

ВОПРОСЫ

КИБЕРБЕЗОПАСНОСТИ

№6²⁰²⁵
(70)

DOI: 10.21681/2311-3456



Процессы деградации систем управления доступом

Уязвимости асимметричных шифров

Квантовая безопасность





СКАНИРОВАНИЕ НА УЯЗВИМОСТИ НИКОГДА НЕ БЫЛО ТАКИМ БЫСТРЫМ!



ГК «Эшелон» представляет новый релиз системы управления уязвимостями Сканер-ВС 7.

Сканер-ВС 7 — система комплексного анализа защищенности, обеспечивающая своевременное выявление уязвимостей в ИТ-инфраструктуре организаций любого масштаба.

Использование Сканер-ВС 7 обеспечивает непрерывный мониторинг защищенности информационных систем, а также повышает эффективность деятельности ИТ-подразделений и служб безопасности.

Сканер-ВС 7 позволяет проводить как специализированные тесты, так и комплексное тестирование защищенности информационных систем, сочетающее сетевые и системные проверки.



Скачать демо-версию Сканер-ВС Base (ограничения: лицензия 4 IP, 2 месяца) можно на сайте продукта.

<https://scanner-vs.ru/>



ВЫСОКАЯ СКОРОСТЬ ПОИСКА УЯЗВИМОСТЕЙ

Высокая скорость поиска уязвимостей с минимальным уровнем ложных срабатываний для Windows, Linux, отечественных ЗОС: Astra Linux и РЕД ОС.



БАЗА УЯЗВИМОСТЕЙ

Ежедневно обновляемая база данных уязвимостей. Источниками данных для формирования агрегированной базы данных уязвимостей являются БДУ ФСТЭК России, NIST NVD, базы уязвимостей вендоров ОС Astra Linux, РЕД ОС, Ubuntu, RedHat, Windows и др.



КРОССПЛАТФОРМЕННОСТЬ

Поддерживаемые среды функционирования: Astra Linux SE 1.6 - 1.8, РЕД ОС 8, «Альт Рабочая станция 10», Windows 10 / 11 / Server 2022 (WSL).



ОНЛАЙН ДОКУМЕНТАЦИЯ

Информационный портал с технической документацией и обучающими материалами



СЕРТИФИКАТЫ

ФСТЭК России №2204 (4УД, ТУ)
Минобороны России №7136

ВОПРОСЫ КИБЕРБЕЗОПАСНОСТИ

НАУЧНЫЙ РЕЦЕНЗИРУЕМЫЙ ЖУРНАЛ

№ 6 (70) 2025 г.

Выходит 6 раз в год

Журнал выходит с 2013 г. (Свидетельство о регистрации ПИ № ФС77-75239). Перерегистрировано Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций 07.03.2019.

Журнал входит в рейтинг научных изданий ВАК в категории К1, индексируется в RSCI, публикует статьи по специальностям 1.2.4 и 2.3.6 – физ.-мат. науки; 2.2.15, 2.3.1, 2.3.5, 2.3.6 – техн. науки

Главный редактор

МАРКОВ Алексей Сергеевич, д. т. н., с. н. с., Москва

Председатель Редакционного совета

ШЕРЕМЕТ Игорь Анатольевич, академик РАН, д. т. н., профессор, Москва

Шеф-редактор

МАКАРЕНКО Григорий Иванович, с. н. с., шеф-редактор, Москва

Редакционный совет

БАСАРАБ Михаил Алексеевич, д. ф.-м. н., Москва

КАЛАШНИКОВ Андрей Олегович, д. т. н., Москва

КРУГЛИКОВ Сергей Владимирович, д. в. н., к. т. н., профессор, Минск, Беларусь

ПЕТРЕНКО Сергей Анатольевич, д. т. н., профессор, Сириус

СТАРДУБЦЕВ Юрий Иванович, д. в. н., профессор, Санкт-Петербург

ЯЗОВ Юрий Константинович, д. т. н., профессор, Воронеж

Редакционная коллегия

БАБЕНКО Людмила Климентьевна, д. т. н., профессор, Таганрог

БАРАНОВ Александр Павлович, д. ф.-м. н., профессор, Москва

ГАРБУК Сергей Владимирович, к. т. н., с. н. с., Москва

ГАЦЕНКО Олег Юрьевич, д. т. н., с. н. с., Санкт-Петербург

ЗЕГЖДА Дмитрий Петрович, член-корреспондент РАН, д. т. н., профессор, Санкт-Петербург

ЗУБАРЕВ Игорь Витальевич, к. т. н., доцент, Москва

КОЗАЧОК Александр Васильевич, д. т. н., Орел

МАКСИМОВ Роман Викторович, д. т. н., профессор, Краснодар

ПАНЧЕНКО Владислав Яковлевич, академик РАН, д. ф.-м. н., профессор, Москва

ПУДОВКИНА Марина Александровна, д. ф.-м. н., профессор, Москва

ЦИРЛОВ Валентин Леонидович, к. т. н., доцент, Москва

ШАХАЛОВ Игорь Юрьевич, ответственный секретарь, Москва

ШЕЛУПАНОВ Александр Александрович, член-корреспондент РАН, д. т. н., профессор, Томск

ШУБИНСКИЙ Игорь Борисович, д. т. н., профессор, Москва

Учредитель и издатель

АО «Научно-производственное объединение «Эшелон»

Над номером работали:

Г. И. Макаренко – шеф-редактор, И. Ю. Шахалов – отв. секретарь, С. С. Игнатов – верстка, Ю. С. Логинова – зам. главного редактора

Подписано к печати 15.12.2025 г.

Общий тираж 120 экз. Цена свободная

Адрес: 107023, Москва, ул. Электrozаводская, д. 24, стр. 1.

E-mail: editor@cyberrus.info, тел.: +7 (985) 939-75-01.

Требования, предъявляемые к рукописям, размещены на сайте: <https://cyberrus.info/>

Подписка на журнал осуществляется в почтовых отделениях по каталогу «Пресса России». Подписной индекс 40707

СОДЕРЖАНИЕ

УПРАВЛЕНИЕ РИСКАМИ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

АНАЛИЗ ПРОЦЕССОВ ДЕГРАДАЦИИ СИСТЕМ УПРАВЛЕНИЯ ДОСТУПОМ, ОСНОВАННЫХ НА МАНДАТНЫХ МОДЕЛЯХ С ВОЗМОЖНОСТЬЮ «ЗАПИСИ ВНИЗ»

Гайдамакин Н. А. 2

МЕТОДЫ И СРЕДСТВА АНАЛИЗА ЗАЩИЩЕННОСТИ

МНОГОУРОВНЕВЫЙ ФРЕЙМВОРК ОБОСНОВАНИЯ ПРОЦЕДУР МОНИТОРИНГА И РЕАГИРОВАНИЯ НА ИНЦИДЕНТЫ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

Грызунов В. В., Шестаков А. В. 14

ПРИМЕНЕНИЕ МЕТОДА АНАЛИЗА ИЕРАРХИЙ И МЕТОДА АНАЛИТИЧЕСКИХ СЕТЕЙ ДЛЯ ОЦЕНКИ ПРЕДПОЧТИТЕЛЬНОСТИ СИСТЕМ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

Аносов Р. С., Аносов С. С. 25

УЯЗВИМОСТИ АСИММЕТРИЧНЫХ ШИФРОВ БЛОКЧЕЙН-ПЛАТФОРМ

Ищукова Е. А., Петренко С. А., Леонтьева Ю. П. 35

СЕТЕВАЯ БЕЗОПАСНОСТЬ

ИСПОЛЬЗОВАНИЯ УЯЗВИМОСТЕЙ ПРОТОКОЛА WEBAUTHN ДЛЯ ПОЛУЧЕНИЯ НЕСАНКЦИОНИРОВАННОГО ДОСТУПА

Панченко А. Р. 48

МЕТОДЫ И СРЕДСТВА АНАЛИЗА ЗАЩИЩЕННОСТИ

ОЦЕНКА ЗАЩИЩЕННОСТИ ACTIVE DIRECTORY С ИСПОЛЬЗОВАНИЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Булгакова Е. В., Богданов Е. А., Кубанков А. Н. 58

ТЕСТИРОВАНИЕ И МОНИТОРИНГ КИБЕРБЕЗОПАСНОСТИ

ПРОБЛЕМНО-ОРИЕНТИРОВАННАЯ СИСТЕМА МОНИТОРИНГА И РЕАГИРОВАНИЯ НА МНОВЕКТОРНЫЕ АТАКИ В ДЕЦЕНТРАЛИЗОВАННОЙ СРЕДЕ ИНТЕРНЕТА ВЕЩЕЙ

Тебуева Ф. Б., Петренко В. И., Сатыбалдина Д. Ж., Огур М. Г., Гусева Т. М. 69

КВАНТОВАЯ БЕЗОПАСНОСТЬ

АЛГОРИТМ ОБНАРУЖЕНИЯ СИГНАЛА СИНХРОНИЗАЦИИ В КВАНТОВЫХ СЕТЯХ

Плёткин А. П. 81

ТЕХНИЧЕСКОЕ РЕГУЛИРОВАНИЕ ОБЛАСТИ БЕЗОПАСНОСТИ

СХЕМА КРИПТОГРАФИЧЕСКОЙ ЗАЩИТЫ ДАННЫХ ОБЛАЧНЫХ ХРАНИЛИЩ С ВОЗМОЖНОСТЬЮ РАСШИРЕНИЯ КВАНТОВОЗАВИСИМЫМИ КЛЮЧАМИ И ПОСТКВАНТОВЫМИ АЛГОРИТМАМИ

Минаков С. С., Тихов С. В., Цупак А. А. 88

НОРМАТИВНО-ТЕХНИЧЕСКИЕ АСПЕКТЫ ОБЕСПЕЧЕНИЯ ИНФОРМАЦИОННОЙ И КИБЕРБЕЗОПАСНОСТИ ОБЪЕКТОВ ЦИФРОВОЙ ЭЛЕКТРОЭНЕРГЕТИКИ

Куликов А. Л., Лоскутов А. А., Зинин В. М. 101

ИДЕНТИФИКАЦИЯ И АУТЕНТИФИКАЦИЯ

МЕТОД РАСПОЗНАВАНИЯ ЛИЧНОСТИ НА ОСНОВЕ БИОАКУСТИЧЕСКОГО СКАНИРОВАНИЯ В СИСТЕМАХ МНОГОФАКТОРНОЙ АУТЕНТИФИКАЦИИ

Частикова В. А., Гуляй В. Г. 116

БЕЗОПАСНЫЙ ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

МЕТОД ОБЕСПЕЧЕНИЯ КИБЕРУСТОЙЧИВОСТИ БЛОКЧЕЙН-ПЛАТФОРМ НА ОСНОВЕ КИБЕРИММУНИТЕТА

Балябин А. А., Петренко С. А. 127

ИССЛЕДОВАНИЕ СИГНАЛЬНЫХ МЕТОДОВ ВЫЯВЛЕНИЯ СОСТАЯТЕЛЬНЫХ АТАК НА НЕЙРОННЫЕ МОДЕЛИ ЧЕРЕЗ ОБЪЯСНИТЕЛЬНУЮ ВИЗУАЛИЗАЦИЮ

Чеколаев Д. А., Шевченко А. В., Алексеев А. К., Трофимов Ю. В., Аверкин А. Н. 140

ПОДХОД К ОЦЕНКЕ КАЧЕСТВА ГЕНЕРАЦИИ СЦЕНАРИЕВ ПЕНТЕСТА ПРИ ПОМОЩИ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ

Конев А. А., Паюсова Т. И. 147

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В ВОЕННОМ ДЕЛЕ: ВОЗМОЖНОСТИ, УГРОЗЫ, ПЕРСПЕКТИВЫ

Ромашкина Н. П. 158

МАШИННОЕ ОБУЧЕНИЕ

ЭКСПЕРИМЕНТ ПО ПРОГНОЗИРОВАНИЮ ОТКАЗОВ УСТРОЙСТВА ПРОМЫШЛЕННОГО ИНТЕРНЕТА ВЕЩЕЙ НА БАЗЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

Марков Г. А. 166

АНАЛИЗ ПРОЦЕССОВ ДЕГРАДАЦИИ СИСТЕМ УПРАВЛЕНИЯ ДОСТУПОМ, ОСНОВАННЫХ НА МАНДАТНЫХ МОДЕЛЯХ С ВОЗМОЖНОСТЬЮ «ЗАПИСИ ВНИЗ»

Гайдамакин Н. А.¹

DOI: 10.21681/2311-3456-2025-6-2-13

Цель статьи: провести теоретический анализ функционирования мандатных систем управления доступом, в которых с целью расширения функциональных возможностей по обработке информации субъектам доступа разрешается осуществление «записей вниз».

Метод исследования: применение аппарата пуассоновских потоков событий для формализации и анализа функционирования мандатных систем управления доступом с возможностью «записей вниз».

Результаты исследования: процесс функционирования мандатной системы управления доступом с возможностью «записей вниз» рассматривается как пуассоновский поток однородных событий возникновения и реализации высокоуровневыми субъектами доступа (пользователями с высоким уровнем доступа) потребностей записей информации в низкоуровневые объекты доступа (файлы, документы с низким уровнем конфиденциальности). В результате каждого такого события после завершения «записи вниз», чтобы не допустить в дальнейшем возможного ознакомления низкоуровневых пользователей с высокоуровневой информацией, уровень конфиденциальности объекта доступа повышается до уровня доступа соответствующего (записывающего) субъекта. Вводится понятие деградации системы управления доступом, которое заключается в том, что при двух уровнях доступа и конфиденциальности (высокий и низкий) со временем все объекты доступа по уровню конфиденциальности становятся высокоуровневыми, и соответственно недоступными по «чтению» для субъектов (пользователей) с низким уровнем доступа.

Анализируется вероятность наступления деградации системы управления доступом в случае стационарности и нестационарности пуассоновского потока событий «записей вниз». При убывающей со временем интенсивности событий «записей вниз» рассматривается необычный на первый взгляд характер изменения со временем вероятности деградации, который можно назвать эффектом «послемаксимового затухания», заключающийся сначала в увеличении со временем вероятности деградации с достижением в определенный момент времени некоторого максимального значения, а затем в постепенном уменьшении вероятности деградации до нулевого значения. При линейном и экспоненциальном характере убывания интенсивности потока событий «записей вниз» получены соотношения для момента времени с максимальной вероятностью деградации, который определяется величиной скорости уменьшения интенсивности потока событий.

Научная новизна: введено и рассмотрено понятие «деградации» мандатных систем управления доступом с возможностью «записей вниз». Процесс функционирования мандатных систем управления доступом с возможностью «записей вниз» в контексте анализа деградации рассмотрен как пуассоновский поток однородных событий. Проанализирован т.н. «эффект «послемаксимового затухания» вероятности деградации при убывающей со временем интенсивности потока событий «записей вниз».

Ключевые слова: мандатный принцип управления доступом, «запись вниз», поток событий «записей вниз», деградация системы управления доступом, эффект «послемаксимового затухания».

Введение

Политика мандатного управления доступом к конфиденциальной информации, регламентированная для применения в компьютерных системах государственных организаций, основывается на исходной формальной модели Д. Белла и Л. ЛаПадулы [1]. Принцип управления доступом в модели Белла ЛаПадулы выражается двумя правилами – «запрет чтения вверх» и «запрет записи вниз», что обеспечивает математически доказуемую невозможность ознакомления пользователем с низким уровнем доступа

(низким уровнем допуска-доверия) с информацией высокого уровня конфиденциальности.

Вместе с тем, «запрет записи вниз» существенно снижает функциональные возможности по обработке информации в компьютерных системах, означая принципиальную невозможность записи (корректировки, редактирования) пользователем с высоким уровнем доступа (например, руководителем организации) в объекты (документы) с низким уровнем конфиденциальности.

¹ Гайдамакин Николай Александрович, доктор технических наук, профессор, профессор кафедры алгебры и фундаментальной информатики Уральского федерального университета имени первого Президента России Б. Н. Ельцина. г. Екатеринбург, Россия. E-mail: n.a.gaidamakin@urfu.ru

Для преодоления данного ограничения был предложен подход [2], иногда не совсем правильно именуемый моделью «Low-Water-Mark»², заключающийся во введении в модель Белла ЛаПадулы дополнительной команды `reset`, которая «поднимает» уровень конфиденциальности объекта доступа до максимального и тем самым становится возможной запись в него любым пользователем, в т.ч. с самым высоким уровнем доступа.

Вместе с тем, нетрудно увидеть, что в практической реализации данный подход может приводить к переходу со временем всех низкоуровневых объектов доступа в высокоуровневые и в результате становится невозможным доступ к ним по чтению для низкоуровневых пользователей, что, в свою очередь, приводит также к существенному ограничению функциональности систем обработки информации. Подобные особенности функционирования мандатных систем управления доступом с возможностью «записей вниз» можно именовать тенденцией к деградации.

С учетом отмеченного выше широкого применения мандатных систем управления доступом и стремления в создании не только безопасных, но и полнофункциональных современных компьютерных систем, представляется актуальной задача теоретического анализа отмеченной тенденции деградации при возможности «записи вниз».

Исходные положения

Как и в классической модели Белла ЛаПадулы будем основываться на субъектно-объектной теоретико-множественной формализации, в рамках которой в контексте анализа безопасности информации компьютерная система представляется совокупностью сущностей двух видов – субъектов доступа $s \in S$ и объектов доступа $o \in O$, множество которых $S \cup O$ отражено на линейно упорядоченное множество уровней безопасности L . В каждый момент времени любой экземпляр $x \in (S \cup O)$ сущностей компьютерной системы (субъект или объект доступа) характеризуется параметром $f_L(x)$, который является элементом множества L , т.е. характеризует определенный уровень безопасности $l \in L$ соответствующего элемента множества $S \cup O$. Уровень безопасности субъектов является аналогом степени (уровня) допуска сотрудников (пользователей), уровень безопасности объектов является аналогом уровня (грифа) конфиденциальности документов.

2 Следует отметить, что в большинстве работ по формальным моделям безопасности (см., например, [3, 4, 5]) под термином «Low-Water-Mark» понимается принцип присвоения наиболее критичным в аспекте нарушения целостности сущностям (объектам и субъектам доступа) наиболее низких уровней в решетке безопасности, трактуемой как решетка уровней конфиденциальности [6]. Данный принцип, как и в целом известная мандатная модель целостности К. Биби [6], является инверсией принципа «High-Water-Mark», при котором наиболее критичным в аспекте конфиденциальности сущностям присваиваются наиболее высокие уровни в решетке безопасности [3, 7].

Критерием безопасности доступов субъектов к объектам является запрет потоков информации «сверху-вниз», т.е. от сущностей с высоким уровнем безопасности к сущностям с более низким уровнем безопасности. Потоки «сверху-вниз» прямо (при чтении высокоуровневого объекта низкоуровневым субъектом) или транзитивно (при записи высокоуровневым субъектом в низкоуровневый объект с последующим его чтением низкоуровневым субъектом) приводят к нарушению одного из 3-х составляющих безопасности информации компьютерных систем – нарушению конфиденциальности.

Соответственно доступ на чтение безопасен тогда и только тогда, когда уровень безопасности субъекта равен или превышает уровень безопасности объекта $f_L(s) \geq f_L(o)$ – т.н. правило NRU (no read up, нельзя читать вверх). Отметим, что в случае $f_L(s) \geq f_L(o)$ реализуется или безопасный «горизонтальный» поток (между равноуровневыми сущностями) или безопасный поток «снизу-вверх», т.е. из низкоуровневого объекта к высокоуровневому субъекту.

Доступ по записи безопасен тогда и только тогда, когда уровень безопасности субъекта равен или меньше уровня безопасности объекта $f_L(s) \leq f_L(o)$ – т.н. правило NWD (no write down, нельзя писать вниз). В этом случае также реализуется безопасный «горизонтальный» поток или безопасный поток «снизу-вверх» от низкоуровневого субъекта в высокоуровневый объект.

В практической реализации модели Белла – ЛаПадулы в операционных системах и СУБД применяется линейная решетка конечного набора уровней безопасности (обычно 3-4 уровня). При этом любая пара «субъект–объект» доступа сравнима по параметру $f_L(x)$. В этом случае множество L можно рассматривать как некую порядковую шкалу безопасности сущностей компьютерной системы.

Вместе с тем использование линейной решетки с небольшим числом уровней безопасности приводит к существенному огрублению доступа, в т.ч. к избыточности прав доступа. Иначе говоря, по правилам NRU и NWD субъекту с определенным уровнем безопасности доступны по чтению и записи любые объекты с таким же уровнем безопасности, и, кроме того, только по чтению все субъекты с более низким, а по записи – все объекты с более высоким уровнем безопасности. Это вынуждает вносить в мандатную модель механизмы дискреционного разграничения доступа посредством дополнительного введения матрицы доступа $A[s, o]$. Строки матрицы доступа $A[s, o]$ соответствуют субъектам доступа (пользователям), столбцы объектам доступа, а в ячейках фиксируются разрешения доступа соответствующих субъектов к соответствующим объектам. В результате в рамках

мандатных полномочий пользователей для чтения $f_L(s) \geq f_L(o)$ и для записи $f_L(s) \geq f_L(o)$ (необходимое условие) доступ субъекта к объекту возможен только при наличии непосредственного разрешения в соответствующей ячейке матрицы доступа (достаточное условие).

Состояние компьютерной системы $V\{L(S), L(O), A[s, o]\}$ определяется совокупностью текущих значений уровней безопасности субъектов $f_L(s) \in L(S)$ и объектов доступа $f_L(o) \in L(O)$, значений ячеек матрицы доступа $A[s, o]$, а также совокупностью доступов, осуществляемых субъектами к объектам. Изменения значений $f_L(s)$, $f_L(o)$ и $A[s, o]$ трактуется как переход системы в новое состояние V^* , что, в свою очередь, обеспечивает новую совокупность доступов субъектов к объектам.

Д. Э. Белл и Л. ЛаПадула доказали т.н. «основную теорему безопасности» (Basic Security Theorem), суть которой заключается в том, что, если изменения значений $f_L(s)$ и $f_L(o)$ приводят в новом состоянии к новым мандатным полномочиям определенных субъектов к определенным объектам, то адекватно должны изменяться соответствующие ячейки матрицы доступа $A[s, o]$, и наоборот. В теореме доказывается, что при таких условиях в любом произвольном состоянии системы V гарантированно отсутствуют доступы, нарушающие правила NRU и NWD, т.е. система функционирует безопасно в смысле критерия недопущения потоков «сверху-вниз», приводящих к нарушению конфиденциальности информации объектов доступа.

Трудно переоценить методологическую роль, которую сыграла модель Белла ЛаПадулы в теории и практике защищенных компьютерных систем. Была математически доказана теоретическая безопасность компьютерных систем, в которых доступ осуществляется по мандатным правилам, и, что, может быть не менее важно в практическом плане, на основе модели обеспечивалась возможность воспроизведения в компьютерных системах, установленных для «бумажных» технологий нормативных «допусковых» (т.е. «мандатных») регламентаций работы с конфиденциальной информацией.

Вместе с тем, как часто бывает в отношении новых подходов и методов, а также из-за определенной абстрактности и схоластичности модели Белла ЛаПадулы, были представлены многочисленные работы по ее критическому анализу [8, 9].

Прежде всего, следует отметить, что полный запрет всех информационных потоков «сверху-вниз» оправдан далеко не всегда, поскольку основывается на неявном предположении, что любой информационный элемент (слово, запись, символ, абзац, страница) документа, т.е. объекта доступа, является

конфиденциальным соответствующего уровня. Очевидно, что во многих случаях часть информации документа является конфиденциальной, а другая нет или конфиденциальной меньшего уровня. Ввиду этого на практике, сотруднику, не допущенному к документам определенного уровня конфиденциальности, может доводиться из таких документов только та часть информации, те выдержки, которые по уровню конфиденциальности соответствуют уровню допуска сотрудника.

Аналогично не вся информация, записываемая в какой-либо объект субъектом доступа, является конфиденциальной уровня безопасности субъекта. Сотрудник может вносить (записывать) данные в рабочую тетрадь или в [разрабатываемый] документ, уровень конфиденциальности которых ниже его уровня допуска, при условии того, что вносимые данные будут иметь соответствующий грифу документа уровень конфиденциальности.

Другим недостатком модели Белла ЛаПадулы, который исследовался еще в работах Дж. МакЛина [8, 9], является отсутствие механизмов и процедур установления и изменения уровней безопасности субъектов и объектов доступа. Дж. МакЛином были введены понятия авторизованной функции перехода и уполномоченных субъектов (для изменения уровней безопасности субъектов или объектов доступа), что существенно снизило схоластичность исходной модели Белла ЛаПадулы и обеспечило условия ее практической реализации в компьютерных системах.

Анализируя вслед за Дж. МакЛином процессы изменения уровней безопасности объектов доступа в мандатных системах, следует отметить еще один существенный недостаток базовой модели Белла-ЛаПадулы, который заключается в отсутствии временного аспекта в понятии «доступ субъекта к объекту» [10]. Иначе говоря, процессы доступа субъектов к объектам рассматриваются как одномоментные («мгновенные») явления, переводящие систему в новое состояние. Таким образом, доступы «не длятся», и соответственно в промежутках между «мгновенно» (точнее вне временного аспекта) осуществленными доступами «ничего не происходит».

В результате системы управления доступом на основе модели Белла-ЛаПадулы представляют собой системы с дискретным временем [10].

Будем считать, что в компьютерной системе действует дискретное время, в каждый момент которого t_i состояние системы V_i характеризуется определенной декомпозицией на множество субъектов S , помеченных соответствующими метками (уровнями) безопасности и осуществляющих доступы на множестве объектов O . В результате доступов формируются информационные потоки и, в т.ч. могут изменяться

уровни безопасности объектов, что переводит в момент времени t_{i+1} систему в новое состояние $V_{t_{i+1}}$. Соответственно уровни безопасности субъектов и объектов доступа в различные моменты времени будем индексировать верхними временными индексами – $f_L^{(t_i)}(s)$, $f_L^{(t_i)}(o)$.

Как отмечалось выше, в рамках мандатного управления доступом в некоторых случаях субъекту s необходимо внести информацию (осуществить запись, т.е. доступ по записи) в объект o , уровень безопасности которого ниже его уровня безопасности – $f_L^{(t_i)}(s) > f_L^{(t_i)}(o)$. В этом случае в целях обеспечения критерия безопасности (недопущение потоков «сверху-вниз») субъект s сначала подает команду $reset(o)$, в результате которой уровень безопасности объекта o поднимается до максимального, и, соответственно, по правилам NRU и NWD становится для субъекта s доступным по записи. По завершению записи, т.е. в новом состоянии $V_{t_{i+1}}$, если уровень безопасности субъекта s не является максимальным, уровень безопасности объекта o снижается до уровня записывающего субъекта $f_L^{(t_{i+1})}(o) = f_L^{(t_{i+1})}(s)$.

Для упрощения анализа будем считать систему статичной по декомпозиции на субъекты доступа. Иначе говоря, количество M субъектов доступа (s_1, s_2, \dots, s_M) и их декомпозиция на группы по уровням безопасности не изменяются во времени. Будем также считать систему статичной по количеству объектов доступа – $N_L^{(t_i)} = N_L^{(t_{i+1})}$, но их декомпозиция на группы по уровням безопасности (грифам конфиденциальности), напротив, может изменяться при осуществлении высокоуровневыми субъектами «записей вниз», по завершению которых низкоуровневые объекты становятся высокоуровневыми. Таким образом исходное количество низкоуровневых объектов в мандатной системе управления доступом с возможностью «записей вниз» с течением времени сокращается.

Также для простоты будем предполагать, что в системе управления доступом два уровня безопасности – низкий l_1 и высокий l_2 ($l_1 < l_2$). Соответственно $N_{l_1}^{(t_0)}$ объектов доступа в начальный момент времени t_0 помечены нижним уровнем безопасности l_1 , а $N_{l_2}^{(t_0)}$ верхним уровнем безопасности l_2 , при этом $N_{l_1}^{(t_0)} + N_{l_2}^{(t_0)} = N$.

Определим понятие состояния деградации системы управления доступом как переход в состояние V_{Degrad} при котором $N_{l_2}^{(t_{Degrad})} = N$, $N_{l_1}^{(t_{Degrad})} = 0$.

Также будем полагать, что уровни безопасности субъектов доступа при переходах системы из одного состояния в другое не меняются $f_L^{(t_i)}(s) = f_L^{(t_{i+1})}(s)$, $i = 0, 1, 2, \dots$

Кроме того, будем полагать, что субъект s с уровнем безопасности l_2 при осуществлении доступа

по записи может вносить в объект доступа o информацию как уровня l_2 , так и уровня l_1 .

Деградация мандатной системы управления доступом как результат простейшего потока однородных событий «записей вниз»

Переход объектов с уровня безопасности l_1 на уровень l_2 происходит в том случае, когда у субъектов с уровнем безопасности l_2 в какой-либо момент времени t_i возникает необходимость записи информации уровня l_1 в объект уровня безопасности $f_L^{(t_i)}(o) = l_1$ («запись вниз»).

Возникновение и осуществление такой потребности «записи вниз» будем трактовать как некое событие. Очевидно, такие события могут происходить в любой момент времени $t_0, t_1, t_2, \dots, t_i$, и в результате мы имеем поток однородных событий, результатом каждого из которых является уменьшение на единицу количества объектов уровня l_1 , что, собственно, и составляет тенденцию к деградации системы.

Если данный поток событий детерминированный, и соответствующее событие с вероятностью, равной 1, возникает в каждый следующий момент времени t_0, t_1, t_2, \dots , то деградация системы наступит в момент времени $t_i = t_{N_{l_1}^{(t_0)}}$, т.е. $i = N_{l_1}^{(t_0)}$.

Если поток событий случайный, то он может быть простейшим потоком [однородных] событий, который характеризуется свойствами ординарности, стационарности и отсутствия последствия.

В нашем случае ординарность будет означать, что в любой момент времени $t_0, t_1, t_2, \dots, t_i$ потребность записи информации уровня l_1 возникает и реализуется только у какого-либо одного субъекта s уровня l_2 в отношении какого-либо одного объекта o уровня l_1 . Возникновение и осуществление аналогичной потребности «записи вниз» у другого пользователя уровня l_2 или в отношении другого объекта уровня l_1 трактуется как следующий (другой) момент дискретного времени функционирования системы управления доступом.

Стационарность и отсутствие последствия будет означать, что вероятность наступления соответствующей потребности не зависит от момента времени t_i и от количества состоявшихся к этому времени k таких событий.

Важно отметить, что свойство отсутствия последствия также обусловлено указанным выше отсутствием временного аспекта доступов в исходной модели мандатного доступа Белла–ЛаПадулы. Иначе говоря, в системе управления доступом обработка событий «записей вниз» происходит «мгновенно» (точнее вне временного аспекта) и поэтому не влияет на поступление (возникновение) и обработку последующих событий.

Простейший пуассоновский поток однородных событий характеризуется постоянной интенсивностью λ возникновения соответствующих событий, определяемой как среднее число событий в единицу времени, т.е. на интервале от t_{i-1} до t_i .

Очевидно, что при гипотезе об ординарности соответствующего потока событий в отмеченном выше смысле $0 < \lambda \leq 1$.

При пуассоновском потоке вероятность появления за i шагов точно k событий определяется известным выражением [11]:

$$P_{t_i}(k) = e^{-\lambda i} \frac{(\lambda i)^k}{k!}. \quad (1)$$

Для анализа процесса деградации удобнее рассмотреть вероятность того, что за i шагов, т.е. на промежутке времени от t_0 до t_i , произойдет не менее $N_{l_1}^{(t_0)}$ соответствующих событий, т.е. ($k \geq N_{l_1}^{(t_0)}$). Противоположное событие ($k < N_{l_1}^{(t_0)}$) является суммой событий, каждое из которых заключается в том, что на промежутке от t_0 до t_i произойдет только точно 0 событий, или только точно 1 событие, или только точно 2 события, или ... только точно $N_{l_1}^{(t_0)} - 1$ событие. Тогда для простейшего пуассоновского потока событий вероятность события $P_{t_i}(k \geq N_{l_1}^{(t_0)})$ определяется следующим соотношением:

$$P_{t_i}(k \geq N_{l_1}^{(t_0)}) = 1 - e^{-\lambda i} \sum_{m=0}^{N_{l_1}^{(t_0)}-1} \frac{(\lambda i)^m}{m!}. \quad (2)$$

На рис. 1. представлены графики вероятностей $P_{t_i}(k \geq N_{l_1}^{(t_0)})$, отражающих процесс деградации мандатной системы управления доступом при гипотезе о простейшем пуассоновском потоке событий «записей вниз» для $N_{l_1}^{(t_0)} = 10$, $N_{l_1}^{(t_0)} = 20$, $\lambda = 1$, $\lambda = 0,5$, $\lambda = 0,1$, соответственно.

Как видно из приведенных на рисунке графиков, процесс деградации при гипотезе о простейшем пуассоновском потоке отражает традиционный вид интегральных функций распределения вероятностей, характеризующий в данном случае то, что с увеличением интервала наблюдения (момента-шага времени t_i) вероятность превышения количества (k) произошедших событий «записей вниз» величины $N_{l_1}^{(t_0)}$ стремится к единице. Иначе говоря, в случае простейшего пуассоновского потока событий «записей вниз» деградация системы управления доступом в любом случае произойдет.

Когда (как скоро) наступит деградация – естественным образом зависит от двух параметров, исходного количества низкоуровневых объектов $N_{l_1}^{(t_0)}$ и интенсивности потока событий λ . Очевидно, что значения этих параметров определяются особенностями предметной области компьютерной системы, точнее говоря, особенностями информации (тематическими, технологическими, финансовыми, политическими и т.д.) низкоуровневых объектов и функциональными особенностями субъектов доступа (процессов пользователей). Определение параметра λ может осуществляться эмпирическим путем на основе анализа статистики возникновения событий «записей вниз».

Деградация мандатной системы управления доступом как результат нестационарного потока однородных событий «записей вниз»

Рассмотренный выше поток однородных событий «записей вниз» является стационарным. Вместе с тем, очевидно в определенных предметных областях у пользователей-субъектов доступа потребность «записей вниз» может изменяться с течением времени – уменьшаться или, наоборот, увеличиваться.

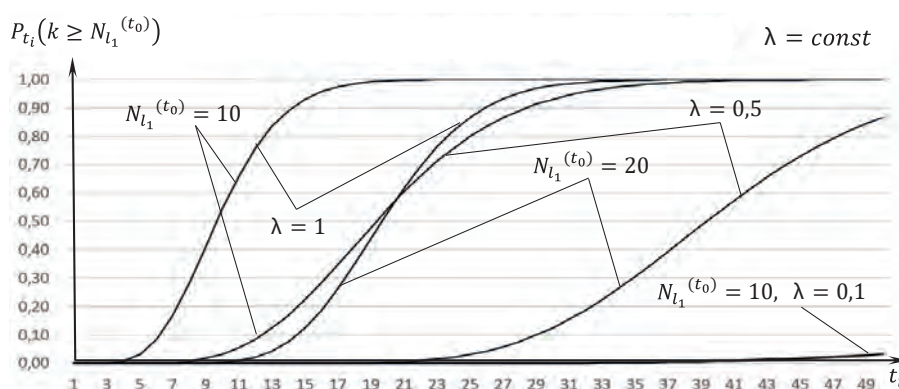


Рис. 1. Зависимость вероятности деградации мандатной системы управления доступом от времени при гипотезе о простейшем пуассоновском потоке событий «записей вниз».

($N_{l_1}^{(t_0)}$ – количество низкоуровневых объектов в начальный момент времени; k – количество произошедших событий «записей вниз» к моменту-шагу времени t_i , т.е. количество низкоуровневых объектов, перешедших к моменту времени t_i в высокоуровневые; λ – интенсивность потока событий, т.е. среднее количество соответствующих событий в единицу времени).

В результате интенсивность (параметр λ) потока событий в таких системах не является константой и изменяется с течением времени.

Оставаясь в предположении об ординарности и отсутствии последствия потока событий «записей вниз» приходим к нестационарному пуассоновскому потоку событий, в котором вероятность деградации системы управления доступом с возможностью «записей вниз» представляется следующим соотношением

$$P_{t_i}(k \geq N_{t_i}^{(t_0)}) = 1 - e^{-\lambda(t_i)t_i} \sum_{m=0}^{N_{t_i}^{(t_0)}-1} \frac{(\lambda(t_i)t_i)^m}{m!}. \quad (3)$$

Характер зависимости интенсивности потока событий $\lambda(t_i)$ от времени может быть разным – линейным, периодическим, экспоненциальным и т.д.

При линейном характере интенсивность потока событий «записей вниз» является функций вида:

$$\lambda(t_i) = \lambda_0 \pm \beta i. \quad (4)$$

С учетом ординарности в отмеченном выше смысле потока событий «записи вниз» и ограничения значений параметра λ диапазоном $[0-1]$ будем полагать $\lambda_0 = 1$ при убывающем во времени характере $\lambda(t_i) = \lambda_0 - \beta i$. При возрастающем характере $\lambda(t_i) = \lambda_0 + \beta i$ положим $\lambda_0 = 0,1$.

На рис. 2 представлены графики вероятности деградации системы управления доступом при возрастающем характере интенсивности потока событий «записей вниз» во времени $\lambda(t_i) = \lambda_0 + \beta i$ и различных значениях параметра β , определяющем скорость возрастания $\lambda(t_i)$.

Из рисунка видно, что зависимость вероятности деградации $P_{t_i}(k \geq N_{t_i}^{(t_0)})$ от времени при возрастающей во времени интенсивности потока событий «записей вниз» $\lambda(t_i) = \lambda_0 + \beta i$ аналогична соответствующей

зависимости при стационарном потоке событий, только отличается «ускоренным» характером.

При убывающем характере интенсивности потока событий «записей вниз» после определенного момента времени ($i = \lambda_0 / \beta$) значения функции $\lambda(t_i) = \lambda_0 - \beta i$ становятся отрицательными и по смыслу параметра $\lambda(\lambda > 0)$ с этого момента приравняются к нулю.

На рис. 3 представлены графики вероятности деградации мандатной системы управления доступом при убывающем характере интенсивности потока событий от времени $\lambda(t_i) = \lambda_0 - \beta i$ и различных значениях параметра β , ($\lambda_0 = 1$), вычисляемые по соотношению:

$$P_{t_i}(k \geq N_{t_i}^{(t_0)}) = 1 - e^{-i(\lambda_0 - \beta i)} \sum_{m=0}^{N_{t_i}^{(t_0)}-1} \frac{i(\lambda_0 - \beta i)^m}{m!}. \quad (5)$$

Как видно из рис. 3, вероятность деградации мандатной системы управления доступом при убывании интенсивности потока событий «записей вниз» во времени имеет существенно иной характер.

Сначала при увеличении интервала наблюдения вероятность деградации, как и в случае стационарного потока, также возрастает, причем при малых значениях параметра β практически до единицы (см. рис. 3), но далее при анализе для больших интервалов наблюдения (анализа) начинает уменьшаться, стремясь к нулю.

Такое странное на первый взгляд поведение объясняется тем, что вероятность $P_{t_i}(k \geq N_{t_i}^{(t_0)}) = 1 - P_{t_i}(k < N_{t_i}^{(t_0)})$ не характеризуется «накопительным» эффектом», а рассматривается в целом по промежутку времени от t_0 до t_i , на котором не должно произойти или только точно 0 событий «записей вниз», или только точно 1 событие, или ... только точно $N_{t_i}^{(t_0)} - 1$ событий. И соответственно, чем больше

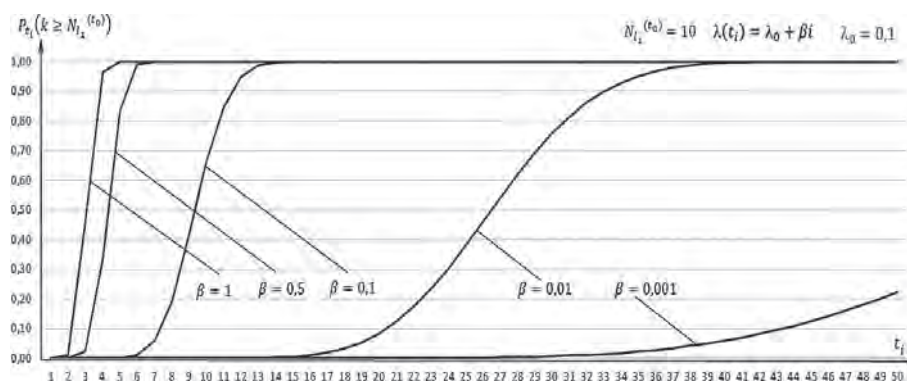


Рис. 2. Зависимость вероятности деградации мандатной системы управления доступом от времени при гипотезе о нестационарном пуассоновском потоке событий «записей вниз» с линейно возрастающей по времени интенсивностью.

($N_{t_i}^{(t_0)}$ – количество низкоуровневых объектов в начальный момент времени; k – количество произошедших событий «записей вниз» к моменту-шагу времени t_i , т.е. количество низкоуровневых объектов, перешедших к моменту времени t_i в высокоуровневые; $\lambda(t_i)$ – интенсивность потока событий, увеличивающаяся с течением времени со значением в нулевой момент времени $\lambda_0 = 0,1$ и со скоростью, определяемой параметром β).

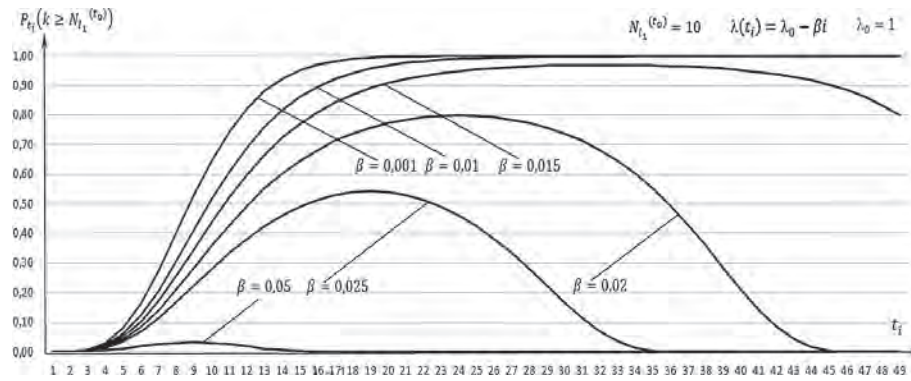


Рис. 3. Зависимость вероятности деградации мандатной системы управления доступом от времени при гипотезе о нестационарном пуассоновском потоке событий «записей вниз» с линейно убывающей по времени интенсивностью.

($N_i^{(t_0)}$ – количество низкоуровневых объектов в начальный момент времени; k – количество произошедших событий «записей вниз» к моменту-шагу времени t_i , т.е. количество низкоуровневых объектов, перешедших к моменту времени t_i в высокоуровневые; $\lambda(t_i)$ – интенсивность потока событий, уменьшающаяся с течением времени со значением в нулевой момент времени $\lambda_0 = 1$ и со скоростью, определяемой параметром β).

период ($t_0 - t_i$), т.е. чем больше t_i (при дискретном времени – i) и, соответственно, чем меньше $\lambda(t_i)$, – тем меньше вероятность совокупности отмеченных событий (только точно 0 событий, только точно 1 событие и т.д.) и соответственно суммы таких вероятностей.

Таким образом, если деградации не произошло в начальные периоды времени, когда интенсивность потока событий «записей вниз» была еще достаточно высокой (не нулевой), то в более поздние моменты времени, когда интенсивность потока событий становится практически нулевой, вероятность деградации также стремится к нулю. Данный эффект можно назвать эффектом «послемаксимума затухания» пуассоновского процесса при убывающей интенсивности событий, который при увеличении интервала наблюдения проявляется тем быстрее, чем большим значением характеризуются скорость убывания интенсивности потока событий (параметр β).

Максимальное значение вероятности деградации характеризует экстремум (максимум) $P_i(k \geq N_i^{(t_0)})$ как функции от момента времени t_i . Значение момента времени t_i , при наблюдении на котором вероятность деградации будет максимальной, можно определить решением уравнения

$$\frac{dP_i(k \geq N_i^{(t_0)})}{di} = 0. \quad (6)$$

Для решения этого уравнения в выражении для $P_i(k \geq N_i^{(t_0)})$ по соотношению (5) обозначим $y(i) = i(\lambda_0 - \beta i)$, $N = N_i^{(t_0)}$ и произведем дифференцирование по i функции

$$f(i) = 1 - e^{-y(i)} \sum_{m=0}^{N_i^{(t_0)}-1} \frac{(y(i))^m}{m!} (y(i))^m / m!.$$

В результате получаем

$$e^{-y(i)} \frac{de^{-y(i)}}{di} \left(\sum_{n=0}^{N-1} \frac{(y(i))^n}{n!} - \sum_{n=1}^{N-1} \frac{n(y(i))^{n-1}}{n!} \right) = 0. \quad (7)$$

Учитывая, что $e^{-y(i)} > 0$, для нахождения 1-го корня имеем

$$\lambda_0 - 2\beta i = 0,$$

откуда

$$i_{\text{экстремум1}} = \frac{\lambda_0}{2\beta}. \quad (8)$$

Второй и третий корни определяются по равенству 0 третьего сомножителя в соотношении (7):

$$\begin{aligned} \left(\sum_{n=0}^{N-1} \frac{(y)^n}{n!} - \sum_{n=1}^{N-1} \frac{n y^{n-1}}{n!} \right) &= \sum_{n=0}^{N-1} \frac{(y)^n}{n!} - \sum_{n=1}^{N-1} \frac{(y)^{n-1}}{(n-1)!} = \\ &= \sum_{n=0}^{N-1} \frac{(y)^n}{n!} - \sum_{m=0}^{N-2} \frac{(y)^m}{m!} = 0, \end{aligned}$$

где $j = n - 1$.

Преобразовывая, далее имеем

$$\sum_{n=0}^{N-1} \frac{(y)^n}{n!} - \sum_{m=0}^{N-2} \frac{(y)^m}{m!} = \frac{(y)^{N-1}}{(N-1)!} = 0,$$

откуда следует

$$i(\lambda_0 - \beta i) = 0.$$

В результате получаем³:

$$i_{\text{экстремум2}} = 0, i_{\text{экстремум3}} = \frac{\lambda_0}{\beta}.$$

Второй и третий корни не рассматриваем по смыслу отмеченных ранее ограничений на интервал значений интенсивности потока событий – $0 < \lambda \leq 1$.

Таким образом, максимум вероятности деградации мандатной системы управления доступом с возможностью «записей вниз» приходится на момент времени $t_i = \frac{\lambda_0}{\beta}$ и не зависит от начального количества

3 Результаты по нахождению точек экстремума функции $P(t_i) (k \geq N_i^{(t_0)})$ при убывающем характере интенсивности потока событий получены совместно с Караваевым А. С.

низкоуровневых объектов $N_{l_i}^{(t_0)}$, что соответствует численным данным на графиках $P_{t_i}(k \geq N_{l_i}^{(t_0)})$, представленным на рис. 3.

Значение вероятности деградации $P_{t_i}(k \geq N_{l_i}^{(t_0)})$ в точке максимума определяется соотношением

$$1 - e^{-\frac{\lambda_0}{2\beta}(\lambda_0 - \beta \frac{\lambda_0}{2\beta})} \sum_{m=0}^{N_{l_i}^{(t_0)}-1} \frac{\left(\frac{\lambda_0}{2\beta}(\lambda_0 - \beta \frac{\lambda_0}{2\beta})\right)^m}{m!}, \quad (9)$$

анализ которого показывает, что единичное значение $P_{t_i}(k \geq N_{l_i}^{(t_0)})$ в точке максимума достигается при нулевом значении скорости убывания интенсивности потока событий «записей вниз» ($\beta = 0$), т.е. при стационарном потоке событий.

Иначе говоря, при очень малых значениях β вероятность деградации мандатной системы управления доступом с возможностью «записей вниз» в момент времени $t_i = \frac{\lambda_0}{2\beta}$ может быть сколь угодно близкой к единице (но не равной ей).

Интересно отметить, что в других приложениях, характеризующихся сходными по смыслу процессами деградации, рассматривается экспоненциальный

характер зависимости $\lambda(t_i)$. В частности, в исследованиях старения и смертности биологических видов широко известно распределение Гомпертца-Мейкхэма [12], которое также применяется в исследованиях «старения» информации [13], в частности старения научной [14] и маркетинговой [15] информации, в анализе надежности и жизненного цикла сложных технических систем [16] и т.д.

В отмеченных предметных областях интенсивность потока событий рассматривается как функция вида

$$\lambda(t_i) = a + be^{\lambda_0 t_i}, \quad (10)$$

что обеспечивает учет как постоянной во времени составляющей процессов деградации (параметр a), так и переменной с течением времени (параметры b и λ_0). При этом параметр λ_0 определяет «скорость» экспоненциального изменения интенсивности потока событий и может быть как положительным (интенсивность потока событий со временем возрастает), так и отрицательным (интенсивность потока событий со временем убывает).

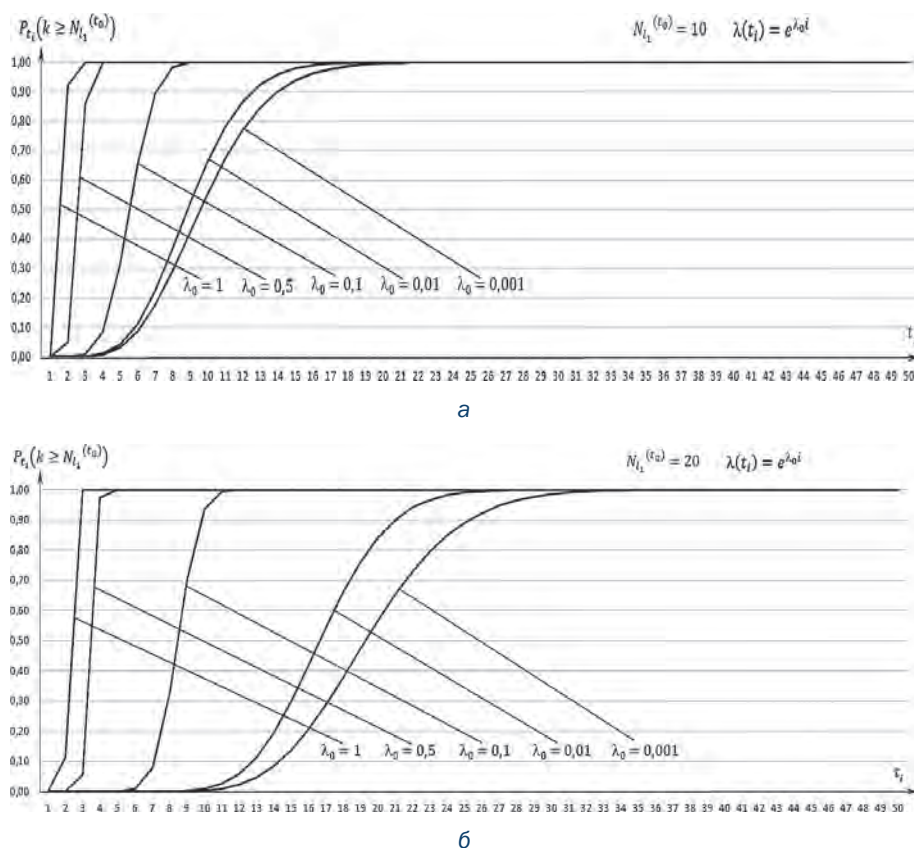


Рис. 4. Зависимость вероятности деградации мандатной системы управления доступом от времени при гипотезе о нестационарном пуассоновском потоке событий «записей вниз» с экспоненциально возрастающей по времени интенсивностью.

($N_{l_i}^{(t_0)}$ – количество низкоуровневых объектов в начальный момент времени; k – количество произошедших событий «записей вниз» к моменту-шагу времени t_i , т.е. количество низкоуровневых объектов, перешедших к моменту времени t_i в высокоуровневые; $\lambda(t_i) = e^{\lambda_0 t_i}$ – интенсивность потока событий, экспоненциально возрастающая с течением времени со скоростью, определяемой параметром λ_0).

Рассмотрим сначала, как и при линейном характере функции $\lambda(t_i)$, деградацию мандатной системы управления доступом при экспоненциально возрастающей с течением времени интенсивностью потока событий «записей вниз». При этом для простоты положим $a = 0$, $b = 1$. Тогда $\lambda(t_i)$ будет характеризоваться простой экспоненциальной зависимостью от времени $\lambda(t_i) = e^{\lambda_0 i}$, $\lambda_0 \geq 0$. Соответственно зависимость вероятности деградации системы управления доступом от времени будет выражаться следующим соотношением:

$$P_{t_i}(k \geq N_{t_i}^{(t_0)}) = 1 - e^{-ie^{\lambda_0 i}} \sum_{m=0}^{N_{t_i}^{(t_0)}-1} \frac{(ie^{\lambda_0 i})^m}{m!}. \quad (11)$$

На рис. 4 представлены графики вероятности деградации мандатной системы управления доступом при экспоненциально возрастающем характере интенсивности потока событий «запись вниз» во времени $\lambda(t_i) = e^{\lambda_0 i}$, различных значениях параметра λ_0 , определяющем скорость возрастания $\lambda(t_i)$, и при начальном количестве низкоуровневых объектов $N_{t_i}^{(t_0)} = 10$ (рис. 4а), $N_{t_i}^{(t_0)} = 20$ (рис. 4б).

Из рис. 4 видно, что, как и в случае с линейно возрастающей во времени интенсивностью, при экспоненциально возрастающей во времени интенсивности потока событий «записей вниз» $\lambda(t_i) = e^{\lambda_0 i}$ характер изменения со временем вероятности деградации системы управления доступом в целом аналогичен характеру при стационарном потоке событий и отличается только большей «ускоренностью» возрастания.

При экспоненциально убывающем характере интенсивности потока событий «записей вниз» вероятность деградации мандатной системы управления доступом определяется выражением

$$P_{t_i}(k \geq N_{t_i}^{(t_0)}) = 1 - e^{-ie^{-\lambda_0 i}} \sum_{m=0}^{N_{t_i}^{(t_0)}-1} \frac{(ie^{-\lambda_0 i})^m}{m!}. \quad (12)$$

где λ_0 – параметр, определяющий скорость убывания интенсивности потока событий «записей вниз» во времени.

