 2019. – С. 1323-1327.

12. Bone D. et al. Intoxicated speech detection by fusion of speaker normalized hierarchical features and GMM supervectors // Twelfth Annual Conference of the International Speech Communication Association. – 2011.

13. Bocklet T., Riedhammer K., Nöth E. Drink and Speak: On the automatic classification of alcohol intoxication by acoustic, prosodic and text-based features //Twelfth Annual Conference of the International Speech Communication Association. 2011.

14. Berninger K., Hoppe J., Milde B. Classification of speaker intoxication using a bidirectional recurrent neural network //Text, Speech, and Dialogue: 19th International Conference, TSD 2016, Brno, Czech Republic, September 12-16, 2016, Proceedings 19. – Springer International Publishing, 2016. – pp. 435-442.

15. Методические указания Министерства здравоохранения СССР от 03.07.1974 «О судебно-медицинской диагностике смертельных отравлений этиловым алкоголем и допускаемых при этом ошибках» [Электронный ресурс] // Законодательство в области судебно-медицинской экспертизы. 1974. URL: http://zakon.forens-med.ru/doc/mz/3_25_101.html (дата обращения: 05.10.2025).

16. Частикова В. А., Жерлицын С. А., Войлова Д. О. Нейросетевая система биометрической идентификации личности по голосу //Вестник Адыгейского государственного университета. Серия 4: Естественно-математические и технические науки. – 2023. – №. 1 (316). – С. 70-79.

17. Badr A. A., Abdul-Hassan A. K. CatBoost Machine Learning Based Feature Selection for Age and Gender Recognition in Short Speech Utterances //International Journal of Intelligent Engineering & Systems. – 2021. – Т. 14. – №. 3.

18. Sulavko A. et al. Biometric-based key generation and user authentication using voice password images and neural fuzzy extractor // Applied System Innovation. 2025.Т. 8. №. 1. С. 13.

References

1. Inivatov, D. P. Analiticheskoe issledovanie problemy biometricheskoj identifikacii i autentifikacii sub#ektov po golosu / D. P. Inivatov // Voprosy zashhity informacii. – 2023. – № 3(142). – S. 28-38. – DOI 10.52190/2073-2600_2023_3_28.

2. Vlijanie psihofiziologicheskogo sostojanija diktora na parametry ego golosa i rezul'taty biometricheskoj autentifikacii po rechevomu parolju / A.E. Sulavko, A.V. Eremenko, R.V. Borisov, D.P. [i dr.] // Mezhdunarodnoe izdanie «Komp'juternye instrumenty v obrazovanii», g. Sankt-Peterburg, 2017 g., s. 29-47.

3. Siddiqui N., Pryor L., Dave R. User authentication schemes using machine learning methods—a review //Proceedings of International Conference on Communication and Computational Technologies: ICCCT 2021. – Springer Singapore, 2021. – pp. 703-723.

4. Sulavko A.E., Inivatov D.P., Stadnikov D.G., Choban A.G. Preobrazovatel' obrazov golosovyh parolej diktorov v kriptograficheskij kljuch na osnove komiteta predvaritel'no obuchennyh svertochnyh nejronnyh setej // Voprosy zashhity informacii. – 2021. - №4. – S. 23-33.

5. Rezvani S. Intoxication detection from audio using deep learning //Bachelor Thesis. – 2019.

6. Schiel F., Heinrich C., Barfüsser S. Alcohol language corpus: the first public corpus of alcoholized German speech //Language resources and evaluation. – 2012. – T. 46. – №. 3. – S. 503-521.

7. Sigmund M., Zelinka P. Analysis of voiced speech excitation due to alcohol intoxication //Information Technology and Control. – 2011. – vol. 40. – no. 2. – pp. 143-150.

8. Terlapu P. V. Intelligent Novel Approach for Identification of Alcohol Consumers using Incremental Hidden Layer Neurons ANN (IHLN-ANN)-Based Model on Vowelized Voice Dataset. – 2023.

9. Shenoi V. V., Kuchibhotla S., Kotturu P. An efficient state detection of a person by fusion of acoustic and alcoholic features using various classification algorithms //International Journal of Speech Technology. – 2020. – T. 23. – pp. 625-632.