На рис. 5 представлены графики вероятности деградации мандатной системы управления доступом

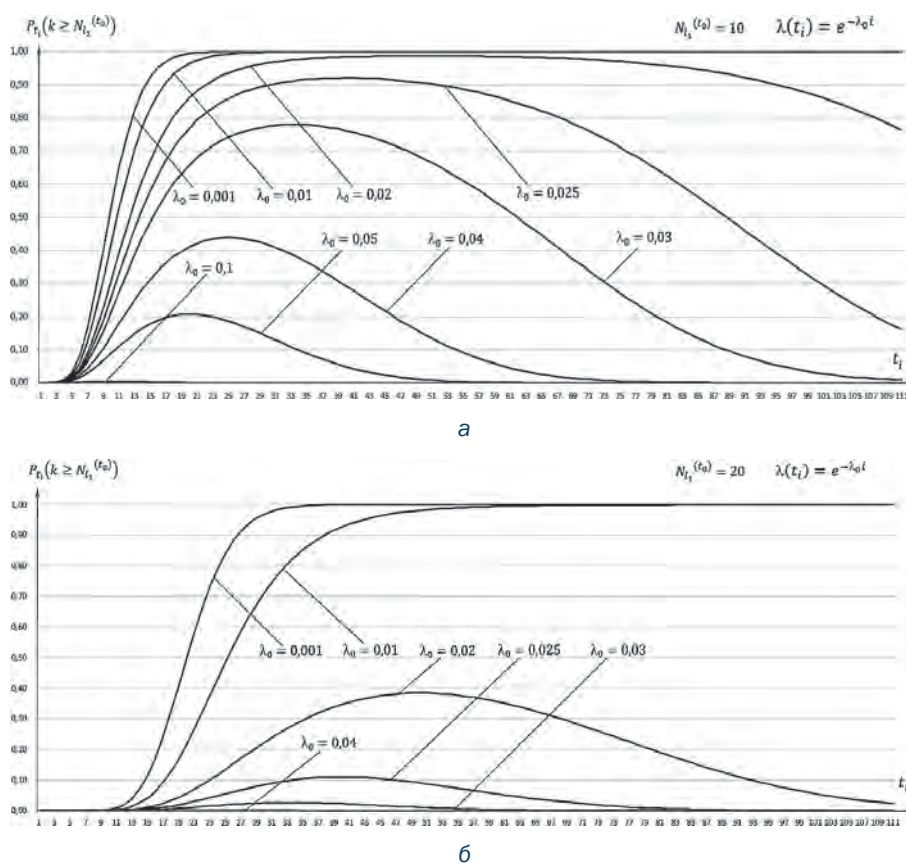


Рис. 5. Зависимость вероятности деградации мандатной системы управления доступом от времени при гипотезе о нестационарном пуассоновском потоке событий «записей вниз» с экспоненциально убывающей по времени интенсивностью.

($N_{t_i}^{(t_0)}$ – количество низкоуровневых объектов в начальный момент времени; k – количество произошедших событий «записей вниз» к моменту-шагу времени t_i , т.е. количество низкоуровневых объектов, перешедших к моменту времени t_i в высокоуровневые; $\lambda(t_i) = e^{\lambda_0 i}$ – интенсивность потока событий, экспоненциально возрастающая с течением времени со скоростью, определяемой параметром λ_0).

при экспоненциально убывающем характере интенсивности потока событий «записей вниз» во времени $\lambda(t_i) = e^{-\lambda_0 t_i}$, различных значениях параметра λ_0 , определяющем скорость убывания $\lambda(t_i)$, и при начальном количестве низкоуровневых объектов $N_{l_i}^{(t_0)} = 10$ (рис. 5а), $N_{l_i}^{(t_0)} = 20$ (рис. 5б).

Из рис. 5 видно, что как и в случае с линейно убывающей интенсивностью потока событий «запись вниз», зависимость от времени вероятности деградации мандатной системы управления доступом характеризуется таким же эффектом «послемаксимового затухания», т.е. наличием максимума и дальнейшим убыванием к нулю при увеличении момента времени наблюдения-анализа t_i .

Производя аналогичным образом дифференцирование по i функции в правой части соотношения (12) и приравнявая результат дифференцирования к нулю, находим точку максимум вероятности деградации $P_{t_i}(k \geq N_{l_i}^{(t_0)})$:

$$t_{imax} = \frac{1}{\lambda_0},$$

что также соответствует численным данным на графиках $P_{t_i}(k \geq N_{l_i}^{(t_0)})$, представленным на рис. 5.

Заключение

Представленный анализ мандатных систем управления доступом с возможностью «записей вниз» показывает в большинстве случаев их «неизбежную» деградацию при тех или иных условиях и параметрах возникновения у пользователей соответствующих потребностей (в «записях вниз»). При этом, однако, при убывающих со временем у пользователей потребностях в «записях вниз» тенденция к деградации характеризуется эффектом «послемаксимового затухания», что может быть использовано в технологических процессах обработки информации или отражать такие процессы.

Тем не менее тенденция к деградации в виде уменьшения со временем количества низкоуровневых объектов из-за их перехода в высокоуровневые проявляется в любом случае.

В результате стремление к расширению функциональности мандатных систем обработки информации

через введение возможности «записей вниз» приводит, в свою очередь, к ограничению функциональности в виде тенденции к деградации, при которой низкоуровневые пользователи (с низким уровнем доступа) со временем не смогут получать доступ по чтению ко всем или к большинству объектов доступа. При этом следует оговориться, что данные выводы характеризуют системы управления доступом, в которых со временем новые низкоуровневые объекты доступа не создаются. В этом отношении порождение новых низкоуровневых объектов может снизить остроту проблемы деградации и соответствующее ограничение функциональности таких систем.

Принципиальным же направлением поиска теоретических и практических решений по устранению проблемы деградации в мандатных системах управления доступом с возможностью «записей вниз» и обеспечению их большей функциональности является разработка механизмов анализа конфиденциальности информации (классификации информации), записываемой пользователями в объекты доступа. При наличии таких доверенных и эффективных механизмов в случае, когда пользователь с высоким уровнем допуска записывает в низкоуровневый объект только низкоуровневую информацию, отпадает необходимость в команде reset и после записи объект остается низкоуровневым, что устраняет тенденцию к деградации.

Однако подобные механизмы на сегодняшний день эффективно и в полной мере реализуются только на основе содержательного анализа информации человеком-пользователем, в первую очередь при присваивании объекту доступа его уровня (метки) конфиденциальности. Тем не менее в DLP системах [17], которые играют в настоящее время важную роль в системах защиты информации, применяются те или иные механизмы анализа (классификации) конфиденциальности информации, «уходящей» за периметр защищаемой системы. Очевидно, подобные «наработки» следует проанализировать на предмет возможности использования в мандатных системах управления доступом при санкционировании запросов «записей вниз».

Литература

1. Bell D. E., LaPadula L. J. Secure Computers Systems: Unified Exposition and Multics Interpretation. Bedford, Mass.: MITRE Corp., 1976. – MTR-2997 Rev. 1.
2. Guo J. K. et al. Applicability of low water-mark mandatory access control security in linux-based advanced networked consumer electronics // First IEEE Consumer Communications and Networking Conference, 2004. CCNC 2004. – IEEE, 2004. – С. 364–369.
3. Landwehr K. E. Formal Model Computer Security // Computer surveys, Vol 13, No 3, September 1981. pp. 247–278.
4. Bishop, M. Introduction to Computer Security / M. Bishop. – Addison- Wesley Professional, 2004. – 785 p.
5. Mohamed A., Auer D., Hofer D., Küng J. A systematic literature review for authorization and access control: definitions, strategies and models // International Journal of WebInformation Systems. Vol. 18 No. 2/3, 2022. pp. 156-180. DOI 10.1108/IJWIS-04-2022-0077.
6. Biba K. J. Integrity considerations for secure computer systems. ESD-TR- 76-372, ESD/AFSC, Hanscom AFB, Bedford, Mass., April 1977 (MITRE MTR-3153, NTIS AD A039324).

7. Weissman C. Security controls in the ADEPT-50 time sharing system. Proc. 1969 AFIPS Fall Jt Computer Conf., vol. 35, AFIPS Press, Arlington, Va., pp. 119–133.
8. McLean J. Security models and information flow. In Proceedings of the 1990 IEEE Symposium on Research in Security and Privacy. IEEE Computer Society Press, 1990.
9. McLean J. The specification and modeling of computer security. Computer, 23(1): 9–16, January 1990.
10. Гайдамакин Н. А. Модель и метрики осведомленности в конфиденциальной информации. Часть 1. Потенциальная осведомленность // Прикладная дискретная математика. 2023. № 61. С. 86–103.
11. Кингман Дж. Пуассоновские процессы. — М.: МЦНМО, 2007. — 136 с.
12. Gavrilov L. A., Gavrilova N. S. The Biology of Life Span: A Quantitative Approach., New York: Harwood Academic Publishe, 1991.
13. Парамонов И. Ю. Модель учета ценности и старения информации при оценивании эффективности функционирования систем информационного обеспечения // Системы управления, связи и безопасности. 2016. № 1. С.238–334.
14. Климов Ю. Н. Исследование закономерностей старения научной информации // Машиностроение, методы исследования и моделирования. 2004. № 4. С. 40–44.
15. Чернявский А. Д. Моделирование скорости распространения маркетинговой информации с учётом изменения ценности информации во времени // Экономический анализ: теория и практика. 2009. № 7. С. 54–60.
16. Ташевский А. Г. Математические модели продолжительности жизненного цикла технических систем // Научно-технические ведомости Санкт-Петербургского государственного политехнического университета. 2014. № 1(190). С. 169–178.
17. Papadimitriou P., Garcia-Molina H. Data Leakage Detection // IEEE transactions on knowledge and data engineering, vol. 23, no. 1, January 2011. pp. 51-63.

ANALYSIS OF THE DEGRADATION OF ACCESS CONTROL SYSTEMS BASED ON MANDATED MODELS WITH THE ABILITY TO «WRITE DOWN»

Gaydamakin N. A.⁴

Keywords: mandatory access control principle, «write down», «write down» event flow, degradation of the access control system, «post-maximum attenuation» effect.

Purpose of the study: to conduct a theoretical analysis of the functioning of mandatory access control systems, in which, in order to expand the functional capabilities for processing information, access subjects are allowed to perform «write down».

Methods of research: application of the Poisson event flow apparatus for formalization and analysis of the functioning of mandatory access control systems with the possibility of "write down".

Result: The process of functioning of the mandatory access control system with the possibility of «write down» is considered as a Poisson flow of homogeneous events of occurrence and implementation by high-level access subjects (users with a high access level) of needs for recording information in low-level access objects (files, documents with a low confidentiality level). As a result of each such event after the completion of the «write down», in order to prevent possible further familiarization of low-level users with high-level information, the confidentiality level of the access object is increased to the access level of the corresponding (writing) subject. The concept of degradation of the access control system is introduced, which consists in the fact that with two levels of access and confidentiality (high and low), over time all access objects in terms of confidentiality become high-level, and accordingly inaccessible for «reading» for subjects (users) with a low access level.

The probability of the onset of degradation of the access control system in the case of stationarity and non-stationarity of the Poisson flow of "write down" events is analyzed. With decreasing intensity of "down write" events over time, an unusual at first glance nature of the change in the probability of degradation over time is considered, which can be called the «post-maximum attenuation» effect, consisting first in an increase in the probability of degradation over time with the achievement of a certain maximum value at a certain point in time, and then in a gradual decrease in the probability of degradation to zero. With a linear and exponential nature of the decrease in the intensity of the flow of «down write» events, relationships are obtained for the time point with the maximum probability of degradation, which is determined by the rate of decrease in the intensity of the flow of events.

References

1. Bell D. E., LaPadula L. J. Secure Computers Systems: Unified Exposition and Multics Interpretation. Bedford, Mass.: MITRE Corp., 1976. – MTR-2997 Rev. 1.
2. Guo J. K. et al. Applicability of low water-mark mandatory access control security in linux-based advanced networked consumer electronics // First IEEE Consumer Communications and Networking Conference, 2004. CCNC 2004. – IEEE, 2004. – С. 364–369.
4. Nikolay A. Gaydamakin, Dr.Sc. of Technical Sciences, Professor, Professor of the Department of Algebra and Fundamental Informatics, Ural Federal University named after the first President of Russia B. N. Yeltsin, Yekaterinburg, Russia. E-mail: n.a.gaidamakin@urfu.ru

3. Landwehr K. E. Formal Model Computer Security // Computer surveys, Vol 13, No 3, September 1981. pp. 247–278.
4. Bishop, M. Introduction to Computer Security / M. Bishop. – Addison–Wesley Professional, 2004. – 785 p.
5. Mohamed A., Auer D., Hofer D., Kűng J. A systematic literature review for authorization and access control: definitions, strategies and models // International Journal of Web Information Systems. Vol. 18 No. 2/3, 2022. pp. 156–180. DOI 10.1108/IJWIS-04-2022-0077.
6. Biba K. J. Integrity considerations for secure computer systems. ESD-TR- 76-372, ESD/AFSC, Hanscom AFB, Bedford, Mass., April 1977 (MITRE MTR–3153, NTIS AD A039324).
7. Weissman C. Security controls in the ADEPT-50 time sharing system. Proc. 1969 AFIPS Fall Jt Computer Conf., vol. 35, AFIPS Press, Arlington, Va., pp. 119–133.
8. McLean J. Security models and information flow. In Proceedings of the 1990 IEEE Symposium on Research in Security and Privacy. IEEE Computer Society Press, 1990.
9. McLean J. The specification and modeling of computer security. Computer, 23(1): 9–16, January 1990.
10. Gaydamakin N.A. Model' i metriki osvedomlennosti v konfidentsial'noy informatsii. Chast' 1. Potentsial'naya osvedomlennost' // Prikladnaya diskretnaya matematika. 2023. № 61. S. 86–103.
11. Kingman Dzh. Puassonovskie protsessy. – M.: MTSNMO, 2007. – 136 s.
12. Gavrilov L. A., Gavrilova N. S. The Biology of Life Span: A Quantitative Approach., New York: Harwood Academic Publishe, 1991.
13. Paramonov I. Yu. Model' ucheta tsennosti i stareniya informatsii pri otsenivanii effektivnosti funktsionirovaniya sistem informatsionnogo obespecheniya // Sistemy upravleniya, svyazi i bezopasnosti. 2016. № 1. S. 238–334.
14. Klimov Yu. N. Issledovaniye zakonomernostey stareniya nauchnoy informatsii // Mashinostroyeniye, metody issledovaniya i modelirovaniya. 2004 №4. C. 40–44.
15. Chernyavskiy A. D. Modelirovaniye skorosti rasprostraneniya marketingovoy informatsii s uchotom izmeneniya tsennosti informatsii vo vremeni // Ekonomicheskiy analiz: teoriya i praktika. 2009. № 7. S. 54–60.
16. Tashevskiy A. G. Matematicheskiye modeli prodolzhitel'nosti zhiznennogo tsikla tekhnicheskikh sistem // Nauchno-tekhnicheskiye vedomosti Cankt-Peterburgskogo gosudarstvennogo politekhnicheskogo universiteta. 2014. № 1(190). C. 169–178.
17. Papadimitriou P., Garcia-Molina H. Data Leakage Detection // IEEE transactions on knowledge and data engineering, vol. 23, no. 1, January 2011. pp. 51–63.



МНОГОУРОВНЕВЫЙ ФРЕЙМВОРК ОБОСНОВАНИЯ ПРОЦЕДУР МОНИТОРИНГА И РЕАГИРОВАНИЯ НА ИНЦИДЕНТЫ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

Грызунов В. В.¹, Шестаков А. В.²

DOI: 10.21681/2311-3456-2025-6-14-24

Цель исследования: обосновать подходы к рациональной организации мониторинга и реагирования на возможные инциденты информационной безопасности.

Метод исследования: информационная система описывается разработанной и апробированной авторами иерархической моделью, включающей обеспечивающий уровень, уровень персонала, а также уровни аппаратного и программного обеспечения. Модель получает на вход множества возможных состояний информационной системы и инцидентов информационной безопасности, допустимые значения вероятностей рисков информационной безопасности и множество допустимых управляющих воздействий. На выходе модели – управляющее воздействие, которое удерживает вероятность риска информационной безопасности в заданном диапазоне.

Полученный результат: многоуровневый фреймворк обеспечивает сквозное описание процесса управления инцидентами информационной безопасности – от обнаружения инцидента до формирования плана улучшений. Он позволяет количественно оценивать эффективность мер реагирования на каждом уровне информационной системы – обеспечивающем, персонала, аппаратном и программном. Ключевое преимущество подхода – возможность компенсировать недостатки одного уровня за счёт усиления другого, что расширяет возможности удерживать вероятность риска информационной безопасности в заданном диапазоне. На основе полученных оценок формируются экономически обоснованные стратегии информационной безопасности. Центральным практическим выводом является принцип «перераспределения риска», который заменяет догматическое требование устранить все уязвимости на целенаправленное, измеримое и рентабельное управление вероятностью риска.

Научная новизна: в отличие от существующих моделей разработанная модель явно учитывает совокупную вероятность риска информационной безопасности и взаимозависимость всех четырёх уровней информационной системы за счёт формирования соответствующего оператора.

Ключевые слова: терминосистема, кибербезопасность, управление инцидентами.

Введение

Процессы мониторинга инцидентов информационной безопасности (ИИБ) и реагирования на ИИБ вынужденно входят в повседневную деятельность компаний, - к этому побуждает и законодательство, и экономические соображения. Результаты анализа нормативных правовых актов (НПА), в части необходимости мониторинга и реагирования на ИИБ, подтверждают несогласованность периодичности мониторинга и реагирования (табл. 1).

Анализ доступной статистики об утечках персональных данных подтверждает экономическими показателями положительный эффект регламентированного тестирования ИС, а регулярного мониторинга угроз ИБ с применением SIEM (Security Information and Event Management), MDR (Managed Detection and Response) и CEM (Continuous Exposure Management) – вероятностными показателями. Средства мониторинга и реагирования на ИИБ

Таблица 1.

Сводные данные из НПА в части мониторинга и реагирования на ИИБ

Документ	Мониторинг / Обнаружение	Регистрация / Отчёт	Реагирование (сроки)
ГОСТ Р 59547 / Р 59709	да	да	да
ГОСТ ИСО/МЭК ТО 18044 2007	Не определено	Немедленно	По итогам формы отчёта
Приказ ФСБ № 282 (КИИ)	Сенсоры / круглосуточно	В течение 3 ч / 24 ч	В течение 48 ч
Положение ЦБ № 822 П (страхование)	Регистрация событий	Описание мер	В сроки регулятора

1 Грызунов Виталий Владимирович, доктор технических наук, доцент, профессор кафедры прикладной математики и информационных технологий Санкт-Петербургского университета ГПС МЧС России. Санкт-Петербург, Россия. ORCID <https://orcid.org/0000-0003-4866-217X>. E-mail: viv1313r@mail.ru

2 Шестаков Александр Викторович, доктор технических наук, старший научный сотрудник, ведущий научный сотрудник Санкт-Петербургского университета ГПС МЧС России. Санкт-Петербург, Россия. ORCID <https://orcid.org/0000-0002-8462-6515>. E-mail: alexandr.shestakov01@yandex.ru

Таблица 2.

Сущности мониторинга и реагирования на возможные ИИБ в НПА

Элемент стандарта	Мониторинг / обнаружение	Реагирование / ответ
ГОСТ 57580.1	Мониторинг событий ИБ и их регистрация	Классификация и организация реакции
ГОСТ 57580.3/4	Мониторинг рисков и событий	Реакция и восстановление после инцидентов
NIST 800 61r3	Фаза Detection & Analysis (SIEM, IDS)	Фаза Response & Recovery
ISO 27035	Обнаружение и оценка событий и инцидентов	Реакция, восстановление, анализ уроков

становятся неотъемлемой частью системы управления ИИБ.

Анализ руководящих документов подтверждает, что мониторинг и реагирование на ИИБ являются частью процесса управления ИИБ (табл. 2).

Цель настоящего исследования – обосновать подходы к рациональной организации мониторинга и реагирования на возможные ИИБ.

Анализ литературы

Формализованный подход к автоматизации обнаружения и реагирования на инциденты ИБ изложен в статье [1]. Основу составляют система поддержки принятия решений (СППР) с моделью на базе теории графов и нечеткой логики, которая соотносит наблюдаемые события с типами инцидентов. Не в полной мере учтён человеческий фактор и множество допустимых состояний системы, которое определено руководящими документами. Модель предполагает наличие оператора, но не рассматривает вопросы его подготовки, квалификации или действий в нештатных ситуациях. Отсутствует анализ возможности интеграции предложенного плана в существующую нормативную базу организации, например, в планы обеспечения непрерывности деятельности.

Комплексный подход к созданию Центра мониторинга ИБ (SOC) предложен в [2]. Описана процессная модель по рекомендациям NIST, структура команды (три линии поддержки). Программно-аппаратная среда рассматривается упрощенно, без учета сложностей гетерогенных инфраструктур (гибридные облака, АСУ ТП) и возможности отказа самих средств мониторинга. Отсутствует описание влияния уровней SOC друг на друга.

Прогрессивная концепция управления ИБ на основе цикла непрерывного детектирования и реагирования (CD/CR), аналогичного DevOps-подходу CI/CD изложена в [3]. Фокус смещается с физической инфраструктуры на абстрактный «управляемый объект», а сценарии реагирования (playbooks) формируются динамически на основе обогащенного контекста инцидента, что позволяет адаптироваться к меняющимся угрозам и проводить глубокий

контекстно-ориентированный анализ атак. Концепция строится на идеализированном предположении о полноте и достоверности собираемых данных, не учитывая «туман войны»: сбои средств мониторинга, нелогируемые техники атак и другие проблемы. Человеческий фактор не учитывается.

Подход к формализации процесса обнаружения инцидентов ИБ на промышленных предприятиях рассмотрен в [4]. Исследование фокусируется на выработке общего математического критерия, представляя мониторинг как вероятностный процесс. На основе теории случайных процессов проанализировано, как различные факторы (сбои оборудования, ошибки ПО, действия персонала) вносят погрешность в измерения и влияют на итоговую вероятность обнаружения. Программно-аппаратная среда рассматривается абстрактно, как источник статистического «шума», без учета всех ее состояний, таких как «полный отказ средств мониторинга».

В контексте «умных городов» рассмотрена проблематика кибербезопасности [5]. Приведён комплексный и актуальный охват темы, затрагивающий технические, организационные и правовые вопросы, однако не содержит детальных технических решений, не формализует связь программно-аппаратного комплекса, персонала и руководящих документов.

Интегрированная система управления киберрисками (i-CSRM) в [6] основана на модели, которая объединяет данные о киберугрозах (CTI), оценку критичности активов с помощью нечёткой логики, прогнозирование рисков на основе машинного обучения и проверку эффективности уже внедрённых средств защиты. Процессы мониторинга и реагирования выступают элементом модели и являются несколько статичными. При этом модель опирается на существующие навыки персонала. Не описано, как модель вписывается во внутренние регламенты организации.

В работе [7] авторы описывают процессы мониторинга и реагирования на ИИБ средствами облачного центра мониторинга (SOC) в Microsoft Azure, в которой развёрнута защищённая виртуальная

сеть (VNet) с межсетевым экраном для веб-приложений (WAF). Инфраструктура интегрирована с SIEM-системой Microsoft Sentinel для мониторинга событий и с Microsoft Defender for Cloud для контроля безопасности и соответствия. Реакция системы тестируется с помощью nmap. Для персонала сделан акцент на модели разделения ответственности между потребителем облака и поставщиком облачных услуг. Работа опирается на фреймворки NIST, CSA и Azure Well-Architected Framework. Реализованное решение позволяет проводить автоматическую оценку на соответствие стандартам GDPR, PCI-DSS, ISO 27001, а конфигурация WAF основана на правилах OWASP.

Экономико-математический подход для обоснования формирования бюджета на ИБ, реализованный на Python, описан в [8]. Программа моделирует состояние бюджета на ИБ, из которого покрываются случайные расходы на устранение инцидентов. В основе лежат положения теории рисков и математики страхования.

Методы исследования

Информационная система (ИС) в настоящем исследовании рассматривается как комплексная организационно-техническая система, имеющая несколько иерархических уровней. Поэтому для её описания целесообразно использовать разработанную и апробированную авторами иерархическую модель информационной системы, включающая обеспечивающий уровень, уровень персонала, а также уровни аппаратного и программного обеспечения.

Метауровни задают требуемые состояния (S) вложенных уровней. Обеспечивающий уровень имеет множество (S^E) состояний, формирует требования к персоналу (P) и к программно-аппаратному комплексу, то есть задаёт требования к множествам состояний (S^P) уровня персонала, уровня (S^{Hard}) аппаратного обеспечения (АО) и уровня (S^{Soft}) программного обеспечения (ПО). Персонал настраивает и обеспечивает функционирование аппаратного и программного обеспечения, то есть задаёт требования к множествам S^{Hard} и S^{Soft} . Аппаратура, в свою очередь, предоставляет заданные ресурсы и среду исполнения для ПО, то есть задаёт требования к S^{Soft} . Например, ПО функционирует в рамках системы команд и ограничений доступа к памяти, заданных аппаратной платформой, а аппаратура эксплуатируется персоналом в соответствии с установленными регламентами.

Примеры состояний каждого уровня ИС

Обеспечивающий уровень – объём финансирования в месяц (s_1^E), требования по повышению квалификации персонала (s_2^E), требование ко времени предоставления отчётов об ИББ (s_3^E).

Уровень персонала – компетенции персонала в сфере ИБ (s_1^P), загрузка в течение рабочего дня (s_2^P), укомплектованность специалистами ИБ (s_3^P).

Уровень АО – наличие резервного канала связи (s_1^{Hard}), включение устройства защиты от ПЭМИН (s_2^{Hard}), применение внешнего сервера для сбора данных от DLP-агентов (s_3^{Hard}).

Уровень ПО – развёрнута SIEM-система (s_1^{Soft}), установлено антивирусное ПО (s_2^{Soft}), применение TLS (s_3^{Soft}).

Декартово произведение состояний каждого уровня формирует общее состояние ИС (S)

$$S = S^E \times S^P \times S^{Hard} \times S^{Soft}. \quad (1)$$

Вложенный уровень влияет на метауровень через обратную связь. Например, выявление ограничения на уровне ПО, не позволяющего организовать работу распределённой в пространстве-времени команды специалистов ИБ, может потребовать распределить работу в пространстве-времени АО. Что, в свою очередь, откорректирует требования к квалификации персонала и инициирует изменения регламентов или финансирования на обеспечивающем уровне.

В настоящем исследовании используется модель системы адаптивного управления на базе операторного уравнения, разработанная авторами в интересах киберполигона МЧС России.

Результаты

Анализ НПА в сфере ИБ подтверждает, что одна из главных целей управления ИИБ – своевременное реагирование для ограничения ущерба (снижения риска ИИБ). Соответственно, процессы мониторинга и реагирования как часть системы управления ИИБ, во-первых, должны работать на эту же цель – снизить риск ИБ, во-вторых, должны быть согласованы между собой.

Риск ИБ (Ψ_a), ущерб (W) и вероятность риска ИБ (P_{risk}) могут рассчитываться таблично, посредством лингвистических переменных (критический, высокий, низкий и т.п.) или аналитически. Существует ряд способов свести к аналитическому виду представление величин в виде лингвистической переменной. Например, для вероятности риска для каждого значения лингвистической переменной определяются границы значения вероятности риска: низкий [0; 0,2), средний [0,2; 0,4), высокий [0,4; 1]. Затем вместо каждого значения лингвистической переменной подставляется среднее значение соответствующего интервала. Поэтому далее без нарушения общности и для повышения наглядности воспользуемся аналитическим способом расчёта риска ИБ:

$$\Psi_a = P_{risk} \cdot W. \quad (2)$$

Обычно специалисты ИБ не влияют на ущерб, возникающий в ходе возникновения ИИБ, потому

что ущерб определяется целевыми процессами ИС. Специалисты ИБ снижают вероятность риска ИБ (P_{risk}).

Риск ИБ – это мера опасности ИБ, а мера безопасности ИС – доверие, как показано в [9]. При этом термин «доверие» в различных источниках применяется без его точного определения, либо приводятся определения в стиле «доверие – это основания для обоснованной уверенности в том, что объект оценки соответствует функциональным требованиям безопасности», которые не дают представления об единицах измерения «доверия», и следовательно, не позволяют измерить «доверие». Примем в состав терминосистемы настоящего исследования следующие определения.

Определение. Доверие – это величина, дополняющая вероятность риска ИБ до единицы

$$D = 1 - P_{risk}. \quad (3)$$

Определение. Безопасная ИС (ИС, которой можно доверять) с точки зрения информационной безопасности является та ИС, в которой отсутствуют ИИБ.

В «безопасной ИС», согласно предложенной модели ИС, должно быть «доверие» ко всем уровням ИС. В общем случае, исходя из практического опыта, существует позитивная корреляция между «событиями», состоящими в том, что есть «доверие на уровнях ИС». Если есть «доверие» на уровне персонала, то вероятность того, что будет «доверие» на других уровнях (АО или ПО, обеспечивающего уровня) увеличивается. Или выделение финансирования (обеспечивающий уровень) повышает вероятность закупки и своевременного обновления программ и оборудования (ПО и АО), финансовой мотивации персонала (уровень персонала).

Пусть x^i – событие, означающее «на уровне i можно доверять» (отсутствует инцидент). Тогда $P(x^i) = D^i$.

Совокупное доверие – это вероятность того, что всем уровням можно доверять одновременно: $D = P(x^E \cap x^P \cap x^{Hard} \cap x^{Soft})$.

Если события x^i положительно коррелированы, то по определению $P(A \cap B) \geq P(A)P(B)$. Распространяя это на n событий, получаем:

$$P(x^i) \geq \prod P(x^i). \quad (4)$$

Следовательно, произведение вероятностей действительно является нижней границей, и если принять в качестве ограничения независимость событий, состоящих в существовании доверия на каждом уровне, то мы получим нижнюю оценку доверия ко всей ИС, что оправдано с точки зрения обеспечения ИБ. Нижняя граница доверия ко всей ИС будет вычисляться по формуле:

$$D = D^E \times D^P \times D^{Hard} \times D^{Soft}. \quad (5)$$

Из (4) следует, что вероятность общего риска ИБ будет равна:

$$P_{risk} = 1 - D = 1 - (1 - P_{risk}^E)(1 - P_{risk}^P) \times (1 - P_{risk}^{Hard})(1 - P_{risk}^{Soft}). \quad (6)$$

Выражение (5) даёт нам верхнюю оценку вероятности общего риска ИБ.

Из (5) следует, что вероятность общего риска ИБ (далее вероятность риска) может быть снижена специалистами ИБ на каждом уровне ИС.

Множество возможных ИИБ в ИС (Q) включает в себя ИИБ для каждого уровня ИС:

Обеспечивающий уровень (Q^E). Урезали финансирование для обновления средств защиты информации (СЗИ), политика ИБ не предусматривает проверку качества резервных копий, отсутствует перечень конфиденциальной информации.

Уровень персонала (Q^P). Не пройдено повышение квалификации персонала на объекте КИИ в течение четырёх лет, пароль администратора известен рядовому пользователю, персонал не осведомлён о том, что опасно «кликать котиков» в электронных письмах.

Уровень аппаратного обеспечения (Q^{Hard}). Активирована закладка в процессоре, вышло из строя устройство защиты от ПЭМИН, выдана недокументированная команда на маршрутизатор.

Уровень программного обеспечения (Q^{Soft}). Распространение шифровальщика, реализовано повышение привилегий учётной записи пользователя, выполнена атака Tiny Fragment Attack.

$$Q = Q^E \times Q^P \times Q^{Hard} \times Q^{Soft}. \quad (7)$$

Управлять ИИБ, исходя из состояния ИС, означает сделать выбор управляющего воздействия из множества ($U_{доп}$) допустимых управляющих воздействий на основе множеств ИИБ и состояний ИС в каждый момент времени (T) функционирования ИС. Другими словами, для каждой комбинации возникших инцидентов, состояний системы и моментов времени существует предписанное управляющее воздействие. Существует множество отображений (A) булеана (B) во множество допустимых управляющих воздействий для каждого уровня ИС:

$$B^{ИС} = \begin{cases} B^E \cup B^P \cup B^{Hard} \cup B^{Soft} \\ B^E = B(S^E \cup Q^E \cup T) \\ B^P = B(S^P \cup Q^P \cup T) \\ B^{Hard} = B(S^{Hard} \cup Q^{Hard} \cup T) \\ B^{Soft} = B(S^{Soft} \cup Q^{Soft} \cup T) \end{cases}, \quad (8)$$

$$A = A^E \times A^P \times A^{Hard} \times A^{Soft}: \begin{cases} B^E \xrightarrow{A^E} U^E \\ B^P \xrightarrow{A^P} U^P \\ B^{Hard} \xrightarrow{A^{Hard}} U^{Hard} \\ B^{Soft} \xrightarrow{A^{Soft}} U^{Soft} \end{cases}. \quad (9)$$

Фактически, множество A содержит субстанциальные закономерности (существенные для достижения цели деятельности ИС [10]), которые описывают процессы в ИС (закрыли порт на межсетевом экране – трафик не идет, сократили бюджет на выплату денежного содержания специалистов – отток кадров). Субстанциальные закономерности могут быть представлены аналитически, статистически, таблично и т.д.

Один из вариантов представления элемента множества допустимых управляющих воздействий $u \in U_{\text{доп}}$ – кортеж с названием действия (u_n) и снижением вероятности риска (Δp):

$$u_i = \langle u_n, \Delta p \rangle_{i \in U}. \quad (10)$$

С одной стороны, включение защитных мер иногда ограничивает выполнение защищаемых бизнес-процессов. С другой стороны, если ограничено выполнение бизнес-процессов, значит нарушен аспект информационной безопасности «доступность». Поэтому формирование интегральной характеристики «снижение вероятности риска» является отдельной задачей и выходит за рамки данного исследования. Полагается, что снижение вероятности риска учитывает влияние на бизнес-процессы компании.

Сопоставление каждому допустимому действию конкретной величины вероятности риска выполняется, например, с помощью метода iSOFT [10], который предполагает на основе морфологического ящика последовательный поиск субстанциальных закономерностей, связывающих величины между собой.

По проблематике получения количественной оценки вероятности риска опубликовано значительное количество научных работ. В исследовании [11], рассмотрена UTEM – технико-экономическая модель с повышенной точностью расчета вероятности инцидентов и потерь в IaaS/PaaS/SaaS. Метод MAGIC предложен в [12] для оценки вероятности инцидента на основе анализа «киберпозиции» организации. В работе [13] обоснован упрощенный алгоритм, который сочетает рейтинги риска, защитных мер и угроз на основе показателей инфраструктуры. Моделирование вероятности и масштаба киберрисков на уровне предприятий с реальными данными из базы операционных убытков с применением статистических методов, в том числе peaks-over-threshold (POT) моделей, copula-моделирования и эмпирического распределения для оценки зависимости между частотой и тяжестью инцидентов, проработано в [14].

Таким образом, проблематика оценки вероятности риска ИБ достаточно хорошо изучены и выходят за рамки данной статьи.

Кортеж (9) может быть дополнен другими показателями: скорость реализации, стоимость реализации, привлекаемые ресурсы и т.д. В этом случае задача становится многокритериальной, и может быть реализована специальным ПО при практическом применении. Многокритериальность добавит в постановку задачи дополнительные ограничения либо может учитываться при формировании множества S , что не влияет на авторский подход и будет учтена в последующем.

Предлагается следующая формулировка задачи управления ИИБ.

Постановка задачи управления ИИБ на основе операторного представления

Дано

$T \ni t$ – множество моментов времени функционирования ИС;

$Q \ni q$ – множество возможных ИИБ;

$S \ni s$ – множество состояний ИС;

P_{risk}^{\min} – минимальное допустимое значение вероятности риска;

P_{risk}^{\max} – максимальное допустимое значение вероятности риска;

$U_{\text{доп}}$ – множество допустимых действий на каждом уровне ИС;

$A \ni a$ – множество отображений булеана во множество действий на каждом уровне ИС.

Требуется

Найти такое управляющее воздействие, которое при возникновении инцидента ИИБ удержит вероятность риска в заданном диапазоне за заданный интервал времени:

$$\forall b \in B^{\text{ИС}} \exists u \in U_{\text{доп}}: P_{\text{risk}} \in [P_{\text{risk}}^{\min}, P_{\text{risk}}^{\max}]. \quad (11)$$

То есть найти оператор, рассчитывающий вероятность риска:

$$R_U(Q, S, A, U_{\text{доп}}) = P_{\text{risk}} \in [P_{\text{risk}}^{\min}, P_{\text{risk}}^{\max}], \quad (12)$$

где R_U – оператор, отображающий множества $Q, S, A, U_{\text{доп}}$ в интервал $[0; 1]$.

Согласно принципу Беллмана, оператор может быть составлен отдельно для каждого уровня ИС и представлять собой композицию операторов для каждого уровня ИС:

$$\begin{aligned} R_U = & R_U^E(Q^E, S^E, A^E, U_{\text{доп}}^E) \circ R_U^P(Q^P, S^P, A^P, U_{\text{доп}}^P) \circ \\ & \circ R_U^{\text{Hard}}(Q^{\text{Hard}}, S^{\text{Hard}}, A^{\text{Hard}}, U_{\text{доп}}^{\text{Hard}}) \circ \\ & \circ R_U^{\text{Soft}}(Q^{\text{Soft}}, S^{\text{Soft}}, A^{\text{Soft}}, U_{\text{доп}}^{\text{Soft}}). \end{aligned} \quad (13)$$

Подставив выражение (5) в выражение (6), и преобразовав в неравенство, получим:

$$P_{\text{risk}}^{\min} \leq 1 - (1 - P_{\text{risk}}^E)(1 - P_{\text{risk}}^P)(1 - P_{\text{risk}}^{\text{Hard}})(1 - P_{\text{risk}}^{\text{Soft}}) \leq P_{\text{risk}}^{\max}. \quad (14)$$

Это означает, что у специалиста ИБ расширяется выбор доступных вариантов в ходе управления ИИБ, так как появляется возможность:

- 1) управлять ИИБ на разных уровнях ИС,
- 2) использовать инструмент обоснования своих действий и затрат на всех уровнях ИС,
- 3) компенсировать невозможность реагирования на одном уровне ИС реагированием на другом уровне ИС.

Фактически, найденный оператор представляет собой набор playbooks, в которых каждое действие имеет свою оценку эффективности, например, процент снижения вероятности риска.

Поскольку метауровни формируют допустимые состояния вложенных уровней, то применение выражения (6) выполняется последовательно от обеспечивающего уровня к уровню ПО, то есть сначала применяются допустимые действия обеспечивающего уровня, затем П, далее уровня АО и уровня ПО.

Если возможно применение нескольких допустимых действий одновременно, то расчёт нового значения вероятности риска выполняется последовательно для каждого допустимого действия. Поскольку выражения (4) и (5) основаны на операции умножения, которая обладает свойством коммутативности, то порядок выбора допустимых действий для расчёта новой вероятности риска не имеет значения.

Контрольный пример применения обоснования

Пусть имеет место многоэтапная атака с использованием фишинга с целью кражи данных. Предельно допустимая вероятность риска $P_{risk} \in [0,3;0,5]$.

Сценарий инцидента (Q):

- начальная компрометация: злоумышленники провели целевую фишинговую рассылку на сотрудников финансового отдела. Один из сотрудников перешел по ссылке и ввел свои учетные данные на поддельной странице входа в корпоративный портал;
- закрепление и разведка: атакующие, используя украденные учетные данные, успешно подключились к корпоративной сети через VPN, не защищенный многофакторной аутентификацией (MFA);
- движение и цель: злоумышленники получили доступ к общему сетевому диску, где хранились финансовые отчеты, и начал готовить их к эксфильтрации (упаковке в архив);
- срабатывание защиты: на этапе эксфильтрации сработала DLP-система, заблокировав передачу архива на внешний облачный сервис (например, ya.ru), и отправила алерт в SIEM.

Задача для специалистов ИБ: инцидент был оперативно локализован (учетная запись скомпрометированного пользователя заблокирована, пароль сброшен, доступ атакующих прерван). Однако теперь перед руководителем ИБ (CISO) стоит задача не просто «закрыть тикет», а представить руководству план действий по снижению вероятности повторения

подобных инцидентов и обосновать необходимые ресурсы. Это основное применение разработанного подхода.

Шаг 1. Оценка исходного состояния рисков (до инцидента) CISO, используя модель, оценивает текущий уровень вероятности риска P_{risk} для каждого уровня ИС. Эти оценки не берутся «с потолка», а основываются на данных аудитов, пентестов и здравом смысле.

Обеспечивающий уровень. $P_{risk}^E = 0,2$. Обоснование: политики ИБ существуют, но формальны. Бюджет на ИБ выделяется по остаточному принципу.

Уровень персонала. $P_{risk}^P = 0,3$. Обоснование: обучение по кибербезопасности (security awareness) не проводилось два года. Результаты последней симуляции фишинга показали высокий click-rate (30 %).

Уровень АО. $P_{risk}^{Hard} = 0,1$. Обоснование: с оборудованием в целом все в порядке, но отсутствуют аппаратные токены для MFA.

Уровень ПО. $P_{risk}^{Soft} = 0,4$. Обоснование: критически важный фактор: на VPN-шлюзе не включен MFA. DLP-система есть, но настроена в режиме «только мониторинг» для многих правил, чтобы не мешать бизнес-процессам.

Рассчитываем общую вероятность риска по формуле (5):

$$P_{risk} = (1 - (1 - 0,2)(1 - 0,3)(1 - 0,1)(1 - 0,4)) = 0,6976.$$

Общая вероятность риска выходит за пределы допуска, что инцидент и подтвердил.

Шаг 2. Формирование множества допустимых действий ($U_{доп}$) и оценка их эффективности (Δp). CISO вместе с командой формирует набор возможных компенсирующих мер (табл. 3).

Ключевой момент: оценка Δp не является экспертным «гаданием», а привязывается к конкретным метрикам и отраслевым данным.

Шаг 3. Анализ сценариев и принятие решения. Теперь CISO может просчитать несколько стратегий и представить их руководству.

Стратегия 1: «Дешево и быстро» (только организационные меры).

Действия: провести обучение ($\Delta p^P = 30\%$), перевести DLP в блокировку ($\Delta p^{Soft} = 40\%$).

Новые значения вероятностей риска:

$$P_{risk}^P = 0,3(1 - 0,3) = 0,21, P_{risk}^{Soft} = 0,4(1 - 0,4) = 0,24, \\ P_{risk} = 1 - (1 - 0,2)(1 - 0,21)(1 - 0,1)(1 - 0,24) = 0,568.$$

Вывод: вероятность риска снижена, но цель $[0,3; 0,5]$ не достигнута.

Стратегия 2: «Комплексная защита» (внедрение MFA).

Действия: выделить бюджет ($\Delta p^E = 5\%$), провести обучение ($\Delta p^P = 30\%$), внедрить MFA ($\Delta p^{Soft} = 90\%$).

Набор допустимых действий

Уровень	$\langle u_n, \Delta p \rangle$		Обоснование для Δp
	Действие (u_n)	Снижение Δp , %	
Обеспечивающий	Утвердить и выделить бюджет на проект внедрения MFA	5	Действие не снижает риск, но разблокирует возможность технических мер. Эффект проявится через снижение рисков на других уровнях. Оценка вклада 5 % за счет формализации процесса.
Персонал	Провести обязательное обучение и фишинг-симуляцию для фин. отдела	30	По данным Verizon DBIR и собственным тестам, эффективное обучение снижает click-rate в 2-3 раза. Если текущий риск персонала на 50 % обусловлен фишингом, то снижение click-rate на 60 % понизит P_{risk}^P на 30 %.
АО	Вариант А: включить MFA для всех VPN-пользователей	90	MFA блокирует >99 % атак, связанных с компрометацией учетных данных. Эффективное действие против этого вектора. Снижает риск до нуля.
ПО	Вариант Б: перевести правила DLP в режим блокировки	40	Не предотвратит вход, но остановит кражу данных. Эффективно для снижения ущерба, но слабо влияет на вероятность успешной атаки. Снижение P_{risk}^{Soft} на 40 %.

Новые значения вероятностей риска:

$$P_{risk}^E = 0,19, P_{risk}^E = 0,21, P_{risk}^{Soft} = 0,04.$$

$$P_{risk} = 1 - (1 - 0,19)(1 - 0,21)(1 - 0,1)(1 - 0,04) = 0,447.$$

Вывод: предложенные меры позволяют снизить вероятность риска до допустимого значения.

Преимущества предложенного подхода для CISO-практик

Язык для общения с бизнесом: вместо «нам нужен MFA, потому что это безопасно», CISO говорит: «Стратегия 1 стоит X условных единиц, и мы остаемся в красной зоне риска. Стратегия 2 стоит Y условных единиц, но она переводит нас в приемлемую желтую зону, снижая вероятность инцидента на Z %. Вот расчеты».

Обоснованность: оценка Δp привязана к измеримым показателям (click-rate, отраслевая статистика), а не к субъективным ощущениям. Защищает от вопроса «А почему вы решили, что это поможет?».

Гибкость: модель показывает, что нет единственного идеального решения. Можно комбинировать менее эффективные, но дешевые меры на разных уровнях, чтобы достичь приемлемой вероятности риска.

Структурированный подход: весь процесс от анализа инцидента до формирования плана улучшений уложен в понятный фреймворк.

Порядок обоснования процедуры мониторинга и реагирования на ИИБ

Этап 1. Оценка исходного состояния («As-Is»)

Определение области моделирования: выбрать критически важный бизнес-процесс или систему для

анализа (например, атлас ЧС, корпоративная почта, сегмент АСУ ТП).

Декомпозиция по уровням: описать выбранный процесс в терминах четырех уровневой модели:

- **обеспечивающий:** «Какие политики, бюджеты, регламенты на нее влияют?»;
- **персонал:** «Кто работает с процессом (пользователи, администраторы, разработчики)?»;
- **аппаратный:** «На каком оборудовании она работает (серверы, СХД, сетевое оборудование)?»;
- **программный:** «Какое ПО используется (ОС, СУБД, прикладное ПО)?».

Оценка текущей вероятности риска (P_{risk}).

Для каждого из четырех уровней оценить текущую вероятность возникновения инцидента (P_{risk}^E , P_{risk}^P , P_{risk}^{Hard} , P_{risk}^{Soft}).

Что делать на практике: использовать данные аудитов, результаты пентестов, статистику инцидентов, отчеты СЗИ, экспертные оценки, сформировать множества S (текущие состояния) и Q (вероятные инциденты).

Связь с моделью статьи: это начальные значения P_{risk} для выражения (5), как в «Шаге 1» контрольного примера.

Этап 2. Определение целевого состояния («To-Be»)

Установка допустимого уровня риска: совместно с руководством и владельцами критического процесса определить целевой диапазон для совокупной вероятности риска $[P_{risk}^{min}, P_{risk}^{max}]$.

Что делать на практике: перевести технические проценты на язык бизнеса с помощью лингвистических

переменных («зеленая», «желтая», «красная» зоны). Связывать вероятность риска со стоимостью ущерба от ИИБ.

Связь с моделью статьи: это правая часть «операторного уравнения» (6).

Этап 3. Формирование набора инструментов (Playbooks)

Разработка каталога допустимых действий ($U_{\text{доп}}$). Для каждого уровня составить список возможных компенсирующих мер.

Что делать на практике: мозговой штурм с командой ИБ и ИТ. Примеры: «внедрить MFA», «провести обучение персонала», «купить новый файрвол», «обновить политику паролей».

Связь с моделью статьи: это формирование множества $U_{\text{доп}}$.

Оценка эффективности мер (Δp). Для каждого допустимого действия из каталога оценить, на сколько процентов оно снизит вероятность риска на своем уровне.

Что делаем на практике: самый сложный этап. Использовать отраслевую статистику (например, отчеты Verizon DBIR, которые говорят об эффективности MFA), результаты пилотных проектов, экспертные оценки.

Связь с моделью статьи: это присвоение значений Δp элементам из $U_{\text{доп}}$, как в табл. 4 контрольного примера. Сформировать множество отображений A .

Этап 4. Анализ сценариев и принятие решения.

Моделирование стратегий реагирования. Взять типичный или недавний инцидент и просчитать несколько вариантов (стратегий) реагирования, комбинируя меры из каталога допустимых действий.

Что делаем на практике: повторять «Шаг 3» из контрольного примера. Считать новую вероятность риска для «быстрого и дешевого» сценария, для «комплексного и дорогого» и т.д.

Связь с моделью статьи: это практическое решение «операторного уравнения» (6) – поиск такой комбинации действий, которая приводит P_{risk} в целевой диапазон.

Выбор и обоснование оптимальной стратегии. Представить руководству расчеты по разным стратегиям с указанием их стоимости и итогового уровня риска. Принять совместное решение.

Этап 5. Внедрение и контроль (цикл PDCA).

Реализация выбранной стратегии. Внедрить утвержденные меры.

Что делаем на практике: разработать SLAs, обновить инструкции, закупить оборудование, провести тренинги, разработать и утвердить руководящие документы.

Мониторинг и переоценка. Отслеживать ключевые показатели эффективности (KPIs), чтобы проверить, был ли достигнут ожидаемый эффект (Δp).

Периодически возвращаться к Этапу 1 и пересматривать оценки P_{risk} , чтобы модель оставалась актуальной. Период пересмотра может входить как параметр допустимого действия, снижающего общую вероятность риска P_{risk} .

Этап 6. Повторить для другого критического бизнес-процесса.

Адекватность предложенного подхода обеспечивается корректностью постановки задачи и использования математического аппарата и проверкой на практике в рамках сегмента киберполигона на базе Санкт-Петербургского университета ГПС МЧС России.

Выводы

Таким образом, предлагаемый подход позволяет интегрировать существующие подходы ко всей системе управления ИИБ и обосновать порядок мониторинга и реагирования на ИИБ.

Фундамент (Baseline): нормативно-правовой подход используется для выполнения обязательных требований регуляторов и закона. Это гигиенический минимум. Используется для формирования множеств T , S , A , $U_{\text{доп}}$ и отчасти множества Q , например, когда за основу берётся БДУ ФСТЭК России.

Ядро стратегии: риск-ориентированный подход определяет, куда направить основные усилия и ресурсы сверх базовых требований. Применяется для расчёта значений текущих и допустимых вероятностей риска, формирования множеств Q , A и $U_{\text{доп}}$.

Механизм реализации: процессный подход (PDCA) обеспечивает непрерывное управление и улучшение системы ИБ. Применяется для формирования множеств S , A и $U_{\text{доп}}$.

Продвинутые уровни: Zero Trust, Cyber Resilience и Threat Intelligence используются зрелыми организациями для построения проактивной и глубоко эшелонированной защиты от сложных современных атак. Применяется для формирования множеств Q , A и $U_{\text{доп}}$.

Ключевым практическим выводом модели является принцип «перераспределения риска». Позволяет отойти от догматичного требования устранить все уязвимости, перейти к экономически обоснованному управлению ИИБ.

Модель дает CISO математический инструмент для доказательства того, что инвестиции в усиление одного уровня (например, в security awareness персонала) могут принести гораздо больший «возврат на безопасность» (Return on Security Investment), чем дорогостоящие и порой безрезультатные попытки устранить трудноразрешимую проблему на другом уровне иерархии.

Предлагаемый подход является преимущественно стратегическим инструментом для CISO

и IR-менеджера для анализа, планирования и обоснования инвестиций в ИБ, для формирования playbooks.

Подход может использоваться как инструмент поддержки принятия решения SOC-аналитиком в момент атаки, в реальном масштабе времени оценивая эффективность различных сценариев реагирования, таким образом упрощая и ускоряя выбор допустимых действий SOC-аналитика. В этом случае

инструмент должен строиться на playbooks, существующих в компании, и расчёты инструмента будут автоматизированы.

В дальнейших исследованиях предусматривается проработка механизмов гармонизации регламентации мониторинга защищенности информационных ресурсов и реагирования на компьютерные инциденты в информационной инфраструктуре организаций с учетом предложенного подхода.

Статья подготовлена в рамках выполнения НИР «Кибермониторинг» по государственному заданию МЧС России (ЕГИСУ НИОКТР №125031703734-4).

Литература

1. Токарев В.Л. Интеллектуальная поддержка обнаружения инцидентов информационной безопасности / В.Л. Токарев, А.А. Сычугов // Моделирование, оптимизация и информационные технологии. 2023. Т. 11, № 1(40). С. 16-17. DOI 10.26102/2310-6018/2023.40.1.006.
2. Киселев А. А., Практика создания центра мониторинга информационной безопасности / А. А. Киселев, И. В. Коротких, В. В. Шотт // Безопасность цифровых технологий. 2022. Т.11, № 1(40). С.39–51.
3. Олейникова А. А. Концепция управления информационной безопасностью на основе цикла непрерывного детектирования и реагирования на инциденты безопасности информации / А. А. Олейникова, В. В. Золотарев // Известия ЮФУ. Технические науки. – 2023. – № 5(235). С. 66–81. DOI 10.18522/2311-3103-2023-5-66-81.
4. Гончаренко С. Н., Лачихина А. Б. Мониторинг инцидентов безопасности геоинформационной системы управления и контроля деятельности промышленного предприятия // Горный информационно-аналитический бюллетень (научно-технический журнал). – 2022. – №. 3. С. 108–116
5. Ma, C. Smart city and cyber-security; technologies used, leading challenges and future recommendations // Energy Reports. 2021. Т. 7. С. 7999–8012.
6. Kure, H. I., Islam, S., Mouratidis, H. An integrated cyber security risk management framework and risk predication for the critical infrastructure protection // Neural Computing and Applications. 2022. Т. 34, №. 18. Pp. 15241–15271.
7. Tuyishime, E. et al. Enhancing cloud security – proactive threat monitoring and detection using a siem-based approach // Applied Sciences. 2023. Т. 13, №. 22. С. 12359.
8. Краковский Ю. М., Киргизбаев В. П. Системный подход к моделированию работ по устранению инцидентов информационной безопасности применительно к корпоративной информационной системе // Современные технологии. Системный анализ. Моделирование. 2025. № 1(85). С. 116–126.
9. Обеспечение информационной безопасности интегрируемых информационных систем на базе доверия / В. В. Грызунов, А. С. Крюков, А. В. Шестаков, И. А. Зикратов // Труды учебных заведений связи. 2024. Т. 10, № 4. С. 110-125. DOI 10.31854/1813-324X-2024-10-4-110-125.
10. Грызунов В. В. Формирование условия гарантированного достижения цели деятельности информационной системой на базе операторного уравнения // Информатизация и связь. 2022. № 4. С. 67–74. DOI 10.34219/2078-8320-2022-13-4-67-74.
11. Bendicho, C. Cyber security in cloud: Risk assessment models // Intelligent Computing: Proceedings of the 2021 Computing Conference, Volume 1 – Cham : Springer International Publishing. 2021. Pp. 471–482. DOI 10.1007/978-3-030-80126-7_32.
12. Battaglioni M. et al. Magic: A method for assessing cyber incidents occurrence //IEEE Access. 2022. Т. 10. Pp. 73458–73473. DOI 10.1109/ACCESS.2022.3190246.
13. Badhwar, R. Simplified Approach to Calculate the Probability of a Cyber Event //The CISO's Next Frontier: AI, Post-Quantum Cryptography and Advanced Security Paradigms. – Cham : Springer International Publishing. 2021. Pp. 353–359. DOI 10.1007/978-3-030-79623-5_15.
14. Zängerle, D., Schiereck, D. Modelling and predicting enterprise-level cyber risks in the context of sparse data availability // The Geneva Papers on Risk and Insurance-Issues and Practice. 2023. Т. 48, № 2. С. 434–462. DOI 10.1057/s41288-022-00282-6.

A MULTI-LEVEL FRAMEWORK FOR JUSTIFYING INFORMATION SECURITY INCIDENT MONITORING AND RESPONSE PROCEDURES

Gryzunov V. V.³, Shestakov A. V.⁴

Keywords: terminology system, cybersecurity, information security incident management.

Purpose of the study: to substantiate approaches for the rational organization of monitoring and response to potential information security incidents.