10. Bonela A. A. et al. Audio-based Deep Learning Algorithm to Identify Alcohol Inebriation (ADLAIA) //Alcohol. – 2023. – T. 109. – S. 49-54.

11. Wang W., Wu H., Li M. Deep neural networks with batch speaker normalization for intoxicated speech detection //2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). – IEEE, 2019. – S. 1323-1327.

12. Bone D. et al. Intoxicated speech detection by fusion of speaker normalized hierarchical features and GMM supervectors // Twelfth Annual Conference of the International Speech Communication Association. – 2011.

13. Bocklet T., Riedhammer K., Nöth E. Drink and Speak: On the automatic classification of alcohol intoxication by acoustic, prosodic and text-based features //Twelfth Annual Conference of the International Speech Communication Association. 2011.

14. Berninger K., Hoppe J., Milde B. Classification of speaker intoxication using a bidirectional recurrent neural network //Text, Speech, and Dialogue: 19th International Conference, TSD 2016, Brno, Czech Republic, September 12-16, 2016, Proceedings 19. – Springer International Publishing, 2016. – pp. 435-442.

15. Metodicheskie ukazanija Ministerstva zdravoohranenija SSSR ot 03.07.1974 «O sudebno-medicinskoj diagnostike smertel'nyh otravlenij jetilovym alkogolem i dopuskaemyh pri jetom oshibkah» [Jelektronnyj resurs] // Zakonodatel'stvo v oblasti sudebno-medicinskoj jekspertizy. 1974. URL: http://zakon.forens-med.ru/doc/mz/3_25_101.html (data obrashhenija: 05.10.2025).

16. Chastikova V. A., Zherlicyn S. A., Vojlova D. O. Nejrosetevaja sistema biometricheskoj identifikacii lichnosti po golosu //Vestnik Adygejskogo gosudarstvennogo universiteta. Serija 4: Estestvenno-matematicheskie i tehnicheskie nauki. – 2023. – №. 1 (316). – S. 70-79.

17. Badr A. A., Abdul-Hassan A. K. CatBoost Machine Learning Based Feature Selection for Age and Gender Recognition in Short Speech Utterances //International Journal of Intelligent Engineering & Systems. – 2021. – T. 14. – №. 3.

18. Sulavko A. et al. Biometric-based key generation and user authentication using voice password images and neural fuzzy extractor // Applied System Innovation. 2025.T. 8. №. 1. S. 13.

ИНИВАТОВ Даниил Павлович, ассистент кафедры «Комплексная защита информации» федерального государственного автономного образовательного учреждения высшего образования «Омский государственный технический университет». 644050, г. Омск, пр. Мира, 11. E-mail: daniilini@mail.ru

INIVATOV Daniil Pavlovich, Assistant of Integrated data protection department of the Federal State Autonomous Educational Institution of Higher Education "Omsk State Technical University". 11 Mira Ave., Omsk, 644050, Omsk. E-mail: daniilini@mail.ru

Материалы к публикации отправлять по адресу E-mail: urvest@mail.ru
в редакцию журнала «Вестник УрФО. Безопасность в информационной сфере».

Или по почте по адресу: Россия, 454080, г. Челябинск, пр. им. Ленина, д. 76, ЮУрГУ,
Издательский центр

ВЕСТНИК УрФО
Безопасность в информационной сфере № 3(57) / 2025
Подписано в печать 22.10.2025. Дата выхода в свет 22.10.2025.
Формат 70×108 1/16. Печать цифровая. Усл.-печ. л. 7,79. Тираж 50 экз.
Заказ XXX/XXX.
Цена свободная.

Отпечатано в типографии Издательского центра ФГАОУ ВО «ЮУрГУ (НИУ)».
454080, г. Челябинск, пр. им. В. И. Ленина, 76, ЮУрГУ, Издательский центр.

Bulletin of the Ural Federal District
Security in the Sphere of Information No. 3(57) / 2025
Signed to print 22.10.2025. Date of publication of the 22.10.2025.
Format 70×108 1/16. Screen printing. Conventional printed sheet 7,79. Circulation – 50 issues.
Order XXX/XXX.
Open price.

Printed in the printing house of the Publishing Center of FGAOU VO «SUSU (NIU)».
SUSU, Publishing Center, 76, Lenina Str., Chelyabinsk, 454080