Methods of research: the information system is described by a hierarchical model, developed and tested by the authors, which includes the management level, the personnel level, as well as the hardware and software levels. The model takes as input the sets of possible states of the information system and information security incidents, the admissible values of information security risk probabilities, and the set of admissible control actions. The model outputs a control action that maintains the information security risk probability within a specified range.

Results: the multi-level framework provides an end-to-end description of the information security incident management process – from incident detection to the formation of an improvement plan. It allows for quantitative assessment of the effectiveness of response measures at each level of the information system – management, personnel, hardware, and software. A key advantage of the approach is the ability to compensate for the shortcomings of one level by strengthening another, which expands the possibilities for keeping the information security risk probability within a specified range. Based on the obtained assessments, economically justified information security strategies are formed. The central practical conclusion is the principle of "risk redistribution," which replaces the dogmatic requirement to eliminate all vulnerabilities with targeted, measurable, and cost-effective management of the risk probability.

Scientific novelty: unlike existing models, the developed model explicitly accounts for the aggregate information security risk probability and the interdependence of all four levels of the information system by forming a corresponding operator.

References

1. Tokarev V. L. Intellekтуальная поддержка обнаружения инцидентов информационной безопасности / V. L. Tokarev, A. A. Sychugov // Моделирование, оптимизация и информационные технологии. 2023. Т. 11, № 1(40). С. 16-17. DOI 10.26102/2310-6018/2023.40.1.006.
2. Kiselev A. A., Praktika sozdaniya centra monitoringa informacionnoj bezopasnosti / A. A. Kiselev, I. V. Korotkih, V. V. Shott // Bezopasnost' cifrovyykh tekhnologiy. 2022. Т.11, № 1(40). С. 39–51.
3. Olejnikova A. A. Konceptiya upravleniya informacionnoj bezopasnost'ju na osnove cikla nepreryvnogo detektirovaniya i reagirovaniya na incidenty bezopasnosti informacii / A. A. Olejnikova, V. V. Zolotarev // Izvestiya JuFU. Tehnicheskie nauki. – 2023. – № 5(235). С. 66–81. DOI 10.18522/2311-3103-2023-5-66-81.
4. Goncharenko S. N., Lachihina A. B. Monitoring incidentov bezopasnosti geoinformacionnoj sistemy upravleniya i kontrolja dejatel'nosti promyshlennogo predpriyatiya // Gornyj informacionno-analiticheskij bjulleten' (nauchno-tehnicheskij zhurnal). – 2022. – №. 3. С. 108–116
5. Ma, C. Smart city and cyber-security; technologies used, leading challenges and future recommendations // Energy Reports. 2021. Т. 7. С. 7999–8012.
6. Kure, H. I., Islam, S., Mouratidis, H. An integrated cyber security risk management framework and risk predication for the critical infrastructure protection // Neural Computing and Applications. 2022. Т. 34, №. 18. Р. 15241–15271.
7. Tuyishime, E. et al. Enhancing cloud security – proactive threat monitoring and detection using a siem-based approach // Applied Sciences. 2023. Т. 13, №. 22. С. 12359.
8. Krakovskij Ju. M., Kirgizbaev V. P. Sistemnyj podhod k modelirovaniyu rabot po ustraneniyu incidentov informacionnoj bezopasnosti primenitel'no k korporativnoj informacionnoj sisteme // Sovremennye tekhnologii. Sistemnyj analiz. Modelirovanie. 2025. № 1(85). С. 116-126.
9. Obespechenie informacionnoj bezopasnosti integriruemyykh informacionnykh sistem na baze doverija / V. V. Gryzunov, A. S. Krjukov, A. V. Shestakov, I. A. Zikratov // Trudy uchebnykh zavedenij svyazi. 2024. Т. 10, № 4. С. 110–125. DOI 10.31854/1813-324X-2024-10-4-110-125.
10. Gryzunov V. V. Formirovanie uslovija garantirovannogo dostizheniya celi dejatel'nosti informacionnoj sistemoy na baze operatornogo uravneniya // Informatizacija i svyaz'. 2022. № 4. С. 67–74. DOI 10.34219/2078-8320-2022-13-4-67-74.
11. Bendicho, C. Cyber security in cloud: Risk assessment models // Intelligent Computing: Proceedings of the 2021 Computing Conference, Volume 1 – Cham : Springer International Publishing. 2021. Р. 471–482. DOI 10.1007/978-3-030-80126-7_32.
3. Vitaly V. Gryzunov, Dr.Sc. of Technical Sciences, Associate Professor, Professor of the Department of Applied Mathematics and Information Technologies of the St. Petersburg University of the State Fire Service of the Ministry of Emergency Situations of Russia. St. Petersburg, Russia. ORCID <https://orcid.org/0000-0003-4866-217X>. E mail: viv1313r@mail.ru
4. Alexander V. Shestakov, Dr.Sc. of Technical Sciences, Senior Researcher, Leading Researcher of the St. Petersburg University of the State Fire Service of the Ministry of Emergency Situations of Russia. St. Petersburg, Russia. ORCID <https://orcid.org/0000-0002-8462-6515>. E mail: alexandr.shestakov01@yandex.ru

12. Battaglioni M. et al. Magic: A method for assessing cyber incidents occurrence //IEEE Access. 2022. Т. 10. Rr. 73458-73473. DOI 10.1109/ACCESS.2022.3190246.
13. Badhwar, R. Simplified Approach to Calculate the Probability of a Cyber Event //The CISO's Next Frontier: AI, Post-Quantum Cryptography and Advanced Security Paradigms. – Cham : Springer International Publishing. 2021. Rr. 353-359. DOI 10.1007/978-3-030-79623-5_15.
14. Zängerle, D., Schiereck, D. Modelling and predicting enterprise-level cyber risks in the context of sparse data availability // The Geneva Papers on Risk and Insurance-Issues and Practice. 2023. Т. 48, № 2. С. 434–462. DOI 10.1057/s41288-022-00282-6.



ПРИМЕНЕНИЕ МЕТОДА АНАЛИЗА ИЕРАРХИЙ И МЕТОДА АНАЛИТИЧЕСКИХ СЕТЕЙ ДЛЯ ОЦЕНКИ ПРЕДПОЧТИТЕЛЬНОСТИ СИСТЕМ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

Аносов Р. С.¹, Аносов С. С.²

DOI: 10.21681/2311-3456-2025-6-25-34

Цель: сравнительная оценка методов анализа иерархий и аналитических сетей и их применимости для исследования систем информационной безопасности.

Метод исследования: системный анализ и принятие решений, в том числе систематизация, структуризация, декомпозиция, композиция; методы теории графов; матричное исчисление.

Результат исследования: показана применимость метода анализа иерархий и метода аналитических сетей для решения задачи выбора одного из альтернативных вариантов системы информационной безопасности. Метод аналитических сетей в сравнении с методом анализа иерархий более ресурсоемок, но позволяет нюансировано исследовать систему информационной безопасности на основе анализа взаимных влияний и зависимостей компонентов системы, обеспечивая высокую степень обоснованности решений, принимаемых по результатам анализа. Определено значение технологических показателей в исследовании системы информационной безопасности, которые, с одной стороны, характеризуют эффективность применяемых в системе технологий с точки зрения показателей качества защищаемой информации, а с другой стороны – степень реализации соответствующих технологий в системе информационной безопасности. Разработана обобщенная классификация информационных технологий и технологий обеспечения информационной безопасности, выполнена структуризация системы процессов, технологий и факторов в контексте системного анализа информационных технологий, что является основой для построения иерархических и сетевых структур при исследовании систем информационной безопасности.

Научная новизна: показано, что, в сравнении с методом анализа иерархий, метод аналитических сетей и его базовые инструменты – сети и суперматрицы – позволяют детально описать реально существующие взаимные влияния различных характеристик и свойств систем информационной безопасности и формализовать задачу оценки предпочтительности указанных систем с учетом таких влияний.

Ключевые слова: вариант системы, показатель качества, суперматрица, вектор влияния, принятие решений.

Введение

Задача формирования, оценки и выбора вариантов системы информационной безопасности (ИБ) является многокритериальной, ее решение требует учета значительного количества разнородных показателей, характеризующих различные аспекты качества [1–4]. Анализ научных публикаций свидетельствует об интенсивной деятельности в области методических подходов к исследованию различных аспектов систем ИБ, а также подходов к системному обобщению отдельных методик и моделей в форме программных комплексов и систем поддержки принятия решений [5]. Однако уровень научно-методического обеспечения обоснования требований и оценки систем ИБ, по мнению авторов, не соответствует уровню развития аналогичной методологии

для сложной машиностроительной продукции и не удовлетворяет требованиям эффективного проектирования³, что в условиях относительно высокой ресурсоемкости жизненного цикла систем ИБ порождает проблему формирования обоснованного бюджета на информационную безопасность.

Технологический аспект информационной безопасности включает как технологии защиты (рис. 1), так и информационные технологии – объекты защиты (рис. 2).

Многоуровневость и функциональная связность информационных технологий обуславливает возможность применять для системного исследования информационной безопасности методический аппарат иерархических и сетевых структур [6–8]:

1 Аносов Роман Сергеевич, кандидат технических наук, доцент, начальник отдела, Военный учебно-научный центр Военно-воздушных сил «Военно-воздушная академия имени профессора Н. Е. Жуковского и Ю. А. Гагарина», г. Воронеж, Россия. E-mail: an_rs@list.ru

2 Аносов Сергей Сергеевич, главный специалист, общество с ограниченной ответственностью «Газпромнефть – Цифровые решения», г. Воронеж, Россия. E-mail: serg-anosov@mail.ru

3 Надежность и эффективность в технике: справочник в 10 т. / Ред. совет: В. С. Авдеевский [и др.] // Т.3. Эффективность технических систем / Под общ. ред. В.Ф. Уткина, Ю.В. Крючкова – М.: Машиностроение, 1988. – 328 с.

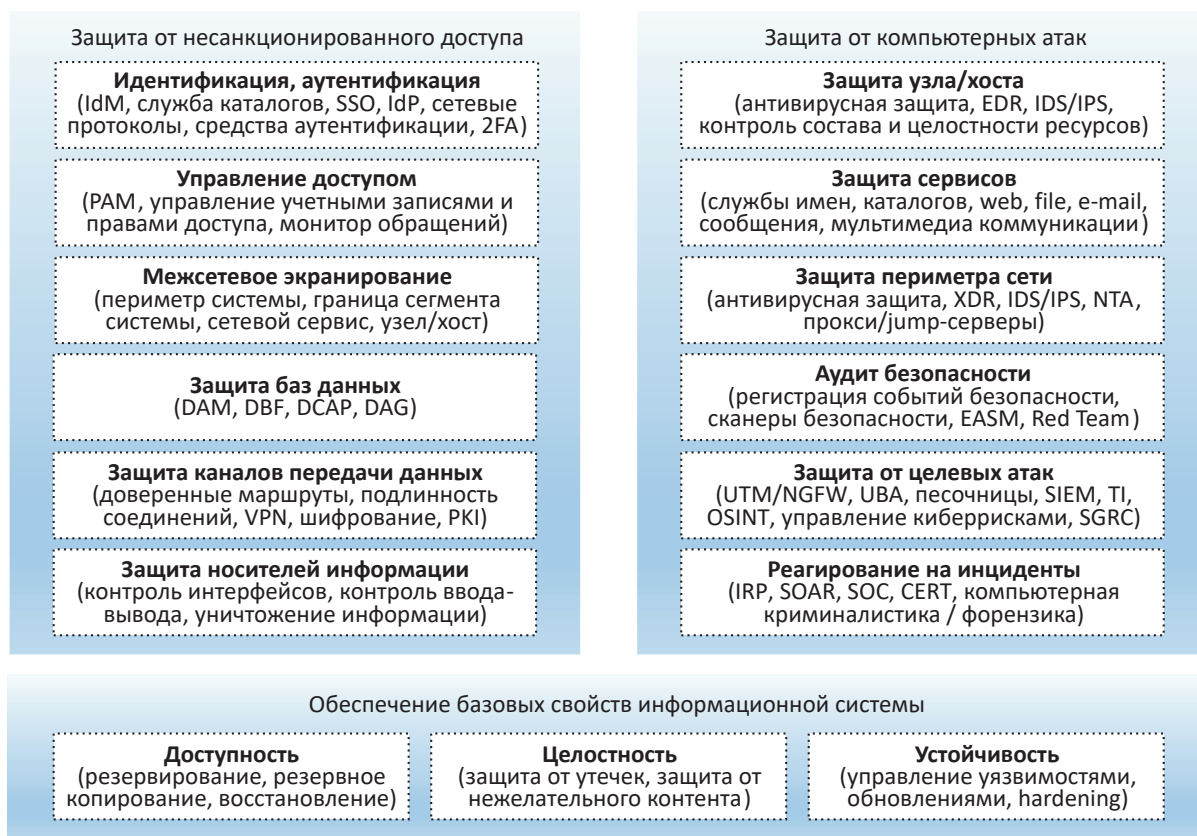


Рис. 1. Типовая обобщенная структура технологий систем ИБ

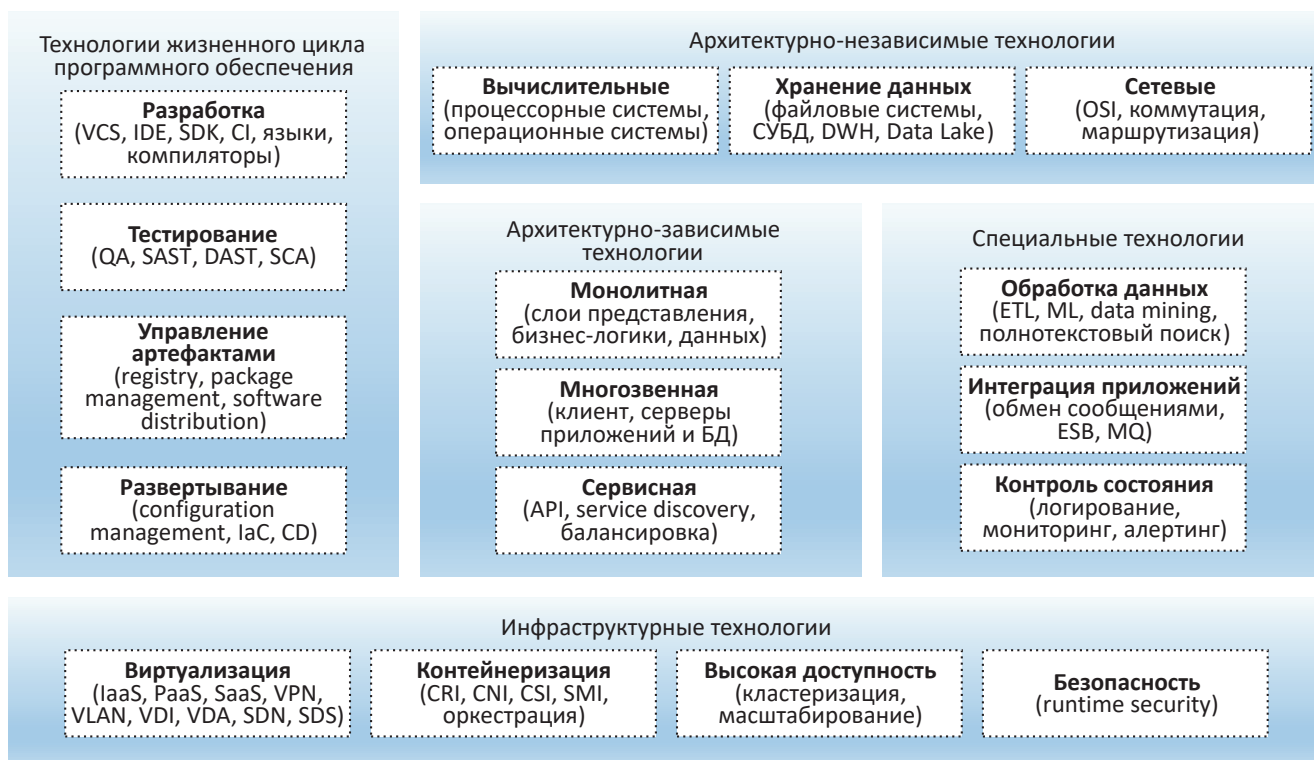


Рис. 2. Типовая обобщенная структура технологий информационных систем

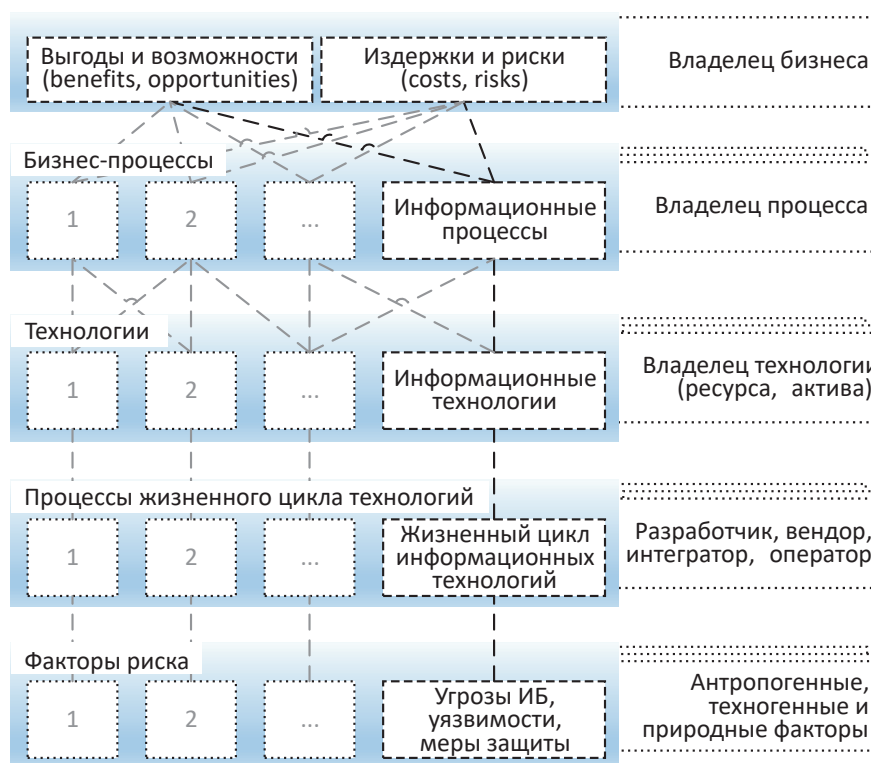


Рис. 3. Стратификация процессов, технологий и факторов в контексте системного анализа информационных технологий

- логико-вероятностный метод оценивания сложных (многоагентных) систем с использованием элементов теории множеств и графов [9];
- метаграмматический подход к структурно-параметрическому синтезу вариантов систем ИБ и метод анализа иерархий Саати для их анализа [10];
- сети Байеса на основе графа атак и когнитивные карты как средства оценки рисков ИБ⁴;
- модель информационного управления агентами в игре с несогласованными представлениями на основе линейных когнитивных карт для построения конфликтно-устойчивой системы ИБ⁵;
- теоретико-игровой подход с использованием когнитивной игры для управления рисками сложной системы на основе распределения ограниченного ресурса между ее элементами [11];
- подход к оценке киберустойчивости компьютерных сетей, основанный на аналитическом моделировании компьютерных атак с применением метода преобразования стохастических сетей⁶.

В контексте системного анализа иерархические и сетевые структуры используются как для декомпозиции систем на подсистемы, так и для их композиции в надсистемы, обеспечивая органичную связь между различными уровнями (рис. 3). В этом смысле приведенные на рисунках 1 и 2 системы технологий рассматриваются как составные компоненты (подсистемы) информационных систем, реализующих или обеспечивающих процессы бизнес-уровня.

Цель настоящей работы – рассмотреть особенности иерархического и сетевого подходов к принятию решений при оценке предпочтительности систем ИБ.

Иерархический подход

Конструктивным экспертно-аналитическим методическим аппаратом, позволяющим проводить относительные оценки предпочтительности вариантов сложных систем ИБ при их проектировании, является метод анализа иерархий (МАИ)⁷. При применении МАИ задача выбора альтернативных вариантов рассматривается как многокритериальная задача принятия решения, где разнородные частные показатели качества вариантов сворачиваются в единый обобщенный показатель в соответствии с их иерархической структурой. На рисунке 4 приведен пример иерархической структуры с **тремя** уровнями:

4 Сравнительный анализ алгоритмов когнитивного моделирования при оценке рисков информационной безопасности / М. Б. Гузаиров, А. М. Вульф-фин, В. М. Картак [и др.] // Труды Института системного анализа Российской академии наук. – 2019. – Т. 69, № 4. – С. 62–69.

5 Куливец С. Г. Модель информационного управления на основе игры на линейной когнитивной карте / С. Г. Куливец, Н. А. Коргин // Управление большими системами: сборник трудов. – 2011. – № 35. – С. 94–113.

6 Оценка киберустойчивости компьютерных сетей на основе моделирования кибератак методом преобразования стохастических сетей / И. В. Котенко, И. Б. Саенко, М. А. Коцыняк, О. С. Лаута // Труды СПИИРАН. – 2017. – № 6(55). – С. 160–184

7 Саати Т. Принятие решений. Метод анализа иерархий: Пер. с англ. М.: Радио и связь, 1993. – 314 с. Саати Т. Кернс К. Аналитическое планирование. Организация систем / Пер. с англ. под ред. И. А. Ушакова. – М.: Радио и связь, 1991. – 224 с.

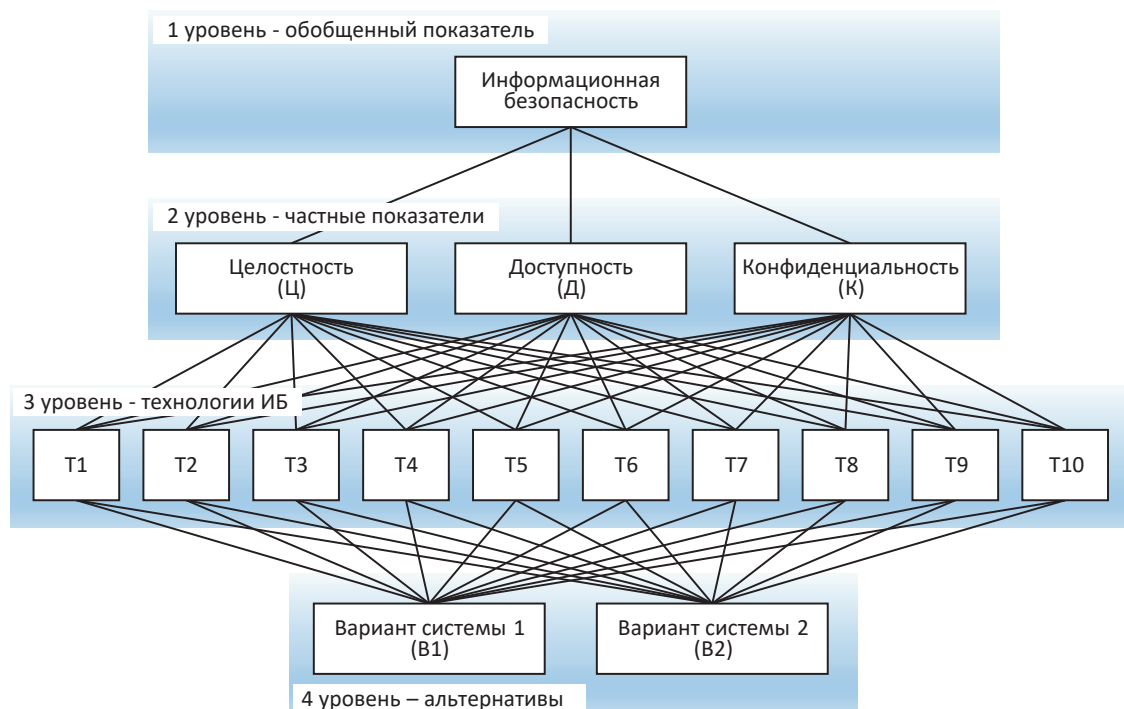


Рис. 4. Иерархическое представление задачи сравнения альтернативных вариантов системы ИБ

- **первый (верхний) уровень** – обобщенный показатель качества;
- **второй уровень** – частные показатели качества;
- **третий уровень** – технологические показатели.

Нижний уровень иерархии представлен вариантами системы (альтернативами).

Подобное сведение взаимозависимых сущностей в иерархическую структуру с единственным обобщенным (глобальным) показателем в вершине позволяет:

- сформировать альтернативы, сбалансированные по совокупности свойств и характеристик в соответствии с их важностью относительно обобщенного показателя (задача синтеза);
- выбрать альтернативу, наиболее предпочтительную по совокупности частных показателей и обладающую наилучшим значением обобщенного показателя (задача анализа).

Возможность обоснованной компенсации низкого (недостаточного) уровня одного из свойств (значения одной из характеристик) создаваемой системы ИБ повышением уровня других ее свойств (значений других характеристик), в том числе конфликтующих с первыми, является несомненным преимуществом МАИ. Во многих сложных технико-экономических задачах смысл заключен в оптимизации не столько абсолютных значений величин, сколько их соотношений.

Процедура оценки альтернативных вариантов по обобщенному показателю включает следующие этапы.

На **первом этапе** элементы каждого уровня иерархии попарно сравниваются между собой по отношению к элементам вышестоящего уровня с использованием шкалы предпочтения (принята шкала от 1 до 10). В случае иерархии, приведенной на рисунке 4, сравнения проводятся относительно третьего, второго и первого уровня иерархии. Результаты попарных сравнений записываются в форме квадратных матриц, примеры которых приведены на рисунках 5–7. Координаты собственных векторов матриц v являются величинами относительной важности (предпочтительности, приоритетности) сравниваемых элементов по отношению к соответствующему элементу вышестоящего уровня иерархии.

ИБ	Ц	Д	К	v
Ц	1	1/5	1/2	0,128
Д	5	1	2	0,595
К	2	1/2	1	0,277

Рис. 5. Матрица сравнений элементов второго уровня иерархии относительно обобщенного показателя (ИБ)

На **втором этапе** для каждого уровня иерархии формируются матрицы, столбцами которых являются собственные векторы матриц попарных сравнений элементов соответствующего уровня. В рассматриваемом случае при сравнении относительно третьего уровня формируется матрица W_{43} :

Ц	<i>T1</i>	<i>T2</i>	<i>T3</i>	<i>T4</i>	<i>T5</i>	<i>T6</i>	<i>T7</i>	<i>T8</i>	<i>T9</i>	<i>T10</i>	ν
<i>T1</i>	1	2	9	9	9	2	3	7	5	3	0,271
<i>T2</i>	1/2	1	9	9	9	5	4	3	5	3	0,245
<i>T3</i>	1/9	1/9	1	1	1	1/4	1/2	3	2	4	0,049
<i>T4</i>	1/9	1/9	1	1	1	1/10	1/10	1/8	5	8	0,033
<i>T5</i>	1/9	1/9	1	1	1	1/5	1/2	4	2	3	0,048
<i>T6</i>	1/2	1/5	4	10	5	1	2	6	4	2	0,146
<i>T7</i>	1/3	1/4	2	10	2	1/2	1	2	6	4	0,107
<i>T8</i>	1/7	1/3	1/3	8	1/4	1/6	1/2	1	7	6	0,055
<i>T9</i>	1/5	1/5	1/2	1/5	1/2	1/4	1/6	1/7	1	2	0,024
<i>T10</i>	1/3	1/3	1/4	1/8	1/3	1/2	1/4	1/6	1/2	1	0,022

Д	<i>T1</i>	<i>T2</i>	<i>T3</i>	<i>T4</i>	<i>T5</i>	<i>T6</i>	<i>T7</i>	<i>T8</i>	<i>T9</i>	<i>T10</i>	ν
<i>T1</i>	1	1	1/10	1/10	1/10	1/2	1	1	1/5	1/5	0,02
<i>T2</i>	1	1	1/10	1/10	1/10	1/2	1	1	1/10	1/10	0,017
<i>T3</i>	10	10	1	3	5	9	9	5	7	5	0,32
<i>T4</i>	10	10	1/3	1	1/5	9	9	1	2	1	0,119
<i>T5</i>	10	10	1/5	5	1	9	9	3	5	7	0,232
<i>T6</i>	2	2	1/9	1/9	1/9	1	1/10	1/10	1/10	1/10	0,014
<i>T7</i>	1	1	1/9	1/9	1/9	10	1	1/10	1/10	1/10	0,019
<i>T8</i>	1	1	1/5	1	1/3	10	10	1	1	1	0,071
<i>T9</i>	5	10	1/7	1/2	1/5	10	10	1	1	1	0,091
<i>T10</i>	5	10	1/5	1	1/7	10	10	1	1	1	0,097

К	<i>T1</i>	<i>T2</i>	<i>T3</i>	<i>T4</i>	<i>T5</i>	<i>T6</i>	<i>T7</i>	<i>T8</i>	<i>T9</i>	<i>T10</i>	ν
<i>T1</i>	1	2	9	9	9	1	1/6	1/5	1/5	1/5	0,068
<i>T2</i>	1/2	1	1	2	2	1	1/10	1/5	1/5	1/5	0,034
<i>T3</i>	1/9	1	1	1	2	1	1/10	1/10	1/10	1/10	0,022
<i>T4</i>	1/9	1/2	1	1	1	3	1/5	1/10	1/10	1/10	0,023
<i>T5</i>	1/9	1/2	1/2	1	1	2	1/5	1/10	1/10	1/10	0,02
<i>T6</i>	1	1	1	1/3	1/2	1	1/3	1/5	1/5	1/5	0,029
<i>T7</i>	6	10	10	5	5	3	1	1	1	1	0,186
<i>T8</i>	5	5	10	10	10	5	1	1	1	1	0,206
<i>T9</i>	5	5	10	10	10	5	1	1	1	1	0,206
<i>T10</i>	5	5	10	10	10	5	1	1	1	1	0,206

Рис. 6. Матрицы сравнений элементов третьего уровня иерархии относительно частных показателей (Ц, Д, К)

T1	<i>B1</i>	<i>B2</i>	ν	T2	<i>B1</i>	<i>B2</i>	ν	T3	<i>B1</i>	<i>B2</i>	ν
<i>B1</i>	1	5	0,833	<i>B1</i>	1	6	0,857	<i>B1</i>	1	7	0,875
<i>B2</i>	1/5	1	0,167	<i>B2</i>	1/6	1	0,143	<i>B2</i>	1/7	1	0,125
T4	<i>B1</i>	<i>B2</i>	ν	T5	<i>B1</i>	<i>B2</i>	ν	T6	<i>B1</i>	<i>B2</i>	ν
<i>B1</i>	1	1/3	0,25	<i>B1</i>	1	9	0,9	<i>B1</i>	1	7	0,875
<i>B2</i>	3	1	0,75	<i>B2</i>	1/9	1	0,1	<i>B2</i>	1/7	1	0,125
T7	<i>B1</i>	<i>B2</i>	ν	T8,9	<i>B1</i>	<i>B2</i>	ν	T10	<i>B1</i>	<i>B2</i>	ν
<i>B1</i>	1	1/4	0,2	<i>B1</i>	1	1	0,5	<i>B1</i>	1	1/4	0,2
<i>B2</i>	4	1	0,8	<i>B2</i>	1	1	0,5	<i>B2</i>	4	1	0,8

Рис. 7. Матрицы сравнений альтернатив (*B1*, *B2*) по технологическим показателям (*T1-T10*)

$$W_{43} = \begin{vmatrix} 0,833 & 0,857 & 0,875 & 0,25 & 0,9 & 0,875 & 0,2 & 0,5 & 0,5 & 0,2 \\ 0,167 & 0,143 & 0,125 & 0,75 & 0,1 & 0,125 & 0,8 & 0,5 & 0,5 & 0,8 \end{vmatrix},$$

относительно второго уровня – матрица W_{32} :

$$W_{32}^T = \begin{vmatrix} 0,271 & 0,245 & 0,049 & 0,033 & 0,048 & 0,146 & 0,107 & 0,055 & 0,024 & 0,022 \\ 0,02 & 0,017 & 0,32 & 0,119 & 0,232 & 0,014 & 0,019 & 0,071 & 0,091 & 0,097 \\ 0,068 & 0,034 & 0,022 & 0,023 & 0,02 & 0,029 & 0,186 & 0,206 & 0,206 & 0,206 \end{vmatrix},$$

относительно первого уровня – матрица W_{21} :

$$W_{21}^T = \begin{vmatrix} 0,128 & 0,595 & 0,277 \end{vmatrix}.$$

На **третьем этапе** путем аддитивной свертки на иерархии (последовательного умножения матриц собственных векторов, построенных для каждого уровня иерархии), получим вектор W_{41} , координаты которого представляют относительные приоритеты сравниваемых вариантов систем ИБ по обобщенному показателю:

$$W_{41}^T = (W_{43} W_{32} W_{21})^T = \begin{vmatrix} 0,61 & 0,39 \end{vmatrix}.$$

Полученный результат трактуется следующим образом: относительная предпочтительность вариантов систем с точки зрения ИБ составляет 0,61 (вариант 1) и 0,39 (вариант 2).

Сетевой подход

Приведенная на рисунке 4 иерархия в ряде случаев оказывается достаточно упрощенной моделью с односторонними связями между элементами соседних уровней. Обратные связи между элементами разных уровней и горизонтальные связи между элементами одного уровня отсутствуют. Во многих задачах принятия решений элементы не поддаются такому линейному упорядочиванию, образуя сеть⁸. На рисунке 8 приведен пример сетевой структуры, элементы e которой сгруппированы в компоненты C (по аналогии с группировкой элементов иерархии по уровням). Стрелками обозначено влияние на уровне компонентов, а влияние элементов e подразумевается. Подобная расширяемость сетевого

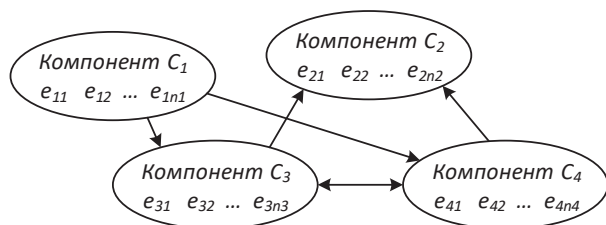


Рис. 8. Сетевая структура

представления позволяет использовать его для анализа вложенных структур, допускающих множество уровней абстракции. В контексте информационных технологий это означает, что компонент C может быть рассмотрен как некоторая технология, а вложенные в него элементы e – как совокупность технологий нижестоящего уровня.

Математическую формализацию задач принятия решений на основе сетевых структур предлагает метод аналитических сетей (МАС), являющийся обобщением метода анализа иерархий. МАС предполагает оценку приоритетности каждого элемента сети относительно всех других элементов по некоторому общему критерию на основе построения блочных матриц, образующих суперматрицу W стохастического типа:

$$W = \begin{matrix} & \begin{matrix} C_1 & C_2 & \dots & C_m \end{matrix} \\ \begin{matrix} C_1 \\ C_2 \\ \dots \\ C_m \end{matrix} & \begin{vmatrix} \begin{matrix} e_{11} & e_{12} & \dots & e_{1n1} \end{matrix} & \begin{matrix} e_{21} & e_{22} & \dots & e_{2n2} \end{matrix} & \dots & \begin{matrix} e_{m1} & e_{m2} & \dots & e_{mnm} \end{matrix} \\ \begin{matrix} e_{11} \\ e_{12} \\ \dots \\ e_{1n1} \end{matrix} & W_{11} & W_{12} & \dots & W_{1m} \\ \begin{matrix} e_{21} \\ e_{22} \\ \dots \\ e_{2n2} \end{matrix} & W_{21} & W_{22} & \dots & W_{2m} \\ \dots & \dots & \dots & \dots & \dots \\ \begin{matrix} e_{m1} \\ e_{m2} \\ \dots \\ e_{mn1} \end{matrix} & W_{m1} & W_{m2} & \dots & W_{mm} \end{vmatrix} \end{matrix},$$

где W_{ij} – матрицы собственных векторов v (векторов приоритетов или векторов влияния), сформированных на основе попарных сравнений между собой элементов одного компонента (диагональные блоки) и элементов различных компонентов (недиагональные блоки); $i = 1 \dots m, j = 1 \dots m$; m – количество компонентов сети.

Каждый столбец матрицы W_{ij} представляет собой вектор влияния v элементов i -го компонента (записанных слева как имена строк) на элемент j -го компонента (записанный сверху как имя столбца). Матрицы W_{ij} стохастические по столбцам, а суперматрица W приводится к стохастическому виду за счет взвешивания блоков W_{ij} в соответствии с приоритетами компонентов сети. Веса блоков,

⁸ Саати Т. Принятие решений при зависимостях и обратных связях. Аналитические сети. Пер. с англ. / Науч. ред. А. В. Андрейчиков, О. Н. Андрейчикова. – М.: ЛЕНАНД, 2019. – 360 с.

в свою очередь, вычисляются на основе попарных сравнений компонентов, записанных в суперматрице слева, по степени влияния на компонент, записанный сверху, на основе критерия более высокого уровня по отношению к критерию, используемому при сравнении элементов. Таким образом формулируются не только влияния между элементами, но и влияния между сущностями более высокого уровня – компонентами сети.

Упомянутое свойство стохастичности суперматрицы W необходимо для сходимости последовательности ее степеней W^k , которая описывает поэтапное распространение влияния в сети вплоть до достижения предельного вектора влияния элементов.

Рассмотрим примеры использования МАС для иерархии и для сети.

Суперматрица для иерархии

Суперматрица W первой степени для рассматриваемой на рисунке 4 иерархии представляет собой запись матриц W_{21} , W_{32} , W_{43} в виде блоков, которые характеризуют влияния элементов одного уровня на элементы другого (соседнего) уровня ($2 \rightarrow 1$, $3 \rightarrow 2$, $4 \rightarrow 3$, где «4» означает уровень альтернатив):

$$W = \begin{pmatrix} 0 & 0 & 0 & 0 \\ W_{21} & 0 & 0 & 0 \\ 0 & W_{32} & 0 & 0 \\ 0 & 0 & W_{43} & A \end{pmatrix}.$$

Нулевые матрицы-блоки W_{11} , W_{22} , W_{33} означают отсутствие горизонтальных связей между элементами отдельного компонента (уровня), W_{12} , W_{23} , W_{34} – отсутствие вертикальных обратных связей между элементами соседних уровней, W_{31} , W_{41} , W_{42} , W_{13} , W_{14} , W_{24} – отсутствие вертикальных связей между элементами уровней, не являющихся соседними. Единица матрица A обеспечивает стохастичность суперматрицы W .

Вторая степень суперматрицы характеризует влияния элементов «через уровень» ($3 \rightarrow 1$, $4 \rightarrow 2$):

$$W^2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ W_{21} & 0 & 0 & 0 \\ 0 & W_{32} & 0 & 0 \\ 0 & 0 & W_{43} & A \end{pmatrix}^2 =$$

$$= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ W_{32} W_{21} & 0 & 0 & 0 \\ 0 & W_{43} W_{32} & A W_{43} & A^2 \end{pmatrix}.$$

Третья степень суперматрицы характеризует влияния «через два уровня» ($4 \rightarrow 1$) и для рассматриваемой трехуровневой иерархии является предельной, то есть суперматрица приобретает устойчивую форму:

$$W^3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ W_{21} & 0 & 0 & 0 \\ 0 & W_{32} & 0 & 0 \\ 0 & 0 & W_{43} & A \end{pmatrix}^3 =$$

$$= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ W_{43} W_{32} W_{21} & A W_{43} W_{32} & A^2 W_{43} & A^3 \end{pmatrix}.$$

Блок $W_{41} = W_{43} W_{32} W_{21}$ в левом нижнем углу суперматрицы W^3 показывает степени влияния альтернатив (элементов четвертого уровня) на величину обобщенного показателя (элемент первого уровня), что в терминах МАИ означает предпочтительность альтернатив.

Суперматрица для сети

Модифицируем приведенную на рисунке 4 иерархию.

На уровне частных показателей (уровне 2):

- добавим связь от элемента «Конфиденциальность» к элементу «Целостность», отражающую положительное влияние уровня конфиденциальности на уровень целостности (ограниченный доступ к информационному ресурсу способствует сохранению его целостности);
- относительное влияние элементов «Конфиденциальность» и «Доступность» на элемент «Целостность» будем характеризовать вектором (1;0).

На уровне технологических показателей (уровне 3):

- добавим связи от элементов Т3 и Т4 к элементу Т5, отражающими положительное влияние одной технологии на другую (например, необходимость для реализации или повышение эффективности);
- допустим равноценное влияние элементов Т3 и Т4 на элемент Т5, характеризующееся вектором (0,5;0,5).

Суперматрица модифицированной иерархии приведена на рисунке 9:

- помимо рассмотренных ранее блоков W_{21} , W_{32} , W_{43} , ненулевыми стали блоки W_{22} и W_{33} , отражающие межэлементные связи на втором и третьем уровне;
- с целью приведения к стохастической форме выполнена нормировка по столбцам (выделены жирным шрифтом) путем умножения элементов столбца на весовой коэффициент блока;
- весовые коэффициенты блоков W_{22} , W_{32} и блоков W_{33} , W_{43} приняты равными 0,5 в допущении равноценности влияния элементов одного уровня на элементы другого уровня и взаимного влияния элементов одного уровня.

Предельная форма суперматрицы приведена на рисунке 10.

		1				3										4	
		ИБ	Ц	Д	К	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	B1	B2
W =	1 ИБ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Ц	0,128	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2 Д	0,595	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	К	0,277	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	T1	0	0,135	0,02	0,068	0	0	0	0	0	0	0	0	0	0	0	0
	T2	0	0,122	0,017	0,034	0	0	0	0	0	0	0	0	0	0	0	0
	T3	0	0,025	0,32	0,022	0	0	0	0	0,25	0	0	0	0	0	0	0
	T4	0	0,017	0,119	0,023	0	0	0	0	0,25	0	0	0	0	0	0	0
	3 T5	0	0,024	0,232	0,02	0	0	0	0	0	0	0	0	0	0	0	0
	T6	0	0,072	0,014	0,029	0	0	0	0	0	0	0	0	0	0	0	0
4	T7	0	0,054	0,019	0,186	0	0	0	0	0	0	0	0	0	0	0	0
	T8	0	0,028	0,071	0,206	0	0	0	0	0	0	0	0	0	0	0	0
	T9	0	0,012	0,091	0,206	0	0	0	0	0	0	0	0	0	0	0	0
	T10	0	0,011	0,097	0,206	0	0	0	0	0	0	0	0	0	0	0	0
	B1	0	0	0	0	0,833	0,857	0,875	0,25	0,45	0,875	0,2	0,5	0,5	0,2	1	0
	B2	0	0	0	0	0,167	0,143	0,125	0,75	0,05	0,125	0,8	0,5	0,5	0,8	0	1

Рис. 9. Суперматрица модифицированной задачи сравнения альтернативных вариантов системы ИБ (степень $k = 1$)

		1				3										4	
		ИБ	Ц	Д	К	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	B1	B2
W ⁴ =	1 ИБ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Ц	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2 Д	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	К	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	T1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	T2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	T3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	T4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3 T5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	T6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	T7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	T8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	T9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	T10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	B1	0.567	0.574	0.627	0.435	0.833	0.857	0.875	0.25	0.731	0.875	0.2	0.5	0.5	0.2	1	0
	B2	0.433	0.426	0.373	0.565	0.167	0.143	0.125	0.75	0.269	0.125	0.8	0.5	0.5	0.8	0	1

Рис. 10. Предельная форма суперматрицы (степень $k = 4$)

Переход от иерархии к сетевой структуре, учитывающей влияние показателя «Конфиденциальность» на показатель «Целостность» и влияние технологий T3, T4 на технологию T5, привел к удлинению маршрутов влияния, увеличению предельной степени k суперматрицы и изменению предпочтительности альтернатив – уменьшению предпочтительности

варианта 1 (с 0,61 до 0,57) и увеличению предпочтительности варианта 2 (с 0,39 до 0,43).

Рассмотренный пример показывает адекватность МАС в более сложных по сравнению с иерархическими структурами случаях. Возможность более детального описания взаимодействий между элементами в сетевой структуре повышает обоснованность решений.

Заключение

1. В основе рассмотренного подхода к оценке системы информационной безопасности находится ее представление в форме расширяемой сети отдельных технологических решений. Такое представление соответствует архитектурной особенности информационных технологий – множественности уровней абстракции – и позволяет организовать процесс экспертизы информационной системы на основе метода аналитических сетей. Базовые абстракции МАС – сети и суперматрицы – корректно моделируют взаимные влияния компонентов и показателей сложной многоуровневой системы ИБ.

2. Существуют задачи, в которых взаимные влияния формализуются введением дополнительных (обратных) связей (циклов) между элементами сети. При наличии циклов предельные степени суперматрицы могут достигать значений $k = 40...60$ и более, а результаты возведения в степень могут не сходиться к одной матрице, например, могут существовать две

чередующиеся предельные формы суперматрицы. Тогда оценивается предел среднего значения последовательности $\lim_{k \rightarrow \infty} (1/N) \sum_{k=1}^N W^k$, называемого цезаровской суммой.

3. Зависимости между элементами сети могут рассматриваться как очевидные – выгоды или издержки, а также иметь вероятностную природу – возможности и риски. При этом приоритеты элементов сети для «невыгодного» влияния оцениваются исходя из принципа: большим приоритетом обладает более «опасный» элемент, приводящий к большему «невыгодному» влиянию. Выгоды, издержки, возможности и риски оцениваются отдельно, на различных структурах, с получением четырех различных ранжирований заданного набора альтернатив. Затем производится мультипликативный или аддитивный синтез приоритетов альтернатив с соответствующей инверсией приоритетов, полученных с применением структур «невыгодного» влияния – издержек и рисков.

Литература

1. Кибербезопасность цифровой индустрии. Теория и практика функциональной устойчивости к кибератакам / Д. П. Зегжда, Е. Б. Александрова, М. О. Калинин [и др.]. – М.: Научно-техническое издательство «Горячая линия-Телеком», 2021. – 560 с.
2. Язов Ю. К. Методология оценки эффективности защиты информации в информационных системах от несанкционированного доступа / Ю. К. Язов, С. В. Соловьев. – Санкт-Петербург: Издательство «Наукоемкие технологии», 2023. – 258 с.
3. Костогрызов А. И. О моделях и методах вероятностного анализа защиты информации в стандартизованных процессах системной инженерии / А. И. Костогрызов // Вопросы кибербезопасности. – 2022. – №6(52). – С. 71-82. – DOI 10.21681/2311-3456-2022-6-71-82.
4. Костогрызов А. И. Обзор стандартизованных риск-ориентированных методов и моделей для обеспечения гарантий качества системы / А. И. Костогрызов // Современные информационные технологии и ИТ-образование. – 2022. – Т. 18, №3. – С. 483-495. – DOI 10.25559/SITITO.18.202203.483-495.
5. Состояние и перспективы развития методического обеспечения технической защиты информации в информационных системах / С. В. Соловьев, М. А. Тарелкин, В. В. Текунов, Ю. К. Язов // Вопросы кибербезопасности. – 2023. – № 1(53). – С. 41-57. – DOI 10.21681/2311-3456-2023-1-41-57.
6. Стародубцев Ю. И. Структурно-функциональная модель киберпространства / Ю. И. Стародубцев, П. В. Закалкин, С. А. Иванов // Вопросы кибербезопасности. – 2021. – № 4(44). – С. 16-24. – DOI 10.21681/2311-3456-2021-4-16-24.
7. Аносов С. С. Оценка информационного риска методом декомпозиции и анализа сетевых и иерархических структур / С. С. Аносов, И. Ю. Шахалов // Безопасные информационные технологии: Сборник трудов Одиннадцатой международной научно-технической конференции, Москва, 06–07 апреля 2021 года. – Москва: Московский государственный технический университет имени Н. Э. Баумана (национальный исследовательский университет) (Москва), 2021. – С. 20–26.
8. Классификация информационных технологий / В. Н. Волкова, А. Ю. Васильев, А. А. Ефремов [и др.] // Прикладная информатика. – 2015. – Т. 10, № 5(59). – С. 124–141.
9. Применение логико-вероятностного метода в информационной безопасности (Часть 1) / А. О. Калашников, К. А. Бугайский, Д. С. Бирин [и др.] // Вопросы кибербезопасности. – 2023. – № 4(56). – С. 23–32. – DOI 10.21681/2311-3456-2023-4-23-32.
10. Метаграмматический подход анализа иерархий для синтеза систем безопасности атомных электростанций / О. И. Атакищев, В. Г. Грибунин, И. Л. Борисенков, М. Н. Лысачев // Вопросы кибербезопасности. – 2023. – № 1(53). – С. 82–92. – DOI 10.21681/2311-3456-2023-1-82-92.
11. Калашников А. О. Управление информационными рисками сложной системы с использованием механизма «когнитивной игры» / А. О. Калашников, Е. В. Аникина // Вопросы кибербезопасности. – 2020. – № 4(38). – С. 2–10. – DOI 10.21681/2311-3456-2020-04-2-10.

APPLICATION OF ANALYTIC HIERARCHY PROCESS AND ANALYTIC NETWORK PROCESS TO ASSESSMENT PRIORITIES OF INFORMATION SECURITY SYSTEMS

Anosov R. S.⁹, Anosov S. S.¹⁰

Keywords: variant of system, quality indicator, supermatrix, influence vector, decision-making.

Purpose of the study: comparative assessment of the analytic hierarchy process and the analytic network process and their applicability to the study of information security systems

Research method: system analysis and decision-making, including systematization, structurization, decomposition, composition; graph theory methods; matrix calculus

Result: applicability of the analytic hierarchy process and the analytic network process for solving the problem of choosing one of the alternative variants of the information security system is shown. The analytic network process is more resource-intensive in comparison with the analytic hierarchy process, but it allows for a nuanced study of the information security system based on the analysis of mutual influences and interdependencies of system components, ensuring a high degree of validity of decisions made based on the analysis results. The significance of technological indicators in the study of the information security system is determined, which, on the one hand, characterize the efficiency of the information security technologies used in the system in the context of protected information quality, and on the other hand, the degree of implementation of the corresponding technologies in the information security system. A generalized classification of information technologies and information security technologies has been developed, the structuring of the system of processes, technologies and factors in the context of the system analysis of information technologies has been carried out, which serves as a foundation for building hierarchical and network structures when researching information security systems.

Scientific novelty: it is shown that, in comparison with the hierarchy-centric model, the network-centric model and its basic tools – networks and supermatrices – allow for a detailed description of the actual mutual influences of various characteristics and properties of information security systems and to formalize the problem of assessment priorities of these systems taking into account such influences.

References

1. Kiberbezopasnost' cifrovoy industrii. Teoriya i praktika funkcional'noj ustojchivosti k kiberatakam / D. P. Zegzhda, E. B. Aleksandrova, M. O. Kalinin [i dr.]. – M.: Nauchno-tehnicheskoe izdatel'stvo «Gorjachaja linija-Telekom», 2021. – 560 s. – ISBN 978-5-9912-0827-7. – EDN BLBTDA.
2. Jazov Ju. K. Metodologija ocenki jeffektivnosti zashhity informacii v informacionnyh sistemah ot nesankcionirovannogo dostupa / Ju. K. Jazov, S. V. Solov'ev. – Sankt-Peterburg: Izdatel'stvo «Naukoemkie tehnologii», 2023. – 258 s. – ISBN 978-5-907618-36-7. – EDN WVCHKW.
3. Kostogryzov A. I. O modeljah i metodah verojatnostnogo analiza zashhity informacii v standartizovannyh processah sistemnoj inzhenerii / A. I. Kostogryzov // Voprosy kiberbezopasnosti. – 2022. – № 6(52). – S. 71–82. – DOI 10.21681/2311-3456-2022-6-71-82. – EDN SGNKNH.
4. Kostogryzov A. I. Obzor standartizovannyh risk-orientirovannyh metodov i modelej dlja obespechenija garantij kachestva sistemy / A. I. Kostogryzov // Sovremennye informacionnye tehnologii i IT-obrazovanie. – 2022. – T. 18, №3. – S. 483–495. – DOI 10.25559/SITITO.18.202203.483-495. – EDN PAPJTZ.
5. Sostojanie i perspektivy razvitiya metodicheskogo obespechenija tehnicheckoj zashhity informacii v informacionnyh sistemah / S. V. Solov'ev, M. A. Tarelkin, V. V. Tekunov, Ju. K. Jazov // Voprosy kiberbezopasnosti. – 2023. – № 1(53). – S. 41–57. – DOI 10.21681/2311-3456-2023-1-41-57. – EDN XAZKQT.
6. Starodubcev Ju. I. Strukturno-funkcional'naja model' kiberprostranstva / Ju. I. Starodubcev, P. V. Zakalkin, S. A. Ivanov // Voprosy kiberbezopasnosti. – 2021. – № 4(44). – S. 16–24. – DOI 10.21681/2311-3456-2021-4-16-24. – EDN BIXMGD.
7. Anosov S. S. Ocenka informacionnogo riska metodom dekompozicii i analiza setevykh i ierarhicheskikh struktur / S. S. Anosov, I. Ju. Shahalov // Bezopasnye informacionnye tehnologii: Sbornik trudov Odinnadcatoj mezhdunarodnoj nauchno-tehnicheskoy konferencii, Moskva, 06-07 aprlja 2021 goda. – Moskva: Moskovskij gosudarstvennyj tehnicheckij universitet imeni N. Je. Baumana (nacional'nyj issledovatel'skij universitet) (Moskva), 2021. – S. 20–26. – EDN MYDQAT.
8. Klassifikacija informacionnyh tehnologij / V. N. Volkova, A. Ju. Vasil'ev, A. A. Efremov [i dr.] // Prikladnaja informatika. – 2015. – T. 10, № 5(59). – S. 124–141. – EDN UQFFCV.
9. Primenenie logiko-verojatnostnogo metoda v informacionnoj bezopasnosti (Chast' 1) / A. O. Kalashnikov, K. A. Bugajskij, D. S. Birin [i dr.] // Voprosy kiberbezopasnosti. – 2023. – № 4(56). – S. 23–32. – DOI: 10.21681/2311-3456-2023-4-23-32. – EDN GIHSBN.
10. Metagrammaticheskij podhod analiza ierarhij dlja sinteza sistem bezopasnosti atomnykh jelektrostantsij / O. I. Atakishhev, V. G. Gribunin, I. L. Borisenkov, M. N. Lysachev // Voprosy kiberbezopasnosti. – 2023. – № 1(53). – S. 82–92. – DOI 10.21681/2311-3456-2023-1-82-92. – EDN XVBJDM.
11. Kalashnikov A.O. Upravlenie informacionnymi riskami slozhnoj sistemy s ispol'zovaniem mehanizma "kognitivnoj igry" / A.O. Kalashnikov, E.V. Anikina // Voprosy kiberbezopasnosti. – 2020. – №4(38). – S. 2-10. – DOI 10.21681/2311-3456-2020-04-2-10. – EDN BGGXVW.

⁹ Roman Anosov, Ph.D., Associate Professor, Military Training and Scientific Center of the Air Force «Air Force Academy named after Professor N. E. Zhukovsky and Yu. A. Gagarin». Voronezh, Russia. E-mail: an_rs@list.ru

¹⁰ Sergey Anosov, Limited Liability Company «Gazpromneft – Digital Solutions». Voronezh, Russia. E-mail: serg-anosov@mail.ru

УЯЗВИМОСТИ АСИММЕТРИЧНЫХ ШИФРОВ БЛОКЧЕЙН-ПЛАТФОРМ

Ищукова Е. А.¹, Петренко С. А.², Леонтьева Ю. П.³

DOI: 10.21681/2311-3456-2025-6-35-47

Цель: выявление потенциально уязвимых мест в реализациях асимметричных алгоритмов криптографии, основанных на использовании эллиптических кривых, применяемых в современных блокчейн-системах.

Методы исследования: основываются на использовании теории информации, теории криптографии и криптоанализа, математического аппарата теории вероятностей и математической статистики, теории построения блокчейн-систем, теории информационной безопасности.

Результаты: рассмотрены основные приемы, которые используются для построения асимметричных шифров, основанных на использовании эллиптических кривых. Рассмотрены алгоритм сложения двух точек, умножения точки на скаляр, определения второй координаты точки. Отдельно кратко рассмотрены свойства генераторов псевдослучайных последовательностей: принципы их построения и их влияние на стойкость асимметричных шифров, в составе которых они используются. В качестве базового алгоритма рассмотрен алгоритм цифровой подписи ECDSA, который используется в составе таких блокчейн-платформ как Bitcoin, Litecoin, Ethereum и многих других.

Научная новизна заключается в рассмотрении ряда кейсов, моделирующих возникновение уязвимостей в асимметричной криптографии, используемой в современных блокчейн-системах. Для каждого кейса выполнено описание проблемы, сформулирована постановка задачи, приведено возможное решение и дана оценка его сложности. Показано, что при правильном использовании математического аппарата шифров, соблюдении требований к выбору стартовых параметров, отсутствии ошибок в программных реализациях, обеспечивается достаточная стойкость.

Ключевые слова: стойкость, алгоритм шифрования, функция хеширования, криптография, криптоанализ, приватный ключ, публичный ключ.

Введение

Блокчейн технологии представляют собой разновидность построения систем распределенного реестра. Их отличительной особенностью является формирование единого связанного списка, в котором каждая следующая запись зависит от предыдущей. Это обеспечивает неизменяемость данных: изменение одной записи невозможно без изменения всех, связанных с ней [1, 2]. Формирование единого связанного списка достигается за счет использования механизмов криптографии [3–16]. В большинстве блокчейн систем используются два основных криптографических инструмента: асимметричная криптография на эллиптических кривых и функции хеширования. Функция хеширования используется для контроля целостности данных, сохраняемых в блокчейне. Также функции хеширования могут быть использованы для выстраивания связей между блоками цепочки, как это сделано, например, в механизме консенсуса Proof-of-Work [2]. Алгоритмы асимметричной криптографии используются

в блокчейне для взаимодействия абонентов в недоверенной среде. У каждого абонента системы есть пара приватный-публичный ключ. Своим приватным ключом может подтвердить совершаемое в системе действие, управлять переводами криптовалютных средств (подтверждать, что данная транзакция назначена ему и он может ее потратить). Кроме того, через публичный ключ обычно определяется адрес пользователя в сети блокчейна. В разных блокчейнах используются разные алгоритмы выработки адреса, но, как правило, он вырабатывается из публичного ключа путем однократного или многократного хеширования с последующей перекодировкой. Известно, что одной из проблем современного блокчейна является проблема масштабируемости. Популярность криптографических алгоритмов на эллиптических кривых обусловлена высокой степенью безопасности при относительно малом размере ключей, что позволяет делать подписи транзакций сравнительно небольшими (в пределах 64 байт).

- 1 Ищукова Евгения Александровна, кандидат технических наук, ведущий научный сотрудник, Научный центр информационных технологий и искусственного интеллекта, Научно-технологический университет «Сириус», Адрес: Россия, 354340, Краснодарский край, Федеральная территория «Сириус», ORCID 0000-0002-6818-1608. E-mail: ischukova.ea@talantiuspeh.ru
- 2 Петренко Сергей Анатольевич, доктор технических наук, профессор, руководитель группы, Научный центр информационных технологий и искусственного интеллекта, Научно-технологический университет «Сириус», Адрес: Россия, 354340, Краснодарский край, Федеральная территория «Сириус», ORCID 0000-0003-0644-1731. E-mail: Petrenko.SA@talantiuspeh.ru
- 3 Леонтьева Юлия Павловна, студент, Институт компьютерных технологий и информационной безопасности, Южный федеральный университет, Россия. ORCID 0009-0006-4778-131X. E-mail: izavodnova@sfedu.ru

В 2022 году Национальный институт стандартов и технологий США NIST объявил о завершении конкурса постквантовой криптографии. Финалистами конкурса стали пять алгоритмов. Два алгоритма (CRYSTALS-KYBER и NTRU) предназначены для шифрования и обмена ключами. Три алгоритма (CRYSTALS-DILITHIUM, FALCON, SPHINCS+) предназначены для формирования и проверки цифровой подписи. Алгоритмы CRYSTALS-DILITHIUM и FALCON основаны на теории решеток, в то время как алгоритм SPHINCS+ сконструирован на основе функций хеширования. В то же время параметры, рекомендованные для данных алгоритмов шифрования, в настоящий момент имеют критически большие размерности в рамках использования в блокчейн-технологиях [17, 18]. Известно, что одной из проблем блокчейна является проблема постоянного роста хранимой базы данных. И при выборе криптографических примитивов важно использовать такие примитивы, которые при небольших размерах обеспечивают надежную защиту. Эллиптическая криптография в современных блокчейн-платформах (например, в Bitcoin, Litecoin and Ethereum) оперирует ключами с размерностью 32 байта и вырабатывает цифровую подпись общим размером 64 байта. В то время как алгоритм CRYSTALS-DILITHIUM имеет публичный ключ размером 1,1 КБ (1024 байт), приватный ключ размером около 2,7 КБ (или 2700 байт) и вырабатывает подпись размером около 2,5 КБ (или 2,440 байт). Алгоритм FALCON имеет публичный ключ размером 1 КБ (800–1000 байт), приватный ключ размером около 1,6 КБ (или 1600 байт) и вырабатывает подпись размером около 887 байт. В алгоритме SPHINCS+ размер публичного ключа может варьироваться от 5 до 20 КБ в зависимости от выбранных параметров, приватный ключ может быть от 20 до 100 КБ в зависимости от конфигурации. Размер подписи для алгоритма SPHINCS+ варьируется в зависимости от выбранной параметризации, но, как правило, составляет около 20 КБ для стандартного уровня безопасности. Как видно из приведенного сравнения, параметры разработанных алгоритмов являются во много раз больше аналогичных параметров асимметричной криптографии на эллиптических кривых. А если учесть, что в одном блоке блокчейна может находиться несколько тысяч транзакций, каждая из которых содержит подпись, то объемы блокчейна возрастут многократно.

В данной статье предлагается рассмотреть ряд кейсов, моделирующих возникновение потенциальных уязвимостей в асимметричной криптографии, используемой в современных блокчейн-системах.

1. Постановка задачи

На сегодняшний день не существует унифицированного подхода к определению криптографической

стойкости. Первые понятия о стойкости шифров заложил К. Шенон. Он же дал определение абсолютно стойкого шифра, криптографические свойства которого не позволяют извлечь статистическую информацию относительно секретных ключей из перехватываемого шифра. Известно, что к абсолютно стойким шифрам на сегодняшний день относится только шифр Вернама. Все остальные криптографические алгоритмы являются условно стойкими. Обычно определяют криптографическую стойкость алгоритма относительно того, сколько времени уйдет на вскрытие сообщения или восстановление ключа, а также какова будет стоимость оборудования, необходимого для проведения анализа. При этом могут учитываться и другие параметры. Например, какой объем памяти необходим для проведения анализа, какое необходимо затратить количество энергии и др. Таким образом, различают условную криптографическую стойкость по времени вычислений (или по количеству совершаемых операций. Исходя из которого можно определить затрачиваемое время) и условную криптографическую стойкость по стоимости вычислений, но также могут быть использованы и дополнительные метрики [19].

Целью настоящей работы является выявление потенциально уязвимых мест в реализациях асимметричных алгоритмов криптографии, основанных на использовании эллиптических кривых, применяемых в современных блокчейн-системах. Для достижения поставленной цели необходимо:

1. Выявить основные алгоритмы, лежащие в основе асимметричных алгоритмов, используемых в современных блокчейн-системах.
2. Сформулировать кейсы, моделирующие возникновение уязвимостей в асимметричной криптографии, используемой в современных блокчейн-системах. Для каждого кейса выполнить описание проблемы, сформулировать постановку задачи, привести возможное решение
3. Для каждого сформулированного кейса дать оценку его сложности и по возможности провести численное моделирование.

2. Объект исследования

Объектом исследования являются асимметричные алгоритмы, использующие эллиптическую криптографию. В первую очередь фокус сосредоточен на алгоритме ECDSA, который используется в составе таких блокчейн-платформ как Bitcoin, Litecoin, Ethereum и многих других.

Также к объектам исследования стоит отнести генераторы псевдослучайных последовательностей (ПСП). В блокчейн-системах генераторы ПСП используются для генерации случайных чисел при выработке ключей для используемого асимметричного

алгоритма шифрования, а также при использовании алгоритмов электронной подписи. От того насколько стойкий генератор ПСП используется при этом, напрямую зависит стойкость самой блокчейн-системы по отношению к взлому (получения доступа к управлению активами пользователя) или подлогу информации.

3. Генераторы ПСП

Под генератором ПСП понимается алгоритм, на выходе которого образуется битовая последовательность на первый взгляд кажущаяся случайной. Псевдослучайность заключается в том, что рано или поздно битовая последовательность начнет повторяться. Длина ПСП до начала ее повторения называется периодом. Стойкие ПСП должны иметь как можно больший период с тем, чтобы аналитику было сложно отличить вырабатываемую последовательность от случайной, а также чтобы не давать аналитику возможности предсказать появление следующих символов вырабатываемой последовательности на основе анализа предыдущих выработанных значений [20]. Для проверки на случайность используют различные статистические тесты [21].

Выделяют два основных способа построения генераторов ПСП – аппаратные и программные. В виду того, что настоящая работа нацелена на исследование в области блокчейн технологий, то в данном случае особый интерес представляют программные реализации генераторов ПСП. Считается, что построение качественных программных генераторов ПСП является задачей более сложной, нежели построение аппаратных генераторов. Обычно в качестве стартовых параметров программные генераторы ПСП могут использовать случайные системные процессы, такие как: системное время; особенности клавиатурного почерка пользователя; данные, вводимые пользователем; параметры операционной системы или содержимое буфера. Здесь важным является применение правила: хороший генератор должен использовать много разных источников случайности, комбинировать их и менять их. Известны случаи, когда, например, использование только одного параметра приводило систему к уязвимости, несмотря на качество самого используемого генератора. Так, в 2019 году была обнаружена уязвимость в менеджере паролей от Лаборатории Касперского (занесена в реестр под номером CVE-2020-27020) в связи с тем, что стартовое значение генератора зависело только от системного времени.

Кейс № 1.

Описание задачи: Оценить качество реализации выбранного генератора ПСП. Определить с какой вероятностью p генератор вырабатывает значения k_1 и k_2 , отстоящие друг от друга не более чем d позиций.

Постановка задачи: Имеется реализация генератора ПСП, выполняющая преобразование $k = Gen()$. Определить вероятность p , с которой $|k_i - k_j| \leq d$, где $i, j = 1, \dots, n$.

Решение:

Вход: Генератор $k = Gen()$; расстояние d ; количество рассматриваемых чисел n .

Выход: Вероятность p ($|k_i - k_j| \leq d$, где $i, j = 1, \dots, n$).

Данный кейс сводится к задаче определения математической статистики распределения формируемых случайным образом чисел по числовой оси. Чем больше будет накопленная статистика, тем точнее будет полученный результат. Для того, чтобы формируемые генератором значения могли рассматриваться как уязвимые, расстояние d между случайно сгенерированными точками должно быть много меньше размерности используемого модуля и доступно для перебора (например, в диапазоне от 232 до 264). При этом вероятность формирования такой разности должна быть не менее 0,5. Будем использовать в алгоритме n точек. Чем больше n , тем точнее определено значение вероятности. При этом количество рассмотренных комбинаций составит $\frac{n(n-1)}{2}$.

Алгоритм 1:

1. Инициализировать переменные
 $k[0] = Gen(); Sum = 0;$
2. Для всех i от 1 до n :
 - 2.1. $k[i] = Gen();$
 - 2.2. Для всех j от 0 до $i-1$:
 - 2.2.1. Если $(|k[i] - k[j]| \leq d)$,
то $Sum = Sum + 1;$
3. $p = \frac{2 \times Sum}{n(n-1)}.$

4. Асимметричная криптография на основе эллиптических кривых

Для асимметричных шифров на основе эллиптической криптографии используются эллиптические кривые двух видов: бинарные эллиптические кривые и эллиптические кривые, ограниченные модулем простого числа p . Дадим определение каждому из видов кривой.

Эллиптические кривые в простом поле ограничены модулем простого числа p , задаются в форме Вейерштрасса, которая имеют следующий вид:

$$y^2 = x^3 + ax + b, \quad (1)$$

где $a, b \in F_p$ и $4a^3 + 27b^2 \neq 0$. При этом все действия с кривой ограничиваются модулем p .

Эллиптические кривые в бинарном поле рассматриваются над конечным полем F_q , где $q = 2^m$ и задаются уравнением вида:

$$y + xy = x^3 + ax^2 + b, \quad (2)$$

где $a, b \in F_p$.



Рис. 1. Виды эллиптических кривых

В зависимости от того, как задаются параметры, различают следующие виды эллиптических кривых: кривая Монтгомери, кривая Коблица, кривая Эдвардса, скрученная кривая Эдвардса, кривая МоТЕ и кривая со случайными параметрами. Схематично весь спектр используемых эллиптических кривых представлен на рис. 1, а более детальное описание приведено в работах [22–25].

В настоящей работе мы ограничимся рассмотрением эллиптических кривых Коблица вида secp256k1, так как именно они применяются в самых известных блокчейн платформах, таких как Bitcoin и Ethereum. Кривая Коблица задается в виде уравнения (1) в случае использования эллиптических кривых в простом поле и в виде уравнения (2) в случае использования эллиптических кривых в бинарном поле при условии, что параметр $a \in F_2$.

В общем случае эллиптические кривые над полем F_p задаются в виде кортежа:

$$T = (p, a, b, G, n, h),$$

где p – большое целое число, определяющее конечное поле F_p ; a и b – параметры уравнений (1) или (2); G – базовая точка эллиптической кривой, которая задается в виде двух координат (x, y) ; n – простое число, определяющее порядок базовой точки G ; h – кофактор подгруппы.

Приведем рекомендованные параметры кортежа T в табл. 1. Для всех рассматриваемых кривых $h = 01$.

Порядок группы определяет количество всех точек на эллиптической кривой в соответствии с заданным модулем p и может быть вычислен в соответствии с теоремой Хассе. Параметр n определяет порядок базовой точки G , то есть показывает количество в циклической подгруппе, порожденной точкой G . Более подробную информацию о параметре n можно найти в стандарте SEC2. В общем случае моделирование изменения параметра порядка группы для кривой вида $y^2 = x^3 + 7$ при малых размерностях модуля p отражен на графике рис. 2 и показывает, что разные точки эллиптической кривой будут обеспечивать разные уровни криптографической стойкости.

Рассмотрим основные математические приемы, которые используются в эллиптической криптографии. В эллиптической криптографии над точками производятся операции сложения и умножения точки на скаляр. При этом умножение точки на скаляр выполняется как многократное сложение разных точек или удвоения одинаковых. В настоящей работе будет рассмотрен классический способ умножения точки на скаляр.

Алгоритм «Сложение двух разных точек эллиптической кривой»

Случай 1. Вход: Эллиптическая кривая в простом поле вида (1). Две точки эллиптической кривой $P(x_1, y_1)$ и $Q(x_2, y_2)$, при этом $P, Q \in F_p$; $P \neq Q$, $x_1 \neq x_2$.

Таблица 1.

Рекомендованные параметры эллиптических кривых

Эллиптич. кривая	Модуль p	Уравнение	G
secp160k1	$2^{160} - 2^{32} - 2^{14} - 2^{12} - 2^9 - 2^8 - 2^7 - 2^3 - 2^2 - 1$	$y^2 = x^3 + 7$	02 3B4C382C E37AA192 A4019E76 3036F4F5 DD4D7EBB
secp224k1	$2^{224} - 2^{32} - 2^{12} - 2^{11} - 2^9 - 2^7 - 2^4 - 2 - 1$	$y^2 = x^3 + 5$	03 A1455B33 4DF099DF 30FC28A1 69A467E9 E47075A9 0F7E650E B6B7A45C
secp256k1	$2^{256} - 2^{32} - 2^9 - 2^8 - 2^7 - 2^6 - 2^4 - 1$	$y^2 = x^3 + 7$	02 79BE667E F9DCBBAC 55A06295 CE870B07 029BCFDB 2DCE28D9 59F2815B 16F81798

Рис. 2. График зависимости отклонения порядка группы n от модуля p

Действия:

1. $\lambda = \frac{y_2 - y_1}{x_2 - x_1}$;
2. $x_3 = \lambda^2 - x_1 - x_2$;
3. $y_3 = \lambda^2(x_1 - x_3) - y_1$.

Выход: Точка $R(x_3, y_3)$, для которой справедливо $R = P + Q$, $R \in F_p$.

Случай 2.

Вход: Эллиптическая кривая в простом поле вида (1). Две одинаковые точки эллиптической кривой $P(x_1, y_1)$ и $P(x_1, y_1)$, при этом $R \in F_p$, $y_1 \neq 0$.

Действия:

1. $\lambda = \frac{3x_1^2 + a}{2y_1}$;
2. $x_3 = \lambda^2 - 2x_1$;
3. $y_3 = \lambda^2(x_1 - x_3) - y_1$.

Выход: Точка $R(x_3, y_3)$, для которой справедливо $R = P + P = 2P$, $R \in F_p$.

Случай 3.

Вход: Эллиптическая кривая в бинарном поле вида (2). Две точки эллиптической кривой $P(x_1, y_1)$ и $Q(x_2, y_2)$, при этом $P, Q \in F_{2^m}$; $P \neq Q$, $x_1 \neq x_2$.

Действия:

1. $\lambda = \frac{y_1 + y_2}{x_1 + x_2}$;
2. $x_3 = \lambda^2 + \lambda + x_1 + x_2 + a$;
3. $y_3 = \lambda(x_1 + x_3) + x_3 + y_1$.

Выход: Точка $R(x_3, y_3)$, для которой справедливо $R = P + Q$; $R \in F_{2^m}$.

Случай 4.

Вход: Эллиптическая кривая в бинарном поле вида (2). Две одинаковые точки эллиптической кривой $P(x_1, y_1)$ и $P(x_1, y_1)$, при этом $P \in F_{2^m}$, $y_1 \neq 0$.

Действия:

1. $x_3 = \frac{x_1^2 + b}{x_1^2}$;
2. $y_3 = x_1^2 + \frac{x_1 + y_1}{x_1}x_3 + x_3$.

Выход: Точка $R(x_3, y_3)$, для которой справедливо $R = P + P = 2P$, $R \in F_{2^m}$.

Случай 5.

Вход: Эллиптическая кривая любого вида (1) или (2). Две точки эллиптической кривой $P(x_1, y_1)$ и $Q(x_2, y_2)$, при этом $P \neq Q$, $x_1 = x_2$.

Выход: Точка $R(x_3, y_3) = P + Q = \infty$ (нейтральный элемент).

Случай 6.

Вход: Эллиптическая кривая любого вида (1) или (2). Две точки эллиптической кривой $P(x_1, y_1)$, при этом $y_1 = 0$.

Выход: Точка $R(x_3, y_3) = 2P = \infty$ (нейтральный элемент).

Случай 7.

Вход: Эллиптическая кривая любого вида (1) или (2). Две точки эллиптической кривой $P(x_1, y_1)$ и $Q = \infty$ (нейтральный элемент).

Выход: Точка $R(x_3, y_3) = P + Q = P + \infty = P$.

Алгоритм «Умножение точки эллиптической кривой на скаляр»

Вход: Эллиптическая кривая в простом поле вида (1). Точка эллиптической кривой $P(x_1, y_1)$, при этом $P \in F_p$; скаляр k – целое число, которое можно представить в виде битовой последовательности $k = (k_{m-1}, k_{m-2}, \dots, k_1, k_0)$.

Действия:

1. Инициализация данных. Точка $R = (0, 0)$. Точка $S = (x_1, y_1)$.
2. Для всех i от 0 до $m - 1$
 - 2.1. Если $k_i = 0$, то
 - 2.1.1 Если $R = 0$, то $R = S$ иначе $R = R + S$
 - 2.2 $S = S + S$

Выход: Точка $R = kP$, $R \in F_p$.

Часто для экономии пространства точка эллиптической кривой сохраняет только одну координату X .

Так как эллиптическая кривая задается квадратным уравнением, то всегда существует две координаты Y для одного заданного X . Для того, чтобы избежать путаницы, перед координатой X помещают дополнительный байт b и вместе они образуют последовательность в 33 байта ($b||X$). Если $b = 02$, то координата Y четная, если $b = 03$, то координата Y нечетная. Для одного и того же X координаты Y всегда будут образовывать пару из четного и нечетного числа в силу того, что операции в поле ограничены нечетным числом. Также первый байт b может быть равен 04. Обычно это означает, что точка сохранена в развернутом виде с обеими координатами.

Алгоритм «Восстановление координаты y для точки эллиптической кривой»

Вход: Эллиптическая кривая в простом поле вида

(1). Координата точки эллиптической кривой ($b||x$).

Действия:

1. $y = \sqrt{(x^3 + ax + b) \bmod p}$.
2. Если $((b = 2) \text{ и } ((y \& 1) = 1))$, то $y = p - y$.

Выход: Точка $R = (x, y)$.

Теперь с использованием введенных операций над точками эллиптических кривых рассмотрим сами криптографические примитивы. В данной части предлагается рассмотреть новый авторский вариант использования эллиптических кривых для передачи данных в зашифрованном виде, а также известный классический алгоритм ECDSA для создания и проверки электронной подписи.

Алгоритм электронной подписи ECDSA

Вход: Эллиптическая кривая в простом поле вида (1) с кортежем параметров $T = (p, a, b, G, n, h)$, общим для двух пользователей системы A и B , сообщение M , $h(M)$ – заданная хеш-функция.

Генерация ключа (Выполняет пользователь A):

1. Генерируется случайное число d_A не больше величины n . Число d_A – приватный ключ пользователя A .
2. Вычисляется точка $Q_A = d_A G$. Точка Q_A – публичный ключ пользователя A .
3. Пользователю B сообщается публичный ключ Q_A .

Подпись (Выполняет пользователь A):

1. Генерируется число k_s не больше величины n . Число k_s – сессионный ключ, каждый раз разный для подписи сообщений M .
2. Вычисляется точку $k_s G = (x_1, y_1)$ и $r = x_1 \bmod n$.
3. Если $r = 0$, то необходимо вернуться к шагу 1
4. Вычисляется $t = k_s^{-1} \bmod n$ (при помощи расширенного алгоритма Евклида).
5. Вычисляется $z = h(M)$, результат представляется в виде большого целого числа.
6. Вычисляется $s = t(z + rd_A) \bmod n$.
7. Если $s = 0$, то необходимо вернуться к шагу 1
8. Пользователю B пересылается сообщение M и подпись к нему в виде пары значений (r, s) .

Выход: Подпись = (r, s) .

Проверка подписи (Выполняет пользователь B):

1. Выполняется проверка, что числа r и s лежат в диапазоне от 1 до $n - 1$.
2. Вычисляется $z = h(M)$, результат представляется в виде большого целого числа.
3. Вычисляет $w = s^{-1} \bmod n$ (при помощи расширенного алгоритма Евклида).
4. Вычисляет $u = zw \bmod n$.
5. Вычисляет $v = rw \bmod n$.
6. Вычисляет точку $(x, y) = uG + vQ_A$.
7. Если $r \neq x \bmod n$, то подпись некорректна.

Выход: Точка $uG + vQ_A = uG + vd_A G = (u + vd_A)G = (zw + rwd_A)G = (z + rd_A)wG = (z + rd_A)s^{-1}G = (z + rd_A)s^{-1}G = st^{-1}s^{-1}G = t^{-1}G = (k_s^{-1})^{-1}G = k_s G$.

Известно, что две самые известные платформы Bitcoin и Ethereum используют одну и ту же эллиптическую кривую secp256k1 с одинаковыми параметрами (табл. 1). Это означает, что данные платформы будут вырабатывать одинаковые пары ключей. В этом легко можно убедиться с использованием любого онлайн генератора ключей для данных платформ (например, www.rfctools.com). На рис. 3 показан результат генерации ключей для обеих платформ. В верхней части окно для генерации ключей платформы Эфириум, а в нижней части – окно для генерации ключей платформы Биткойн. Видно, что при одном и том же заданном параметре d сгенерированы одинаковые ключи с той лишь разницей, что для платформы Биткойн ключ представлен в сжатом виде и имеет первый добавленный байт 02, означающий, что координата y будет иметь четное значение.

Проблема: с точки зрения криптографической стойкости для блокчейн систем первоочередную роль играют способы извлечения или подбора приватного ключа пользователя системы, а также сценарии подмены подписи в сообщениях.

Описание общей задачи: В общем случае задача дискретного логарифмирования сводится к следующему. Задана эллиптическая кривая с кортежем параметров $T = (p, a, b, G, n, h)$. Для двух точек P и Q необходимо найти такое целочисленное d , для которого $dP = Q$. Подходы, которые применяются к анализу во многом зависят от параметров самой эллиптической кривой. Наиболее эффективным методом анализа на сегодняшний день является ро-метод Полларда, время выполнения которого оценивается как $\frac{1}{r} \sqrt{\frac{pn}{2}}$, где n – порядок точки эллиптической кривой, а r – число параллельных процессоров для проведения вычислений [22]. Задача дискретного логарифмирования для эллиптических кривых с рекомендованными параметрами (табл. 1) является на сегодняшний день сложно вычислимой и не применимой на практике. Тем не менее, рассмотрим несколько ситуаций, когда ошибки в реализациях

0- Private ECDSA Key: Ethereum

9967C44629C6A02E43C607D2C40F2317BE50CF9093034D5200CD3E07130C0A8C

(any random 256-bit number from 0x1 to 0xFFFF FFFF FFFF FFFF FFFF FFFF FFFF FFFF BAAE DCE6 AF48 A03B BFD2 5E8C D036 4140)

more info: Secp256k1

☒ Auto next steps

1- Public ECDSA Key:

b5c0a81d908bc8ac40f6af45c862ab75ebbb09160c0611f1aea0202c0bd6759d40893cc187a8e5d9296fd911f95f366280a9355783532fe88d

0- Private ECDSA Key (aka Bitcoin private key):

9967C44629C6A02E43C607D2C40F2317BE50CF9093034D5200CD3E07130C0A8C

(any random 256-bit number from 0x1 to 0xFFFF FFFF FFFF FFFF FFFF FFFF FFFF FFFF BAAE DCE6 AF48 A03B BFD2 5E8C D036 4140)

more info: Secp256k1

Public key type:

☒ Compressed

☐ Uncompressed

☒ Auto next steps

1- Public ECDSA Key (aka Bitcoin public key):

02B5C0A81D908BC8AC40F6AF45C862AB75EBBB09160C0611F1AEA0202C0BD6759

Рис. 3. Пример генерации одинаковых адресов для платформ Биткоин и Эфириум

или неправильное использование параметров системы может привести к уязвимостям.

Кейс № 2.

Описание задачи: Определить приватный ключ пользователя, если известен его публичный ключ, а в реализации алгоритма выработки ключей используется слабая ПСП.

Постановка задачи: генерация пары ключей выполняется в соответствии с алгоритмом цифровой подписи ECDSA с заданным кортежем параметров $T = (p, a, b, G, n, h)$. Известен публичный ключ пользователя $AQ_A = (x, y)$. Известно, что реализация генератора ПСП $k = \text{Gen}()$ генерирует параметры k , для которых $p(|k_i - k_j| \leq d) \geq 0,5$. Требуется определить приватный ключ пользователя A .

Решение:

Вход: генератор $k = \text{Gen}()$; расстояние d ; публичный ключ пользователя A

$$Q_A = (x, y).$$

Выход: приватный ключ пользователя $A d_A$.

В данном случае можно предположить, что если сформировать новый приватный ключ, то он с вероятностью 0,5 будет отстоять от ключа пользователя A не более чем на d точек. Сложность работы алгоритма будет напрямую зависеть от величины d .

Алгоритм 2:

1. Сгенерировать $k = \text{Gen}()$.
2. Вычислить точку $Q_1 = k * G$;

3. Если $Q_1 = Q_A$, то $d_A = k$; прервать работу алгоритма.
4. Для всех i от 1 до d :
5. $Q_1 = Q_1 + G$;
6. Если $Q_1 = Q_A$, то $d_A = k + i$; прервать работу алгоритма.
7. Если ключ не найден, то вернуться к шагу 1.

Результаты:

Все эксперименты для каждого набора параметров проводились по 100 раз, после чего бралось усредненное значение затраченного времени. Здесь и далее для других экспериментов программный код написан на языке Python 3.9, испытания проводились на ПК с процессором AMD Ryzen 5 3500U with Radeon Vega Mobile Gfx 2,10 GHz. Эксперимент выполнялся для эллиптической кривой стандарта secp256k1, параметры для кортежа которой определены в табл. 1. Результаты эксперимента представлены в табл.2.

Кейс № 3.

Описание задачи: определить приватный ключ пользователя, если известно, что в реализации алгоритма подписи ECDSA допущена ошибка и используется постоянное значение сеансового ключа k_s .

Постановка задачи: Генерация пары ключей выполняется в соответствии с алгоритмом цифровой подписи ECDSA с заданным кортежем параметров $T = (p, a, b, G, n, h)$. Известно, что значение k_s является

Таблица 2.

Количественные характеристики результатов определения приватного ключа при слабой ПСП для рекомендованных параметров кривой secp256k1

№ п/п	d	Время подбора, сек	Кол-во попыток	d_A в 16-ричной форме	Q_A в сжатой форме
1	5	0,0069	9	9b27f4269de7d343 103ff00ec251ceff 7c09109df7ef9cc6 ab7203f2b8bb382b	0345ded15ee5f5e5da 33f29c8e9970c9f4 4b036581d66b996f 4d56cf3378882667
2	50	0,0423	54	0f0593537d2c49e6 896e197f9f6cfe29 5e4373df9aaf3021 a55ec26ad04c4e4b	032256596f03a2b3c0 a331e147db42e8e6 7fdc5a1f036e1f8d 89f8df0ff8f822da
3	1000	0,7556	1076	5a9fc8eb89957a96 beea4cfb27856a3d 58a5e7ad94a3f4ea 4dbd3194e7f53b43	038309b865ded35114 23860f20fda7a365 d2e5e6cf0a054d1c a379dee08d4d2a1
4	100000	83,3632	101204	c6cbcbf0736f8193 dbe6c09a09b31955 de385328e4094f40 67d10b6d98d03b7d	0341b90854b5aa5650 90ad60c5e469e18a 803f06d8ad84a86f 6792437116f46cad

одинаковым для всех сообщений M . Известны два сообщения $M1$ и $M2$ и соответствующие им значения подписей (r_1, s_1) и (r_2, s_2) соответственно. Требуется определить приватный ключ пользователя A .

Решение:

Вход: Сообщения $M1$ и $M2$ и их подписи (r_1, s_1) и (r_2, s_2) соответственно.

Выход: приватный ключ пользователя A d_A .

Так как значение сессионного ключа не меняется, то координата r для точек $M1$ и $M2$ будет вычислена одинаково: $r = r_1 = r_2 = k_s G$.

Для сообщения $M1$ параметр s_1 вычисляется как $s_1 = k_s^{-1}(h(M1) + rd_A)$.

Для сообщения $M2$ параметр s_2 вычисляется как $s_2 = k_s^{-1}(h(M2) + rd_A)$.

Умножив оба уравнения на k_s , получим:

$$\begin{aligned}s_1 k_s &= h(M1) + rd_A; \\ s_2 k_s &= h(M2) + rd_A.\end{aligned}$$

Тогда:

$$\begin{aligned}s_1 k_s - s_2 k_s &= h(M1) + rd_A - h(M2) + rd_A; \\ k_s(s_1 - s_2) &= h(M1) - h(M2); \\ k_s &= (h(M1) - h(M2))(s_1 - s_2)^{-1}.\end{aligned}$$

Так как значения $s_1, s_2, M1, M2$ известны, то мы легко можем определить параметр k_s . Найдя его, из формулы $s_1 k_s = h(M1) + rd_A$ определим значение секретного ключа d_A :

$$d_A = (s_1 k_s - h(M1))r^{-1}.$$

Результаты.

Для данного кейса было проведено два эксперимента. В первом эксперименте использовалось уравнение для кривой secp256k1, однако все параметры

были взяты в маленьких размерностях, для проверки работоспособности предложенного кейса (табл. 3). Для второго эксперимента был взят кортеж параметров из табл. 1, соответствующий кривой secp256k1 (табл. 4). Генерация постоянного сеансового ключа реализована его одноразовой выработкой для последующего использования при подписании всех сообщений. В качестве сообщений взяты два текста $M1 = \text{«Message1»}$, $M2 = \text{«Message2»}$. Строковые значения сообщений $M1$ и $M2$ преобразуются в последовательность байтов соответствующих символов в кодировке UTF-8, после чего вычисляется хэш с помощью алгоритма SHA-256:

UTF-8 «Message1» = '0x4d65737361676531'.
 $h(M1) = 960c9384c0db44a860f1309fa04b2d2ec34370ee0148838b659f65d75bf1c85c$.
 UTF-8 «Message2» = '0x4d65737361676532'.
 $h(M2) = 6b5bcdd8bc2e343fe28e6061e5f5659f85f83cc84350488b71b28298bb0c7b15$.

В случае первого эксперимента для небольших значений модуля p в качестве хеш-значения использовался результат $h(M) \bmod p$. На примере кортежа $T_3 = (163, 0, 7, (105;150), 138, 1)$ и сообщений $M1 = \text{«Message1»}$, $M2 = \text{«Message2»}$ разберём процесс восстановления приватного ключа.

Случайно сгенерированный сеансовый ключ:
 $k_s = 105$.

Публичный ключ $Q_A = (70, 113)$.

Координата r для сообщений $M1$ и $M2$: $r_1 = r_2 = k_s G = 6$.

Таблица 3.

Параметры и результаты определения приватного ключа при постоянном значении сеансового ключа k_s

№ п/п	$T = (p, a, b, G, n, h)$	Среднее время работы программы
1	$T_1 = (7, 0, 7, (2;6), 6, 1)$	0,0001185
2	$T_2 = (61, 0, 7, (30;44), 60, 1)$	0,0001597
3	$T_3 = (163, 0, 7, (105;150), 138, 1)$	0,0001798
4	$T_4 = (349, 0, 7, (97;163), 312, 1)$	0,0002110
5	$T_5 = (433, 0, 7, (236;293), 396, 1)$	0,0002934

Таблица 4.

Результаты моделирования вычисления приватного ключа при постоянном значении сеансового ключа k_s

Сеансовый ключ k_s	24cc15aa81a8affdf09ac591c2f8bb2fed22c79b16bf4cdfd6e242ee1f8ddb1
Параметр r_1 подписи сообщения $M1$	c6a25157b952d0ab79dca9694bda8431eb68bc74509ce4efbf70f21441c0da6
Параметр s_1 подписи сообщения $M1$	e90a808b407fa02eb380f63b98e66827332d591006907c73baae722148c39ad8
Параметр r_2 подписи сообщения $M2$	c6a25157b952d0ab79dca9694bda8431eb68bc74509ce4efbf70f21441c0da6
Параметр s_2 подписи сообщения $M2$	7d5923769e75baf2fe1cc377275bb87b30473cbea534749ca6d04d2dc2a14f59
Восстановление секретного ключа d_A	ee87f2154fe7b2bc24f3befec93676d100db075354e5f92e4aec698ca779e112
Среднее время работы программы	0,1915586

Для сообщения $M1$ параметр

$$s_1 = k_s^{-1}(h(M1) + rd_A) = 134.$$

Для сообщения $M2$ параметр

$$s_2 = k_s^{-1}(h(M2) + rd_A) = 111.$$

Вычисление приватного ключа

$$d_A = (s_1 k_s - h(M1))r^{-1} = 127.$$

Кейс № 4.

Описание задачи. Подменить подпись сообщения $M1$ на подпись сообщения $M2$ в алгоритме ECDSA, если известно, что в реализации алгоритма подписи допущена ошибка и генерация сеансового ключа k_s генерирует параметры $k = Gen()$, для которых $p(|k_i - k_j| \leq d) \geq 0,5$.

Постановка задачи. Генерация пары ключей выполняется в соответствии с алгоритмом цифровой подписи ECDSA с заданным кортежем параметров $T = (p, a, b, G, n, h)$. Известно, что $p(|k_i - k_j| \leq d) \geq 0,5$. Известно сообщение $M1$ и его подпись (r_1, s_1) . Требуется для сообщения $M2$ сформировать подпись (r_2, s_2) от лица пользователя A без знания его секретного ключа.

Решение:

Вход: Генератор $k = Gen()$; расстояние d ; сообщение $M1$ и его подпись (r_1, s_1) .

Выход: Подпись (r_2, s_2) для сообщения $M2$.

Первая часть работы алгоритма схожа с Алгоритмом 2 для Кейса № 2. Ожидается, что для новой подписи параметр k_s с вероятностью 0,5 будет отстоять от того, который был использован для подписи сообщения $M1$ не более чем на d точек. Сложность работы алгоритма будет напрямую зависеть от величины d . Тогда с помощью Алгоритма 3 можно определить параметр k_s известной подписи.

Алгоритм 3:

1. Сгенерировать $k = Gen()$.
2. Вычислить точку $Q1 = k * G = (x_2, y_2)$;
3. Если $r_1 = x_2$, то $k_s = k$; прервать работу алгоритма.
4. Для всех i от 1 до d :
5. $Q1 = Q1 + G = (x_{i+2}, y_{i+2})$;
6. Если $r_1 = x_{i+2}$, то $k_s = k + i$; прервать работу алгоритма.
7. Если k_s не найден, то вернуться к шагу 1.

В случае успешного срабатывания Алгоритма 2, найденный параметр k_s для подписи сообщения $M1$

будет использован для формирования подписи сообщения $M2$. Для сообщения $M1$ подпись сформирована на следующем образом:

$$r_1 = k_s G$$

$$s_1 = k_s^{-1}(h(M1) + rd_A)$$

Задача заключается в том, чтобы сформировать подпись для сообщения $M2$ вида:

$$r_2 = r_1 = k_s G$$

$$s_2 = k_s^{-1}(h(M2) + rd_A)$$

Таким образом получается, что надо вычислить только значение s_2 . Преобразуем уравнения для s_1 и s_2 следующим образом:

$$s_1 k_s = h(M1) + rd_A;$$

$$s_2 k_s = h(M2) + rd_A.$$

Тогда:

$$s_1 k_s - s_2 k_s = h(M1) + rd_A - h(M2) - rd_A;$$

$$k_s(s_1 - s_2) = h(M1) - h(M2);$$

$$s_1 - s_2 = (h(M1) - h(M2))k_s^{-1};$$

$$s_2 = s_1 - (h(M1) - h(M2))k_s^{-1}.$$

Так как значения s_1 , $M1$, $M2$ и k_s известны, то легко можно определить параметр s_2 . Подпись (r_2 , s_2) сформирована без применения приватного ключа пользователя A и будет корректно проверена при использовании публичного ключа пользователя A .

Результаты.

Для реализации поставленной задачи были взяты четыре различных значения расстояния $d = 5, 50, 1000, 100000$, входные сообщения $M1, M2$, и кортеж параметров эллиптической кривой стандарта

Таблица 5.

Параметры и результаты подделки подписи сообщения при ошибке в алгоритме генерации сеансового ключа k_s

d	Время работы программы, сек	Кол-во попыток перебора k_s	d_A	Q_A	k_s
5	0,0070	8	7bffa3bf1e22cee4f70c62f7c79e671dc007a3754b29e4b232cd7ccb4cc77dfb	0331eb2c9de1cb9779ed3995fcfa68c9a1062c6a206f204de9d35b74883e2dc135	ec1deee6be60e11a5da9bb0c16f199a7d1be103ad066a77abb45be0c2a74bc66
50	0,0750	86	577cd96fe3b2f95bdebf8fe367419860be1a8dce235c489621382e6513f9c73	036c734bd04a7a3d941f0bfe187b60d17ea88d32568d98c6203d7d741d88c49312	9fff0bc70e3a40469d29e37ea5424b56f6620946ecff7782722d4ce8497debb
1000	1,0680	1172	cc17968b9e781a78564013feee370cf1d0a8c02a8733bc4f873af4352c1bdc21	03a333bd403738329db03194e8081ae6fe237c98e1473d7d424584860fec206a8a	46caae5f76866ed82cdb82cae0a9da3d1eb2e8f42f076537e91ed0657f1fbd4e
100000	364,3640	249230	5aa8f20f98389035722db3a5adf51d6e6c13c8abc380a551f050a1f751a55baa	03f31b53b04e172c957a981c0a70f22c8ba1a984dcf9c7b7c4e1d0689fea4261a6	a23fcee41475299c796b4f96d6887fc988d47d12d4fd7a039538f5f7fa1ae6f

Таблица 6.

Параметры и результаты подделки подписи сообщения при ошибке в алгоритме генерации сеансового ключа k_s

d	r_1	s_1	r_2	s_2
5	6d9cc47faeff7678d1c53da125b7f4d89286b646dbcb1e64852689b54410495	9d641b1f781584950bd858fcb04fa939da6c96e26dcb8044d0a1633a53fba374	26d2f6ce9154631e4a5be8abacb15c36f40e73750a7ab02a45d495f09960fa55	ae1ed24d3cf7cda29572a34242dfe4a861539b8f14c97e1d57c9f4446dcb0e56
50	2f2f93d07da96f86b1f8cdab9ccf710c3e1904dc20851d308fc675c538f67031	bd9859c8394f1e6a661804f555c3a3d037ada9747e1a0ee736ec90496b15b4ec	753ebb733d7e79aad8ec74be34409da6d836c33dbedacc8744feb12b605506f	1c3016afd047551f3e5333d2fab2261394e3f978657567f578bf06539f16ba38
1000	51917a1a80feb1862c18c6516f17d426bb3211f9f52cc3d5762e110e9b08c753	9be668e11e48f942bbf77cbdd2f8a2c7baa451670515be8252fa838cd1b8c22c	85e8a767ac51c383ad29fe8b5cba7b0bc106d9b0133a3e8f02a499dc35892efd	1d17e28fd8c485587425c66c50eea3efb54465d4aad25091bb07fcf1f3681ee
100000	726f60201ce9a52ec5b4db8fb89d29b91122ab7057034cd5ca75c7b3409924df	8a660b2e4cc856d2ed15529fd253a576071fe3b1086f148db55a6005ad38423d	12adf440b5dcb6f649529b7088e9375e0a5d5840261e1ad0cbe722c46dd7d883	1cf760c24bce61f76c6f73e0eb94457d313695e6a9645f7deaab10173bcd2104

secp256k1 из табл. 1. Результаты экспериментов представлены в табл. 5 и 6.

Генерация сеансового ключа реализована его единоразовой выработкой для последующего использования при подписании всех сообщений, расстояние d , как величина возможного отклонения, в процессе подбора ключа задавалось различное для каждого эксперимента. С увеличением расстояния d уменьшается вероятность подбора корректного сеансового ключа, так как перебор будет увеличивать диапазон допустимых значений.

Выводы

В работе рассмотрены потенциально уязвимые места в использовании асимметричных алгоритмов шифрования в современных блокчейн-системах. Рассмотренные кейсы моделируют ситуации, которые могут возникнуть в результате неправильного

использования стартовых параметров криптосистемы или при неверной реализации вспомогательных компонентов. Таких, как например, генераторы псевдослучайных последовательностей. Для каждого рассмотренного кейса приведено математическое решение сформулированной проблемы, определена вычислительная сложность и проведены эксперименты, в том числе с использованием параметров, рекомендованных для эллиптической кривой secp256k1.

Достоверность предлагаемого научного подхода подтверждается применением общенаучных методов исследования, достаточным информационным обеспечением, а также корректным применением методов криптографии, в том числе в построении формульных доказательств и выводов и экспериментальным подтверждением работоспособности выведенных формул.

Результаты получены при финансовой поддержке проекта «Технологии противодействия ранее неизвестным квантовым киберугрозам», реализуемого в рамках государственной программы федеральной территории «Сириус» «Научно-технологическое развитие федеральной территории «Сириус» (Соглашение №23-03 от 27.09.2024 г.).

Литература

1. Kakarlapudi P. V., Mahmoud Q. H. A Systematic Review of Blockchain for Consent Management. Healthcare. 2021; 9(2):137. <https://doi.org/10.3390/healthcare9020137>.
2. Ищукова Е. А., Панасенко С. П., Романенко К. С., Салманов В. Д. Криптографические основы блокчейн-технологий. – М.: ДМК Пресс, 2022. – 302 с. ISBN: 978-5-97060-865-4.
3. Zhang H., Jiang W., Ding J. A Blockchain Network Admission Control Mechanism Using Anonymous Identity-Based Cryptography. Applied Sciences. 2025; 15(1):130. <https://doi.org/10.3390/app15010130>.
4. Chin Y.-C., Hsu C.-L., Lin T.-W., Tsai K.-Y. A Hierarchical Blockchain System for Social Economy Services. Electronics. 2024; 13(20):4004. <https://doi.org/10.3390/electronics13204004>.
5. Joni S. A., Rahat R., Tasnin N., Ghose P., Uddin M. A., Ayoade J. Hybrid-Blockchain-Based Electronic Voting Machine System Embedded with Deepface, Sharding, and Post-Quantum Techniques. Blockchains. 2024; 2(4):366–423. <https://doi.org/10.3390/blockchains2040017>.
6. Kim H., Kim W., Kang Y., Kim H., Seo H. Post-Quantum Delegated Proof of Luck for Blockchain Consensus Algorithm. Applied Sciences. 2024; 14(18):8394. <https://doi.org/10.3390/app14188394>.
7. Gu H., Shang J., Wang P., Mi J., Bhattacharjya A. A Secure Protocol Authentication Method Based on the Strand Space Model for Blockchain-Based Industrial Internet of Things. Symmetry. 2024; 16(7):851. <https://doi.org/10.3390/sym16070851>.
8. Thantharate P., Thantharate A. ZeroTrustBlock: Enhancing Security, Privacy, and Interoperability of Sensitive Data through ZeroTrust Permissioned Blockchain. Big Data and Cognitive Computing. 2023; 7(4):165. <https://doi.org/10.3390/bdcc7040165>.
9. Thanalakshmi P., Rishikesh A., Marion Marceline J., Joshi GP, Cho W. A Quantum-Resistant Blockchain System: A Comparative Analysis. Mathematics. 2023; 11(18):3947. <https://doi.org/10.3390/math11183947>.
10. Wenhua Z., Qamar F., Abdali T.-AN., Hassan R., Jafri S. T. A., Nguyen Q. N. Blockchain Technology: Security Issues, Healthcare Applications, Challenges and Future Trends. Electronics. 2023; 12(3):546. <https://doi.org/10.3390/electronics12030546>.
11. Di Scala A. J., Gangemi A., Romeo G., Vernetti G. Special Subsets of Addresses for Blockchains Using the secp256k1 Curve. Mathematics. 2022; 10(15):2746. <https://doi.org/10.3390/math10152746>.
12. Longo R., Mascia C., Meneghetti A., Santilli G., Tognolini G. Adaptable Cryptographic Primitives in Blockchains via Smart Contracts. Cryptography. 2022; 6(3):32. <https://doi.org/10.3390/cryptography6030032>.
13. Bellés-Muñoz M., Whitehat B., Baylina J., Daza V., Muñoz-Tapia J. L. Twisted Edwards Elliptic Curves for Zero-Knowledge Circuits. Mathematics. 2021; 9(23):3022. <https://doi.org/10.3390/math9233022>.
14. Martínez V. G., Hernández-Álvarez L., Encinas L. H. Analysis of the Cryptographic Tools for Blockchain and Bitcoin. Mathematics. 2020; 8(1):131. <https://doi.org/10.3390/math8010131>.
15. Sala M., Sogorno D., Taufer D. A Small Subgroup Attack on Bitcoin Address Generation. Mathematics. 2020; 8(10):1645. <https://doi.org/10.3390/math8101645>.
16. Марков А. С. Кибербезопасность и информационная безопасность как бифуркация номенклатуры научных специальностей // Вопросы кибербезопасности. 2022, № 1(47), с. 2–9.

17. Petrenko A. S., Petrenko S. A. Basic Algorithms Quantum Cryptanalysis. The journal «Cybersecurity Issues», 2023, no. 1(53), pp. 100–115. doi: 10.21681/2311-3456-2023-1-100-115.
18. Petrenko A. S. Applied Quantum Cryptanalysis (scientific monograph). River Publishers, 2023, 256 pp. ISBN 9788770227933. doi: 10.1201/9781003392873.
19. Марков А. С. Важная веха в безопасности открытого программного обеспечения // Вопросы кибербезопасности, 2023, № 1(53), С.2–12.
20. Pandey D. K., Nicolosi A. R. Pseudorandom Function from Learning Burnside Problem. Mathematics. 2025; 13(7):1193. <https://doi.org/10.3390/math13071193>.
21. Ishchukova, E., Borlakov, R. Reliability of Information Conversion When Encrypting Graphic Images. In: Raza, Z., Babenko, M., Sajid, M., Lapina, M., Zolotarev, V. (eds) AISMA-2023: International Workshop on Advanced Information Security Management and Applications. AISMA 2023. Lecture Notes in Networks and Systems, vol. 1207. Springer, Cham, 2024. https://doi.org/10.1007/978-3-031-77229-0_10.
22. Jebrane J., Chhaybi A., Lazaar S., Nitaj A. Elliptic Curve Cryptography with Machine Learning. Cryptography. 2025; 9(1):3. <https://doi.org/10.3390/cryptography9010003>.
23. Martinez-Diaz I., Ali R., Jamil M. K. On the Search for Supersingular Elliptic Curves and Their Applications. Mathematics. 2025; 13(2):188. <https://doi.org/10.3390/math13020188>.
24. Aljaedi A., Rashid M., Jamal S. S., Alharbi A. R., Alotaibi M. An Optimized Flexible Accelerator for Elliptic Curve Point Multiplication over NIST Binary Fields. Applied Sciences. 2023; 13(19):10882. <https://doi.org/10.3390/app131910882>.
25. Lone P. N., Singh D., Stoffová V., Mishra D. C., Mir U. H., Kumar N. Cryptanalysis and Improved Image Encryption Scheme Using Elliptic Curve and Affine Hill Cipher. Mathematics. 2022; 10(20):3878. <https://doi.org/10.3390/math10203878>.
26. Sattar B., Sadkhan A. Proposed Developments of Pollards Rho Method for Attacking the ECDLP // 2021 7th International Engineering Conference «Research & Innovation amid Global Pandemic» (IEC). DOI: 10.1109/IEC52205.2021.9476119.

VULNERABILITIES OF ASYMMETRIC CIPHERS OF BLOCKCHAIN PLATFORMS

Ishchukova E. A.⁴, Petrenko S. A.⁵, Leonteva I. P.⁶

Keywords: mandatory access control principle, «write down», «write down» event flow, degradation of the access control system, «post-maximum attenuation» effect.

Purpose: the aim of this work is to identify potential vulnerabilities in the implementations of asymmetric ciphers based on the elliptic curves, applied in modern blockchain systems.

Method: the research methods are based on the use of information theory, the theory of cryptography and cryptanalysis, the mathematical apparatus of probability theory and mathematical statistics, the theory of constructing blockchain systems, and the theory of information security.

Results: the paper considers the main techniques used to construct asymmetric ciphers based on the use of elliptic curves. The algorithm for adding two points, multiplying a point by a scalar, and determining the second coordinate of a point is considered. The properties of pseudorandom sequence generators are briefly considered separately: the principles of their construction and their impact on the stability of asymmetric ciphers in which they are used. The ECDSA digital signature algorithm, which is used in blockchain platforms such as Bitcoin, Litecoin, Ethereum, and many others, is considered as a basic algorithm.

The scientific novelty lies in the consideration of a number of cases simulating the emergence of vulnerabilities in asymmetric cryptography used in modern blockchain systems. For each case, a description of the problem is made, a statement of the task is formulated, a possible solution is given and an assessment of its complexity is given. It is shown that with the correct use of the mathematical apparatus of ciphers, compliance with the requirements for the selection of starting parameters, the absence of errors in software implementations, sufficient stability is ensured.

References

1. Kakarlapudi P.V., Mahmoud Q.H. A Systematic Review of Blockchain for Consent Management. Healthcare. 2021; 9(2):137. <https://doi.org/10.3390/healthcare9020137>.
2. Ishchukova E. A., Panasenkov S. P., Romanenko K. S., Salmanov V. D. Kriptograficheskie osnovy blokchejn-tehnologij. – M.: DMK Press, 2022. – 302 s. ISBN: 978-5-97060-865-4.

4 Evgeniya A. Ishchukova, Ph.D. (in Tech.), Leading researcher, Scientific Center for Information Technologies and Artificial Intelligence, Sirius University of Science and Technology, Address: Olimpiyskiy ave. b.1, Sirius, Sirius Federal Territory, Krasnodar region, Russia, 354340. ORCID 0000-0002-6818-1608. E-mail: ishchukova.ea@talantiuspeh.ru

5 Sergei A. Petrenko, Dr.Sc. (of Tech.), Professor, Team Leader, Scientific Center for Information Technologies and Artificial Intelligence, Sirius University of Science and Technology, Address: Olimpiyskiy ave. b.1, Sirius, Sirius Federal Territory, Krasnodar region, Russia, 354340. ORCID 0000-0003-0644-1731. E-mail: Petrenko.SA@talantiuspeh.ru

6 Yuliya P. Leontieva, Student, Institute of Computer Technologies and Information Security, Southern Federal University. ORCID 0009-0006-4778-131X, E-mail: izavodnova@sfedu.ru

3. Zhang H., Jiang W., Ding J. A Blockchain Network Admission Control Mechanism Using Anonymous Identity-Based Cryptography. *Applied Sciences*. 2025; 15(1):130. <https://doi.org/10.3390/app15010130>.
4. Chin Y.-C., Hsu C.-L., Lin T.-W., Tsai K.-Y. A Hierarchical Blockchain System for Social Economy Services. *Electronics*. 2024; 13(20):4004. <https://doi.org/10.3390/electronics13204004>.
5. Joni S. A., Rahat R., Tasnin N., Ghose P., Uddin M. A., Ayoade J. Hybrid-Blockchain-Based Electronic Voting Machine System Embedded with Deepface, Sharding, and Post-Quantum Techniques. *Blockchains*. 2024; 2(4):366-423. <https://doi.org/10.3390/blockchains2040017>.
6. Kim H., Kim W., Kang Y., Kim H., Seo H. Post-Quantum Delegated Proof of Luck for Blockchain Consensus Algorithm. *Applied Sciences*. 2024; 14(18):8394. <https://doi.org/10.3390/app14188394>.
7. Gu H., Shang J., Wang P., Mi J., Bhattacharjya A. A Secure Protocol Authentication Method Based on the Strand Space Model for Blockchain-Based Industrial Internet of Things. *Symmetry*. 2024; 16(7):851. <https://doi.org/10.3390/sym16070851>.
8. Thantharate P., Thantharate A. ZeroTrustBlock: Enhancing Security, Privacy, and Interoperability of Sensitive Data through ZeroTrust Permissioned Blockchain. *Big Data and Cognitive Computing*. 2023; 7(4):165. <https://doi.org/10.3390/bdcc7040165>.
9. Thanalakshmi P., Rishikesh A., Marion Marceline J., Joshi GP, Cho W. A Quantum-Resistant Blockchain System: A Comparative Analysis. *Mathematics*. 2023; 11(18):3947. <https://doi.org/10.3390/math11183947>.
10. Wenhua Z., Qamar F., Abdali T.-AN., Hassan R., Jafri S. T. A., Nguyen Q. N. Blockchain Technology: Security Issues, Healthcare Applications, Challenges and Future Trends. *Electronics*. 2023; 12(3):546. <https://doi.org/10.3390/electronics12030546>.
11. Di Scala A. J., Gangemi A., Romeo G., Vernetti G. Special Subsets of Addresses for Blockchains Using the secp256k1 Curve. *Mathematics*. 2022; 10(15):2746. <https://doi.org/10.3390/math10152746>.
12. Longo R., Mascia C., Meneghetti A., Santilli G., Tognolini G. Adaptable Cryptographic Primitives in Blockchains via Smart Contracts. *Cryptography*. 2022; 6(3):32. <https://doi.org/10.3390/cryptography6030032>.
13. Bellés-Muñoz M., Whitehat B., Baylina J., Daza V., Muñoz-Tapia J. L. Twisted Edwards Elliptic Curves for Zero-Knowledge Circuits. *Mathematics*. 2021; 9(23):3022. <https://doi.org/10.3390/math9233022>.
14. Martínez V. G., Hernández-Álvarez L., Encinas L. H. Analysis of the Cryptographic Tools for Blockchain and Bitcoin. *Mathematics*. 2020; 8(1):131. <https://doi.org/10.3390/math8010131>.
15. Sala M., Sogioro D., Taufer D. A Small Subgroup Attack on Bitcoin Address Generation. *Mathematics*. 2020; 8(10):1645. <https://doi.org/10.3390/math8101645>.
16. Markov A. S. Kiberbezopasnost' i informacionnaja bezopasnost' kak bifurkacija nomenklatury nauchnyh special'nostej // *Voprosy kiberbezopasnosti*. 2022, № 1(47), c. 2–9.
17. Petrenko A. S., Petrenko S. A. Basic Algorithms Quantum Cryptanalysis. The journal «Cybersecurity Issues», 2023, no. 1(53), pp. 100–115. doi: 10.21681/2311-3456-2023-1-100-115.
18. Petrenko A. S. Applied Quantum Cryptanalysis (scientific monograph). River Publishers, 2023, 256 pp. ISBN 9788770227933. doi: 10.1201/9781003392873.
19. Markov A. S. Vazhnaja veka v bezopasnosti otkrytogo programmnogo obespechenija // *Voprosy kiberbezopasnosti*, 2023, № 1(53), S. 2–12.
20. Pandey D. K., Nicolosi A. R. Pseudorandom Function from Learning Burnside Problem. *Mathematics*. 2025; 13(7):1193. <https://doi.org/10.3390/math13071193>.
21. Ishchukova, E., Borlakov, R. Reliability of Information Conversion When Encrypting Graphic Images. In: Raza, Z., Babenko, M., Sajid, M., Lapina, M., Zolotarev, V. (eds) AISMA-2023: International Workshop on Advanced Information Security Management and Applications. AISMA 2023. Lecture Notes in Networks and Systems, vol. 1207. Springer, Cham, 2024. https://doi.org/10.1007/978-3-031-77229-0_10.
22. Jebrane J., Chhaybi A., Lazaar S., Nitaj A. Elliptic Curve Cryptography with Machine Learning. *Cryptography*. 2025; 9(1):3. <https://doi.org/10.3390/cryptography9010003>.
23. Martinez-Diaz I., Ali R., Jamil M. K. On the Search for Supersingular Elliptic Curves and Their Applications. *Mathematics*. 2025; 13(2):188. <https://doi.org/10.3390/math13020188>.
24. Aljaedi A., Rashid M., Jamal S. S., Alharbi A. R., Alotaibi M. An Optimized Flexible Accelerator for Elliptic Curve Point Multiplication over NIST Binary Fields. *Applied Sciences*. 2023; 13(19):10882. <https://doi.org/10.3390/app131910882>.
25. Lone P. N., Singh D., Stofková V., Mishra D. C., Mir U. H., Kumar N. Cryptanalysis and Improved Image Encryption Scheme Using Elliptic Curve and Affine Hill Cipher. *Mathematics*. 2022; 10(20):3878. <https://doi.org/10.3390/math10203878>.
26. Sattar B. Sadkhan A Proposed Developments of Pollards Rho Method for Attacking the ECDLP // 2021 7th International Engineering Conference «Research & Innovation amid Global Pandemic» (IEC). DOI: 10.1109/IEC52205.2021.9476119.



ИСПОЛЬЗОВАНИЯ УЯЗВИМОСТЕЙ ПРОТОКОЛА WEBAUTHN ДЛЯ ПОЛУЧЕНИЯ НЕСАНКЦИОНИРОВАННОГО ДОСТУПА

Панченко А. Р.¹

DOI: 10.21681/2311-3456-2025-6-48-57

Цель статьи: исследование текущего стандарта беспарольной аутентификации FIDO2 на наличие уязвимостей и подтверждение отсутствия у него доказательной устойчивости.

Методы исследования: анализ текущего стандарта беспарольной аутентификации на наличие уязвимостей. Реализация и эксплуатация обнаруженной уязвимости. Данная уязвимость базируется на использовании модифицированной атаки «человек посередине» с использованием вредоносного программного обеспечения и социальной инженерии.

Результаты исследования: проведен анализ текущего стандарта беспарольной аутентификации. В рамках данного анализа была сформирована формальная схема работы протокола WebAuthn. Была обнаружена уязвимость, позволяющая получить доступ к аккаунту легитимного пользователя, защищенного текущим стандартом FIDO2. Для реализации данной уязвимости использовалась модифицированная атака «человек посередине», в рамках которой был реализован сервер, занимающийся пересылкой сообщений во время проведения атаки. Было реализовано вредоносное программное обеспечение типа троянский конь. Данное вредоносное программное обеспечение предварительно было помещено в систему жертвы, где оно выдавало себя за браузер жертвы, с которым связался легитимный сервер. На стороне злоумышленника была реализована программа, считывающая запрос аутентификации в системе атакующего. Данная программа используется для переопределения функции во время процесса аутентификации и последующей пересылки запроса на сервер. В рамках реализации данной уязвимости и ее последующего выполнения было подтверждено отсутствие доказательной устойчивости у текущего стандарта беспарольной аутентификации FIDO2. Был предложен способ потенциальной защиты от данной уязвимости. Так же был предложен способ модификации сбора данных на этапе регистрации, для последующего уведомления об потенциальной атаке жертвы во время процедуры подписания.

Научная новизна: обнаруженная уязвимость подтверждает отсутствие доказательной устойчивости у текущего стандарта беспарольной аутентификации FIDO2. На данный момент все браузеры используют WebAuthn это означает, что данной уязвимости подвержены все современные браузеры. Предложен потенциальный способ защиты от данной уязвимости и способ предупреждения жертвы в рамках уже проводимой атаки.

Ключевые слова: беспарольная аутентификация, ассиметричное шифрование, Self-XSS, Python, FIDO2, CTAP2.1, Authorization Gesture.

Введение

В современном мире все люди пользуются паролями. Все исследования последнего времени демонстрируют, что пароли устарели уже давно. «В июне 2024 года «Лаборатория Касперского» проанализировала 193 млн паролей, обнаруженных в публичном доступе на даркнет-ресурсах. Результаты исследования показали, что: почти половину из них (45 %, или 87 млн) мошенники смогут подобрать менее чем за минуту; большинство проанализированных паролей могут быть легко скомпрометированы с помощью умных алгоритмов; только 23 % (44 млн) комбинаций оказались достаточно стойкими: на их взлом ушло бы больше года» [1, с. 3]. Подсчет энтропии для паролей разной длины показывает, что 8 символов даже при самой большой мощности алфавита имеет значение энтропии 47,6 [1]. «Если вес пароля от 34 до 67,

то пароль относится к категории «Хороший», а если более 67, то пароль считается надежным.» [1, с. 49]. Так же ситуацию осложняет человеческий фактор, связанный тем, что люди не стремятся делать сложные пароли самостоятельно, так как они слишком сложны для запоминания, хоть и более безопасные. «Так, по информации securitylab в 2008 году 84 % случаев взломов систем безопасности компаний и персональных страниц причиной была именно слабая парольная защита»². Данный тезис подтверждается в статье [3, с. 79]: «Пароли, самостоятельно устанавливаемые пользователями, бывают качественными крайне редко. Сложность паролей, генерируемых различными информационными системами, на практике часто оказывается обесцененной самими работниками предприятия, записывающими

1 Панченко Александр Романович, аспирант, Южный Федеральный Университет «ЮФУ», Институт компьютерных технологий и информационной безопасности, г. Таганрог, Россия. ORCID: <https://orcid.org/0009-0001-4720-5164>. E-mail: alpanchenko@sfsedu.ru

2 Снегуров А. В., Чакрян В. Х. Анализ устойчивости ко взлому современных механизмов парольной защиты операционных систем // Восточно-Европейский журнал передовых технологий. 2011 – No 10. – С. 27–29.

и сохраняющими выданные им пароли». Иногда для лучшего запоминания сложного пароля люди используют «информативные последовательности символов»³, что тоже не самый надежный способ защиты данных. Важным моментом является тот факт, что небольшая модификация семантически значимой последовательности символов может превратить её в достаточно стойкий или даже очень стойкий пароль. Пример такого превращения фразы в пароль показан в статье [4, с. 159]. Суровая реальность показывает, что данным методом люди зачастую пренебрегают. При этом число сервисов, учетных записей и аккаунтов с каждым годом неуклонно растёт, и для каждого из них требуется свой, в идеальном варианте, уникальный пароль. Но практика показывает, что часто один пароль используется во множестве сервисов и аккаунтов либо с минимальными изменениями, либо вообще без них. «Мы обнаружили, что повторное использование и модификация паролей являются очень распространенным явлением (наблюдается у 52 % пользователей)»⁴. «Также очень важно не использовать одинаковые пароли для разных учётных записей, ведь если один пароль будет скомпрометирован, это может повлечь за собой взлом и других учётных записей, где соответствующий пароль используется» [5]. А если учитывая факт про сложность паролей, то ситуация становится критической [6, с. 66]. Статистика фишинга подтверждает всю критичность ситуации на данный момент: «Что касается глобальных тенденций, то в 2022 году APWG зафиксировала около 4,7 миллиона фишинговых атак. С 2019 года наблюдается ежегодное увеличение числа атак более чем на 150 %» [7, с. 3]. Важно упомянуть еще и технические аспекты, связанные с ограничениями при создании пароля в разных ОС. Численные характеристики данного аспекта показаны в статье [8]. Проанализировав все минусы парольной аутентификации можно прийти к выводу, что на данный момент направление, связанное с беспарольной аутентификацией [9, 10, с. 144] – переходом от фактора знания (пароля), к фактору владения является наиболее перспективным в сложившихся условиях.

Постановка задачи

В настоящей работе рассматривается схема и принцип работы стандарта FIDO2. Данные об текущем стандарте были получены из документации стандарта с официального сайта FIDO, а также из статьи

[11]. Информация о протоколе WebAuthn была получена из официального сайта консорциума W3C, а также из статей [12–14]. Важно заметить, что предыдущие анализы безопасности WebAuthn показывали его достаточно высокую криптостойкость в рамках стандарта FIDO2 [15, с. 453]. Принцип работы протокола CTAP2.1 был получен из официальной документации FIDO, а также из статей [16, 17]. В данной статье он не затрагивается так как не подвержен атаке, но он является частью стандарта. Отдельно рассматривается уязвимость, которую возможно осуществить на текущую версию протокола WebAuthn. Данная работа рассматривает уязвимость с точки зрения совокупности недостатков таких как WebAuthn, который является JavaScript API, и уязвимости, связанной с архитектурой стандарта, в которой присутствует возможность использования нелегитимных программ в качестве клиента. Данная атака реализовывалась автором самостоятельно, на момент написания статьи автору не попадались статьи, описывающие данную уязвимость.

Цель данного исследования заключается в подтверждении отсутствия доказательной устойчивости у текущего стандарта беспарольной аутентификации FIDO2 для последующей разработки безопасного протокола беспарольной аутентификации с использованием технологии Passkey [18, с. 203] на основе выявленных уязвимостей.

Описание стандарта FIDO2

Объектом исследования является стандарт беспарольной аутентификации FIDO2. В рамках работы FIDO2 лежат алгоритмы хеширования и принцип ассиметричного шифрования. Стойкие алгоритмы хеширования являются залогом стойкости криптосистем и текущий стандарт в этом случае не исключение. В источнике [19] первая глава посвящена алгоритмам хеширования.

В рамках работы стандарта FIDO2 используется ассиметричная криптография в варианте электронной подписи. Для этого открытый ключ отправляется в сервис и сохраняется на его сервере. Закрытый ключ надёжно хранится на пользовательском устройстве аутентификации (токене). Основным способом взаимодействия пользователя и аутентификатора является Authorization Gesture [20] или «Разрешающий жест» или «Жест авторизации». Под данным понятием подразумевается физическое взаимодействие пользователя с устройством аутентификации в рамках таких процедур как регистрация и аутентификация. Выполняя данный «жест», пользователь подтверждает, что разрешает выполнение процедуры. Данное действие может включать проверку пользователя, если используемое устройство аутентификации способно на это. Например, биометрическая

3 Тюрин К. А., Семин Р. В. Анализ стойкости парольных фраз на основе информационной энтропии // Известия ЮФУ. Технические науки. 2015 – No 5. – С. 18–27.

4 Chun Wang, Steve T.K. Jan, Hang Hu, Douglas Bossart, Gang Wang. The Next Domino to Fall: Empirical Analysis of User Passwords across Online Services // CODASPY '18: Proceedings of the Eighth ACM Conference on Data and Application Security and Privacy. 2018. – pp. 196-203. – DOI:10.1145/3176258.3176332.

проверка (автоматическое распознавание индивидов на основе их биологических и поведенческих характеристик) или простой ввод PIN-кода, или может включать простую проверку присутствия пользователя (простое нажатие на кнопку на аутентификаторе). Данный стандарт состоит из двух протоколов: WebAuthn и CTAP2.1.

Протокол веб-аутентификации

WebAuthn – это интерактивный протокол между тремя сторонами: токеном (ассоциируется с пользователем, который выполняет «жест авторизации»), клиентом (в рамках данного протокола это может быть приложение или веб-сервис) и сервером. Токен связан с открытым ключом подтверждения, который предварительно зарегистрирован на сервере. Протокол определяет два типа взаимодействий: регистрацию и аутентификацию. В рамках данной статьи рассмотрим лишь аутентификацию, которая и подвергается атаке.

Формальная схема процесса работы протокола WebAuthn во время выполнения аутентификации представлена на рисунке 1.

Во время выполнения процесса аутентификации используются следующие алгоритмы: Achallenge, Acommand, Aresponse, Acheck.

Инициация аутентификации происходит на стороне сервера алгоритмом Achallenge. В ходе его работы снова создается случайная строка битов rs длиной l или более. Сервер не передается клиенту (браузеру) по каналу связи идентификатор пользователя uid , который был создан ранее, во время регистрации пользователя. В итоге клиенту отправляется сообщение cr , содержащее в себе идентификатор сервера idS и rs .

На стороне клиента происходит проверка идентификатора сервера и разрыв связи в случае несоответствия. В итоге работы алгоритма Acommand происходит формирование сообщения аутентификации M_a , в котором содержится идентификатор сервера и хешированная случайная строка, полученная от сервера.

На стороне токена алгоритм Aresponse извлекает учетные данные, связанные с идентификаторами сервера, из контекста регистрации токена. Затем увеличивает счетчик n на единицу. Далее происходит формирование подписи и ответного сообщения, а также обновление регистрационного состояния токена.

В конце на сервере запускается алгоритм Acheck, который по идентификатору учетных данных cid извлекает контекст регистрации на сервере. После этого происходит проверка если проверка успешна, то регистрационное состояние обновляется на сервере.

Атака на протокол WebAuthn

В рамках основной задачи проводилось исследование протокола и его формализованной формы на наличие уязвимостей. В ходе анализа, а также изучения информации из свободных источников была обнаружена потенциальная уязвимость, которая позволяет получить доступ к учетной записи с включенной беспарольной аутентификацией по стандарту FIDO2. Данная уязвимость подтверждает отсутствие доказательной устойчивости у текущего стандарта. В качестве аутентификатора (токена) выступал персональный компьютер с включенным windows hello. Запрос осуществляется в рамках аутентификации на веб-сервисе.

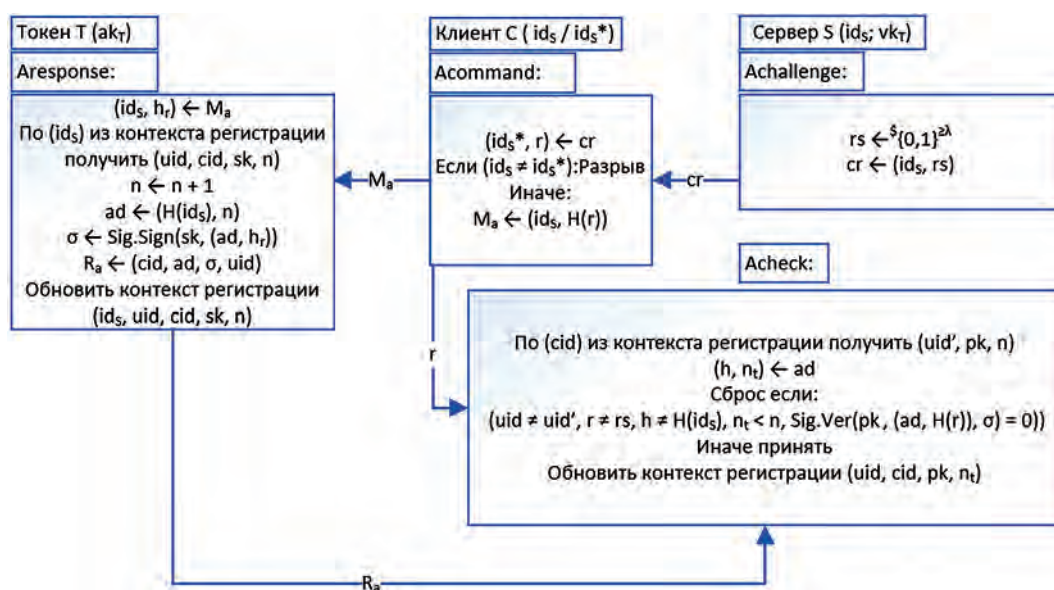


Рис. 1. Протокол WebAuthn (аутентификация)

В рамках выполнения данной атаки нужно обговорить несколько важных моментов и только после этого перейти к рассмотрению сценария атаки, его реализации и результатам её выполнения.

Данная атака проводится на уже зарегистрированного пользователя. Сам пользователь мог ранее быть зарегистрирован на основании классического логина и пароля, но позже к его аккаунту была подключена и беспарольная аутентификация на основе стандарта FIDO2. Проверяющая сторона (RP) осуществляет проверку пользователя на основе криптографических подходов – цифровой подписи. В рамках стандарта, если пользователь зарегистрирован сразу беспарольным способом или включил его позднее, можно сказать, что у него уже есть пара криптографических ключей. Открытый ключ и учетные данные уже есть на сервере RP, а закрытый ключ никогда не покидает токен пользователя. При аутентификации пользователь запрашивает сервер RP, который извлекает идентификаторы учетных данных пользователя, запросившего аутентификацию. После этого формируются параметры, которые должны быть переданы клиенту (браузеру) пользователя, через протокол https в ответ на вызов API (WebAuthn). Далее в рамках работы протокола WebAuthn происходит проверка идентификатора сервера, от имени которого пришли параметры. Если все верно, то происходит вызов функции `navigator.credentials.get()`. Данная функция передает информацию по протоколу STAP2.1 на аутентификатор. Обычно для подписи запроса используется алгоритм ECDSA^{5,6}.

Оценка уязвимости

Нужно понимать, что веб-аутентификация это JavaScript API, это значит, что WebAuthn будет следовать всем правилам работы данного языка программирования. А именно переопределению функций⁷. Переопределение функции в JavaScript – это способность дочернего класса или подкласса предоставить реализацию конкретного метода, который уже определен в родительском классе или его суперклассе. В данном случае это подразумевает, что можно запустить пользовательскую реализацию функции `navigator.credentials.get()` в виде «межсайтового скриптинга» (XSS) [20]. Данное действие должно переопределить собственную реализацию этой же функции в оригинальном WebAuthn. В итоге вызов WebAuthn не будут функционировать так, как это задумывалось.

Пример такого переопределения показан на рисунке 2.

```
> class cred
{
  static get(options)
  {
    console.log("Пример переопределения функции");
  }
}
class navigator
{
  navigator.credentials=cred;
}
< class cred
{
  static get(options)
  {
    console.log("Пример переопределения функции");
  }
}
```

Пример переопределения функции VM3937:5

Рис. 2. Пример переопределения функции

В рамках атаки злоумышленник может выполнить «самостоятельный межсайтовый скриптинг» (Self-XSS) в своей собственной системе, изменив функциональность WebAuthn для себя. В результате, мера защиты, которую, например, Google использует на странице входа в свой аккаунт, становится бесполезной, так как предупреждение видит только злоумышленник, совершающий это намеренно. Предупреждение показано на рисунке 3.

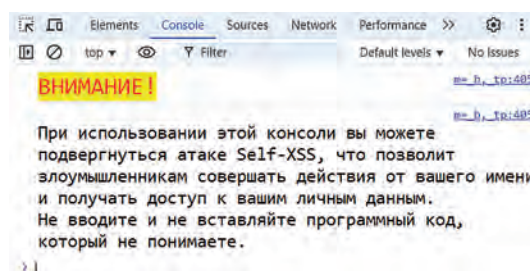


Рис. 3. Предупреждение в консоли на официальной странице google

Сценарий атаки

Рассмотрим сценарий самой атаки.

1. Злоумышленник должен развернуть сервер, который будет находиться в том же участке сети, что и легитимный пользователь. Данный сервер будет осуществлять все пересылку между злоумышленником и вредоносным ПО на компьютере жертвы. Важно заметить, что для попадания в локальную сеть жертвы могут использоваться недостатки или уязвимости из других веб-приложений. «По данным компании Positive Technologies 17 % от общего числа атак связаны с уязвимостями и недостатками защиты веб-приложений, которые могут быть использованы для проникновения в локальный сетевой периметр организации или распространения вредоносного программного обеспечения» [21].

5 Michael Braun, Anton Kargl. A Note on Signature Standards // IACR Cryptology ePrint Archive 2007/357, 2007.

6 Johnson, D., Menezes, A. & Vanstone, S. The Elliptic Curve Digital Signature Algorithm (ECDSA). IJIS 1, 36–63 (2001). <https://doi.org/10.1007/s102070100002>.

7 Bloch Joshua, Guy L.Steele, Jr. Effective Java: programming language guide. – edition 2. print. May 2001. – 180 pp.



Рис. 4. Сценарий атаки

2. Злоумышленник запускает приложение, который копирует в буфер обмена код нужной ему Self-XSS. Этот код далее необходимо вставить в консоль браузера перед вызовом WebAuthn. После того, как код будет вставлен произойдет переопределение navigator.credentials.get(), параметры из которой будут считаны приложение, запущенным на машине злоумышленника. Эти данные будут закодированы в формате Base64 и после этого будут пересланы на облачный сервер.

3. Вредоносным ПО в системе жертвы запрашивает у сервера данные в формате Base64. Затем декодирует данные в формат, который соответствует WebAuthn. После чего запрашивает токен для подписания данного запроса. Тут в дело вступает социальная инженерия, о которой уже говорил ранее. После того как запрос подписан он заново кодируется и отправляется обратно на сервер. Сам сервер после этого отправит ответ на систему атакующего, точнее на его приложение. Нужно добавить небольшую ремарку. Способ попадания вредоносной программы может быть самым разнообразным. «По статистике, веб-браузер является одним из основных инструментов доставки вредоносных программ на компьютеры пользователей» [22]. Потенциальным способом защиты в этом случае может быть фаззинг-тестирование. В статье [22] рассмотрены разные подходы к фаззингу. Для передачи вредоносного ПО на систему жертвы может использоваться DNS-туннелирование. В статье [23] отражены основные особенности работы DNS-туннелирования как способа доставки вредоносного ПО на узел, защищенный межсетевым экраном.

4. Далее приложение декодирует Base64 и возвращает его вредоносной реализации navigator.credentials.get() которую запускали при помощи Self-XSS из буфера обмена в самом начале атаки. После этого вредоносная реализация декодирует его обратно в ArrayBuffers и возвращает на сервер RP.

Сценарий атаки в виде схемы показан на рисунке 4.

Выполнение атаки

В рамках ранее описанного сценария проведем выполнение атаки и зафиксируем ее результаты. Первое, что нужно сделать – развернуть сервер, который будет пересылать сообщения между участниками атаки. Данный сервер является основным в данной атаке, но функционал его достаточно прост – пересылка сообщений между ПК атакующего и ПК жертвы. Он, как и все остальные компоненты эксплойта,



Рис. 5. Схема работы сервера между ПК атакующего и жертвы

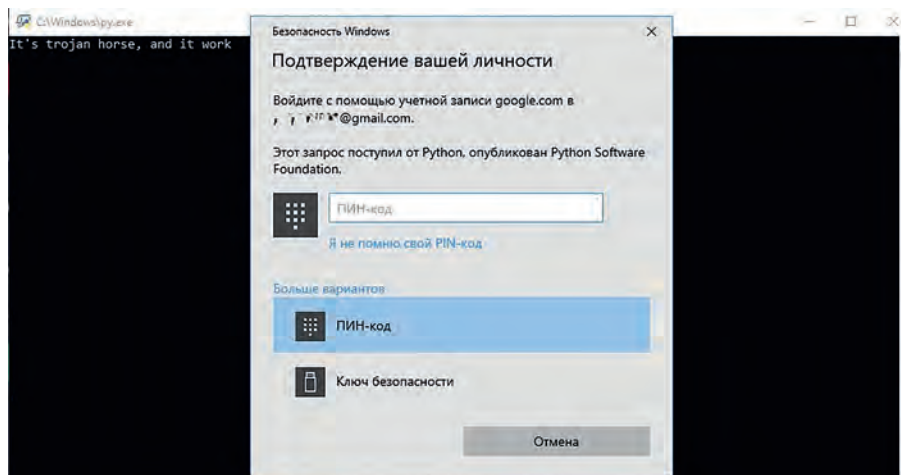


Рис. 6. Запрос «жеста авторизации»

был реализован автором. Сам сервер был реализован на основе Apache2 с использованием `mod_wsgi`, для использования скрипта, написанного на python с использованием библиотеки flask и расширением CORS. Пример блок-схемы работы сервера показан на рисунке 5.

Далее злоумышленник запускает на своей стороне программу, которая должна будет считать данные от вредоносной реализации. По понятным соображениям публиковать код эксплойта автор не должен, но принцип работы объяснен будет. Данную реализацию посредством Self-XSS запускает сам атакующий. Важно уточнить, что сделать это нужно перед запуском работы протокола WebAuthn.

Данные, полученные от вредоносной реализации функции WebAuthn, будут закодированы в нужный формат и пересланы на сервер, который уже был ранее запущен. После этого сервер отправит закодированные данные на вредоносное ПО в системе жертвы. Данное ПО декодирует данные в формат необходимый для подписи и, имитируя из себя клиента, сделает запрос к аутентификатору. В рамках работы протокола СТАР2.1 будет выведен запрос на совершение «жеста авторизации», тут жертва, поддавшаяся атаке социальной инженерии, введет свой pin-код. Пример запроса «жеста авторизации» токена после запроса от вредоносного программного обеспечения типа троянский конь, показана на рисунке 6.

После того как будет выполнен «жест авторизации» и запрос будет подписан, вредоносное ПО закодирует запрос в нужный формат и отправит его обратно на сервер. Код по понятным причинам опубликован быть не может, поэтому будет приведена блок-схема алгоритма работы вредоносного ПО (рис. 7).

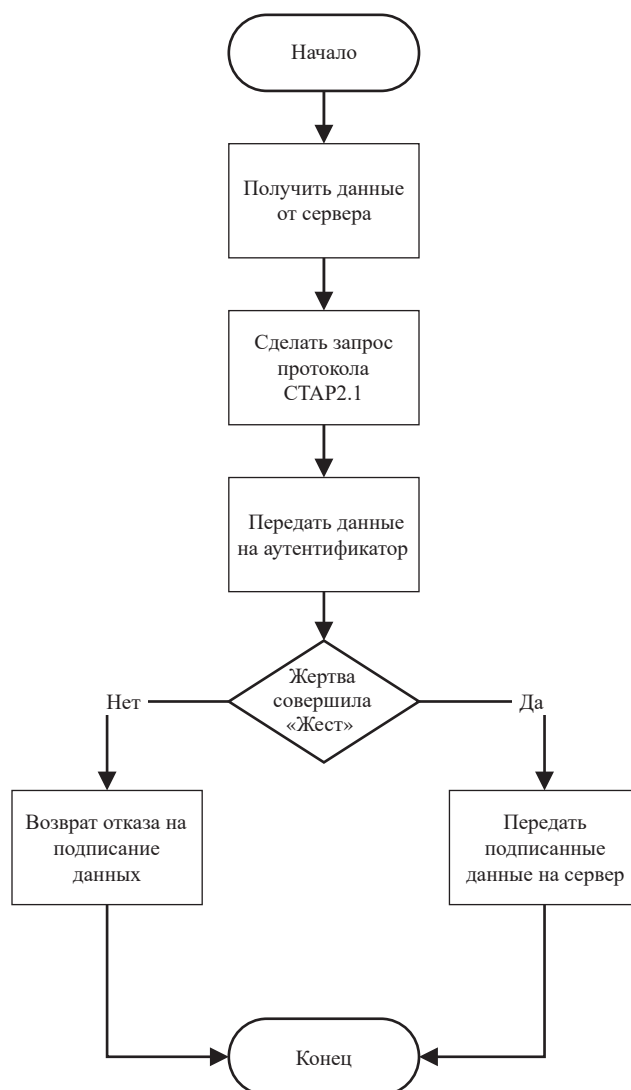


Рис. 7. Схема работы вредоносного ПО

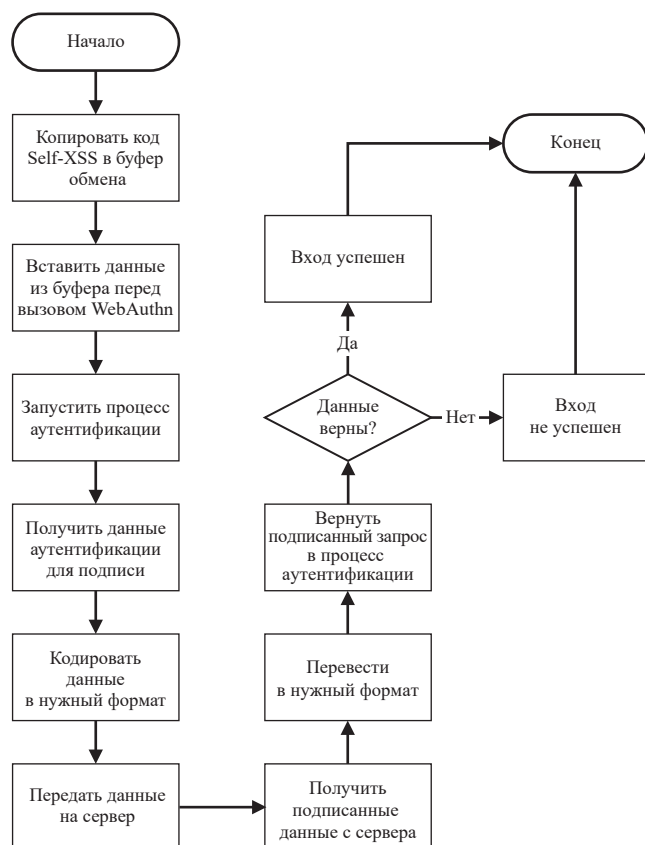


Рис. 8. Схема работы программы на ПК атакующего

После этого приложение на компьютере атакующего получит данные с сервера, переведет их в нужный формат и вернет во вредоносную реализацию, которая ранее была запущена. В конце вредоносная реализация вернет данные в формат WebAuthn и отправит их на сервер RP. После этого сервер проверит подпись и убедившись в том, что ключ и подпись соответствуют тем, что были с момент регистрации

легитимного пользователя, предоставит доступ. Пример блок-схемы алгоритма работы программы на ПК атакующего показан на рисунке 8.

В итоге, в рамках экспериментов, был получен доступ к учетной записи Google, которая использовала беспарольную аутентификацию. Подтверждение успешной аутентификации показан на рисунке 9. Код 200 означает успех.

Потенциальные способы защиты

На данный момент рассматривается несколько потенциальных вариантов, если не закрытия данной уязвимости, то хотя бы усложнения ее реализации. Основой атаки выступает возможность переопределения функции в языке JavaScript. Это означает, что большинство браузеров подвержены данной уязвимости. Предлагаемое автором решение не является истинной в последней инстанции, а лишь рекомендация по улучшению реализации WebAuthn. При длительном рассмотрении будет найдены способы обхода данного изменения, но данное изменение позволяет усложнить подготовку атаки. Пример уязвимой реализации WebAuthn показан на рисунке 10.

На девятнадцатой строке вызывается `navigator.credentials.get()`, которая может быть переопределена злоумышленником за счет Self-XSS. Пример такого переопределения ранее был показан на рисунке 2. Способ защиты следующий. Известно, что скрипты, которые находятся в теге `head` выполняются в первую очередь. И не асинхронные функции в теге `head` всегда будут выполняться до вызова любой другой функции `onload` или до выполнения Self-XSS в консоли. Эта особенность была бы полезна для устранения уязвимости, описанной в данной статье. В рекомендуемой реализации WebAuthn можно предложить разработчику создать некую

```

[02/Mar/2025 20:09:02] "OPTIONS /getoptions?site=https://accounts.google.com HTTP/1.1" 200 -
[02/Mar/2025 20:09:18] "POST /getoptions?site=https://accounts.google.com HTTP/1.1" 200 -
  
```

Рис. 9. Ответ сервера RP

```

11 <script>
12   function start()
13   {
14     fetch('/api/authenticate/begin' , {
15       method: "POST",
16     }).then(function(response) {
17       if(response.ok) return response.arrayBuffer();
18       throw new Error('No credential available to authenticate! ');
19     }).then(CBOR.decode). then(function(options) {
20       return navigator.credentials.get(options);
21     }).then ( function(assertion) {
22       return fetch('/api/authenticate/complete', {
23         method: "POST",
24         headers: [{"Content-Type": "application/cbor"}
25         body: CBOR.encode({
26           "credentialId": new Uint8Array(assertion.rawId),
27           "authenticatorData": new Uint8Array(assertion.response.authenticatorData),
28           "clientDataJSON": new Uint8Array(assertion.response.clientDataJSON) ,
29           "signature": new Uint8Array(assertion.response.signature)
30         })
31       });
32     });
33   }
  
```

Рис. 10. Пример уязвимой реализации WebAuthn


```

15 function start()
16   fetch('/api/authenticate/begin' {,
17     method: "POST",
18   }).then(function(response) {
19     if(response.ok) return response.arrayBuffer();
20     throw new Error('No credential available to authenticate! ');
21   }).then(CBOR.decode).then(function(options) {
22     return unknowntotheattackerfunction(options);
23   }).then ( function(assertion) {
24     return fetch('/api/authenticate/complete', {
25       method: "POST",
26       headers: {"Content-Type": "application/cbor"},
27       body: CBOR.encode({
28         "credentialId": new Uint8Array(assertion.rawId),
29         "authenticatorData": new Uint8Array(assertion.response.authenticatorData),
30         "clientDataJSON": new Uint8Array(assertion.response.clientDataJSON),
31         "signature": new Uint8Array(assertion.response.signature)

```

Рис. 11. Модифицированная версия реализации WebAuthn

резервную копию функции navigator.credentials.get() с таким именем, которое может быть неизвестно для злоумышленника. Данное имя нужно сделать или динамическим, или случайным. Или данное имя может быть скрыто, чтобы злоумышленник не знал его. Последний вариант возможно сделать, с помощью такой строки кода: const unknowntotheattackerfunction=navigator.credentials.get.bind(navigator.credentials), расположенной в теге head, тем самым зафиксировав её выполнение до консоли.

В основной части веб-страницы вместо «navigator.credentials.get()» нужно использовать вызов функции «unknowntotheattackerfunction()». Пример модифицированной версии реализации WebAuthn показан на рисунке 11.

Данная модификация позволит усложнить реализацию атаки так как указатель на исходную функцию уже находится в браузере и не может быть переопределен, если имя скрыто. Данная защита не панацея, а лишь усложнение атаки для злоумышленника.

Отдельно можно рассмотреть скорее не способ защиты, а способ предупреждения пользователя. Так как текущая версия стандарта FIDO2 сильно полагается на «жест авторизации», то необходимо расширить данные, которые собираются на этапе регистрации, а затем отправляются на подпись пользователю при входе. Нужно фиксировать «fingerprint» браузера, с которого происходила регистрация и в момент входа сверять его с тем, что хранится на сервере. Если они разные, то нужно записать в «параметры» функции navigator.credentials.get() флаг с тем, что данные не совпали и это потенциально не легитимный пользователь. Далее данный флаг будет обработан токеном и будет выдано предупреждение: «Обнаружен вход с неизвестного устройства. Если это не вы, то прервите текущие операции входа!» Данная атака не затрагивает протокол CTAP2.1, а лишь социальной инженерией заставляет пользователя поверить в то, что вход осуществляет он. И пользователь не видит никаких предупреждений и может не насторожиться.

Заключение

В настоящей работе проведено исследование текущего стандарта беспарольной аутентификации FIDO2 на наличие уязвимостей. В рамках обнаруженной уязвимости было проведена разработка, развертывание всех необходимых компонентов. После чего экспериментально была подтверждена возможность эксплуатации данной уязвимости на актуальной версии стандарта FIDO2. Данная уязвимость позволяет заявить о возможности несанкционированного доступа к учетным данным пользователя во время процесса аутентификации. Уязвимость оказывает влияние на безопасность систем аутентификации.

В рамках данного исследования было подтверждено отсутствие доказательной устойчивости у текущего стандарта. Обнаруженная уязвимость позволяет получить доступ к аккаунту, защищенному по текущему стандарту беспарольной аутентификацией. При реализации данной уязвимости сложностью для атакующего в ней является заражение компьютера жертвы вредоносным программным обеспечением типа троянский конь и последующая атака социальной инженерии.

В рамках стандарта беспарольной аутентификации FIDO2 считается, что «жест авторизации», который делает пользователь, и информация, которая отобразится на токене, позволит человеку отличить настоящий запрос, сделанный им, от мошеннической схемы. Данное предположение можно было бы считать верными, если бы не статистика успешных мошеннических атак по всему миру за последние несколько лет. Есть предложения касательно способа потенциальной защиты от данной уязвимости, а также способ сбора дополнительных данных на этапе регистрации для уведомления пользователя в процессе атаки. Все данные способы, с точки зрения автора, скорее временные заплатки. Для полноценной защиты нужно рассматривать другую архитектуру для будущего протокола, которую автор надеется предложить в будущем.

Литература

1. Современные направления применения комбинаторики в области защиты персональных данных / М. О. Тенякина, М. В. Богданова, К. И. Быкова, К. А. Сакалова // Международный научно-исследовательский журнал. – 2024. – № 9(147). – DOI: 10.60797/IRJ.2024.147.1.
2. Салита, Д. С. Методы оценки надежности парольных систем / Д. С. Салита, А. А. Удовик // Проблемы правовой и технической защиты информации. – 2020. – № 8. – С. 47–51.
3. Информационная безопасность: парольная защита / И. А. Рябов, М. Г. Койцан, А. А. Кузнецов, В. В. Ермолаева // Тенденции развития науки и образования. – 2023. – № 93-8. – С. 76–79. – DOI 10.18411/trnio-01-2023-401.
4. Назаров, Д. М. Методика создания надежного пароля для обеспечения экономической безопасности в условиях цифровизации / Д. М. Назаров // Известия Санкт-Петербургского государственного экономического университета. – 2022. – № 1(133). – С. 155–160.
5. Кочанова А. Г. Надёжные пароли: как их создать и чем они полезны // Вестник науки. 2023. № 6(63).
6. Абидарова А. А. Анализ надежности паролей для обеспечения информационной безопасности // Известия Тульского государственного университета. Технические науки. 2021. – № 8. – С. 66–68.
7. Селиверстов В. В, Корчагин С. А. Анализ актуальности и состояния современных фишинг-атак на объекты критической информационной инфраструктуры // Инженерный вестник Дона. 2024. – № 6. – С. 216–229.
8. Степкин, Б. А. Как требования к паролю влияют на его безопасность / Б. А. Степкин, С. В. Малахов // Скиф. Вопросы студенческой науки. – 2021. – № 4(56). – С. 83–86.
9. Lyastani, S. G., Schilling, M., Neumayr, M., Backes, M., & Bugiel, S. (2020). Is FIDO2 the kingslayer of user authentication? A Comparative Usability Study of FIDO2 Passwordless Authentication. 2022 IEEE Symposium on Security and Privacy (SP), 268–285. <https://doi.org/10.1109/sp40000.2020.00047>
10. Докучаев В А., Мытенков С. С., Рахмани Д. Д., Сафонов И. А. Анализ уязвимостей и рисков традиционных парольных систем в контексте корпоративных распределенных систем и критически важных инфраструктур // Экономика и качество систем связи. 2025. № 36. С. 135–147.
11. Mitra, A., & Ghosh, A. (2024). FIDO2: A comprehensive study on passwordless authentication. International Journal of Engineering Research and Applications, 14(7), 58–63. <https://doi.org/10.9790/9622-14075863>.
12. Dourado, M. R., Gestal, M., & Vázquez-Naya, J. M. (2020). Implementing a web application for W3C WebAuthn protocol testing. MDPI, 5. <https://doi.org/10.3390/proceedings2020054005>.
13. Bindel, N., Gama, N., Guasch, S., Ronen, E. (2023). To Attest or Not to Attest, this is the Question – Provable Attestation in FIDO2. In: Guo, J., Steinfeld, R. (eds) Advances in Cryptology – ASIACRYPT 2023. ASIACRYPT 2023. Lecture Notes in Computer Science, vol. 14443. Springer, Singapore. https://doi.org/10.1007/978-981-99-8736-8_10.
14. Hanzlik, L., Loss, J., & Wagner, B. (2023). Token meets Wallet: Formalizing Privacy and Revocation for FIDO2. 2022 IEEE Symposium on Security and Privacy (SP), 1491–1508. <https://doi.org/10.1109/sp46215.2023.10179373>.
15. Gudipati, R. R. (2025). Demystifying fido: a technical deep dive into modern authentication // International journal of information technology and management information systems, 16(2), 452–466. https://doi.org/10.34218/ijitmis_16_02_029.
16. Bindel, N., Cremers, C., & Zhao, M. (2023b). FIDO2, CTAP 2.1, and WebAuthN 2: Provable Security and Post-Quantum Instantiation. 2022 IEEE Symposium on Security and Privacy (SP). <https://doi.org/10.1109/sp46215.2023.10179454>.
17. Guan, J., Li, H., Ye, H., Zhao, Z. (2022). A Formal Analysis of the FIDO2 Protocols. In: Atluri, V., Di Pietro, R., Jensen, C. D., Meng, W. (eds) Computer Security – ESORICS 2022. ESORICS 2022. Lecture Notes in Computer Science, vol 13556. Springer, Cham. https://doi.org/10.1007/978-3-031-17143-7_1.
18. Dr. A.Shaji George. (2024). The Dawn of Passkeys: Evaluating a Passwordless Future. Partners Universal Innovative Research Publication (PUIRP), 02(01), 202–220. <https://doi.org/10.5281/zenodo.10697886>.
19. Криптографические основы блокчейн-технологий / Е. А. Ищукова, С. П. Панасенко, К. С. Романенко, В. Д. Салманов. – Москва: ООО «ДМК Пресс. Электронные книги», 2022. – 301 с.
20. Barbosa, M., Boldyreva, A., Chen, S., Warinschi, B. (2021). Provable Security Analysis of FIDO2. In: Malkin, T., Peikert, C. (eds) Advances in Cryptology – CRYPTO 2021. CRYPTO 2021. Lecture Notes in Computer Science(), vol 12827. Springer, Cham. https://doi.org/10.1007/978-3-030-84252-9_5.
21. Боровков В. Е. Методы защиты веб-приложений от злоумышленников / В. Е. Боровков, П. Г. Ключарев // Вопросы кибербезопасности. – 2023. – № 5(57). – С. 89–99. – DOI 10.21681/2311-3456-2023-5-89-99.
22. Козачок, А. В. Подходы к оценке поверхности атаки и фаззингу веб-браузеров / А. В. Козачок, Д. А. Николаев, Н. С. Ерохина // Вопросы кибербезопасности. – 2022. – № 3(49). – С. 32–43. – DOI 10.21681/2311-3456-2022-3-32-43.
23. Москвичев А. Д. Использование DNS-туннелирования для передачи вредоносного программного обеспечения / А. Д. Москвичев, К. С. Москвичева // Вопросы кибербезопасности. – 2022. – № 4(50). – С. 91–99. – DOI 10.21681/2311-3456-2022-4-91-99.

USES OF WEBAUTHN PROTOCOL VULNERABILITIES TO OBTAIN UNSANCTIONED ACCESS

Panchenko A. R.⁸

Keywords: password-free authentication; asymmetric encryption; Self-XSS; Python; FIDO2; CTAP2.1; Authorization Gesture.

Purpose of the study: investigation of the current FIDO2 password-free authentication standard for vulnerabilities and confirmation of its lack of evidence-based stability.

Methods of research: analysis of the current password-free authentication standard for vulnerabilities. Implementation and exploitation of the discovered vulnerability. This vulnerability is based on the use of a modified man-in-the-middle attack using malicious software and social engineering.

⁸ Alexander R. Panchenko, postgraduate student, Southern Federal University «SFedU», Institute of Computer Technologies and Information Security, Taganrog, Russia. ORCID: <https://orcid.org/0009-0001-4720-5164>. E-mail: alpanchenko@sfedu.ru

Research results: the analysis of the current standard of password-free authentication is carried out. As part of this analysis, a formal scheme of the WebAuthn protocol was formed. A vulnerability has been discovered that allows access to the account of a legitimate user protected by the current FIDO2 standard. To implement this vulnerability, a modified man-in-the-middle attack was used, in which a server was implemented that transmits messages during the attack. Malicious Trojan horse type software was implemented. This malicious software was previously placed in the victim's system, where it posed as the victim's browser, which was contacted by a legitimate server. On the attacker's side, a program was implemented that reads the authentication request in the attacker's system. This program is used to override the function during the authentication process and then forward the request to the server. As part of the implementation of this vulnerability and its subsequent implementation, the lack of evidence-based stability of the current FIDO2 password-free authentication standard was confirmed. A method of potential protection against this vulnerability has been proposed. A method was also proposed to modify data collection at the registration stage, for subsequent notification of a potential victim attack during the signing procedure.

Scientific novelty: the discovered vulnerability confirms the lack of evidence-based stability of the current FIDO2 password-free authentication standard. Currently, all browsers use WebAuthn, which means that all modern browsers are affected by this vulnerability. A potential way to protect against this vulnerability and a way to prevent the victim from being attacked is proposed.

References

1. Sovremennye napravleniya primeneniya kombinatoriki v oblasti zashhity personal'nyh dannyh / M. O. Tenjachkina, M. V. Bogdanova, K. I. Bykova, K. A. Sakalova // Mezhdunarodnyj nauchno-issledovatel'skij zhurnal. – 2024. – № 9(147). – DOI: 10.60797/IRJ.2024.147.1.
2. Salita, D. S. Metody ocenki nadezhnosti parol'nyh sistem / D. S. Salita, A. A. Udovik // Problemy pravovoj i tehnicheckoj zashhity informacii. – 2020. – № 8. – S. 47–51.
3. Informacionnaja bezopasnost': parol'naja zashhita / I. A. Rjabov, M. G. Kojcan, A. A. Kuznecov, V. V. Ermolaeva // Tendencii razvitiya nauki i obrazovanija. – 2023. – № 93-8. – S. 76–79. – DOI 10.18411/trnio-01-2023-401.
4. Nazarov, D. M. Metodika sozdaniya nadezhnogo parolja dlja obespechenija jekonomicheskoy bezopasnosti v uslovijah cifrovizacii / D. M. Nazarov // Izvestija Sankt-Peterburgskogo gosudarstvennogo jekonomicheskogo universiteta. – 2022. – № 1(133). – S. 155–160.
5. Kochanova A. G. Nadjozhnye paroli: kak ih sozdat' i chem oni polezny // Vestnik nauki. 2023. № 6(63).
6. Abidarova A. A. Analiz nadezhnosti parolej dlja obespechenija informacionnoj bezopasnosti // Izvestija Tul'skogo gosudarstvennogo universiteta. Tehniceskie nauki. 2021. – № 8. – S. 66–68.
7. Seliverstov V. V, Korchagin S. A. Analiz aktual'nosti i sostojanija sovremennyh fishing-atak na ob'ekty kriticeskoj informacionnoj infrastruktury // Inzhenernyj vestnik Dona. 2024. – № 6. – S. 216–229.
8. Stepkin, B. A. Kak trebovanija k parolju vlijajut na ego bezopasnost' / B. A. Stepkin, S. V. Malahov // Skif. Voprosy studencheskoj nauki. – 2021. – № 4(56). – S. 83–86.
9. Lyastani, S. G., Schilling, M., Neumayr, M., Backes, M., & Bugiel, S. (2020). Is FIDO2 the kingslayer of user authentication? A Comparative Usability Study of FIDO2 Passwordless Authentication. 2022 IEEE Symposium on Security and Privacy (SP), 268–285. <https://doi.org/10.1109/sp40000.2020.00047>
10. Dokuchaev V A., Mytenkov S. S., Rahmani D. D., Safonov I. A. Analiz ujazvimostej i riskov tradicionnyh parol'nyh sistem v kontekste korporativnyh raspredelennyh sistem i kriticeski vaznyh infrastruktur // Jekonomika i kachestvo sistem svjazi. 2025. № 36. S. 135-147.
11. Mitra, A., & Ghosh, A. (2024). FIDO2: A comprehensive study on passwordless authentication. International Journal of Engineering Research and Applications, 14(7), 58–63. <https://doi.org/10.9790/9622-14075863>.
12. Dourado, M. R., Gestal, M., & Vázquez-Naya, J. M. (2020). Implementing a web application for W3C WebAuthn protocol testing. MDPI, 5. <https://doi.org/10.3390/proceedings2020054005>.
13. Bindel, N., Gama, N., Guasch, S., Ronen, E. (2023). To Attest or Not to Attest, this is the Question – Provable Attestation in FIDO2. In: Guo, J., Steinfeld, R. (eds) Advances in Cryptology – ASIACRYPT 2023. ASIACRYPT 2023. Lecture Notes in Computer Science, vol 14443. Springer, Singapore. https://doi.org/10.1007/978-981-99-8736-8_10.
14. Hanzlik, L., Loss, J., & Wagner, B. (2023). Token meets Wallet: Formalizing Privacy and Revocation for FIDO2. 2022 IEEE Symposium on Security and Privacy (SP), 1491–1508. <https://doi.org/10.1109/sp46215.2023.10179373>.
15. Gudipati, R. R. (2025). Demystifying fido: a technical deep dive into modern authentication // International journal of information technology and management information systems, 16(2), 452–466. https://doi.org/10.34218/ijitmis_16_02_029.
16. Bindel, N., Cremers, C., & Zhao, M. (2023b). FIDO2, CTAP 2.1, and WebAuthN 2: Provable Security and Post-Quantum Instantiation. 2022 IEEE Symposium on Security and Privacy (SP). <https://doi.org/10.1109/sp46215.2023.10179454>.
17. Guan, J., Li, H., Ye, H., Zhao, Z. (2022). A Formal Analysis of the FIDO2 Protocols. In: Atluri, V., Di Pietro, R., Jensen, C.D., Meng, W. (eds) Computer Security – ESORICS 2022. ESORICS 2022. Lecture Notes in Computer Science, vol 13556. Springer, Cham. https://doi.org/10.1007/978-3-031-17143-7_1.
18. Dr. A.Shaji George. (2024). The Dawn of Passkeys: Evaluating a Passwordless Future. Partners Universal Innovative Research Publication (PUIRP), 02(01), 202–220. <https://doi.org/10.5281/zenodo.10697886>.
19. Kriptograficheskie osnovy blokchejn-tehnologij / E. A. Ishhukova, S. P. Panasenko, K. S. Romanenko, V. D. Salmanov. – Moskva: OOO «DMK Press. Jelektronnye knigi», 2022. – 301 s.
20. Barbosa, M., Boldyreva, A., Chen, S., Warinschi, B. (2021). Provable Security Analysis of FIDO2. In: Malkin, T., Peikert, C. (eds) Advances in Cryptology – CRYPTO 2021. CRYPTO 2021. Lecture Notes in Computer Science(), vol 12827. Springer, Cham. https://doi.org/10.1007/978-3-030-84252-9_5.
21. Borovkov V.E. Metody zashhity veb-prilozhenij ot zloumyshlennikov / V. E. Borovkov, P. G. Kljucharev // Voprosy kiberbezopasnosti. – 2023. – № 5(57). – S. 89–99. – DOI 10.21681/2311-3456-2023-5-89-99.
22. Kozachok, A. V. Podhody k ocenke poverhnosti ataki i fazzingu veb-brauzerov / A. V. Kozachok, D. A. Nikolaev, N. S. Erohina // Voprosy kiberbezopasnosti. – 2022. – № 3(49). – S. 32-43. – DOI 10.21681/2311-3456-2022-3-32-43.
23. Moskvichev A. D. Ispol'zovanie DNS-tunnelirovanija dlja peredachi vredonosnogo programmogo obespechenija / A. D. Moskvichev, K. S. Moskvicheva // Voprosy kiberbezopasnosti. – 2022. – № 4(50). – S. 91-99. – DOI 10.21681/2311-3456-2022-4-91-99.

ОЦЕНКА ЗАЩИЩЁННОСТИ ACTIVE DIRECTORY С ИСПОЛЬЗОВАНИЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Булгакова Е. В.¹, Богданов Е. А.², Кубанков А. Н.³

DOI: 10.21681/2311-3456-2025-6-58-68

Цель исследования: применение методов искусственного интеллекта, в частности машинное обучение и нейронные сети, для оценки защищённости системы Microsoft Active Directory и выявления факторов, влияющих на уровень безопасности корпоративных сетей. Работа направлена на разработку и сравнение алгоритмов, способных прогнозировать степень уязвимости пользователей и подсистем Active Directory.

Методы исследования: исследование основывается на методах машинного обучения, включающие метрические алгоритмы (линейная регрессия, метод ближайших соседей, дерево решений, случайный лес, градиентный бустинг) и нейронные сети, реализованные в среде Jupyter Notebook с применением библиотек pandas, sklearn и keras. На основе подготовленного датасета выполнена стандартизация и нормализация параметров, отражающих конфигурацию пользователей Active Directory. Для проверки эффективности алгоритмов проведено сравнение по критериям точности прогнозирования и среднеквадратичной ошибки.

Результаты исследования: проведён анализ факторов, влияющих на защищённость корпоративной инфраструктуры Active Directory, включая тип операционной системы, длину и срок действия пароля, уровень привилегий, параметры делегирования и наличие предварительной аутентификации Kerberos. На основе подготовленного датасета были реализованы и протестированы различные алгоритмы машинного обучения. Результаты показали, что модель дерева решений продемонстрировала наилучшие показатели — точность прогнозирования 0,96 и среднеквадратичную ошибку 0,091, что свидетельствует о её высокой эффективности в задаче оценки защищённости Active Directory. В дополнение была разработана модель нейронной сети и подтверждена её способность корректно обрабатывать параметры Active Directory и определять степень безопасности пользователей этой системы. Полученные результаты указывают на перспективность применения технологий искусственного интеллекта для автоматизации анализа уязвимостей и прогнозирования рисков информационной безопасности в корпоративных сетях.

Научная новизна: научная новизна исследования заключается в разработке и апробации интегрированного подхода к оценке защищённости Active Directory на основе машинного обучения и нейронных сетей. Предложено использование интеллектуальных моделей для прогнозирования уровня безопасности пользователей с учётом комплексных параметров инфраструктуры Active Directory, что позволяет формировать автоматизированную систему раннего предупреждения об уязвимостях корпоративных сетей.

Ключевые слова: Active Directory, информационная безопасность, искусственный интеллект, машинное обучение, нейронные сети, дерево решений, анализ защищённости, уязвимости, корпоративные сети

Введение

В настоящее время организации все чаще внедряют гибридные форматы работы. Важнейшей компонентой таких инфраструктур является служба каталогов Active Directory (AD), отвечающая за централизованное администрирование учетных записей пользователей, ресурсов и групп.

Увеличение сложности корпоративных ИТ-систем влечет за собой повышение требований к их безопасности. Риск несанкционированного доступа, применение устаревшего программного обеспечения создают условия для компрометации Active Directory.

В условиях постоянного роста объемов данных, традиционные методы анализа и аудита теряют в результативности. Это связано с человеческим фактором и с ограниченной масштабируемостью

существующих инструментов. Использование технологий искусственного интеллекта открывает новые перспективы для автоматизации оценки состояния безопасности инфраструктуры Active Directory, давая возможность выявлять отклонения и векторы атак злоумышленников.

В статье будут рассмотрены подходы к оценке защищённости Active Directory с использованием методов искусственного интеллекта.

Параметры рассматриваемой базы данных

Для анализа системы были определены основные параметры, которые влияют на безопасность корпоративной сети, и присвоены им значения (вес), которые будут влиять на конечное значение безопасности, целевым признаком будет относительная

1 Булгакова Елена Валерьевна, кандидат юридических наук, доцент, заместитель заведующего кафедрой информационной безопасности по научной работе, Финансовый университет при Правительстве РФ, г. Москва, Россия. E-mail: koordinator-proekta@mail.ru

2 Богданов Евгений Александрович, Ph.D., заведующий кафедрой информационной безопасности, Финансовый университет при Правительстве РФ, г. Москва, Россия. E-mail: eabogdanov@fa.ru

3 Кубанков Александр Николаевич, доктор военных наук, кандидат технических наук, профессор, профессор кафедры информационной безопасности, Финансовый университет при Правительстве РФ, г. Москва, Россия. E-mail: kan9991@gmail.com

защищенность системы. Более высокие значения веса – больше шанс компрометации системы. Таблица с весами для каждого параметра (табл. 1):

Таблица 1.
Параметры датасета с весами каждого значения

Параметр	Значение	Вес
Operating_System	Windows XP	30
	Windows 7	10
	Windows 10	5
	Windows 11	0
	Ubuntu	0
Relationship	Admins	10
	Peers	5
	Exec	5
	Users	5
	Active	3
Expiring_pass	True	0
	False	3
Pass_Length	Short	5
	Medium	3
	Long	0
Change_pass	True	0
	False	3
Auth_Kerberos	True	0
	False	10
Updating	True	0
	False	3
Privilege	High	50
	Medium	30
	Low	20
Delegation_Param	Const	0
	Resource	3
	Unconst	5
Safety	0	–
	1	

Описание параметров:

- Operating System отвечает за выбранную операционную систему.
- Relationship – параметр, показывающий, кем является пользователь:

- Admins – администраторы систем Active Directory;
- Peers – пользователи, близкие по полномочиям и по местоположению в сети к администраторам;
- Exec – сотрудники, имеющие ограниченный доступ к правам администратора;
- Users – обычные пользователи.
- Active – сотрудники, работающие за устройством на данный момент.
- Expiring_pass – параметр, показывающий просрочен пароль или нет.
- Pass_Length – длина пароля; параметр важен, так как пароль может быть получен методом полного перебора [1, 2].
- Change_pass – менял ли пользователь пароль.
- Auth_Kerberos – требует ли учетная запись предварительную аутентификацию Kerberos [3].
- Updating – показатель, отвечающий за обновление операционной системы.
- Privilege – доступность к файлам во всей структуре.
- Delegation_Param – делегирование функций.
- Safety – оценка защищенности системы AD (целевой параметр). 0 – система защищена, иначе – 1.

Подготовка данных в датасете

Для подготовки базы данных, с которой мы будем взаимодействовать, воспользуемся программным обеспечением Jupyter Notebook [4]. Для корректной работы с датасетом загружаем необходимые библиотеки, в первую очередь библиотеку Pandas [5].

Далее загружаем датасет и находим средние значения для каждого параметра (рис. 1).

```
#Средние значения для каждого параметра
df.mean()

Operating_System    6.955
Expiring_pass       1.473
Change_pass         1.374
Auth_Kerberos       4.740
Pass_Length         2.722
Updating            1.926
Privilege           30.520
Delegation_Param    2.585
Relationship        4.574
Safety              0.557
dtype: float64
```

Рис. 1. Средние значения для каждого параметра

Стандартизируем и нормализуем данные для удобной работы, здесь понадобятся соответственно StandardScaler и MinMaxScaler [6] (рис. 2, 3).

Представим наши данные в виде графиков (рис. 4).

```
scaler = StandardScaler()
pd.set_option('display.float_format', lambda x: '%0.3f' % x)
std_df = scaler.fit_transform(df)
origin_df = scaler.inverse_transform(std_df)
```

std_df

```
array([[ 2.61347413, -0.98215912,  1.08784471, ...,  1.18223728,
         0.28652464,  0.89181396],
       [ 0.34532561,  1.01816496, -0.91924885, ..., -1.26545896,
         0.28652464, -1.1213101 ],
       [ 2.61347413, -0.98215912, -0.91924885, ...,  1.18223728,
         0.28652464,  0.89181396],
       ...,
       [ 2.61347413, -0.98215912,  1.08784471, ..., -1.26545896,
         0.28652464,  0.89181396],
       [-0.22171152, -0.98215912, -0.91924885, ..., -1.26545896,
         0.28652464, -1.1213101 ],
       [-0.78874865, -0.98215912, -0.91924885, ...,  0.20315879,
         0.28652464, -1.1213101 ]])
```

Рис. 2. Стандартизация данных в датасете

```
mm = MinMaxScaler()
mm_df = mm.fit_transform(df)
origin_df = mm.inverse_transform(mm_df)
```

mm_df

```
array([[1.      , 0.      , 1.      , ..., 1.      , 0.28571429,
        1.      ],
       [0.33333333, 1.      , 0.      , ..., 0.      , 0.28571429,
        0.      ],
       [1.      , 0.      , 0.      , ..., 1.      , 0.28571429,
        1.      ],
       ...,
       [1.      , 0.      , 1.      , ..., 0.      , 0.28571429,
        1.      ],
       [0.16666667, 0.      , 0.      , ..., 0.      , 0.28571429,
        0.      ],
       [0.      , 0.      , 0.      , ..., 0.6     , 0.28571429,
        0.      ]])
```

Рис. 3. Нормализация данных в датасете

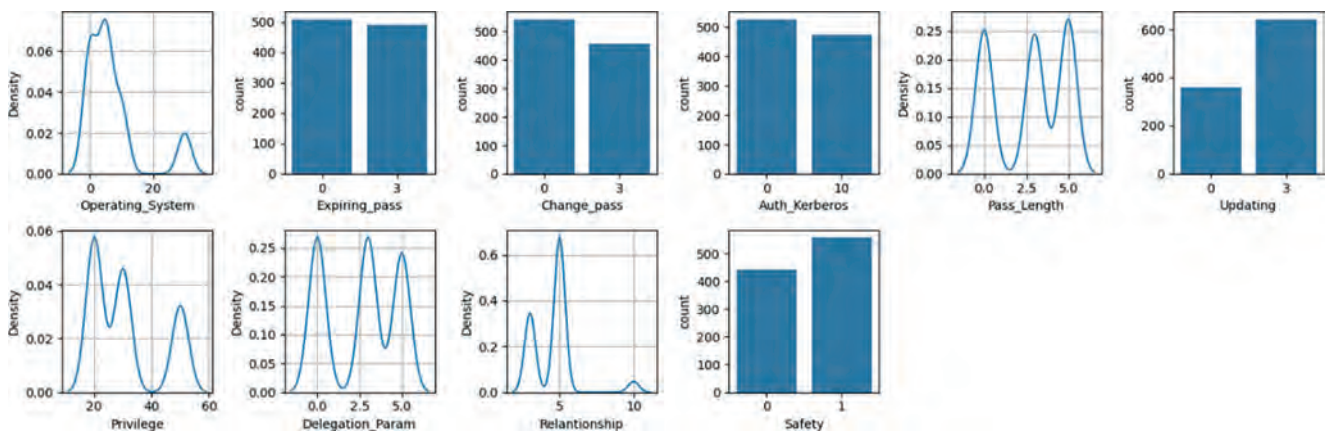


Рис. 4. Графики исходных данных датасета

Реализация метрических алгоритмов

Для реализации метрических алгоритмов нам понадобится библиотека sklearn [7] (рис. 5).

```
lr = LinearRegression() #Линейная регрессия
knnr = KNeighborsRegressor() #Метод ближайших соседей
dt = DecisionTreeRegressor() #Дерево решений
rf = RandomForestRegressor() #Случайный лес
gbr = GradientBoostingRegressor() #Градиентный бустинг
```

Рис. 5. Ввод метрических алгоритмов в систему

Разделим датасет на две отдельных части, в первой будут столбцы, содержащие параметры для оценивания Active Directory, а во второй находятся данные, которые отображают защищенность системы на основе значений первой [8] (рис. 6).

```
y = df['Safety']
x = df.drop('Safety', axis = 1)
```

Рис. 6. Разделение датасета на две части

Для корректной работы алгоритмов машинного обучения разделяем значения на тестовую и обучающую выборки [9] (рис. 7).

```
#Создание обучающей и тестовой выборки
from sklearn.model_selection import train_test_split
xtrain,xtest,ytrain,ytest = train_test_split(x,y,test_size = 0.30, random_state = 42)
```

Рис. 7. Разделение данных на обучающую и тестовую выборки

Выведем графики, обозначим среднюю квадратическую ошибку и эффективность для каждой модели, воспользовавшись библиотеками matplotlib [10] и sklearn.metrics (рис. 8–17). Графики представляют собой соотношение плотности функций от параметра защищенности (табл. 1).

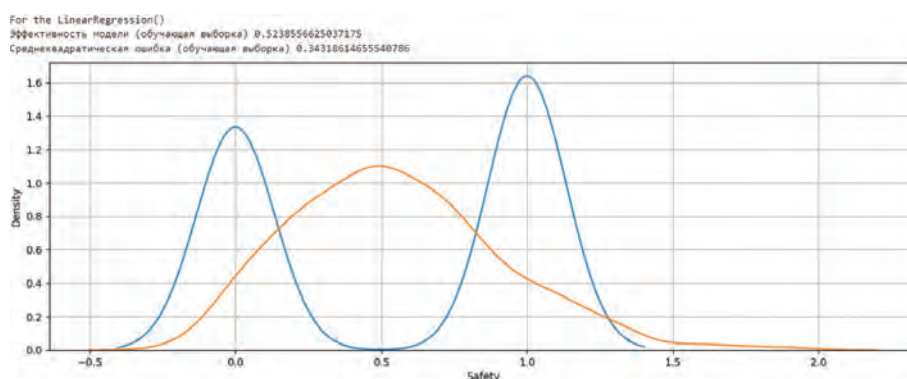


Рис. 8. График линейной регрессии для обучающей выборки

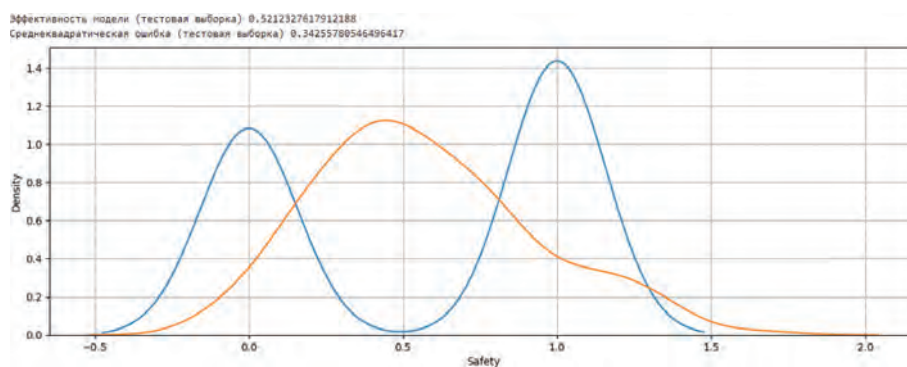


Рис. 9. График линейной регрессии для тестовой выборки

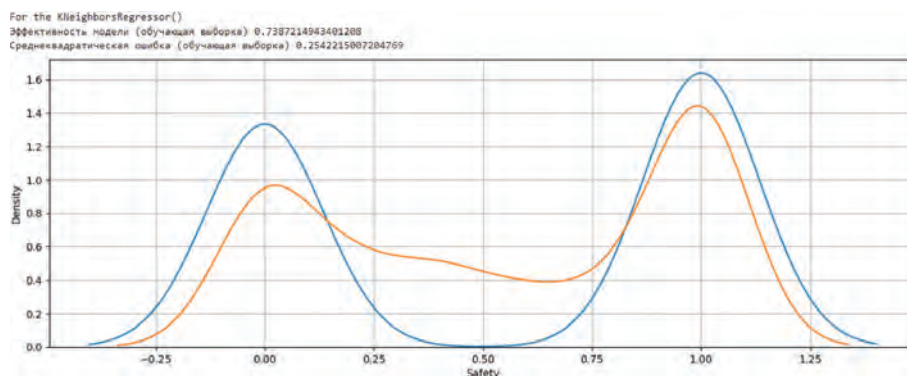


Рис. 10. График метода ближайших соседей для обучающей выборки

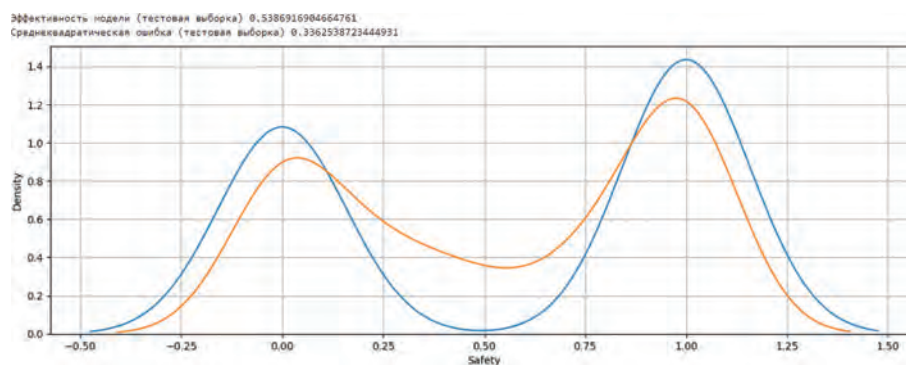


Рис. 11. График метода ближайших соседей для тестовой выборки

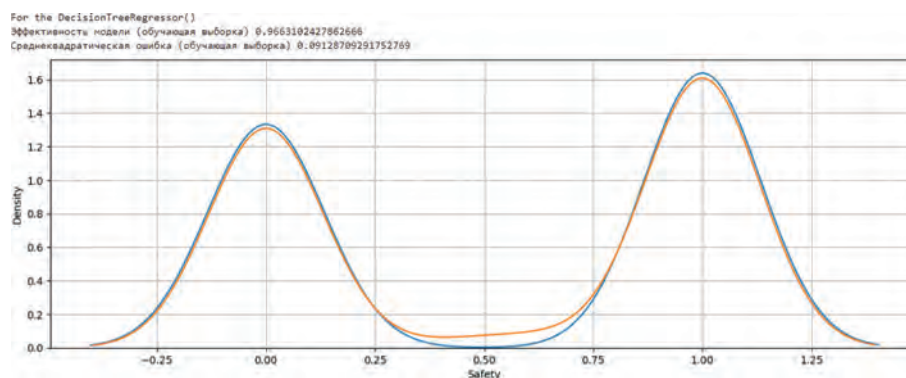


Рис. 12. График дерева решений для обучающей выборки

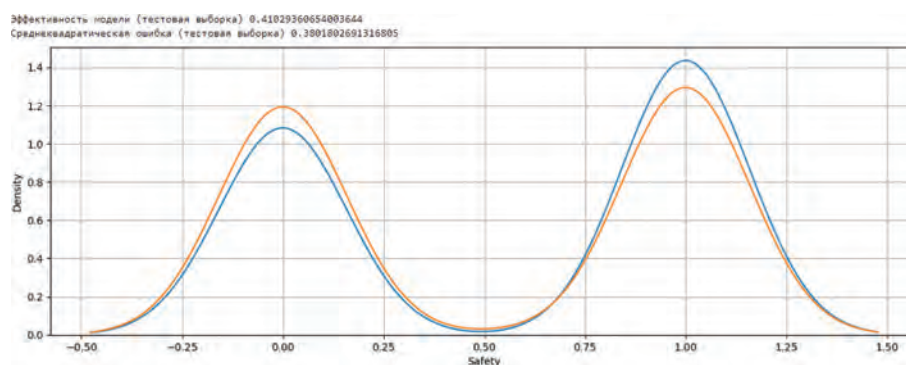


Рис. 13. График дерева решений для тестовой выборки

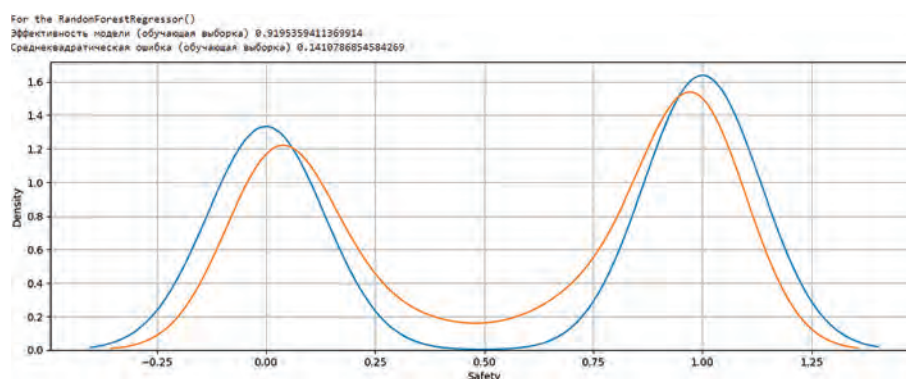


Рис. 14. График случайного леса для обучающей выборки

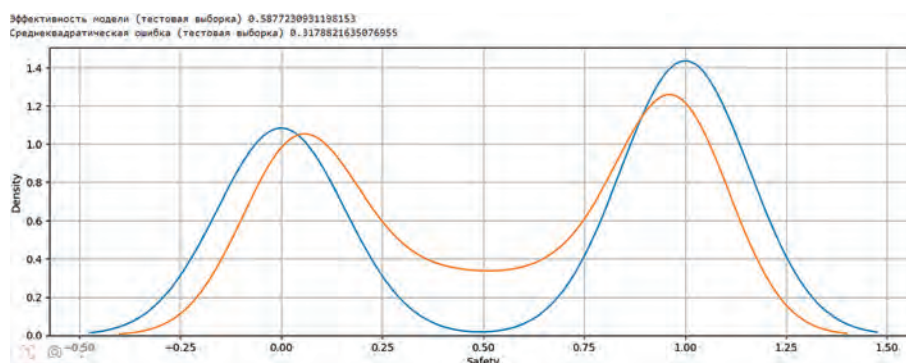


Рис. 15. График случайного леса для тестовой выборки

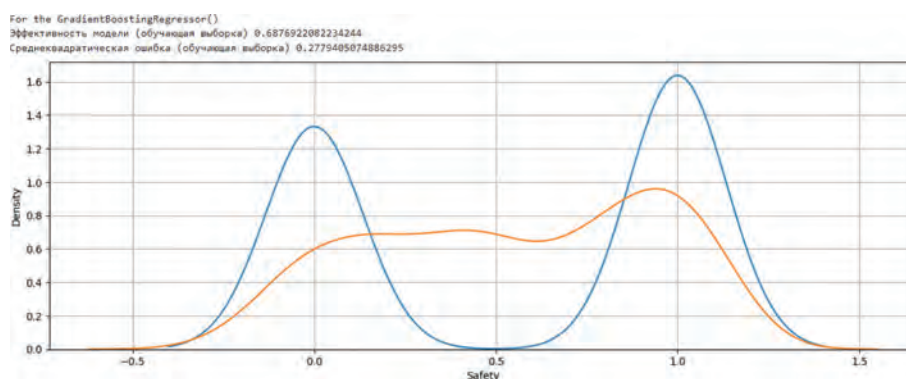


Рис. 16. График градиентного бустинга для обучающей выборки

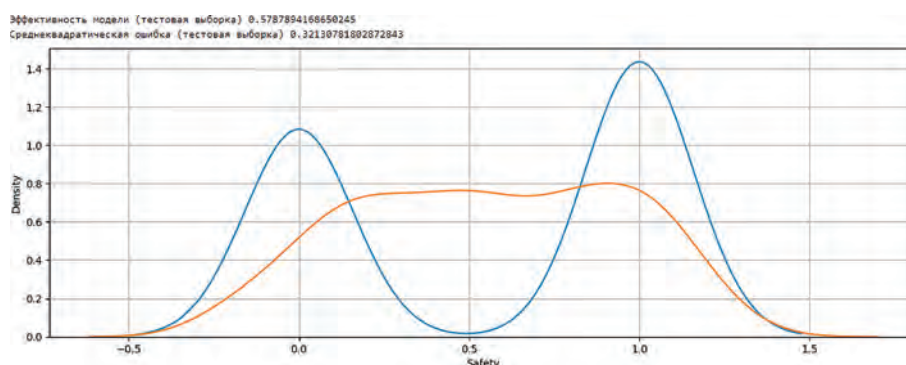


Рис. 17. График градиентного бустинга для тестовой выборки

Как видно из графиков, среди представленных методов МО дерево решений наиболее полно описывает зависимость целевого параметра Safety (защищенность системы AD) от входных параметров. Обратная ситуация возникает с линейной регрессией и градиентным бустингом, которые эту зависимость объясняют плохо.

Реализация работы нейронной сети

Для полноты исследования реализуем модель нейронной сети, которая на основе введенных данных сможет вычислить, защищен ли пользователь от различных атак злоумышленников, соответственно оценивая защищенность Active Directory.

Проведем те же действия над данными, как и с метрическими алгоритмами, разделим их на две части, сначала на входные и выходные параметры, затем на обучающую и тестовую выборки.

Сначала используем Keras (рис. 18):

```
import tensorflow
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense
```

Рис. 18. Импорт Keras

Для добавления слоев в первую очередь инициализируем нейронную сеть в системе. Добавим объект нейронной сети типа sequential (рис. 19).

```
classifier = Sequential()
```

Рис. 19. Инициализация нейронной сети

```
classifier.add(Dense(units = 6, activation = 'relu', input_dim = 9, kernel_initializer = 'uniform'))
```

Рис. 20. Интеграция входного и первого скрытого слоев

```
classifier.add(Dense(units = 6, activation = 'relu', kernel_initializer = 'uniform'))
```

Рис. 21. Интеграция второго скрытого слоя

Добавим входной слой и первый скрытый слой, использующиеся при обработке данных (рис. 20). Слой создается с нелинейной функцией активации «relu», kernel_initializer = 'uniform' является параметром по умолчанию и создает вес слоя [11, 12].

После чего добавляем второй скрытый слой, установив такие же значения, как и при создании первого слоя (рис. 21).

Заключительным слоем будет выходной, в нем будет присутствовать один нейрон [13].

Проводим тренировку нейронной сети, которая будет реализована на основе тренировочной выборки (рис. 22).

```
classifier.fit(X_trainNN, y_trainNN, batch_size = 10, epochs = 100)
```

Epoch	Time	Accuracy	Loss
4/100	0s 3ms/step	0.8230	0.5240
5/100	0s 3ms/step	0.8358	0.5424
6/100	0s 3ms/step	0.8242	0.5025
7/100	0s 3ms/step	0.8757	0.4382
8/100	0s 3ms/step	0.8632	0.4206
9/100	1s 3ms/step	0.8950	0.3717
10/100	0s 3ms/step	0.9087	0.3473
11/100	0s 3ms/step	0.8807	0.3482
12/100	0s 3ms/step	0.9117	0.3108
12/100	0s 3ms/step	0.8981	0.3206

Рис. 22. Тренировка нейронной сети

После тренировки нейронной сети получаем систему, позволяющую определять защищенность того или иного пользователя в корпоративной сети.

Нейронная сеть была сформирована из входного, выходного и двух скрытых слоев в каждом слое было по 6 нейронов, всего использовалось 100 эпох. Тестовая выборка составляла 0.3 от всех исходных данных, что является наиболее оптимальным. При наличии большего количества, это значение будет варьироваться.

Тестирование нейронной сети

После обучения системы выведем прогноз для тестовой выборки, в которой содержатся получившиеся

значения и выводы для каждого набора параметров (рис. 23).

```
y_predNN = classifier.predict(X_testNN)
y_predNN
```

Time
10/10 0s 14ms/step

```
array([[0.97422767],
       [0.63163245],
       [0.80098426],
       [0.03812398],
       [0.13650697],
       [0.6287609 ],
       [0.99998623],
       [0.01737837],
       [0.01010334],
       [0.62079155],
       [0.15893725],
       [0.99894065],
       [0.9999503 ],
       [0.11676516],
       [0.982782 ],
       [0.05276635],
```

Рис. 23. Прогноз для тестовой выборки

Получившиеся значения представим в виде булевых значений (защищенная система или нет). Для разделения используется значение 0,5 и выводы больше него означают, что система не защищена (рис. 24).

Вводим матрицу несоответствий [14, 15]. Для реализации используется метод confusion_matrix, который подсчитывает все данные, а также тестовая выборка (рис. 25).

Получившиеся значения в большинстве отображаются как правильные (ошибок I и II рода относительно немного), это означает, что алгоритм работает корректно.

Проведем анализ для отдельно взятых сгенерированных пользователей для того, чтобы посмотреть правильность работы алгоритма [16] (рис. 26).

```
y_predNN = (y_predNN > 0.5)
y_predNN
```

```
array([[ True],
       [ True],
       [ True],
       [False],
       [False],
       [ True],
       [ True],
       [False],
       [False],
       [ True],
       [False],
       [ True],
       [ True],
       [False],
       [ True],
       [False],
       [False],
       [ True]])
```

Рис. 24. Прогноз в виде булевых значений

```
cm = confusion_matrix(y_testNN, y_predNN)
cm
```

```
array([[107, 23],
       [ 28, 142]], dtype=int64)
```

Рис. 25. Формирование матрицы несоответствий

Первый пользователь работает на ОС Windows 7, у него просроченный короткий пароль, который не изменялся в течении 100 дней, система не обновлялась с первичной установки, ограниченный доступ к функциям администратора, ресурсное делегирование, имеет активную сессию (рис. 27).

```
new_prediction1 = classifier.predict(scaler.transform(np.array([[10,3,0,0,5,3,30,3,3]])))
new_prediction2 = classifier.predict(scaler.transform(np.array([[0,0,0,0,5,3,20,3,3]])))

1/1 ————— 0s 98ms/step
1/1 ————— 0s 70ms/step
```

Рис. 26. Код ввода случайных значений для пользователей

```
new_prediction1
```

```
array([[0.8170041]], dtype=float32)
```

Рис. 27. Вывод решения о первом пользователе

```
new_prediction2
```

```
array([[0.02366159]], dtype=float32)
```

Рис. 28. Вывод решения о втором пользователе

Второй пользователь работает на ОС Windows 11, система не обновлялась, короткий пароль, нет доступа к функциям администратора, ресурсное делегирование, имеет активную сессию (рис. 28).

Возможные параметры, которые могут быть добавлены в систему

Добавление новых параметров является неотъемлемой частью заполнения датасета, так как в статье представлены лишь некоторые параметры, которые имеют наибольший вес во всем объеме настроек. Злоумышленники обладают огромным набором утилит для влияния и проникновения в систему, увеличив количество входных данных появляется возможность минимизировать шанс ошибки.

Большинство параметров имеет важное значение для безопасности пользователя, более детальный анализ показывает, что имеет смысл добавить список новых настроек, которые помогут в решении основной задачи. Все это может влиять как на пользователя, так и на сервер, к которому он относится. Разберем подробнее отдельно возможные варианты параметров.

Accountdisable – определяет активность учетной записи, является ли данный пользователь отключенным или нет, есть ли у него возможность управлять аккаунтом. В компании может содержаться огромное количество УЗ без работающего владельца. Такие записи в отдельных предприятиях не проверяются и остаются на многие годы без проверок.

Trusted_To_Auth_For_Delegation – параметр освещающий о том, что учетная запись делегируется, в данном случае служба использует права пользователя и выполняет проверку от его лица в других удаленных местах сети (рис. 29).

Также для администраторов следует ввести дополнительный параметр, который будет отвечать за доступность участников к сети, подключаться и видеть контроллер домена. В папке C:\Windows\NTDS\ находится файл ntds.dit, который хранит информацию

Delegation is a security-sensitive operation, which allows services to act on behalf of another user.

- ☐ Do not trust this computer for delegation
- ☐ Trust this computer for delegation to any service (Kerberos only)
- ☒ Trust this computer for delegation to specified services only
- ☐ Use Kerberos only
- ☒ Use any authentication protocol

Рис. 29. Включенный параметр
Trusted_To_Auth_For_Delegation

о всей локальной сети Active Directory. Чем больше пользователей подключены к контроллеру, тем больше шанс пропустить злоумышленника, который скомпрометирует данные.

Заключение

В условиях роста числа удалённых пользователей традиционные методы администрирования Active Directory, уязвимости, связанные с использованием устаревших операционных систем, слабых паролей, создают серьёзные риски компрометации и требуют применения более интеллектуальных методов анализа.

В результате исследования был сформирован датасет, включающий параметры, влияющие на защищённость пользователей Active Directory, и реализованы

различные методы машинного обучения. Сравнительный анализ показал, что с задачей оценки защищённости лучше всего справился алгоритм дерева решений, обеспечив высокую точность прогнозов и низкий уровень ошибок, при этом хуже всего справились алгоритмы линейной регрессии и градиентного бустинга.

Также внимание уделено разработке и обучению нейронной сети. Было показано, как современные инструменты машинного обучения и анализа данных позволяют относительно просто проектировать модели (в частности, нейронных сетей), которые способны обрабатывать большие объёмы данных и выполнять предсказания с высокой точностью.

Результаты работы показывают перспективность применения технологий ИИ для анализа информационных систем и прогнозирования рисков информационной безопасности в режиме реального времени. Так, внедрение методов машинного обучения в процесс оценки защищённости Active Directory позволяет существенно повысить точность и скорость анализа, минимизировать влияние человеческого фактора и обеспечить более высокий уровень устойчивости корпоративной инфраструктуры к современным киберугрозам.

Литература

1. Review of prevention schemes for man-in-the-middle (MITM) attack in vehicular ad hoc networks. Al-Shareeda M. A. International Journal of Engineering and Management Research. 2020. – 6 p. – DOI 10.31033/ijemr.10.3.23.
2. Правоторова А. Ю. Оценка стойкости паролей: сравнительное исследование эффективности длины и сложности / А. Ю. Правоторова // Наука. Инновации. будущее – 2025 : Сборник статей II Международной научно-практической конференции, Петрозаводск, 15 мая 2025 года. – Петрозаводск: Международный центр научного партнерства «Новая Наука» (ИП Ивановская И.И.), 2025. – С. 125–132. – EDN MEPRXB.
3. Khalil Nabab Pinjari. LSTM-Enabled Big Data Security Framework Integrating Kerberos Authentication on AWS for Robust Cloud Protection / Khalil Nabab Pinjari, Abu Zar Muhammad, Yogesh Kumar Sharma // Nanotechnology Perceptions. – 2024. – Vol. 20, No. 7. – DOI 10.62441/nano-ntp.v20i7.4414. – EDN ZHQLIB.
4. Абрамова Е. В. Возможности Google Colab и Jupyter Notebook для решения задач искусственного интеллекта / Е. В. Абрамова, Л. А. Максименко // Регулирование земельно-имущественных отношений в России: правовое и геопространственное обеспечение, оценка недвижимости, экология, технологические решения. – 2023. – № 1. – С. 23–29. – DOI 10.33764/2687-041X-2023-1-23-29. – EDN CYLSKT.
5. Ильичев В. Ю. Анализ массивов данных с использованием библиотеки Pandas для Python / В. Ю. Ильичев, Е. А. Юрик // Научное обозрение. Технические науки. – 2020. – № 4. – С. 41–45. – EDN BKGJHM.
6. Сударинов Г. В. Использование библиотеки Pandas для анализа данных / Г. В. Сударинов, И. А. Ашмаров // Мир образования - образование в мире. – 2023. – № 1(89). – С. 184–188. – DOI 10.51944/20738536_2023_1_184. – EDN NWHNRO.
7. Florescu D. A Machine Learning Based Software Pipeline to Pick the Variable Ordering for Algorithms with Polynomial Inputs / D. Florescu, M. England // Lecture Notes in Computer Science. – 2020. – Vol. 12097 LNCS. – P. 302–311. – DOI 10.1007/978-3-030-52200-1_30. – EDN JDDNYT.
8. Retnoningsih, E. Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python / E. Retnoningsih, R. Pramudita // Bina Insani ICT Journal. – 2020. – Vol. 7, No. 2. – P. 156. – DOI 10.51211/biict.v7i2.1422. – EDN TNZXLG.
9. Булгакова Е. В. Проблема точности и объяснимости при внедрении искусственного интеллекта в системы управления информацией и событиями безопасности / Е. В. Булгакова, Д. С. Дойников, А. Н. Кубанков // Наукоемкие технологии в космических исследованиях Земли. – 2025. – Т. 17, № 3. – С. 35–41. – DOI 10.36724/2409-5419-2025-17-3-35-41. – EDN QBTNNT.
10. Магеррамов И. М. Нечеткие подходы к решению задач классификации / И. М. Магеррамов, Т. С. Александрова // Интеллектуальные ресурсы - региональному развитию. – 2021. – № 2. – С. 76–82. – EDN XRMACX.
11. Feng, G. Common Python Data Analysis Method Based on Deep Learning / G. Feng // Journal of Physics: Conference Series. – 2021. – Vol. 2037, No. 1. – P. 012132. – DOI 10.1088/1742-6596/2037/1/012132. – EDN YQAYRU.
12. Abdulkadirov, R. Survey of Optimization Algorithms in Modern Neural Networks / R. Abdulkadirov, P. Lyakhov, N. Nagornov // Mathematics. – 2023. – Vol. 11, No. 11. – P. 2466. – DOI 10.3390/math11112466. – EDN BAUENY.
13. Tao, Ch. Applications of Bayesian Neural Networks in Outlier Detection / Ch. Tao // Big Data. – 2023. – Vol. 11, No. 5. – P. 369–386. – DOI 10.1089/big.2021.0343. – EDN DXZRFS.

14. Iiduka, H. Training deep neural networks using conjugate gradient-like methods / H. Iiduka, Y. Kobayashi // Electronics. – 2020. – Vol. 9, No. 11. – P. 1–25. – DOI 10.3390/electronics9111809. – EDN TMEEXV.
15. Трунов Е. Е. Обнаружение угроз безопасности информации с использованием глубоких нейронных сетей в компьютерных сетях в режиме реального времени / Е. Е. Трунов, С. Г. Ключев // Моделирование, оптимизация и информационные технологии. – 2022. – Т. 10, № 3(38). – С. 12–13. – DOI 10.26102/2310-6018/2022.38.3.011. – EDN MNLGVN.
16. Nakhushiev, R. S. Application of the neural networks for cryptographic information security / R. S. Nakhushiev, N. V. Sukhanova // Proceedings of the 2020 IEEE International Conference «Quality Management, Transport and Information Security, Information Technologies», IT and QM and IS 2020, Yaroslavl, 07–11 сентября 2020 года. – Yaroslavl, 2020. – P. 421–423. – DOI 10.1109/ITQMIS51053.2020.9322981. – EDN FALTNR.

ASSESSMENT OF ACTIVE DIRECTORY SECURITY USING ARTIFICIAL INTELLIGENCE

Bulgakova E. V.⁴, Bogdanov E. A.⁵, Kubankov A. N.⁶

Keywords: Active Directory, information security, artificial intelligence, machine learning, neural networks, decision tree, security assessment, vulnerabilities, corporate networks.

Purpose of the study: the study aims to apply artificial intelligence methods, including machine learning and neural networks, to assess the security of Microsoft Active Directory and identify factors affecting the security level of corporate networks. The work focuses on developing and comparing algorithms capable of predicting the vulnerability levels of Active Directory users and subsystems.

Methods of research: the study is based on machine learning methods, including metric algorithms (linear regression, nearest neighbour method, decision tree, random forest, gradient boosting) and neural networks, implemented in the Jupyter Notebook environment using the pandas, sklearn and keras libraries. Based on the prepared dataset, the parameters reflecting the configuration of Active Directory users were standardised and normalised. To verify the effectiveness of the algorithms, a comparison was made based on the criteria of prediction accuracy and root mean square error.

Results: an analysis was conducted of factors affecting the security of the Active Directory corporate infrastructure, including operating system type, password length and validity period, privilege level, delegation settings, and the presence of Kerberos pre-authentication. Various machine learning algorithms were implemented and tested based on the prepared dataset. The results showed that the decision tree model demonstrated the best performance, with a prediction accuracy of 0.96 and a root mean square error of 0.091, indicating its high effectiveness in assessing Active Directory security. In addition, a neural network model was developed and its ability to correctly process Active Directory parameters and determine the security level of users of this system was confirmed. The results obtained indicate the promise of using artificial intelligence technologies to automate vulnerability analysis and predict information security risks in corporate networks.

Scientific novelty: the scientific novelty of the research lies in the development and testing of an integrated approach to assessing Active Directory security based on machine learning and neural networks. The use of intelligent models is proposed to predict the level of user security, taking into account the complex parameters of the Active Directory infrastructure, which allows the formation of an automated early warning system for corporate network vulnerabilities.

References

1. Review of prevention schemes for man-in-the-middle (MITM) attack in vehicular ad hoc networks. Al-Shareeda M. A. International Journal of Engineering and Management Research. 2020. – 6 p. – DOI 10.31033/ijemr.10.3.23.
2. Pravotorova A. Ju. Ocenka stojkosti parolej: sravnitel'noe issledovanie jeffektivnosti dliny i slozhnosti / A. Ju. Pravotorova // Nauka. Innovacii. budushhee – 2025 : Sbornik statej II Mezhdunarodnoj nauchno-prakticheskoy konferencii, Petrozavodsk, 15 maja 2025 goda. – Petrozavodsk: Mezhdunarodnyj centr nauchnogo partnerstva «Novaja Nauka» (IP Ivanovskaja I.I.), 2025. – S. 125–132. – EDN MEPXNB.
3. Khalil Nabab Pinjari. LSTM-Enabled Big Data Security Framework Integrating Kerberos Authentication on AWS for Robust Cloud Protection / Khalil Nabab Pinjari, Abu Zar Muhammad, Yogesh Kumar Sharma // Nanotechnology Perceptions. – 2024. – Vol. 20, No. 7. – DOI 10.62441/nano-ntp.v20i7.4414. – EDN ZHQLIB.
4. Abramova E. V. Vozmozhnosti Google Colab i Jupyter Notebook dlja reshenija zadach iskusstvennogo intellekta / E. V. Abramova, L. A. Maksimenko // Regulirovanie zemel'no-imushhestvennyh otnoshenij v Rossii: pravovoe i geoprostranstvennoe obespechenie, ocenka nedvizhimosti, jekologija, tehnologicheskie reshenija. – 2023. – № 1. – S. 23–29. – DOI 10.33764/2687-041X-2023-1-23-29. – EDN CYLSKT.
5. Il'ichev V. Ju. Analiz massivov dannyh s ispol'zovaniem biblioteki Pandas dlja Python / V. Ju. Il'ichev, E. A. Jurik // Nauchnoe obozrenie. Tehnicheskie nauki. – 2020. – № 4. – S. 41–45. – EDN BKGJHM.
- 4 Elena V. Bulgakova, Ph.D. in Law, associate professor, deputy chair of the department of information security for scientific research, Financial University under the Government of the Russian Federation, Moscow, Russia. E-mail: koordinatork-proekta@mail.ru
- 5 Evgeny A. Bogdanov Ph.D., chair of the department of information security, Financial University under the Government of the Russian Federation, Moscow, Russia. E-mail: Eabogdanov@fa.ru
- 6 Alexander N. Kubankov, Dr.Sc. in Military Sciences, professor, professor of the department of information security, Financial University under the Government of the Russian Federation, Moscow, Russia. E-mail: kan9991@gmail.com

6. Sudarikov G. V. Ispol'zovanie biblioteki Pandas dlja analiza dannyh / G. V. Sudarikov, I. A. Ashmarov // Mir obrazovaniya - obrazovanie v mire. – 2023. – № 1(89). – S. 184–188. – DOI 10.51944/20738536_2023_1_184. – EDN NWHNRO.
7. Florescu D. A Machine Learning Based Software Pipeline to Pick the Variable Ordering for Algorithms with Polynomial Inputs / D. Florescu, M. England // Lecture Notes in Computer Science. – 2020. – Vol. 12097 LNCS. – P. 302–311. – DOI 10.1007/978-3-030-52200-1_30. – EDN JDDNYT.
8. Retnoningsih, E. Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python / E. Retnoningsih, R. Pramudita // Bina Insani ICT Journal. – 2020. – Vol. 7, No. 2. – P. 156. – DOI 10.51211/biict.v7i2.1422. – EDN TNZXLG.
9. Bulgakova E. V. Problema tochnosti i ob#jasnimosti pri vnedrenii iskusstvennogo intellekta v sistemy upravleniya informaciej i sobytijami bezopasnosti / E. V. Bulgakova, D. S. Dojnikov, A. N. Kubankov // Naukoemkie tehnologii v kosmicheskikh issledovaniyah Zemli. – 2025. – T. 17, № 3. – S. 35–41. – DOI 10.36724/2409-5419-2025-17-3-35-41. – EDN QBTNNT.
10. Magerramov I. M. Nechetkie podhody k resheniju zadach klassifikacii / I. M. Magerramov, T. S. Aleksandrova // Intellekturnye resursy – regional'nomu razvitiyu. – 2021. – № 2. – S. 76–82. – EDN XRMACX.
11. Feng, G. Common Python Data Analysis Method Based on Deep Learning / G. Feng // Journal of Physics: Conference Series. – 2021. – Vol. 2037, No. 1. – P. 012132. – DOI 10.1088/1742-6596/2037/1/012132. – EDN YQAYRU.
12. Abdulkadirov, R. Survey of Optimization Algorithms in Modern Neural Networks / R. Abdulkadirov, P. Lyakhov, N. Nagornov // Mathematics. – 2023. – Vol. 11, No. 11. – P. 2466. – DOI 10.3390/math11112466. – EDN BAUENY.
13. Tao, Ch. Applications of Bayesian Neural Networks in Outlier Detection / Ch. Tao // Big Data. – 2023. – Vol. 11, No. 5. – P. 369–386. – DOI 10.1089/big.2021.0343. – EDN DXZRFS.
14. Iiduka, H. Training deep neural networks using conjugate gradient-like methods / H. Iiduka, Y. Kobayashi // Electronics. – 2020. – Vol. 9, No. 11. – P. 1–25. – DOI 10.3390/electronics9111809. – EDN TMEEXV.
15. Trunov E. E. Obnaruzhenie ugroz bezopasnosti informacii s ispol'zovaniem glubokih neyronnyh setej v komp'yuternyh setjah v rezhime real'nogo vremeni / E. E. Trunov, S. G. Kljuev // Modelirovanie, optimizacija i informacionnye tehnologii. – 2022. – T. 10, № 3(38). – S. 12–13. – DOI 10.26102/2310-6018/2022.38.3.011. – EDN MNLGVN.
16. Nakhushhev, R. S. Application of the neural networks for cryptographic information security / R. S. Nakhushhev, N. V. Sukhanova // Proceedings of the 2020 IEEE International Conference «Quality Management, Transport and Information Security, Information Technologies», IT and QM and IS 2020, Yaroslavl, 07–11 sentjabrja 2020 goda. – Yaroslavl, 2020. – P. 421–423. – DOI 10.1109/ITQMIS51053.2020.9322981. – EDN FALTNR.



ПРОБЛЕМНО-ОРИЕНТИРОВАННАЯ СИСТЕМА МОНИТОРИНГА И РЕАГИРОВАНИЯ НА МНОГОВЕКТОРНЫЕ АТАКИ В ДЕЦЕНТРАЛИЗОВАННОЙ СРЕДЕ ИНТЕРНЕТА ВЕЩЕЙ

Тебуева Ф. Б.¹, Петренко В. И.², Сатыбалдина Д. Ж.³, Огур М. Г.⁴, Гусева Т. М.⁵

DOI: 10.21681/2311-3456-2025-6-69-80

Цель исследования: повышение эффективности мониторинга и реагирования на многовекторные атаки в децентрализованной среде Интернета вещей за счёт интеграции федеративного обучения, глубоких автоэнкодеров и распределённого реестра IOTA. Приоритеты включают точное обнаружение атак, минимизацию ложных срабатываний, снижение времени реакции и сохранение конфиденциальности данных.

Метод исследования: разработана проблемно-ориентированная система, объединяющая локальный мониторинг на IoT-узлах с автоэнкодерами для выявления аномалий, федеративное обучение с алгоритмом FedAvg для коллективного обновления моделей, а также децентрализованное распространение оповещений через распределённый реестр IOTA. Система реализует защищённый обмен модельными параметрами, цифровую подпись сообщений и асинхронное реагирование через сеть публикации/подписки.

Результат исследования: экспериментальные исследования на реальных данных N-Balot с имитацией многовекторных атак показали высокую точность обнаружения (около 95%), достижение F1-меры свыше 94%, при уровне ложных срабатываний около 4%. Время реакции системы не превышало 5 секунд, что существенно улучшает оперативность противодействия атакам. Федеративное обучение обеспечило устойчивое повышение качества модели с учётом распределённости и гетерогенности данных. Архитектура доказала масштабируемость, отказоустойчивость и способность эффективно выявлять комплексные угрозы на разных уровнях системы.

Практическая ценность решения заключается в возможности его внедрения в промышленных IoT, умных городах и медицинских сетях для повышения кибербезопасности с сохранением приватности и снижением нагрузок на сеть.

Научная новизна состоит в комплексном синтезе федеративного обучения, глубоких автоэнкодеров и технологии распределённых реестров для эффективного мониторинга многовекторных атак в децентрализованных IoT-средах. Предложенный подход сочетает преимущества распределённого обучения и блокчейн-механизмов для достижения высокой адаптивности, точности и безопасности в условиях быстрорастущих и разнообразных IoT-инфраструктур.

Вклад авторов: Тебуева Ф. Б. предложила концепцию и общую структуру исследования, сформулировала ключевые гипотезы и руководила проведением экспериментов; Петренко В. И. разработал математическую модель системы, предложил алгоритмы федеративного обучения и участвовал в создании архитектуры комплексной системы мониторинга и реагирования; Сатыбалдина Д. Ж. отвечала за методологию глубокого обучения, в частности, разработку и оптимизацию моделей автоэнкодеров для обнаружения аномалий; Огур М. Г. провёл экспериментальное моделирование, подготовил набор данных и осуществлял сбор и анализ экспериментальных результатов; Гусева Т. М. занималась реализацией механизмов взаимодействия компонентов системы через распределённый реестр IOTA, а также подготовкой текстовой части исследования и оформлением публикации.

Ключевые слова: Интернет вещей; многовекторные атаки; обнаружение вторжений; федеративное обучение; автоэнкодер; IOTA; блокчейн; аномалия; мониторинг безопасности.

Введение

Интернет вещей (Internet of Things, IoT) с его интеллектуальными приложениями и услугами, в настоящее время охватил ключевые области в нашей повседневной жизни, включая промышленность, медицину, сельское хозяйство / сельское хозяйство, умные города, умные дома, интеллектуальный

транспорт, интеллектуальные автономные транспортные средства, роботов и модульных роботов (например, беспилотные летательные аппараты, беспилотные наземные транспортные средства и беспилотные подводные аппараты) [1]. Ограниченность вычислительных ресурсов устройств и их

- 1 Тебуева Фарица Биляловна, доктор физико-математических наук, доцент, профессор кафедры вычислительной математики и кибернетики, ФГАОУ ВО «Северо-Кавказский федеральный университет». г. Ставрополь, Россия. ORCID: <https://orcid.org/0000-0002-7373-4692>. E-mail: fariza.teb@gmail.com
- 2 Петренко Вячеслав Иванович, кандидат технических наук, доцент, заведующий кафедрой организации и технологии защиты информации, ФГАОУ ВО «Северо-Кавказский федеральный университет». г. Ставрополь, Россия. <https://orcid.org/0000-0003-4293-7013>. E-mail: vipetrenko@ncfu.ru
- 3 Сатыбалдина Дина Жагыпаровна, кандидат физико-математических наук, директор НИИ Информационной безопасности и криптологии, НАО «Евразийский национальный университет имени Л. Н. Гумилева». г. Астана, Республика Казахстан. <https://orcid.org/0000-0003-0291-4685>. E-mail: satybaldina_dzh@enu.kz
- 4 Огур Максим Геннадьевич, старший преподаватель кафедры вычислительной математики и кибернетики, ФГАОУ ВО «Северо-Кавказский федеральный университет». г. Ставрополь, Россия. <https://orcid.org/0000-0002-2387-0901>. E-mail: ogur26@gmail.com
- 5 Гусева Татьяна Михайловна, ассистент кафедры организации и технологии защиты информации, ФГАОУ ВО «Северо-Кавказский федеральный университет». г. Ставрополь, Россия. E-mail: tatyana.petrova.96@bk.ru

распределённый характер формируют новые уязвимости, угрожающие конфиденциальности, целостности и доступности данных [2]. Традиционные централизованные методы кибербезопасности часто оказываются неэффективными в условиях распределённой архитектуры IoT и жёстких ограничений по энергопотреблению и вычислительной мощности [3]. Как следствие, фиксируется увеличение числа успешных атак на IoT-инфраструктуры [4]. Особую опасность представляют многовекторные атаки, в которых злоумышленники комбинируют различные методы воздействия (например, одновременное проведение DDoS-атаки и внедрение вредоносного кода) [5]. Подобные скоординированные воздействия сложно обнаружить и нейтрализовать, поскольку они затрагивают различные уровни системы (сетевой, прикладной, данных) и могут маскировать друг друга [6]. Это обуславливает необходимость разработки комплексных систем мониторинга безопасности, способных в реальном времени выявлять аномалии разнородной природы.

Ключевым инструментом защиты IoT-сетей являются системы обнаружения вторжений (IDS). Наиболее перспективными представляются IDS, идентифицирующие отклонения от нормального профиля работы устройств и сетевого трафика [7]. С ростом объёмов данных и усложнением угроз для анализа всё активнее применяются методы машинного обучения, в частности глубокого обучения [8]. Глубокие нейронные сети и автоэнкодеры демонстрируют высокую точность выявления ранее неизвестных атак за счёт обнаружения скрытых паттернов аномального поведения [9]. Например, в работе [10] с помощью глубокого автоэнкодера достигнуто высокоточное обнаружение ботнет-атак на IoT-устройства. Однако большинство современных IDS для IoT основаны на централизованном сборе и анализе данных: информация с устройств передаётся на сервер, где обучается общая модель классификации трафика. Такой подход имеет два существенных недостатка: (1) угрозу конфиденциальности и утечек данных из-за централизованного хранения и передачи чувствительной информации; (2) высокую нагрузку на сеть и задержки, что критично для ресурсограниченных распределённых сред IoT [11].

Федеративное обучение (Federated Learning, FL) позволяет решить эти проблемы путём переноса процесса обучения моделей на конечные устройства с последующей агрегацией локальных обновлений [12]. При использовании FL исходные данные не покидают устройство: каждый узел обучает локальную модель на своих данных, а на сервер передаются только обновления параметров (градиенты или веса модели). Сервер-агрегатор объединяет

их (например, с помощью алгоритма FedAvg) для обновления глобальной модели, которая затем рассылается участникам [13]. Этот подход обеспечивает сохранение конфиденциальности данных и снижение сетевой нагрузки, позволяя использовать знания множества распределённых источников. FL уже применяется для IDS в IoT. В частности, в [14] разработана федеративная самообучающаяся IDS, адаптирующаяся к типу устройства, которая показала точность обнаружения атак ботнета Mirai на уровне 95,6 % с задержкой ~257 мс и минимальным количеством ложных срабатываний. Другие исследования также подтверждают, что федеративные IDS по точности сопоставимы с централизованными, обеспечивая при этом приватность данных. Так, в [15] показано, что совместное обучение простой нейронной сети на устройствах с использованием FedAvg даёт метрики (точность, полнота, F1-score), comparable с централизованной моделью при обнаружении атак, но без передачи исходных данных.

Параллельно растёт интерес к использованию технологий блокчейн и распределённых реестров для повышения надёжности систем безопасности IoT [16]. Традиционный блокчейн обеспечивает неизменяемость и отслеживаемость записей о событиях безопасности, позволяя узлам доверять зафиксированным в реестре предупреждениям об атаках [17]. Однако классические блокчейн-платформы (например, Ethereum) страдают от ограниченной масштабируемости, низкой скорости транзакций и высокой энергоёмкости алгоритмов консенсуса, что затрудняет их применение в IoT-сетях [18]. В связи с этим в последние годы появились альтернативные распределённые реестры, ориентированные на IoT. Один из наиболее перспективных вариантов – платформа IOTA, использующая направленный ациклический граф (Tangle) вместо традиционной цепочки блоков [19]. IOTA сохраняет ключевые свойства блокчейна (неизменяемость, прозрачность транзакций), но устраняет комиссионные сборы и обеспечивает высокую пропускную способность и быстроедействие за счёт параллельной обработки операций [20]. Исследования показывают, что IOTA существенно превосходит традиционный блокчейн по масштабируемости и энергоэффективности, что делает её одним из немногих практических решений для ресурсограниченных IoT-устройств [21]. Авторы [22] в своём обзоре отмечают эволюцию IoT-экосистем от блокчейна к IOTA, подчёркивая её преимущества в контексте IoT-приложений.

Учитывая указанные тенденции, актуальной научной задачей является синтез федеративного обучения, интеллектуального аномального мониторинга и распределённого реестра в единую систему защиты

IoT. Отдельные компоненты уже продемонстрировали свою эффективность: автоэнкодеры обнаруживают сложные и малозаметные атаки [23], федеративное обучение обеспечивает адаптивность и приватность [24], а реестры на основе блокчейн-технологий повышают доверие между узлами и устойчивость к компрометации журналов событий [17]. Однако интеграция этих компонентов в комплексную систему мониторинга многовекторных атак исследована недостаточно. Существующие решения, как правило, фокусируются на одной из составляющих (например, только на распределённом обнаружении или только на регистрации событий в блокчейне) и не учитывают специфику многовекторных сценариев атак.

В последние годы активно внедряются новые методы обеспечения безопасности, направленные на противодействие современным киберугрозам в условиях быстрорастущей экосистемы IoT. К ним относятся принципы Zero Trust, интеллектуальные системы обнаружения и реагирования (EDR/XDR), а также применение искусственного интеллекта и машинного обучения для анализа и предотвращения атак [4]. Например, в [24] предложено объединить блокчейн с федеративным обучением для IDS в медицинских IoT-сетях, достигнув точности ~97–98 % на реальных наборах данных. Al Sadi и др. разработали систему P-IOTA, в которой сетевые контроллеры SDN выявляют DDoS-атаки и отправляют оповещения в реестр IOTA для глобального оповещения. Тем не менее, остаётся открытым вопрос проектирования единой проблемно-ориентированной системы, способной в реальном времени обнаруживать и нейтрализовать многовекторные атаки в IoT-среде за счёт совместного использования FL, методов глубокого обучения и распределённого реестра.

Целью настоящего исследования является повышение эффективности системы мониторинга и реагирования на многовекторные атаки в децентрализованной IoT-среде за счёт минимизации совокупной ошибки обнаружения атак и времени реакции на них.

1. Постановка задачи мониторинга и реагирования на многовекторные атаки в децентрализованной среде Интернета вещей

Особенностью предлагаемого подхода является использование федеративного обучения (FL), которое позволяет обучать модели обнаружения атак непосредственно на устройствах без передачи исходных данных, обеспечивая таким образом приватность информации и устойчивость системы к новым, неизвестным видам атак. В статье [25] предложена модель угроз информационной безопасности агентов в децентрализованной среде Интернета вещей, формализующая сценарии атак на информационную безопасность доверенного взаимодействия.

Введем обозначения переменных и параметров:

N – общее число IoT-узлов (участников FL);

$D_i = \{x_{i,1}, x_{i,2}, \dots, x_{i,n_i}\}$ – набор локальных данных (трафик, события) на узле i ;

w_i^t – локальные параметры модели (веса автоэнкодера) на узле i на t -м раунде FL;

w_{glob}^t – глобальные параметры модели после централизованного или агрегированного обновления на t -м раунде;

$f(\cdot; w)$ – функция автоэнкодера (модель обнаружения аномалий с параметрами w);

$L(x; w)$ – функция потерь (ошибка реконструкции) автоэнкодера;

α – доля (или вес) вклада каждого узла при глобальном обновлении (FedAvg);

θ_{alert} – порог аномальности, при превышении которого срабатывает тревога;

$y_{i,j} \in \{0,1\}$ – метка события в D_i : 1 – атака, 0 – норма;

τ – максимальное допустимое время для срабатывания и распространения предупреждения (констрейнт на быстроедействие);

$S = \{s_1, s_2, \dots, s_M\}$ – множество вариантов многовекторных атак, где каждая атака s_m представляет собой комбинацию нескольких векторов (например, DoS + спуфинг + перебор паролей и др.);

$D_i^{s_m}$ – локальный набор данных на узле i , относящийся к атаке s_m , либо нормальным состояниям;

$L^{s_m}(x; w)$ – функция потерь (ошибка реконструкции) автоэнкодера для данных, поражённых атакой s_m ;

β_m – вес, отражающий критичность (приоритетность) обнаружения многовекторной атаки s_m ;

$T_{detect}^{s_m}, T_{alert}^{s_m}$ – время обнаружения и оповещения для атаки s_m .

Целевая функция задачи имеет вид:

$$\min_{w_{glob}} \sum_{i=1}^N \frac{n_i}{\sum_k n_k} \left(\sum_{m=1}^M \beta_m \cdot L_i^{s_m}(w_{glob}) \right) + \lambda \cdot \sum_{m=1}^M \gamma_m \cdot (T_{detect}^{s_m} + T_{alert}^{s_m}), \quad (1)$$

где $L_i^{s_m}(w) = \frac{1}{|D_i^{s_m}|} \sum_{x \in D_i^{s_m}} L^{s_m}(x; w)$ – усреднённая функция потерь по данным под многовекторной атакой s_m ; $\beta_m \geq 0$ – приоритет обнаружения отдельных многовекторных атак, чтобы уделить особое внимание наиболее критичным; $\gamma_m \geq 0$ – вес времени реакции для каждой атаки; $\lambda \geq 0$ – параметр балансировки между точностью обнаружения и скоростью реакции.

Задача имеет ограничения:

1. Ограничения приватности (Privacy-preserving):

$\forall i, j: D_i^{s_m} \cap D_j^{s_m} = \emptyset$, и $D_i^{s_m}$ не передаются вне узла, (2)

то есть исходные данные остаются локальными, передача происходит только агрегированных параметров моделей:

$$w_{glob}^{t+1} = \sum_{i=1}^N \alpha_i w_i^t, \sum_i \alpha_i = 1, \alpha_i \geq 0. \quad (3)$$

2. Ограничения вычислительных ресурсов узлов.

Параметры модели $f(\cdot; w_i^t)$ должны удовлетворять локальным ресурсным ограничениям:

$$C_{time}(f_i) \leq C_{time}^{max}, C_{memory}(f_i) \leq C_{memory}^{max} \quad (4)$$

где $C_{time}(f_i)$ – время вычисления модели на устройстве i ; $C_{memory}(f_i)$ – объём памяти, необходимый под модель; C_{time}^{max} , C_{memory}^{max} – максимальные допустимые значения, заданные техническими характеристиками устройств.

3. Ограничения на скорость обнаружения и реагирования.

Для каждого типа атаки s_m суммарное время обнаружения и оповещения должно удовлетворять:

$$T_{detect}^{s_m} + T_{alert}^{s_m} \leq T_{max} \quad (5)$$

где $T_{detect}^{s_m}$ – время с момента начала атаки s_m до её обнаружения системой; $T_{alert}^{s_m}$ – время распространения предупреждения по сети IoT.

4. Ограничения на качество детекции для каждого типа многовекторной атаки.

Для обеспечения надёжного обнаружения многовекторных атак пороги тревог $\theta_{alert}^{s_m}$ должны поддерживать минимальные значения метрик:

■ минимальный уровень Recall (полнота):

$$Recall^{s_m} = \frac{TP^{s_m}}{TP^{s_m} + FN^{s_m}} \geq R_{min}^{s_m} \quad (6)$$

■ максимальный уровень ложных срабатываний (False Positive Rate):

$$FPR^{s_m} = \frac{FP^{s_m}}{FP^{s_m} + TN^{s_m}} \leq FPR_{max}^{s_m} \quad (7)$$

где TP^{s_m} , FN^{s_m} , FP^{s_m} , TN^{s_m} – соответственно истинно-положительные, ложные отрицательные, ложные положительные и истинно-отрицательные результаты для атаки s_m ; $R_{min}^{s_m}$, $FPR_{max}^{s_m}$ – целевые минимальные/максимальные значения для метрик.

2. Используемые методы и технологии

2.1. Расширенный алгоритм федеративного обучения FedAvg

Федеративное усреднение с агрегацией импульса FedAvg является передовым методом, разработанным для повышения производительности систем обнаружения вторжений IoT. Этот метод расширяет традиционный подход федеративного усреднения, включая член импульса в процесс агрегации [2].

Федеративное обучение FedAvg обеспечивает плавную и быструю сходимость глобальной модели обнаружения вторжений, даже при наличии шумных или разнородных клиентских обновлений, значительно повышая точность и эффективность. Правило обновления выражается следующим образом:

$$\theta_t = \theta_{t-1} = \eta \times (\alpha \times \sum_i \nabla \text{Loss}(\theta_i) + (1 - \alpha) \times \text{PrevAggGrad}), \quad (8)$$

где θ_t – обновленные параметры глобальной модели в раунде t ; θ_{t-1} – глобальные параметры модели из предыдущего раунда; η – скорость обучения, которая контролирует величину обновлений; α – параметр импульса, уравнивающий влияние текущего градиента и прошлых суммарных градиентов; $\nabla \text{Loss}(\theta_i)$ – локальный градиент функции потерь, вычисленный участвующим клиентом i ; PrevAggGrad – агрегированный градиент из предыдущего раунда.

Эта формулировка расширяет стандартный алгоритм FedAvg, вводя член импульса $\nabla \text{Loss}(\theta_i)$, который помогает уменьшить колебания и сгладить процесс оптимизации. Включение $(1 - \alpha) \times \text{PrevAggGrad}$ гарантирует, что историческая информация о градиенте способствует глобальному обновлению, обеспечивая стабильность и более быструю сходимость, особенно в средах данных не независимых и одинаково распределенных.

Расширенный алгоритм федеративного обучения FedAvg описывает метод агрегации на стороне сервера с импульсом для FL. Алгоритм направлен на повышение сходимости и устойчивости путем включения импульса в агрегацию градиентов от участвующих сущностей. Этот параметр импульса улучшает способность ориентироваться в сложных ландшафтах оптимизации и эффективно распространять информацию между федеративными раундами обучения. Алгоритм начинается с инициализации глобальных параметров модели (θ_0), параметра импульса (α) и скорости обучения (η). Затем он выполняет несколько раундов федеративного обучения до достижения сходимости. В каждом раунде алгоритм инициализирует агрегированный градиент (AggGrad) нулем. Он собирает локальные параметры модели (θ_i) из каждой участвующей сущности и вычисляет локальный градиент (Grad_i) на основе градиента функции потерь ($\nabla \text{Loss}(\theta_i)$). Эти локальные градиенты накапливаются для обновления агрегированного градиента (AggGrad).

Для применения импульса алгоритм обновляет агрегированный градиент, объединяя предыдущий агрегированный градиент (PrevAggGrad), взвешенный по $(1 - \alpha)$, с текущим агрегированным градиентом (AggGrad), взвешенным по α . Эта корректировка сохраняет историческую информацию из предыдущих раундов, помогая сгладить влияние зашумленных или флуктуирующих градиентов.

После агрегации глобальные параметры модели (θ_t) обновляются путем вычитания произведения скорости обучения (η) и агрегированного градиента (AggGrad) из предыдущих глобальных параметров

модели (θ_{t-1}). Обновлённые параметры затем распределяются обратно участвующим сущностям для следующего раунда обучения. На протяжении всей работы алгоритма PrevAggGrad сохраняет агрегированный градиент из предыдущего раунда для обеспечения непрерывности импульса. Этот итерационный процесс продолжается до тех пор, пока не будут выполнены критерии сходимости, что завершает обучение.

2.2. Глубокие автоэнкодеры

Глубокие автоэнкодеры обучаются восстанавливать входные данные, сжимая их в компактное представление (кодировку) и затем декодируя обратно. Основная идея заключается в том, что автоэнкодер хорошо восстанавливает нормальные данные, но плохо справляется с аномалиями, что позволяет выявлять отклонения на основе ошибки реконструкции.

Автоэнкодер состоит из двух частей:

1. Кодировщик (энкодер) – преобразует входные данные в скрытое представление меньшей размерности.
2. Декодировщик (декодер) – восстанавливает исходные данные из скрытого представления.

Обучение автоэнкодера проводится на нормальных данных, чтобы минимизировать ошибку восстановления. После обучения для новых данных вычисляется ошибка реконструкции: если она превышает заданный порог, событие считается аномальным. Это позволяет обнаруживать неизвестные атаки без необходимости их предварительного описания.

Функция кодирования (энкодер) представляет собой отображение

$$f_{\theta_e} : \mathbb{R}^d \rightarrow \mathbb{R}^h, \quad (9)$$

где θ_e – параметры кодировщика, d – размерность входных данных, h – размерность скрытого представления ($h < d$).

Функция декодирования (декодер) представляет собой отображение

$$g_{\theta_d} : \mathbb{R}^h \rightarrow \mathbb{R}^d, \quad (10)$$

где θ_d – параметры декодировщика.

Автоэнкодер обучается путем минимизации различия между исходными данными и восстановленными данными. Основное требование – точное воспроизведение исходных данных на выходе.

Обучение происходит по следующим этапам:

Этап 1. Ошибка реконструкции.

Для входного вектора x автоэнкодер вычисляет восстановленный выход:

$$\hat{x} = g_{\theta_d} f_{\theta_e}(x). \quad (11)$$

Ошибка реконструкции определяется как среднеквадратичное отклонение:

$$L(x; \Theta) = \|x - \hat{x}\|^2, \quad (12)$$

где $\Theta = \{\theta_e, \theta_d\}$ – параметры модели.

Этап 2. Обучение.

Автоэнкодер обучается на множестве нормальных данных $X_{\text{train}} = \{x_{\text{norm}}^i\}$ путем минимизации средней ошибки:

$$L(\Theta) = \frac{1}{|X_{\text{train}}|} \sum_{x \in X_{\text{train}}} \|x - g_{\theta_d}(f_{\theta_e}(x))\|^2. \quad (13)$$

Этап 3. Обнаружение аномалий.

Для нового входного вектора x_{new} вычисляется ошибка реконструкции $e_{x_{\text{new}}}$

$$e_{x_{\text{new}}} > \theta, \quad (14)$$

где θ – порог аномальности, событие помечается как аномальное. Порог θ выбирается так, чтобы вероятность ложных срабатываний α была минимальной

$$\Pr_{x \sim P_{\text{norm}}}(e(x) > \theta) = \alpha.$$

3. Предлагаемая проблемно-ориентированная система

3.1. Алгоритм предлагаемой системы

Для решения поставленной задачи (1)–(7) разработан комбинированный метод, сочетающий распределённое обучение модели обнаружения аномалий и децентрализованное распространение оповещений. На рисунке 1 представлена общая схема предлагаемой системы. Она включает три основных процесса, происходящие непрерывно и параллельно: локальный мониторинг на узлах IoT, федеративное обновление модели обнаружения, децентрализованное оповещение и реагирование.

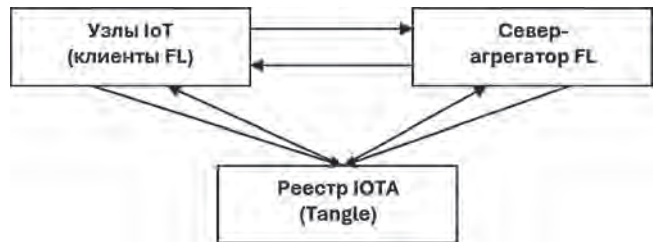


Рис. 1. Общая схема предлагаемой системы

1. Локальный мониторинг на узлах IoT.

Каждый узел (устройство) выполняет постоянный сбор данных о своём состоянии и трафике (например, сетевые потоки, системные логи, показания датчиков и пр.). Эти данные в режиме реального времени подаются на вход локальной модели обнаружения аномалий – лёгкого автоэнкодера, заранее обученного распознавать нормальное поведение устройства. Модель рассчитывает степень аномалии каждого наблюдения: если ошибка реконструкции превышает заданный порог θ , событие помечается

как потенциальная атака. Порог θ выбирается на основе статистики ошибок на обучающей выборке (например, θ соответствует квантилю 99-го процента ошибок на нормальных данных, чтобы обеспечить низкий уровень ложных срабатываний).

2. Федеративное обновление модели обнаружения.

Для адаптации к новым угрозам и изменениям в поведении устройств узлы периодически участвуют в обновлении модели автоэнкодера. Процесс координируется центральным агрегатором по следующему алгоритму:

1. Агрегатор рассылает актуальные веса глобальной модели W всем узлам.
2. Каждый узел проводит локальное обучение на новых данных, включая нормальные и аномальные сессии с пониженным весом.
3. Локальные веса W_i или градиенты ΔW_i передаются агрегатору.
4. Агрегатор усредняет полученные обновления.

Такой подход позволяет модели обучаться на распределённых данных всех устройств, повышая её обобщающую способность. Федеративное обучение сокращает объём передаваемых данных до параметров модели (несколько КБ вместо МБ исходных логов) и обеспечивает сохранение приватности, поскольку исходные данные не покидают устройства.

3. Децентрализованное оповещение и реагирование.

При срабатывании детектора аномалий узел формирует сигнал тревоги с идентификатором устройства, временной меткой, типом активности и хэшем подозрительных данных. Сигнал публикуется в распределённый реестр IOTA через транзакцию в Tangle, обеспечивая быстрое и безкомиссионное распространение. Подписанные узлы получают уведомление и применяют локальные меры: фильтрацию трафика, изоляцию устройств или оповещение администратора. Все события фиксируются в реестре, формируя неизменяемый журнал. Система устойчива к отказам: данные сохраняются в Tangle и не могут быть удалены злоумышленником.

Алгоритм работы предлагаемой системы имеет вид:

Алгоритм 1. Federated Anomaly Detection and Response in IoT

Вход: Порог аномалии θ (14), начальные веса модели W_0 , интервал федеративных обновлений T .

Выход: Обновляемая глобальная модель обнаружения; транзакции-оповещения об атаках в реестре.

Этап 1. Инициализация:

- считывать новые данные x (пакет трафика, измерение и т.п.);
- вычислить $\hat{x} = AE(x)$ проходом через автоэнкодер;

- рассчитать ошибку $e = \|x - \hat{x}\|^2$ (12), (14);
- если $e > \theta$: пометить событие как аномалию и перейти к шагу 3; иначе продолжить мониторинг.

Этап 2. Локальная реакция на аномалию (узел-детектор):

- сформировать сообщение об атаке m с детализацией (ID узла, время, признаки аномалии);
- отправить транзакцию в реестр IOTA с сообщением m (функция PublishToTangle(m));
- продолжить мониторинг (шаг 2).

Этап 3. Глобальное оповещение (каждый узел, асинхронно):

- при получении из реестра нового сообщения m об атаке: проверить подпись и целостность;
- выполнить преднастроенные меры: например, обновить локальные правила фильтрации, пометить соответствующие данные как вредоносные, уведомить администратора.

Этап 4. Федеративное обновление модели (агрегатор, каждые T времени):

- разослать текущее состояние модели $W^{(t)}$ на все узлы;
- для каждого узла i : обучить копию модели на локальных данных (например, за последние T интервала), получить обновлённые веса $W_i^{(t+1)}$; отправить их агрегатору;
- по получении всех (или большинства) обновлений вычислить новое глобальное состояние: $W^{(t+1)} = \frac{1}{\sum_i n_i} \sum_i n_i W_i^{(t+1)}$;
- установить $t := t + 1$.

Этап 5. Обновление локальных моделей: По мере получения обновлённых глобальных весов $W^{(t+1)}$ узлы обновляют свои автоэнкодеры (начиная новый цикл мониторинга с улучшенной моделью).

Алгоритм обеспечивает цикличное улучшение модели и параллельное реагирование на инциденты. Фаза реагирования (этапы 3–4) выполняется асинхронно и значительно быстрее обучения: сигнал тревоги распространяется через IOTA с задержкой в секунды, тогда как обучение модели происходит в фоне с интервалом порядка часа. Это позволяет быстро локализовать и ограничить атаку до обновления глобальной модели.

3.2. Архитектура предлагаемой системы

Архитектура системы включает ключевые компоненты:

1. Узлы IoT – гетерогенные устройства с агентом безопасности, выполняющим сбор диагностических данных и локальный запуск автоэнкодера для обнаружения аномалий. При срабатывании генерируется оповещение. Узлы реализуют федеративное обучение, имеют защищённое хранилище ключей для подписания сообщений.

2. Центральный сервер-агрегатор – координатор федеративного обучения, хранящий глобальную модель и собирающий обновления. Размещается в облаке или на периферии. Обеспечена защита сервера, но система сохраняет базовую функциональность при его отказе. Сервер взаимодействует с IOTA, публикует глобальные сообщения и обновления.

3. Распределённый реестр IOTA – децентрализованная сеть для хранения и распространения сообщений об атаках без комиссий. Узлы IoT и сервер подключаются к сети по схеме публикация/подписка. Может использоваться публичный Tangle или выделенный кластер.

4. Администратор/аналитический центр – обеспечивает мониторинг, управление параметрами системы, доступ к истории атак. Не является обязательным, но важен для эксплуатации.

Взаимодействие построено по двум осям: обучение (узлы ↔ сервер-агрегатор) и реагирование (узлы ↔ узлы через реестр), что обеспечивает многоуровневую защиту от локального обнаружения до глобальной координации и оповещения.

Ключевыми процессами являются:

1. Обмен параметрами модели (FL) – агент на узле выступает клиентом FL, соединяется с сервером через защищённый канал (TLS). Протокол предусматривает аутентификацию клиентов и проверку подписей глобальной модели для предотвращения атак с подменой модели.

2. Публикация и получение оповещений (IOTA) – модуль Tangle Client отправляет зашифрованные сообщения об атаках в сеть IOTA с использованием общего тега («IoTSecAlert»). Подписанные узлы получают уведомления и инициируют локальные меры защиты, обеспечивая быструю реакцию на распространение угроз.

3. Хранилище и анализ данных – преимущественно онлайн-мониторинг с возможностью сбора агрегированных статистик и результатов обнаружения. Журнал транзакций IOTA служит неизменяемым хранилищем истории инцидентов. Сервер или администратор могут использовать данные для отчётности, дообучения и прогнозирования атак.

Ниже описан сценарий, иллюстрирующий работу системы при возникновении атаки:

0. Узлы А, В, С запущены, начальная модель автоэнкодера обучена. Обмен параметрами с сервером происходит каждые 30 минут.

1. Злоумышленник начинает многовекторную атаку на узел А – массированный трафик и эксплуатация уязвимости прошивки.

2. Узел А фиксирует аномалию: автоэнкодер (13) выдаёт высокую ошибку реконструкции (9), (10), (12). Генерирует и публикует сообщение об атаке в IOTA.

3. Узлы В и С получают уведомление, блокируют подозрительный IP по правилам брандмауэра, предотвращая распространение атаки.

4. Сервер-агрегатор регистрирует инцидент и при необходимости запускает внеочередной раунд федеративного обучения. Часто вмешательство не требуется – узлы уже отреагировали.

5. Атака продолжается, злоумышленник пытается заразить узел В, но трафик блокируется. Узел А частично изолирован.

6. Узел В обнаруживает аномалию другим вектором, публикует своё сообщение. Узлы А и С обновляют защитные меры. Вся сеть обменивается информацией, снижая эффект неожиданности атаки.

7. После инцидента участники запускают федеративное обучение на собранных данных, включая аномалии, повышая чувствительность модели. Обновлённая модель распространяется по узлам, замыкая цикл обучения.

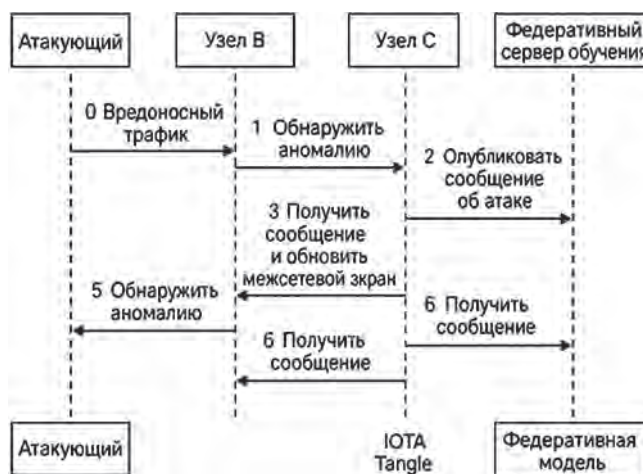


Рис. 2. Сценарий, иллюстрирующий работу системы при возникновении атаки

3.3. Метрики эффективности предложенной системы

Формально, качество обнаружения настраивается и оценивается следующими показателями. Обозначим: TP – число корректно обнаруженных атак (True Positives), FN – атак, которые система не обнаружила (False Negatives), FP – ложных тревог (False Positives), TN – корректных отрицаний (True Negatives).

1. Доля верно классифицированных ситуаций (атака / норма) среди всех

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (15)$$

При сильном дисбалансе (гораздо больше нормальных случаев) эта метрика не столь информативна, поэтому вводят дополнительные:

2. Точность прогноза атаки, характеризующая надёжность срабатываний (сколько ложных тревог)

$$\text{Precision} = \frac{TP}{TP + FP} \quad (16)$$

3. Полнота, чувствительность (доля обнаруженных атак)

$$\text{Recall} = \frac{TP}{TP + FN} \quad (17)$$

4. F1-мера

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (18)$$

5. Доля ложных тревог среди всех нормальных событий

$$\text{FPR} = \frac{FP}{FP + TN} \quad (19)$$

4. Экспериментальные исследования

4.1. Экспериментальная установка и методология

Экспериментальная оценка системы мониторинга и реагирования проводилась в виртуальной IoT-среде с моделированием многовекторных атак. Для этого был использован специализированный стенд, включающий 9 виртуальных IoT-узлов различных типов устройств и сервер-агрегатор. Аппаратные ресурсы узлов соответствовали типовым характеристикам периферийных устройств, а сетевая инфраструктура воспроизводила параметры беспроводных IoT-сетей.

В качестве данных применён набор N-BalIoT, содержащий трафик девяти видов устройств в нормальном режиме и под воздействием атак. Из 115 признаков были отобраны 43 наиболее релевантных, а каждому узлу был назначен конкретный тип устройства для обеспечения гетерогенности поведения.

В эксперименте рассматривались два типа атак: классические одновекторные и синтезированные многовекторные с одновременным воздействием на несколько узлов. Для моделирования аутентификации добавлен признак количества неудачных попыток входа.

Перед тестированием проведена предварительная тренировка автоэнкодера на нормальных данных с установкой порога аномалий, обеспечивающего низкий уровень ложных срабатываний. Эксперимент длился 30 минут с периодическим обновлением глобальной модели в ходе шести раундов федеративного обучения.

Для сравнительного анализа применялись три конфигурации систем обнаружения: централизованная, федеративная и локальные модели. Во время эксперимента фиксировались ключевые показатели эффективности, включая точность и полноту обнаружения, время реагирования, а также нагрузку

на сеть и вычислительные ресурсы. Использовались синтетические метки атак для обеспечения точности оценки.

4.2. Результаты эксперимента

Результаты экспериментов представлены в табл. 1 и на рис. 3, 4. Были сформированы результаты по следующим аспектам: качество обнаружения атак, эффективность реагирования, производительность и накладные расходы, а также влияние федеративного обучения.

Таблица 1.

Итоги обнаружения (среднее по узлам)
по метрикам качества (16)–(19)

Подход	Precision, %	Recall, %	F1 score, %	FPR, %
Централизованный	95,8	95,1	95,5	1,2
Федеративный	94,7	94,0	94,3	3,9
Локальные	90,5	88,2	89,3	5,4

Экспериментальные исследования показали высокую эффективность предложенного подхода. Федеративный метод достиг F1-меры 94,3 %, близкой к централизованному решению (95,5 %) и значительно превосходящей изолированные локальные модели (89,3 %). Уровень ложных срабатываний в федеративной системе составил 3,9 %, что соответствует требованиям IoT. Полнота (recall) улучшена за счёт обобщения знаний о различных атаках в распределённой сети.

Среднее время обнаружения атаки – 1,8 с, доставка оповещений через реестр IOTA – 2,7 с, суммарное время реакции не превышает 4,5 с, что существенно быстрее развития атак. Объём передаваемых данных менее 1,2 МБ за 30 минут, загрузка процессора – 15–25 %, потребление памяти не превышает 50 МБ на узел.

В ходе федеративного обучения точность обнаружения повысилась с 92 % до 95 % к шестому раунду, что подтверждает адаптивность и эффективность коллективного обучения. Кривые Accuracy и Loss на рисунке 3 показывают, что федеративная модель приближается к централизованной по качеству, превышая локальные модели по точности восстановления.

Рисунок 4 иллюстрирует пример временной линии многовекторной атаки: показаны моменты срабатывания на узлах и доставки сигналов – как раз видно, что сигнал от узла А успел дойти до других до того, как их атаки начались, существенно снизив последствия.

Полученные результаты экспериментальных исследований свидетельствуют о достижении поставленных целей работы. Разработанная система

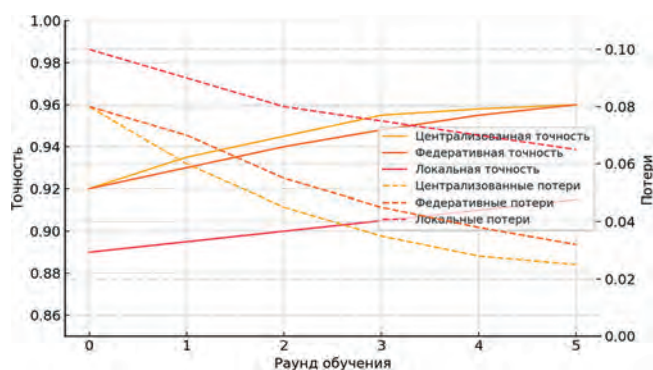


Рис. 3. Кривые изменения Accuracy (16) и Loss (ошибки реконструкции)



Рис. 4. Временная линия многовекторной атаки

продемонстрировала способность к эффективному обнаружению различных типов атак с показателем F1-меры, превышающим 94 %, при времени реакции менее 5 секунд. Важным достижением является приближение эффективности предложенного децентрализованного подхода к показателям централизованных систем при сохранении преимуществ распределенной архитектуры.

Несмотря на высокие показатели, достигнутые в контролируемых условиях, практическое внедрение системы требует решения ряда дополнительных задач. Реальный трафик IoT-устройств обладает повышенной зашумленностью и вариативностью, что может вызвать рост числа ложных срабатываний. Для снижения этого влияния целесообразно применять адаптивные пороговые значения и совершенствовать модели обнаружения аномалий. Кроме того, вопросы безопасности самой системы требуют особого внимания. Цифровые подписи обеспечивают защиту от несанкционированного доступа, однако возможны атаки на модель обучения со стороны скомпрометированных устройств. Для решения этих проблем необходимы механизмы византийской отказоустойчивости и эффективные системы управления доверием.

Заключение

В работе предложена и исследована проблемно-ориентированная система мониторинга и реагирования на многовекторные атаки в децентрализованных IoT-средах. Разработанное решение интегрирует три ключевых компонента: федеративное обучение на основе алгоритма FedAvg, глубокие автоэнкодеры для обнаружения аномалий и распределенный реестр IOTA для координации реагирования. Такой комплексный подход позволяет одновременно решать проблемы распределенности данных, обнаружения неизвестных угроз и обеспечения доверия в отсутствие централизованных органов. Архитектура системы включает три уровня: устройства с локальными моделями обнаружения, сервер-агрегатор для координации обучения и распределенный реестр для обмена информацией об атаках. Представленная математическая модель формализует процессы обучения, обнаружения и реагирования, а также определяет критерии эффективности системы.

Экспериментальные исследования на наборе данных N-BalIoT подтвердили высокую эффективность предложенного подхода. Федеративная система продемонстрировала точность обнаружения атак на уровне 95 %, что сопоставимо с централизованными решениями. Показатель F1-меры улучшен на 5 процентных пунктов по сравнению с изолированными локальными системами. Время реакции системы не превышает 5 секунд, а ресурсные затраты соответствуют ограничениям IoT-устройств. Теоретическая значимость работы заключается в демонстрации возможности synergies технологий машинного обучения и распределенных реестров для безопасности IoT. Практическая ценность состоит в разработке архитектуры, применимой в промышленных IoT, умных городах и медицинских сетях.

Перспективные направления дальнейших исследований включают: расширение класса detectable атак, повышение устойчивости системы к targeted атакам, оптимизацию производительности для микроконтроллеров, а также интеграцию механизмов reinforcement learning для активного противодействия угрозам. Проведенное исследование подтверждает возможность создания эффективных распределенных систем безопасности для Интернета вещей, сочетающих высокую точность обнаружения с сохранением конфиденциальности данных и отказоустойчивостью.

Исследование выполнено при финансовой поддержке Российского научного фонда, проект №24-21-00481 по теме «Методы противодействия многовекторным атакам на децентрализованные системы Интернета вещей».

Литература

1. Yaacoub J.-P. A., Noura H. N., Salman O. Security of federated learning with IoT systems: issues, limitations, challenges, and solutions // *Internet of Things and Cyber-Physical Systems*. 2023. Vol. 3. P. 155–179. DOI: 10.1016/j.iotcps.2023.04.001.
2. Khraisat A., Alazab A., Jan T. Federated learning for intrusion detection in IoT environments: a privacy-preserving strategy // *Discover Internet of Things*. 2025. Vol. 5, № 1. Article 17 p. DOI: 10.1007/s43926-025-00169-7.
3. Olanrewaju-George B., Pranggono B. Federated learning-based intrusion detection system for the internet of things using unsupervised and supervised deep learning models // *Cyber Security and Applications*. 2025. Vol. 3, December. Article 100068. DOI: 10.1016/j.csa.2024.100068.
4. Karunamurthy A., Vijayan K., Kshirsagar P. R. et al. An optimal federated learning-based intrusion detection for IoT environment // *Sci Rep*. 2025. Vol. 15, Article 8696. DOI: 10.1038/s41598-025-93501-8.
5. Rampone G., Ivaniv T., Rampone S. A hybrid federated learning framework for privacy-preserving near-real-time intrusion detection in IoT environments // *Electronics*. 2025. Vol. 14, № 7. Article 1430. DOI: 10.3390/electronics14071430.
6. Meidan Y., Bohadana M., Mathov Y., Mirsky Y., Shabtai A., Breitenbacher D., Elovici Y. N-Balot – network-based detection of IoT botnet attacks using deep autoencoders // *IEEE Pervasive Computing*. 2018. Vol. 17, № 3. P. 12–22. DOI: 10.1109/MPRV.2018.03367731.
7. Anand R. V., Magesh G., Alagiri I. et al. Design of an improved model using federated learning and LSTM autoencoders for secure and transparent blockchain network transactions // *Sci Rep*. 2025. Vol. 15, Article 1615. DOI: 10.1038/s41598-024-83564-4.
8. Nguyen V.-D., Diro A., Chilamkurti N., Heyne W., Phan K. T. Novel blockchain-enabled federated learning scheme for IoT anomaly detection // *IEEE Access*. DOI: 10.1109/11070312.
9. Friha O., Ferrag M. A., Benbouzid M., Berghout T., Kantarci B., Choo K.-K. R. 2DF-IDS: decentralized and differentially private federated learning-based intrusion detection system for industrial IoT // *Computers & Security*. 2023. Vol. 127. Article 103097. DOI: 10.1016/j.cose.2023.103097.
10. Begum K., Mozumder M. A. I., Joo M., Kim H. BFLIDS: blockchain-driven federated learning for intrusion detection in IoMT networks // *Sensors*. 2024. Vol. 24, № 14. Article 4591. DOI: 10.3390/s24144591.
11. Yang E., Jeong S., Seo C. Harnessing feature pruning with optimal deep learning based DDoS cyberattack detection on IoT environment // *Scientific Reports*. 2025. Vol. 15. DOI: 10.1038/s41598-025-02152-2.
12. Saranya K., Valarmathi A. A multilayer deep autoencoder approach for cross layer IoT attack detection using deep learning algorithms // *Scientific Reports*. 2025. Vol. 15, Article 10246. DOI: 10.1038/s41598-025-93473-9.
13. Regan C., Nasajpour M., Parizi R. M., Pouriyeh S., Dehghantanha A., Choo K.-K. R. Federated IoT attack detection using decentralized edge data // *Machine Learning with Applications*. 2022. Vol. 8. Article 100263. DOI: 10.1016/j.mlwa.2022.100263.
14. Khan A. A., Waseem M., Alshamrani N., Alharbi M., Alhazmi A. S., Zohdy A. M., Alattas F. A., Al Ghamdi A. Learning-based methods for cyber attacks detection in IoT systems: a survey on methods, analysis, and future prospects // *Electronics*. 2022. Vol. 11, № 9. Article 1502. DOI: 10.3390/electronics11091502.
15. Ferrag M. A., Friha O., Maglaras L., Janicke H., Shu L. Security of federated learning with IoT systems: issues, limitations, challenges, and solutions // *Internet of Things*. 2023. T. 22. C. 100222. DOI: 10.1016/j.iot.2023.100222.
16. Alshaikhli M., Elfouly T., Elharrouss O., Mohamed A., Ottakath N. Evolution of internet of things from blockchain to IOTA: a survey // *IEEE Access*. 2022. Vol. 10. P. 844–866. DOI: 10.1109/ACCESS.2021.3138353.
17. Shalabi K., Abu Al-Haija Q., Al-Fayoumi M. A. A blockchain-based intrusion detection/prevention systems in IoT network: a systematic review // *Procedia Computer Science*. 2024. Vol. 236. P. 410–419. DOI: 10.1016/j.procs.2024.05.048.
18. Alharthi H., Alshehri S., Kalkatawi M. Revolutionizing IoT security: a blockchain and federated learning-based anomaly detection system // In: *Proceedings of the 2024 7th Artificial Intelligence and Cloud Computing Conference (AICCC '24)*. 2024. P. 565–572. DOI: 10.1145/3719384.3719466.
19. Al Sadi A., Mazzocca C., Melis A., Montanari R., Prandini M., Romandini N. P-IOTA: a cloud-based geographically distributed threat alert system that leverages P4 and IOTA // *Sensors*. 2023. Vol. 23, № 6. Article 2955. DOI: 10.3390/s23062955.
20. Lazzarini R., Tianfield H., Charissis V. Federated learning for IoT intrusion detection // *AI*. 2023. Vol. 4. №. 3. P. 509–530. DOI: 10.3390/ai4030028.
21. Ferrag M. A., Friha O., Maglaras L., Janicke H., Shu L. On the performance of federated learning algorithms for IoT // *IoT*. 2023. T. 3, № 2. C. 273–284. DOI: 10.3390/iot3020016.
22. Alsaedi A., Moustafa N., Tari Z., Mahmood A., Anwar A. Security and privacy-enhanced federated learning for anomaly detection in IoT infrastructures // *Journal of Information Security and Applications*. 2021. Vol. 58. Article 102413. DOI: 10.1016/j.jisa.2021.102413.
23. Khan M. A., Waseem M., Alshamrani N., Alharbi M., Alhazmi A. S., Zohdy A. M., Alattas F. A., Al Ghamdi A. Security considerations for Internet of Things: a survey // *SN Computer Science*. 2020. Vol. 1, no. 4. Article 193. DOI: 10.1007/s42979-020-00201-3.
24. Ceccarelli M., Zecchini M., Brighente A., Conti M. GitHub – MMw-Unibo/FETA: enabling federated learning at the edge through the IOTA Tangle // *Future Generation Computer Systems*. 2024. P. 17–29. DOI: 10.1016/j.future.2023.10.014.
25. Тебуева Ф. Б., Рябцев С. С., Огур М. Г., Андреев И. А., Горяйнов С. А. Модель угроз информационной безопасности агентов в децентрализованной среде Интернета вещей, формализующая сценарии атак на информационную безопасность доверенного взаимодействия // *Кузнечно-штамповочное производство. Обработка материалов давлением*. 2024. № 11. С. 220–232.

PROBLEM-ORIENTED SYSTEM FOR MONITORING AND RESPONDING TO MULTIVECTOR ATTACKS IN A DECENTRALIZED INTERNET OF THINGS ENVIRONMENT

Tebueva F. B.⁶, Petrenko V. I.⁷, Satybaldina D. Zh.⁸, Ogur M. G.⁹, Guseva T. M.¹⁰

Keywords: Internet of Things; multivector attacks; intrusion detection; federated learning; autoencoder; IOTA; blockchain; anomaly; security monitoring.

Objective: to enhance the effectiveness of monitoring and responding to multivector attacks in a decentralized Internet of Things (IoT) environment by integrating federated learning, deep autoencoders, and the distributed IOTA ledger. The priorities include accurate attack detection, minimizing false positives, reducing response time, and preserving data privacy.

Method: a problem-oriented system was developed, combining local monitoring on IoT nodes with autoencoders for anomaly detection, federated learning using the FedAvg algorithm for collective model updates, and decentralized alert dissemination via the distributed IOTA ledger. The system implements secure exchange of model parameters, digital message signing, and asynchronous response through a publish/subscribe network.

Results: experimental studies on the real N-Balot dataset simulating multivector attacks demonstrated high detection accuracy (approximately 95%), achieving an F1-score above 94%, with false positive rates around 4%. The system's response time did not exceed 5 seconds, significantly improving operational reaction to attacks. Federated learning provided steady improvement in model quality considering data distribution and heterogeneity. The architecture proved scalable, fault-tolerant, and capable of effectively detecting complex threats across multiple system levels.

Practical value: the solution is implementable in industrial IoT, smart cities, and medical networks to enhance cybersecurity while maintaining privacy and reducing network load.

Scientific novelty: the study presents a comprehensive synthesis of federated learning, deep autoencoders, and distributed ledger technology for effective monitoring of multivector attacks in decentralized IoT environments. The proposed approach combines the advantages of distributed learning and blockchain mechanisms to achieve high adaptability, accuracy, and security in rapidly growing and diverse IoT infrastructures.

References

1. Yaacoub J.-P. A., Noura H. N., Salman O. Security of federated learning with IoT systems: issues, limitations, challenges, and solutions // Internet of Things and Cyber-Physical Systems. 2023. Vol. 3. P. 155–179. DOI: 10.1016/j.iotcps.2023.04.001.
2. Khraisat A., Alazab A., Jan T. Federated learning for intrusion detection in IoT environments: a privacy-preserving strategy // Discover Internet of Things. 2025. Vol. 5, № 1. Article 17 p. DOI: 10.1007/s43926-025-00169-7.
3. Olanrewaju-George B., Pranggono B. Federated learning-based intrusion detection system for the internet of things using unsupervised and supervised deep learning models // Cyber Security and Applications. 2025. Vol. 3, December. Article 100068. DOI: 10.1016/j.csa.2024.100068.
4. Karunamurthy A., Vijayan K., Kshirsagar P. R. et al. An optimal federated learning-based intrusion detection for IoT environment // Sci Rep. 2025. Vol. 15, Article 8696. DOI: 10.1038/s41598-025-93501-8.
5. Rampone G., Ivaniv T., Rampone S. A hybrid federated learning framework for privacy-preserving near-real-time intrusion detection in IoT environments // Electronics. 2025. Vol. 14, №7. Article 1430. DOI: 10.3390/electronics14071430.
6. Meidan Y., Bohadana M., Mathov Y., Mirsky Y., Shabtai A., Breitenbacher D., Elovici Y. N-Balot – network-based detection of IoT botnet attacks using deep autoencoders // IEEE Pervasive Computing. 2018. Vol. 17, № 3. P. 12–22. DOI: 10.1109/MPRV.2018.03367731.
7. Anand R. V., Magesh G., Alagiri I. et al. Design of an improved model using federated learning and LSTM autoencoders for secure and transparent blockchain network transactions // Sci Rep. 2025. Vol. 15, Article 1615. DOI: 10.1038/s41598-024-83564-4.
8. Nguyen V.-D., Diro A., Chilamkurti N., Heyne W., Phan K. T. Novel blockchain-enabled federated learning scheme for IoT anomaly detection // IEEE Access [электронный ресурс]. DOI: 10.1109/11070312. URL: <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=11070312> (дата обращения – 03.09.2025).
9. Friha O., Ferrag M. A., Benbouzid M., Berghout T., Kantarci B., Choo K.-K. R. 2DF-IDS: decentralized and differentially private federated learning-based intrusion detection system for industrial IoT // Computers & Security. 2023. Vol. 127. Article 103097. DOI: 10.1016/j.cose.2023.103097.
10. Begum K., Mozumder M. A. I., Joo M., Kim H. BFLIDS: blockchain-driven federated learning for intrusion detection in IoMT networks // Sensors. 2024. Vol. 24, № 14. Article 4591. DOI: 10.3390/s24144591.
6. Fariza B. Tebueva, Dr.Sc. of Physical and Mathematical Sciences, Associate Professor, Professor of the Department of Computational Mathematics and Cybernetics, North Caucasus Federal University. Stavropol, Russia. ORCID: <https://orcid.org/0000-0002-7373-4692>. E-mail: fariza.teb@gmail.com
7. Vyacheslav I. Petrenko, Ph.D. of Technical Sciences, Associate Professor, Head of the Department of Organization and Technology of Information Security, North Caucasus Federal University. Stavropol, Russia. <https://orcid.org/0000-0003-4293-7013>. E-mail: vipetrenko@ncfu.ru
8. Dina Z. Satybaldina, Ph.D. of Physical and Mathematical Sciences, Director of the Research Institute of Information Security and Cryptology, L.N. Gumilyov Eurasian National University, Astana, Republic of Kazakhstan. <https://orcid.org/0000-0003-0291-4685>. E-mail: satybaldina_dzh@enu.kz
9. Maxim G. Ogur, Senior Lecturer of the Department of Computational Mathematics and Cybernetics, North Caucasus Federal University, Stavropol, Russia. <https://orcid.org/0000-0002-2387-0901>. E-mail: ogur26@gmail.com
10. Tatyana M. Guseva, Assistant of the Department of Organization and Technology of Information Security, North Caucasus Federal University, Stavropol, Russia. E-mail: tatyana.petrova.96@bk.ru

11. Yang E., Jeong S., Seo C. Harnessing feature pruning with optimal deep learning-based DDoS cyberattack detection on IoT environment // *Scientific Reports*. 2025. Vol. 15. DOI: 10.1038/s41598-025-02152-2.
12. Saranya K., Valarmathi A. A multilayer deep autoencoder approach for cross layer IoT attack detection using deep learning algorithms // *Scientific Reports*. 2025. Vol. 15, Article 10246. DOI: 10.1038/s41598-025-93473-9.
13. Regan C., Nasajpour M., Parizi R. M., Pouriyeh S., Dehghantanha A., Choo K.-K. R. Federated IoT attack detection using decentralized edge data // *Machine Learning with Applications*. 2022. Vol. 8. Article 100263. DOI: 10.1016/j.mlwa.2022.100263.
14. Khan A. A., Waseem M., Alshamrani N., Alharbi M., Alhazmi A. S., Zohdy A. M., Alattas F. A., Al Ghamdi A. Learning-based methods for cyber attacks detection in IoT systems: a survey on methods, analysis, and future prospects // *Electronics*. 2022. Vol. 11, № 9. Article 1502. DOI: 10.3390/electronics11091502.
15. Ferrag M. A., Friha O., Maglaras L., Janicke H., Shu L. Security of federated learning with IoT systems: issues, limitations, challenges, and solutions // *Internet of Things*. 2023. T. 22. C. 100222. DOI: 10.1016/j.iot.2023.100222.
16. Alshaikhli M., Elfouly T., Elharrouss O., Mohamed A., Ottakath N. Evolution of internet of things from blockchain to IOTA: a survey // *IEEE Access*. 2022. Vol. 10. P. 844–866. DOI: 10.1109/ACCESS.2021.3138353.
17. Shalabi K., Abu Al-Haija Q., Al-Fayoumi M. A. A blockchain-based intrusion detection/prevention systems in IoT network: a systematic review // *Procedia Computer Science*. 2024. Vol. 236. P. 410–419. DOI: 10.1016/j.procs.2024.05.048.
18. Alharthi H., Alshehri S., Kalkatawi M. Revolutionizing IoT security: a blockchain and federated learning-based anomaly detection system // In: *Proceedings of the 2024 7th Artificial Intelligence and Cloud Computing Conference (AICCC '24)*. 2024. P. 565–572. DOI: 10.1145/3719384.3719466.
19. Al Sadi A., Mazzocca C., Melis A., Montanari R., Prandini M., Romandini N. P-IOTA: a cloud-based geographically distributed threat alert system that leverages P4 and IOTA // *Sensors*. 2023. Vol. 23, № 6. Article 2955. DOI: 10.3390/s23062955.
20. Lazzarini R., Tianfield H., Charissis V. Federated learning for IoT intrusion detection // *AI*. 2023. Vol. 4. №. 3. P. 509–530. DOI: 10.3390/ai4030028.
21. Ferrag M. A., Friha O., Maglaras L., Janicke H., Shu L. On the performance of federated learning algorithms for IoT // *IoT*. 2023. T. 3, № 2. C. 273–284. DOI: 10.3390/iot3020016.
22. Alsaedi A., Moustafa N., Tari Z., Mahmood A., Anwar A. Security and privacy-enhanced federated learning for anomaly detection in IoT infrastructures // *Journal of Information Security and Applications*. 2021. Vol. 58. Article 102413. DOI: 10.1016/j.jisa.2021.102413.
23. Khan M. A., Waseem M., Alshamrani N., Alharbi M., Alhazmi A. S., Zohdy A. M., Alattas F. A., Al Ghamdi A. Security considerations for Internet of Things: a survey // *SN Computer Science*. 2020. Vol. 1, no. 4. Article 193. DOI: 10.1007/s42979-020-00201-3.
24. Ceccarelli M., Zecchini M., Brighente A., Conti M. GitHub – MMw-Unibo/FETA: enabling federated learning at the edge through the IOTA Tangle [электронный ресурс]. URL: <https://github.com/MMw-Unibo/FETA> (дата обращения – 03.09.2025).
25. Tebueva F. B., Ryabtsev S. S., Ogur M. G., Andreev I. A., Goryainov S. A. Information Security Threat Model for Agents in Decentralized Internet of Things Environment, Formalizing Attack Scenarios on Trusted Interaction Information Security // *Kuznechno-shtampovnoe proizvodstvo. Obrabotka materialov davleniem*. 2024. No. 11. pp. 220–232. (in Russian).



АЛГОРИТМ ОБНАРУЖЕНИЯ СИГНАЛА СИНХРОНИЗАЦИИ В КВАНТОВЫХ СЕТЯХ

Плёткин А. П.¹

DOI: 10.21681/2311-3456-2025-6-81-87

Цель исследования: разработка и исследование алгоритма обнаружения оптического сигнала для синхронизации станций системы квантового распределения ключей с повышенной защищенностью от несанкционированного доступа.

Методы исследования: вероятностное распределение, статистический анализ, однофотонная регистрация.

Результаты исследования: обоснована важность процесса временной синхронизации, которая реализуется посредством высокоточного обнаружения оптического сигнала. Исследован алгоритм обнаружения оптического сигнала для метода синхронизации с повышенной защищенностью от несанкционированного доступа. Предложен усовершенствованный алгоритм анализа временной области распространения синхросигнала с использованием лавинных фотодетекторов. Показано, что предложенный алгоритм позволяет использовать лавинные фотодетекторы в режиме одиночного счета фотонов. Проведен анализ временных характеристик разработанного алгоритма синхронизации и представлено аналитическое выражение для расчета временной задержки детектирования, которое обеспечивает последовательный анализ временных окон с учетом перестраиваемого времени восстановления фотодетектора. Аналитическое выражение может использоваться для инженерных расчетов при проектировании системы КРК. Предложенный алгоритм значительно снижает вероятность несанкционированного доступа к процессу синхронизации и позволяет с заданной точностью определить временные параметры сигнального окна, что является критически важным для последующей работы системы квантового распределения ключей.

Научная новизна: предложен алгоритм обнаружения оптического сигнала в процессе синхронизации, который отличается повышенной защищенностью от несанкционированного доступа. Представлено аналитическое выражение для инженерных расчетов временной задержки детектирования в процессе синхронизации.

Ключевые слова: защищенность, синхронизация, квантовое распределение, однофотонность, оптический импульс.

Введение

Квантовая криптография позволяет обеспечить безусловную защищенность данных методами квантовой физики [1–3]. Физическим воплощением квантовой криптографии являются системы квантового распределения ключей – сложные технологические устройства, в идеализированном варианте позволяющие обеспечить абсолютную криптостойкость передаваемых данных между пользователями. Целью таких систем является создание у двух устройств одинаковой последовательности случайных символов – одноразового блокнота [4, 5]. Исследования действующих систем квантового распределения ключей (СКРК) показывают наличие несовершенств в технической реализации. Такие несовершенства могут являться уязвимостями, позволяющими получить несанкционированный доступ [6]. Важным процессом в работе систем квантового распределения ключей является синхронизация разнесенной в пространстве аппаратуры. Современные квантовые коммуникации реализуются через технологию квантового распределения ключей (КРК) [7]. В базовой конфигурации «точка-точка» два абонента обмениваются квантовыми сигналами по оптическому каналу связи. Эта простая схема лежит в основе построения более сложных квантовых сетей, включая

магистральные. Однако простейшая топология имеет существенные ограничения, главное из которых – максимальная дальность передачи сигнала. Ограничения обусловлены прежде всего физическими свойствами оптического волокна и особенностями работы квантовых протоколов.

Главная задача квантового распределения ключей в магистральных сетях – обеспечить безопасную передачу ключей между удалёнными узлами. Однако существующие технологии сталкиваются с рядом фундаментальных и технических ограничений. Например, ограничение дальности передачи из-за затухания в оптическом волокне. Квантовые сигналы быстро затухают в оптоволокне и уже на расстоянии 100–150 км сигнал становится слишком слабым для детектирования. При этом классические оптические усилители неприменимы, так как они разрушают квантовые состояния, а квантовые повторители невозможно использовать из-за теоремы неклонирования (No-Cloning Theorem). В современных магистральных сетях применяются доверенные промежуточные узлы (ДПУ) – защищённые аппаратные комплексы, где ключи расшифровываются и повторно шифруются. Реализация сетей на основе ДПУ сталкивается с проблемой уязвимости к физическим

¹ Плёткин Антон Павлович, кандидат технических наук, доцент, Южный федеральный университет. г. Таганрог, Россия. E-mail: pljonkin@sfnedu.ru

атакам. Кроме того, каждый новый доверенный узел увеличивает общие риски сети. Активные исследования ведутся в направлении сетей с недоверенными узлами (НДУ). Такие сети работают под управлением квантовых протоколов (MDI-QKD, TF-QKD) на основе запутанных фотонов и считается, что безопасность самой сети не зависит от надежности НПУ. Одной из нерешенных задач в квантовых коммуникациях является проблема «последней мили» (last mile). Это финальный этап доставки квантовых ключей от магистральной сети к конечному пользователю. «Последняя миля» — один из ключевых технологических барьеров для массового внедрения квантовой криптографии.

Большинство исследований сегодня сфокусированы на разработке новых квантовых протоколов и методах предотвращения несанкционированного доступа к системам КРК [8–11]. Лишь малая часть научных статей посвящена вопросам аутентификации, идентификации и синхронизации в квантовых сетях [12, 13]. В данной работе мы описываем важность временной синхронизации при квантовом распределении ключей. Временная синхронизация — критически важный процесс, обеспечивающий корректное детектирование квантовых сигналов и дальнейший безопасный обмен ключами. Точность синхронизации напрямую влияет на эффективность обнаружения фотонов, подавление шумов в квантовом канале, скорость генерации ключей. Синхронизация не влияет на криптографическую стойкость самого протокола, однако, в исследованиях [14–15] показано, что доступ к процессу синхронизации дает злоумышленнику несанкционированный доступ к работе системы КРК. Последнее позволяет злоумышленнику влиять² на работу системы, вносить управляемые помехи, оставаясь незамеченным [16].

Алгоритм обнаружения оптического импульса

До начала работы квантового протокола станции системы КРК должны провести процедуры аутентификации и идентификации [17, 18]. Роль синхронизации при квантовом распределении заключается, в том числе, в корректном сопоставлении сигналов. В протоколах типа BB84 или MDI фотоны передаются в строго определённые временные интервалы и без высокоточной синхронизации приёмник не сможет отличить полезный сигнал от темновых отсчётов детектора или фонового шума. Таким образом, для минимизации ошибок алгоритмы синхронизации стараются добиться пикосекундной точности. В процессе проектирования алгоритмов синхронизации сокращение временного окна детектирования (Δt) способствует снижению вероятности ложных срабатываний фотодетектора.

Вероятность ложных срабатываний (темновых отсчётов) лавинного фотодетектора P_{dark} определяется как вероятность регистрации сигнала при отсутствии фотонов. Для лавинных фотодетекторов она вычисляется по формуле:

$$P_{dark} = R_{dark} \cdot \Delta t, \quad (1)$$

где R_{dark} — частота темновых отсчетов.

В реальных системах КРК следует учитывать вероятность так называемого эффекта послеимпульсов $P_{afterpulse}$, который возникает из-за перезаряда детектора. При этом уточненная модель полной вероятности ложного срабатывания будет иметь вид

$$P_{false} = 1 - (1 - P_{dark}) \cdot (1 - P_{afterpulse}). \quad (2)$$

Высокий уровень P_{false} значительно увеличивает уровень квантовых ошибок (QBER) при формировании ключевой последовательности. Например, для протокола BB84 верхний предел значения QBER составляет 11 %. Наиболее эффективными способами снижения P_{false} являются температурный режим ОЛФД и снижение временного интервала детектирования Δt . Приведем пример экспоненциальной зависимости темновых отсчетов детектора от температуры:

$$R_{dark}(T) = R_0 \cdot e^{-\frac{E_g}{k_B \cdot T}}, \quad (3)$$

где E_g — ширина запрещенной зоны, R_0 — константа материала, k_B — постоянная Больцмана, T — температура.

Эмпирическая формула вероятности возникновения послеимпульсов может быть представлена в следующем виде:

$$R_{afterpulse} = n \cdot \left(1 - e^{-\frac{\tau_d}{\tau_r}}\right) \cdot e^{-\frac{V_e}{V_0}}, \quad (4)$$

где n — технологический коэффициент, V_0 — напряжение, V_e — приложенное напряжение, τ_d — мертвое время детектора, τ_r — постоянная времени восстановления.

Таким образом, оба параметра существенно влияют на скорость генерации ключевой последовательности R_{key} :

$$R_{key} = \frac{R_{pulse} \cdot n_{det} \cdot (1 - R_{false})^N}{2}, \quad (5)$$

где N — число детекторов, n_{det} — эффективность детектирования.

Для коммерческого внедрения КРК (особенно в вопросе «последней мили») критически важно улучшать методы синхронизации, делая их доступными и устойчивыми к помехам. Рассмотрим процесс синхронизации на примере коммерческого образца системы КРК. В реализованных системах квантового распределения ключей распространенным решением является многофотонный режим синхронизации,

² Pljonkin A. Synchronization in quantum key distribution systems / A. Pljonkin, K. Romyantsev, P. K. Singh // Cryptography. — 2017. — No. 1. — P. 18.

при котором фотодетекторы работают в линейном режиме, а сам синхроимпульс представляет собой оптический сигнал высокой интенсивности. Методы обнаружения оптического сигнала могут быть реализованы по двухпроходной [19, 20] или однопроходной схеме [21, 22]. Другой метод подразумевает наличие выделенного волокна непосредственно под синхронизацию. Существуют реализации алгоритмов, основанные на мультиплексировании длин волн и временных отрезков. В классическом варианте однопроходной схемы источник излучения формирует периодическую последовательность оптических импульсов и направляет их в квантовый канал (оптическое волокно, соединяющее приемную и передающую станции системы КРК) [23–25]. На приемной стороне фотодетектор анализирует поступающий сигнал. Детектирование происходит пошаговым стробированием интервалов на всей временной оси. Максимальный период следования выбирается из расчета максимального времени, которое требуется импульсу на преодоление оптического пути. Период следования разбивается на временные интервалы, каждый из которых многократно анализируется (на предмет наличия сигнала). Итогом синхронизации должно быть выделение временного интервала длительностью не более 50 пс. Такая точность достигается путем разбиения временных интервалов с наибольшим числом зафиксированных срабатываний на более короткие по времени. При классическом методе синхронизации в системах квантовой связи не применяются алгоритмы защиты и контроля излучения, поэтому злоумышленнику не составляет труда получить доступ к квантовому каналу и использовать полученные данные для внесения управляемых помех в работу системы КРК.

Нами предложен алгоритм обнаружения оптического сигнала синхронизации в квантовых сетях, основанный на слабофотонной передаче и контролируемой мощности излучения. Число фотонов в синхроимпульсе при этом не превышает 10. В таком случае процесс синхронизации технически не отличается от работы квантового протокола. Модель оптического сигнала в предлагаемом алгоритме описывается как поток фотонов, подчиняющийся пуассоновской статистике, мощность которого в момент времени t выражается формулой

$$P(t) = N \cdot h\nu \cdot f(t), \quad (6)$$

где N – число фотонов в импульсе, $h\nu$ – энергия, $f(t)$ – форма импульса.

Так как для детектирования слабофотонного сигнала применяются лавинные фотодетекторы, то вероятностную модель обнаружения необходимо рассматривать для наличия сигнала ($P(n|H_1)$) и наличия только темнового тока ($P(n|H_0)$):

и

$$P(n|H_1) = \frac{(\eta N)^n e^{-\eta N}}{n!} \quad (7)$$

$$P(n|H_0) = \frac{(R_{dark} T)^n e^{-R_{dark} T}}{n!}. \quad (8)$$

Так как в исследуемом алгоритме синхронизации лавинные фотодетекторы функционируют в режиме Гейгера (одионого счета фотонов), то последовательный анализ временных интервалов невозможен. Наиболее важной характеристикой в процессе обнаружения оптического сигнала при синхронизации системы квантового распределения ключей является время восстановления. Этот параметр определяет период неактивности однофотонного лавинного фотодиода после регистрации фотоэлектрона или импульса темнового тока. В современных детекторах значение времени восстановления может программно настраиваться [26]. При этом минимальное время восстановления позволяет увеличить скорость счёта, но повышает вероятность послеимпульсов. Напротив, максимальное значение снижает уровень шумов, но ограничивает частоту детектирования. Таким образом, оптимальный выбор времени восстановления критически важен для баланса между чувствительностью и уровнем шумов детектора. Этот параметр требует тщательной настройки в зависимости от конкретных условий работы СКРК (длина линии, уровень затухания, требования к скорости генерации ключей).

Как уже упоминалось, ключевая задача синхронизации заключается в точном определении момента прихода фотонных импульсов на ОЛФД. Для этого выполняется измерение общей длины квантового канала, учитывающей как протяженность волоконно-оптической линии связи, так и длину оптических трактов внутри системы КРК. Отметим, что для первичного анализа длины можно использовать рефлектометрический метод, который позволит определить ориентировочную длину квантового канала с точностью до нескольких метров. Такой точности недостаточно для синхронизации, но это позволит существенно сократить время анализа временного кадра.

Если источник излучения и фотодетекторы расположены в одной станции (такой подход актуален для топологии «ДПУ – конечный пользователь»), то ослабление сигнала на основе анализа ВОЛС можно осуществлять как на стороне получателя, так и в начале передачи на стороне отправителя. Оптический импульс с длиной волны 1550 нм формируется источником излучения на станции отправителя. Импульс проходит через оптический канал до станции получателя. Предположим, что на обратном пути сигнал ослабляется до слабофотонного уровня и возвращается на ОЛФД станции отправителя. При каждой посылке импульса на ОЛФД подается стробирующий

сигнал, переводящий детектор в режим регистрации одиночных фотонов. Система вычисляет и фиксирует временную задержку стробирующего сигнала и детектор активируется на строго заданный временной интервал, соответствующий ожидаемому окну прихода сигнала. Режим счета одиночных фотонов в ОЛФД активируется с временными интервалами, превышающими время восстановления. Такой подход позволяет анализировать несколько временных окон в рамках одного временного кадра. Факт регистрации (или отсутствия регистрации) фотоэлектронов или темновых отсчетов в отдельных временных окнах не влияет на общую продолжительность анализа временного кадра в режиме обнаружения сигнального окна. Это позволяет сохранять стабильность работы системы независимо от характера поступающих сигналов. На рисунке 1 представлена пространственно-временная диаграмма, иллюстрирующая процедуру сканирования временного периода в ходе синхронизации.

Предположим, что длина ВОЛС между станцией системы КРК в ДПУ и станцией, расположенной у конечного пользователя, составляет 50 км. Тогда, максимальный период следования T будет рассчитываться исходя из расстояния в 100 км (двухпроходная схема распространения) и скорости распространения оптического сигнала в волокне. Для примера рассмотрим популярную модель ОЛФД InGaAs/InP ID Quantique id230. Подобные фотодетекторы

применяются в системах КРК и имеют настраиваемое время восстановления в пределах от 1 до 100 мкс. Возьмем среднее значение $\tau_d = 50$ мкс, которое является оптимальным в работе при температуре -30°C . Зная параметр τ_d , можем установить длительность временного кадра T_k , при соблюдении условия $T_k > \tau_d$. Пусть $T_k = 80$ мкс. Длительность временного окна $nw = 2$ нс. Длительность оптического импульса $\tau_p = 1$ нс. Вышеперечисленные параметры позволяют вычислить суммарное количество временных окон в периоде T , во временном кадре T_k и, соответственно, число временных кадров в периоде. После каждой посылки оптического импульса ОЛФД переводится в режим одиночного счета фотонов с учетом временной задержки детектирования Z_{nm} . За один период следования алгоритм анализирует одно временное окно nw в одном временном кадре T_k . После анализа всех временных кадров периода T_0 начинается последовательный просмотр временных окон периода T_1 , но значение временной задержки Z_{1n} в каждом кадре теперь увеличивается на длительность временного окна. При анализе каждого nw происходит фиксация фотоэлектронов или темновых отсчетов. Таким образом производится последовательный анализ всей временной области. В результате первого этапа синхронизации алгоритм обнаруживает временные окна ns , в которых были зафиксированы срабатывания фотодетектора и выделяет интервал с максимальным значением.

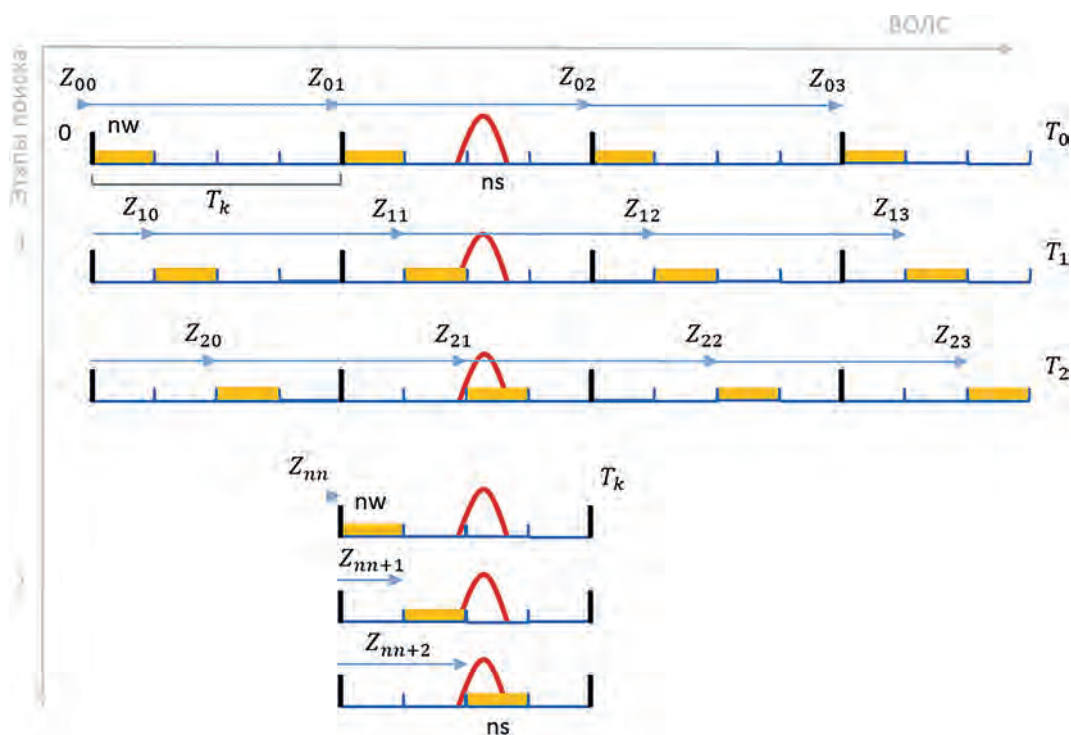


Рис. 1. Временная диаграмма поиска оптического сигнала

Отметим, что объем выборки в каждом временном окне составляет несколько сотен раз. Последнее сделано для того, чтобы минимизировать вероятность ложного срабатывания и однозначно отличить сигнал от темновых отсчетов. На втором этапе синхронизации алгоритм выделяет временной кадр T_k , в котором расположено ns с максимальным числом срабатываний. Отдельно следует выделить случай, когда ns расположены на границе двух временных кадров. В этом случае выделяется два смежных T_k . Далее происходит последовательный анализ всех временных окон в выделенных T_k . Алгоритм поиска на втором этапе аналогичен первому. В результате второго этапа синхронизации обнаруживается временное окно ns с максимальным числом зарегистрированных фотоэлектронов. На следующем этапе происходит дробление обнаруженного временного окна на временные интервалы длительностью по 10 пс. Результатом процесса синхронизации является обнаруженный временной интервал длительностью 10 пс.

Отметим, что в процессе работы квантового протокола требуется периодическая калибровка длины ВОЛС. Последнее связано с физическими изменениями в волокне из-за воздействия внешних факторов. Подстройка может реализовываться каждый раз после нескольких итераций формирования квантового ключа. В таком случае нет необходимости проводить полный анализ всего периода T_n , а достаточным будет реализация уточняющего этапа синхронизации.

Аналитическое выражение для расчета временной задержки детектирования алгоритма обнаружения сигнального временного интервала в процессе синхронизации станций СКРК при топологии «ДПУ – конечный пользователь» имеет вид (9)

$$Z_{nn} = \frac{T}{nT_n} \cdot (a_n - 1) + nw \cdot (T_n - 1), \quad (9)$$

где nT_n – число временных кадров в периоде T ; a_n – порядковый номер активации ОЛФД; T_n – порядковый номер периода.

Выводы и дискуссия

В статье поднимается вопрос синхронизации для сегмента магистральной сети квантовых коммуникаций. Обоснована важность процесса временной синхронизации, которая реализуется посредством высокоточного обнаружения оптического сигнала. Исследован алгоритм обнаружения оптического сигнала для метода синхронизации с повышенной защищенностью от несанкционированного доступа. Предложен усовершенствованный алгоритм анализа временной области распространения синхросигнала с использованием лавинных фотодетекторов. Показано, что предложенный алгоритм позволяет использовать лавинные фотодетекторы в режиме одиночного счета фотонов. Проведен анализ временных характеристик разработанного алгоритма синхронизации и представлено аналитическое выражение для расчета временной задержки детектирования, которое обеспечивает последовательный анализ временных окон с учетом перестраиваемого времени восстановления фотодетектора. Аналитическое выражение может использоваться для инженерных расчетов при проектировании системы КРК. Предложенный алгоритм значительно снижает вероятность несанкционированного доступа к процессу синхронизации и позволяет с заданной точностью определить временные параметры сигнального окна, что является критически важным для последующей работы системы квантового распределения ключей.

Исследование выполнено за счет гранта Российского научного фонда № 25-29-00007, <https://rscf.ru/project/25-29-00007/>

Литература

1. Subramani S., Svn S. K. Review of security methods based on classical cryptography and quantum cryptography // *Cybernetics and Systems*. – 2025. – Т. 56. – №. 3. – P. 302–320. DOI: <https://doi.org/10.1080/01969722.2023.2166261>.
2. Portmann C., Renner R. Security in quantum cryptography // *Reviews of Modern Physics*. – 2022. – Т. 94. – №. 2. – С. 025008. DOI: <https://doi.org/10.1103/RevModPhys.94.025008>.
3. Grasselli F. Quantum cryptography // *Quantum science and technology*. Cham: Springer. – 2021. DOI: 10.1007/978-3-030-64360-7.
4. Kumari A. B. et al. One time pad encryption technique in cryptography // *International Journal of Computational Learning & Intelligence*. – 2023. – Т. 2. – №. 1. – P. 1–7.
5. Al-Smadi A. M. et al. Files cryptography based on one-time pad algorithm // *International Journal of Electrical and Computer Engineering (IJECE)*. – 2021. – Т. 11. DOI: 10.11591/ijece.v11i3.pp2335-2342.
6. Паршуткин А. В. Повышение защищенности информации от утечки через побочные электромагнитные излучения / А. В. Паршуткин, М. Р. Неаскина // *Вопросы кибербезопасности*. – 2022. – № 3(49). – С. 82–89. DOI 10.21681/2311-3456-2022-3-82-89.
7. Chen Y. A. An integrated space-to-ground quantum communication network over 4,600 kilometres // *Nature*. – 2021. – Vol. 589, No. 7841. – P. 214–219. DOI: <https://doi.org/10.1038/s41586-020-03093-8>.
8. Деев А. Д. Квантовые коммуникации через атмосферные (космические) каналы связи / А. Д. Деев, А. А. Калинин, С. П. Кулик // *Интернет изнутри*. – 2024. – № 20. – С. 43–47.

9. Зякин Е. В. Перспективные протоколы КРК для оптической связи в свободном пространстве / Е. В. Зякин, А. В. Молоканов, К. М. Чуриков // Новые технологии. Наука, техника, педагогика. – 2024. – С. 141–148.
10. Петренко, А. С. Метод построения постквантовых алгоритмов ЭЦП с двумя скрытыми группами / А. С. Петренко // Вопросы кибербезопасности. – 2025. – № 2(66). – С. 52–63. DOI 10.21681/2311-3456-2025-2-52-63.
11. Nadlinger D. P. et al. Experimental quantum key distribution certified by Bell's theorem // Nature. – 2022. – Т. 607. – №. 7920. – С. 682–686. DOI: <https://doi.org/10.1038/s41586-022-04941-5>.
12. Lin D. High performance frame synchronization for continuous variable quantum key distribution systems // Optics Express. – 2015. – Vol. 23, No. 17. – P. 22190–22198. DOI: <https://doi.org/10.1364/OE.23.022190>.
13. Calderaro L. et al. Fast and simple qubit-based synchronization for quantum key distribution // Physical Review Applied. – 2020. – Т. 13. – №. 5. – С. 054041. DOI: <https://doi.org/10.1103/PhysRevApplied.13.054041>.
14. Williams J. et al. Implementation of quantum key distribution and quantum clock synchronization via time bin encoding // Quantum Computing, Communication, and Simulation. – SPIE, 2021. – Т. 11699. – P. 16–25. DOI: <https://doi.org/10.1117/12.2581862>.
15. Cochran R. D., Gauthier D. J. Qubit-based clock synchronization for QKD systems using a Bayesian approach // Entropy. – 2021. – Т. 23. – №. 8. – P. 988. DOI: <https://doi.org/10.3390/e23080988>.
16. Nonclassical attack on a quantum keydistribution system / A. Pljonkin, D. Petrov, L. Sabantina, K. Dakhkilgova // Entropy. – 2021. – Vol. 23, No. 5.
17. Сабанов А. Г. Идентификация и аутентификация в цифровом мире / А. Г. Сабанов, А. А. Шелупанов. – М.: Горячая Линия – Телеком. – 2022.
18. Civelli S. et al. Optical identification for user authentication in quantum key distribution systems // IET Conference Proceedings CP839. – Stevenage, UK : The Institution of Engineering and Technology, 2023. – Т. 2023. – №. 34. – P. 815–818. DOI: <https://doi.org/10.1049/icp.2023.2346>.
19. Krawec W. O. Security of a High Dimensional Two-Way Quantum Key Distribution Protocol // Advanced Quantum Technologies. – 2022. – Т. 5. – №. 10. – С. 2200024. DOI: <https://doi.org/10.1002/qute.202200024>.
20. Zheng X., Zhao Z. Quantum key distribution with two-way authentication // Optical and Quantum Electronics. – 2021. – Т. 53. – №. 6. – P. 304. DOI: <https://doi.org/10.1007/s11082-021-02845-8>.
21. Патент 2667755 РФ, МПК H04L9/08. Система релятивистской квантовой криптографии / Кравцов К.С. и др. (РФ). – № 2017117184; заявл. 05.17.2017; опубл. 24.09.2024.
22. Lavie E., Lim C. C. W. Improved coherent one-way quantum key distribution for high-loss channels // Physical Review Applied. – 2022. – Т. 18. – №. 6. – С. 064053. DOI: <https://doi.org/10.1103/PhysRevApplied.18.064053>.
23. Pljonkin A. et al. The Study of Synchronization in Quantum Key Distribution System // Futuristic Trends in Network and Communication Technologies: Third International Conference. – Springer Singapore, 2021. – P. 68–80. DOI: https://doi.org/10.1007/978-981-16-1483-5_7.
24. Румянцев К. Е. Вероятностные характеристики алгоритма обнаружения синхросигналов на основе выбора смежной пары сегментов с максимальным суммарным отсчётом / К. Е. Румянцев, П. Д. Миронова // Известия ЮФУ. Технические науки. – 2023. – № 3 (233). – С. 96–107.
25. Миронова П. Д. Алгоритм обнаружения синхросигналов на основе выбора смежной пары сегментов с максимальным суммарным отсчётом // Фундаментальные и прикладные аспекты компьютерных технологий и информационной безопасности: Сборник статей Всерос. науч.-техн. конф. – Таганрог: ЮФУ, 2023. – P. 52–53.
26. Signorelli F. et al. InGaAs/InP SPAD detecting single photons at 1550 nm with up to 50 % efficiency and low noise // 2021 IEEE International Electron Devices Meeting (IEDM). – IEEE, 2021. – P. 20.3. 1–20.3. 4. DOI: 10.1109/IEDM19574.2021.9720559.

ALGORITHM FOR DETECTING SYNCHRONIZATION SIGNAL IN QUANTUM NETWORKS

Pljonkin A. P.³

Keywords: security, synchronization, quantum distribution, single-photon, optical pulse.

Purpose of the study: development and research of an algorithm for detecting an optical signal for synchronizing stations of a quantum key distribution system with increased protection against unauthorized access.

Methods of research: probability distribution, statistical analysis, single-photon detection.

Results: the importance of the time synchronization process, which is implemented through high-precision detection of an optical signal, is substantiated. An optical signal detection algorithm for a synchronization method with increased protection against unauthorized access is investigated. An improved algorithm for analyzing the time domain of synchronization signal propagation using avalanche photodetectors is proposed. It is shown that the proposed algorithm allows using avalanche photodetectors in the single photon counting mode. An analysis of the time characteristics of the developed synchronization algorithm is carried out and an analytical expression for calculating the time delay of detection is presented, which provides a sequential analysis of time windows taking into account the tunable recovery time of the photodetector. The analytical expression can be used for engineering calculations when designing a QKD system. The proposed algorithm significantly reduces the likelihood of unauthorized access to the synchronization process and allows determining the time parameters of the signal window with a given accuracy, which is critical for the subsequent operation of the quantum key distribution system.

3 Anton P. Pljonkin, Ph.D. of Technical Sciences, Associate Professor, Southern Federal University. Taganrog, Russia. E-mail: pljonkin@sfsedu.ru

Scientific novelty: an algorithm for detecting an optical signal during synchronization is proposed, which is characterized by increased protection against unauthorized access. An analytical expression for engineering calculations of the delay of detection during synchronization is presented.

References

1. Subramani S., Svn S. K. Review of security methods based on classical cryptography and quantum cryptography // Cybernetics and Systems. – 2025. – T. 56. – №. 3. – P. 302-320. DOI: <https://doi.org/10.1080/01969722.2023.2166261>.
2. Portmann C., Renner R. Security in quantum cryptography // Reviews of Modern Physics. – 2022. – T. 94. – №. 2. – C. 025008. DOI: <https://doi.org/10.1103/RevModPhys.94.025008>.
3. Grasselli F. Quantum cryptography // Quantum science and technology. Cham: Springer. – 2021. DOI: 10.1007/978-3-030-64360-7.
4. Kumari A. B. et al. One time pad encryption technique in cryptography // International Journal of Computational Learning & Intelligence. – 2023. – T. 2. – №. 1. – P. 1–7.
5. Al-Smadi A. M. et al. Files cryptography based on one-time pad algorithm // International Journal of Electrical and Computer Engineering (IJECE). – 2021. – T. 11. DOI: 10.11591/ijece.v11i3.pp2335-2342.
6. Parshutkin A., Neaskina M. Increasing the security of information from leakage through side electromagnetic emissions / Voprosy kiberbezopasnosti. – 2022. – № 3(49). – P. 82–89. DOI 10.21681/2311-3456-2022-3-82-89.
7. Chen Y. A. An integrated space-to-ground quantum communication network over 4,600 kilometres // Nature. – 2021. – Vol. 589, No. 7841. – P. 214–219. DOI: <https://doi.org/10.1038/s41586-020-03093-8>.
8. A. D. Deev, A. A. Kalinkin, S. P. Kulik. Kvantovye kommunikacii cherez atmosferynye (kosmicheskie) kanaly svyazi // Internet iznutri. – 2024. – № 20. – Pp. 43–47.
9. E. V. Zyakin, A. V. Molokanov, K. M. Churikov. Promising QKD protocols for optical communications in free space // New Technologies. Science, Engineering, Pedagogics: Proceedings of the All-Russian Scientific-Practical Conference, Moscow, 2024, pp. 141–148.
10. Petrenko A. S. Method for constructing post-quantum algorithms of eds with two hidden groups / Voprosy kiberbezopasnosti. – 2025. – № 2(66). – P. 52–63. DOI 10.21681/2311-3456-2025-2-52-63.
11. Nadlinger D. P. et al. Experimental quantum key distribution certified by Bell's theorem // Nature. – 2022. – T. 607. – №. 7920. – P. 682–686. DOI: <https://doi.org/10.1038/s41586-022-04941-5>.
12. Lin D. High performance frame synchronization for continuous variable quantum key distribution systems // Optics Express. – 2015. – Vol. 23, No. 17. – P. 22190–22198. DOI: <https://doi.org/10.1364/OE.23.022190>.
13. Calderaro L. et al. Fast and simple qubit-based synchronization for quantum key distribution // Physical Review Applied. – 2020. – T. 13. – №. 5. – P. 054041. DOI: <https://doi.org/10.1103/PhysRevApplied.13.054041>.
14. Williams J. et al. Implementation of quantum key distribution and quantum clock synchronization via time bin encoding // Quantum Computing, Communication, and Simulation. – SPIE, 2021. – T. 11699. – P. 16–25. DOI: <https://doi.org/10.1117/12.2581862>.
15. Cochran R. D., Gauthier D. J. Qubit-based clock synchronization for QKD systems using a Bayesian approach // Entropy. – 2021. – T. 23. – №. 8. – P. 988. DOI: <https://doi.org/10.3390/e23080988>.
16. Nonclassical attack on a quantum keydistribution system / A. Pljonkin, D. Petrov, L. Sabantina, K. Dakhkilgova // Entropy. – 2021. – Vol. 23, No. 5.
17. Sabanov A. G., Shelupanov A. A. Identification and authentication in the digital world. Moscow, Hot Line-Telecom, 2022.
18. Civelli S. et al. Optical identification for user authentication in quantum key distribution systems // IET Conference Proceedings CP839. – Stevenage, UK : The Institution of Engineering and Technology, 2023. – T. 2023. – №. 34. – P. 815–818. DOI: <https://doi.org/10.1049/icp.2023.2346>.
19. Krawec W. O. Security of a High Dimensional Two-Way Quantum Key Distribution Protocol // Advanced Quantum Technologies. – 2022. – T. 5. – №. 10. – C. 2200024. DOI: <https://doi.org/10.1002/qute.202200024>.
20. Zheng X., Zhao Z. Quantum key distribution with two-way authentication // Optical and Quantum Electronics. – 2021. – T. 53. – №. 6. – P. 304. DOI: <https://doi.org/10.1007/s11082-021-02845-8>.
21. Pat. 2667755 RF, MPK H04L9/08. Sistema relyativistskoj kvantovoj kriptografii / Kravcov K. S. i dr. (RF). – № 2017117184; yayavl. 05.17.2017; opubl. 24.09.2024.
22. Lavie E., Lim C. C. W. Improved coherent one-way quantum key distribution for high-loss channels // Physical Review Applied. – 2022. – T. 18. – №. 6. – C. 064053. DOI: <https://doi.org/10.1103/PhysRevApplied.18.064053>.
23. Pljonkin A. et al. The Study of Synchronization in Quantum Key Distribution System // Futuristic Trends in Network and Communication Technologies: Third International Conference. – Springer Singapore, 2021. – 3. 68–80. DOI: https://doi.org/10.1007/978-981-16-1483-5_7.
24. Romyancev K. E. Veroyatnostnye xarakteristiki algoritma obnaruzheniya sinxrosignalov na osnove vybora smezhnoj pary segmentov s maksimal'nym summarnym otschyotom / K. E. Romyancev, P. D. Mironova // Izvestiya YuFU. Texnicheskie nauki. – 2023. – № 3 (233). – P. 96–107.
25. Mironova P. D. Algoritm obnaruzheniya sinxrosignalov na osnove vybora smezhnoj pary segmentov s maksimal'nym summarnym otschyotom // Fundamental'nye i prikladnye aspekty komp'yuternyx tehnologij i informacionnoj bezopasnosti: Sbornik statej Vseros. nauch.-texn. konf. – Taganrog: YuFU, 2023. – P. 52–53.
26. Signorelli F. et al. InGaAs/InP SPAD detecting single photons at 1550 nm with up to 50 % efficiency and low noise // 2021 IEEE International Electron Devices Meeting (IEDM). – IEEE, 2021. – P. 20.3. 1–20.3. 4. DOI: 10.1109/IEDM19574.2021.9720559.



СХЕМА КРИПТОГРАФИЧЕСКОЙ ЗАЩИТЫ ДАННЫХ ОБЛАЧНЫХ ХРАНИЛИЩ С ВОЗМОЖНОСТЬЮ РАСШИРЕНИЯ КВАНТОВОЗАВИСИМЫМИ КЛЮЧАМИ И ПОСТКВАНТОВЫМИ АЛГОРИТМАМИ

Минаков С. С.¹, Тихов С. В.², Цупак А. А.³

DOI: 10.21681/2311-3456-2025-6-88-100

Цель исследования: разработка схемы криптографической защиты информации в облачных хранилищах с использованием стандартизированных и перспективных режимов блочных шифров, постквантовых алгоритмов шифрования и квантовозависимых ключей.

Метод(ы) исследования: системный анализ угроз безопасности информации при её обработке и хранении с использованием прикладных протоколов доступа к облачным хранилищам. Синтез криптографических механизмов и форматов для построения автоматизированной гибридной криптографической схемы обработки информации на стороне потребителя облачных услуг с использованием симметричных и асимметричных шифров.

Результат(ы) исследования: представлено развитие криптографической схемы «Утро-1» для обеспечения защиты информации в прикладных протоколах доступа облачных хранилищ. Описаны форматы, функции и логика шифрования, проведены практические испытания программной реализации. Даны пояснения по условиям использования в криптографической схеме постквантовых алгоритмов и/или квантовозависимых ключей.

Научная новизна: получены новые научно-технические результаты в области защиты данных в облачных хранилищах с использованием российских и перспективных иностранных криптографических средств и методов; практическая ценность состоит в развитии криптографической схемы защиты информации от НСД с использованием гибридной композиции симметричных и асимметричных шифров.

Ключевые слова: шифрование, облачный сервис, сетевой доступ, прикладной протокол, безопасность информации, выработка ключей.⁴⁵⁶⁷

Введение

Современный уровень и задачи развития систем информатизации, цифровая трансформация государственного управления привели к переводу на автоматизированную и интеллектуальную обработку большинства процессов управления и обработки информации в госуправлении, негосударственных организациях, на предприятиях промышленности и транспорта. Из таких систем сложилась и нормативно закреплена сфера критической информационной инфраструктуры (КИИ).

При этом значительный пласт автоматизированных и информационных систем, в том числе и информационные системы персональных данных, теперь разворачиваются в сторонних центрах обработки и хранения данных (далее – ЦОД). Более того, в рамках реализации нескольких государственных программ Правительством Российской Федерации одобрены концепции перевода государственных

информационных ресурсов в ЦОДы⁴ и развёртывания государственной единой облачной платформы (платформа ГЕОП «Гособлако»), с 2019 г. начаты соответствующие эксперименты⁵ и к 2021 г. создана единая цифровая платформа Российской Федерации «ГосТех» (платформа ГосТех^{6,7}).

При использовании сторонних для защищаемой системы сервисов облачных вычислений существенно изменяется модель угроз и нарушителя, во многом утрачивается подконтрольность обрабатываемой и хранимой информации, что зачастую не позволяет наделять сторону облачных сервисов требуемыми гарантиями доверия и считать достаточными рубежи защиты только на базе криптографических протоколов взаимодействия с облачным хранилищем (например, протоколы TLS, IPSec).

Таким образом, необходимо при построении сегментов системы с различным уровнем доверия

1 Минаков Сергей Сергеевич, старший научный сотрудник, ФГКНУ «Академия криптографии Российской Федерации». г. Москва, Россия. E-mail: ss_minakov@mail.ru

2 Тихов Станислав Вячеславович, ведущий специалист ООО НТП «Криптософт». г. Пенза, Россия. E-mail: tik.stanislav2015@yandex.ru

3 Цупак Алексей Александрович, доктор физико-математических наук, доцент, доцент кафедры «Математика и суперкомпьютерное моделирование» ФГБОУ «Пензенский государственный университет». г. Пенза, Россия. E-mail: altsupak@yandex.ru

4 Распоряжение Правительства Российской Федерации от 07.10.2015 № 1995-р (ред. от 18.10.2018).

5 Постановление Правительства РФ от 28.08.2019 № 1114.

6 Распоряжение Правительства РФ от 21.10.2022 № 3102-р.

7 «Концепция обеспечения информационной безопасности единой цифровой платформы Российской Федерации «ГосТех» (утв. приказом Минцифры России от 12.01.2023, № 7).

и при наличии сторонних облачных сервисов переходить от криптографического протокола к криптографической схеме [1], в которой сторона, реализующая техническую службу облачного хранилища, не может гарантировать конфиденциальность и целостность данных пользователя, обеспечивать подтверждение корректности реализаций механизмов обеспечения защиты информации при реализации облачных вычислений.

Рассмотрим вариант криптографической защиты с использованием прикладных протоколов доступа и развитие криптографической схемы «Утро-1» [1], уточним криптографические механизмы защиты данных, передаваемых в облачные сервисы категории Data Storage as a Service (далее – STaaS / DsaaS), для повышения эффективности защиты информации при использовании облачных вычислений и стандартизированных интерфейсов и протоколов управления облачными данными CDMI (англ. – Cloud Data Management Interface), включая прикладные протоколы AWS Simple Storage Service (Amazon Web Services S3 protocol), WebDAV и *Calendaring Extensions to WebDAV (CalDAV protocol)* для доступа к ресурсам облачных хранилищ [2].

Внедрение новых суперкомпьютеров и квантовых вычислителей [3–5] позволяют говорить о конкретной технической реализации эффективных квантовых алгоритмов на квантовых компьютерах или массовой эмуляции их при распределенных вычислениях [6]. В данной статье авторы рассматривают два подхода к противодействию «квантовой угрозе» (атаки на асимметричные и симметричные блочные алгоритмы шифрования) в рамках задачи о защищенном облачном хранении данных: использование квантовозависимой ключевой информации, полученной с использованием систем квантового распределения ключей и использование методов «постквантовой» криптографии наряду с классическими стандартизованными блочными шифрами.

Структуры данных в прикладных протоколах доступа к облачным хранилищам файлов

Рассмотрим структуры данных в трех типовых прикладных протоколах доступа к облачному хранилищу WebDAV, S3 и CalDAV. Отметим, что ни WebDAV, ни S3, ни CalDAV не являются полноценными сетевыми протоколами, а скорее представляют собой наборы расширений и дополнений к протоколу HTTP. Тем не менее, опуская тонкости терминологии, будем использовать слово «протокол» (наряду с «технологией» и «интерфейсом») применительно к этим трем технологиям доступа к облачному хранилищу.

Являясь расширением протокола прикладного уровня HTTP, протоколы доступа WebDAV, AWS S3, CalDAV имеют много общего в части базовой технологии передачи и получения данных. В то же время,

они различаются между собой в части используемых методов и структур данных в HTTP-запросах.

В облачных сервисах хранения данных **технология WebDAV** используется в качестве сетевой файловой системы, а также в качестве протокола для удаленного доступа к ее объектам (файл-контейнерам) и управления ими. К основным возможностям WebDAV относят: механизм блокировок на доступ к объектам хранилища, средства для создания, редактирования и удаления свойств (метаданных) объектов, а также использование именованных областей, представляющих коллекцию объектов и выполняющих функцию, аналогичную каталогам в файловой системе.

Согласно спецификации RFC 4918 для удаленного доступа и управления объектами хранилища протокол WebDAV предоставляет «унаследованные» от HTTP команды PUT, GET и DELETE, а также семь новых команд: PROPFIND, PROPPATCH, MKCOL, COPY, MOVE, LOCK и UNLOCK. Команды PUT и GET применяются, соответственно, для записи и чтения данных в/из файл-контейнер(а). Остальные команды предназначены для управления уже размещенными в хранилище объектами (переименование или перемещение, разграничение прав доступа и др.). Таким образом, потенциально уязвимые (содержащие конфиденциальные сведения) данные пользователя передаются только с командами PUT и GET.

Запрос на запись, например, текстовой строки «Hello world» (в кодировке UTF-8) в файл-контейнер *Obj* облачного хранилища имеет вид

```
PUT /Obj HTTP/1.1
Host: www.example.com
Content-Type: text/plain
Content-Length: 12
Hello world
```

Первые четыре строки – это типовые заголовки HTTP-запроса, а последняя строка представляет собой тело запроса с данными для записи в файл-контейнер *Obj*. При успешном выполнении запроса сервер возвращает ответ вида

```
HTTP/1.1 201 Created
Content-Length: 12
Date: Sat, 1 May 2025 09:32:12 GMT
```

Запрос на чтение данных из файл-контейнера *Obj* имеет вид

```
GET /Obj HTTP/1.1
Host: www.example.com
Accept: */*
```

При успешном выполнении запроса сервер отвечает:

```
HTTP/1.1 200 OK
Content-Type: text/plain;
charset=utf-8
Content-Length: 12
Hello world
```

Для успешного выполнения представленных выше запросов, последние должны включать некоторые дополнительные заголовки. К их числу относится заголовок *Authorization*, содержащий информацию для аутентификации стороны, отправляющей запрос облачному сервису.

В представленных примерах тела запросов содержат текстовые данные. В общем же случае они могут содержать произвольные бинарные данные, в частности, любые датаграммы, инкапсулирующие данные пользователя. Это позволяет реализовать «защищенную версию» WebDAV (с размещением данных в файл-контейнерах в зашифрованном виде) без внесения каких-либо изменений или дополнений в оригинальный протокол доступа к облачному хранилищу.

Протокол S3 предназначен для доступа к облачным хранилищам, построенным по одноименной технологии, – так называемым, *S3-хранилищам*. Последние представляют собой плоские (не иерархические) хранилища, где все объекты являются равнозначными и другой, встроенной системы разграничения доступа между ними нет.

Каждый объект в S3-хранилище состоит из трех компонент: уникального идентификатора (строка, служащая для однозначной идентификации объекта и прямого доступа к нему по URL-адресу), метаданных (дополнительные атрибуты объекта, такие, как размер, тип файла и др.) и содержимого. Для удобной работы объекты хранилища можно объединить в так называемые *бакеты* (англ. *buckets*). Бакет получает уникальный идентификатор, входящий в URL каждого объекта, включенного в бакет.

Для управления объектами хранилища протокол S3 предоставляет обширный интерфейс, включающий несколько сотен различных операций (HTTP-запросов). Как и в случае с протоколом WebDAV, конфиденциальные данные пользователя передаются в S3-хранилище лишь в запросах на запись (*PutObject*) и чтение (*GetObject*) данных в/из объекта хранилища, тогда как остальные запросы оперируют либо со служебными, либо с уже размещенными в хранилище данными.

Авторы не приводят здесь форматную структуру запросов *PutObject* и *GetObject*, поскольку они полностью совпадают с форматами запросов *PUT* и *GET* протокола WebDAV, соответственно, за исключением набора HTTP-заголовков, необходимых для успешного выполнения запроса.

Протокол CalDAV определяет стандартный способ доступа, управления и обмена информацией о календарях и расписаниях в формате *iCalendar*. В отличие от протоколов WebDAV и S3, позволяющих размещать в хранилище произвольные бинарные

данные пользователя, протокол CalDAV предназначен для работы с информацией, имеющей специальный формат, определенный в RFC 5545.

Данные о календарях и расписаниях представляются объектом *iCalendar* вида:

```
BEGIN:VCALENDAR
icalbody
END:VCALENDAR
```

Первая и последняя строки являются константными и указывают на начало и конец объекта *iCalendar*. Тело объекта *icalbody* состоит из последовательности свойств календаря и одного или нескольких компонент календаря. Свойства календаря – это атрибуты, которые применяются к объекту календаря в целом. *Компоненты календаря* – это наборы свойств, которые выражают определенную семантику календаря. Так, компоненты календаря могут определять событие, задачу, запись в журнале и информацию о свободном/занятом времени.

Объект *iCalendar* организован в виде отдельных строк, называемых в соответствии с RFC 5545 *контентными строками* (англ. *content lines*), длиной не более 75 байт, разделенных последовательностью символов CRLF. Каждая такая строка соответствует тому или иному свойству календаря или компоненты календаря и имеет вид «name:value», где *name* – название свойства, а *value* – его значение. Ниже представлен пример объекта *iCalendar* из документа RFC 5545:

```
BEGIN:VCALENDAR
VERSION:2.0
PRODID://CalDav client//EN
BEGIN:VEVENT
UID:123
DTSTART:20250101T000000Z
DTEND:20250101T010000Z
SUMMARY: New Year Party
END:VEVENT
END:VCALENDAR
```

Вторая и третья строки – это свойства календаря: *VERSION* указывает на версию спецификации формата, *PRODID* указывает идентификатор приложения, генерирующего данные календаря. Строки с четвертой по девятую соответствуют компоненте календаря *VEVENT*, представляющей событие с четырьмя свойствами: *UID* – его уникальный идентификатор, *DTSTART* и *DTEND* указывают дату и время начала и окончания события, соответственно, *SUMMARY* содержит описание события.

Свойства объектов *iCalendar* строго типизированы. Спецификация RFC 5545 определяет четырнадцать типов значений, в том числе бинарные, текстовые и числовые типы данных, адреса пользователей календаря, несколько типов данных для хранения даты и времени и др. Так, для идентификации значений,

содержащих точную календарную дату и время суток, используется, как правило, тип DATE-TIME. Формат значений DATE-TIME основан на представлении дат и времени, описанном в стандарте ГОСТ Р 7.0.64-2018 (ISO 8601:2004). В примере выше свойства DSTART и DTEND хранят данные типа DATE-TIME.

Строгая типизированность свойств объектов iCalendar делает невозможным применение к некоторым из них криптографических преобразований; это справедливо, например, для свойств, содержащих дату и время, периодичность наступления события (тип данных RECUR). Предлагаемые в данной работе механизмы криптографической защиты данных календаря учитывают эти ограничения.

Рассмотрим протокол CalDAV. Как следует из его названия, CalDAV является производным от WebDAV и для хранения объектов iCalendar использует иерархическую структуру DAV-хранилища. Согласно спецификации RFC 4791, календарь CalDAV представляет собой DAV-коллекцию (каталог) с определенной структурой; каждая такая коллекция-календарь включает набор так называемых *ресурсов объекта календаря* (англ. *calendar object resources*) – файл-контейнеров, которые содержат свойства (набор контентных строк) либо всего календаря, либо одной из его компонент. При этом CalDAV, как протокол доступа, инкапсулирует организацию данных в хранилище и предоставляет инструменты для доступа и управления данными календаря на уровне объектов iCalendar, а не отдельных файлов-ресурсов объектов iCalendar. Так, например, запрос на добавление в календарь нового события имеет вид

```
PUT /calendars/events/
    /qwue23489.ics HTTP/1.1
Host:www.example.com
Content-Type: text/calendar
Content-Length: 234
BEGIN:VCALENDAR
VERSION:2.0
PRODID:-//CalDAV client//EN
BEGIN:VEVENT
UID:123
DTSTAMP:20241111T000000Z
DSTART:20250101T000000Z
DTEND:20250101T010000Z
SUMMARY: New Year Party
END:VEVENT
END:VCALENDAR
```

Как видно из этого примера, тело запроса представляет объект iCalendar. В случае успешного выполнения запроса служба облачного сервиса CalDAV создает в каталоге-календаре файл-контейнер и записывает в него свойства новой компоненты (события) календаря.

Построение гибридной криптографической схемы с использованием сертификатов безопасности

Предлагаемая авторами криптографическая схема (КС) защиты данных, размещаемых в облачных хранилищах, является гибридной в том смысле, что она включает как композиции симметричных алгоритмов шифрования для выполнения преобразований и имитозащиты данных, так и механизмы асимметричной криптографии для защищенного обмена ключевым материалом [2]. В части вариантов реализации симметричной криптографии на основе блочных шифров предлагаемая криптосхема во многом наследует КС «Утро» (вариант «Утро-1»), подробно изложенной в работе [1].

Композиция алгоритмов асимметричных криптографических механизмов в схеме авторов работы строится вокруг протокола Диффи-Хеллмана, но реализованного на эллиптических кривых.

Для изложения предлагаемой криптографической схемы удобно рассмотреть типовой сценарий использования облачного хранилища: сторона А обменивается данными со стороной Б, для чего создает в облачном хранилище новый файл-контейнер *Obj* и размещает в нем данные в защищенном виде.

Предполагается, что сторона А на момент размещения данных в облачном хранилище имеет валидный (верифицированный доверенной третьей стороной) сертификат безопасности стороны Б, содержащий открытый ключ $u \cdot P$ последней.

Размещение данных в файл-контейнере облачного хранилища

Шаг 1. Для каждого нового файл-контейнера облачного хранилища данных и, в частности, файл-контейнера для *Obj*, генерируется пара 256-битных ключей: мастер-ключ K_E для выполнения криптографических преобразований данных и мастер-ключ K_M для выполнения имитозащиты данных.

Шаг 2. При помощи секретного ключа x стороны А и открытого ключа $u \cdot P$ стороны Б формируется ключ шифрования KEK_{VKO} в соответствии с выбранным алгоритмом согласования ключей VKO .

Шаг 3. Используя полученный на шаге 2 ключ шифрования KEK_{VKO} , осуществляется процедура экспорта материала ключей K_E и K_M согласно выбранному алгоритму экспорта ключей. В результате получается зашифрованный материал ключей K_E и K_M , а также значения имитозащитных вставок $MAC(K_E)$, $MAC(K_M)$, которые затем используются в процедуре импорта для проверки корректности восстановления (расшифрования) ключевого материала.

Шаг 4. Размещаемый на облаке файл-контейнер *Obj* представляется в виде набора из N логических последовательно нумерованных равных (возможно, за исключением последнего) фрагментов данных. Пусть Obj_k обозначает k -ый фрагмент файл-контейнера

Obj. При этом размер отдельного фрагмента данных определяется требованиями, предъявляемыми к максимальному объему информации, который может быть обработан на одном ключе. На основе мастер-ключей K_E и K_M формируются два множества (базиса) производных ключей $\{K_E^{kl}\}_{k=1}^N$ и $\{K_M^{kl}\}_{k=1}^N$, соответственно. Ключ K_E^k предназначен для выполнения криптографических преобразований фрагмента Obj_k . Ключ K_M^k предназначен для вычисления значения имитозащитной вставки фрагмента Obj_k .

Шаг 5. Сторона А зашифровывает каждый фрагмент Obj_k , $k = (1, N)$ – выбранным алгоритмом блочного шифрования на соответствующем ключе из базиса $\{K_E^{kl}\}_{k=1}^N$ и вычисляет значение имитозащитной вставки $MAC(Obj_k)$ на соответствующем ключе из базиса $\{K_M^{kl}\}_{k=1}^N$.

Шаг 6. Сторона А последовательно загружает в облачное хранилище зашифрованные фрагменты Obj_k , а также значения $MAC(Obj_k)$. Формат файл-контейнера с защищенными данными обсуждается ниже.

Шаг 7. Полученные на шаге 3 экспортированные (зашифрованные) материалы мастер-ключей K_E и K_M и значения $MAC(K_E)$, $MAC(K_M)$ также загружаются стороной А в облачное хранилище данных. Способы хранения указанных данных в облачном хранилище обсуждаются ниже.

Получение данных из файл-контейнера облачного хранилища

Шаг 1. Сторона Б загружает из облачного хранилища экспортированные (зашифрованные) стороной А ключи K_E и K_M и значения $MAC(K_E)$, $MAC(K_M)$.

Шаг 2. При помощи секретного ключа у стороны Б и открытого ключа $x \cdot P$ стороны А формируется ключ шифрования KEK_{VKO} , идентичный тому, что был получен на шаге 2 схемы размещения данных в файл-контейнере *Obj*.

Шаг 3. Используя KEK_{VKO} , осуществляется процедура импорта ключей K_E и K_M согласно выбранному алгоритму экспорта/импорта ключей. В частности, выполняется проверка совпадения вычисленных на ключе KEK_{VKO} (либо производном от KEK_{VKO} ключе) имитовставок восстановленных материалов ключей K_E и K_M и имитовставок $MAC(K_E)$, $MAC(K_M)$, загруженных на шаге 1 из облака.

Шаг 4. Сторона Б последовательно загружает из облачного хранилища зашифрованные фрагменты Obj_k файл-контейнера *Obj*, а также значения имитозащитных вставок $MAC(Obj_k)$.

Шаг 5. Расшифрование фрагмента Obj_k производится в обратном порядке только после успешной проверки совпадения вычисленной на соответствующем ключе имитовставки такого фрагмента Obj_k и имитовставки, полученной из облачного хранилища на шаге 4.

Используемые в схеме криптографические функции

Для выработки материала K_E и K_M в качестве источников псевдослучайных последовательностей R можно использовать псевдослучайную функцию PRF_TLS_GOSTR3411_2012_256 с длиной выхода 256 бит, определенную в рекомендациях по стандартизации Р 50.1.113-2016, или алгоритм выработки псевдослучайной последовательности R длины 256 бит, определенный в рекомендациях по стандартизации Р 1323565.1.006-2017. При этом энтропийные данные, передаваемые в эти функции в качестве параметров, предлагается получать при помощи аппаратного ДСЧ, биологического ДСЧ, или квантового ДСЧ.

В качестве секретного и открытого ключей, применяемых в изложенной выше схеме в части защищенного обмена материалами ключей K_E и K_M между несколькими пользователями, могут использоваться ключи подписи и проверки подписи согласно алгоритму ГОСТ Р 34.10-2018 с параметрами эллиптических кривых, определенными в Р 50.1.114-2016.

Для получения одинакового для сторон А и Б ключа шифрования KEK_{VKO} используется алгоритм согласования ключей VKO_GOSTR3410_2012_256, описанный в Р 50.1.113-2016, при этом параметр UKM алгоритма должен иметь длину не менее 64 бит ввиду того, что секретный и открытый ключи участников обмена являются длительными периодическими.

Для криптографических преобразований и имитозащиты материалов ключей K_E и K_M могут использоваться алгоритм экспорта/импорта ключей, описанный в Р 50.1.113-2016, или алгоритмы экспорта KExp15 и импорта KImp15, определенные в Р 1323565.1.017 – 2018.

Код аутентификации сообщения используется для решения задач имитозащиты передаваемых в облачное хранилище данных и должен формироваться одним из режимов алгоритма блочного шифрования $E_k^{mode}(T)$, например, ключевой функцией хеширования OMAC-ASPRKM, определенной в ГОСТ 34.13-2018, либо функцией хеширования HMAC, определенной в Р 50.1.113-2016. Значения $MAC_k(Obj_k)$ разных фрагментов файл-контейнера должны вычисляться на разных ключах K_M^k , производных от мастер-ключа K_M , причем материалы ключей K_M и K_E должны быть различными. Таким образом, имитозащита данных требует увеличения вдвое размерности базиса ключей для раздельного использования ключей шифрования и ключей имитозащиты фрагментов данных в связи с существенно различной стойкостью соответствующих криптографических алгоритмов. Альтернативным решением является использование AEAD-режима работы алгоритма блочного шифрования, который обеспечивает так называемое

аутентифицируемое шифрование: шифрование и имитозащиту блока данных на одном ключе. К таким режимам относятся, например, GCM, MGM, MGM2 [7, 8].

Криптографические преобразования данных, размещаемых в файл-контейнере облачного хранилища, рекомендуется реализовать в соответствии с алгоритмами блочного шифрования «Магма» и «Кузнечик», определенными в ГОСТ Р 34.12-2018, в одном из следующих режимов работы: в режиме гаммирования с преобразованием ключа CTR-АСРКМ, определенном в ГОСТ 34.13-2018, в режиме DEC (Disk Encryption with Counter Mode), описанном в рекомендациях по стандартизации Р 1323565.1.042–2022 (некоторые примечательные свойства данного режима работы блочного шифра рассмотрены в работе [9]), или новом (находящемся в процедуре стандартизации) российском режиме блочных шифров ХЕН, перспективным для защиты данных на системных разделах и носителях и в облачных хранилищах [10]. В качестве значения вектора инициализации iv при выполнении процедур блочного шифра в соответствующих режимах применительно к k -ому фрагменту файл-контейнера можно использовать строку (ее байтовое представление), полученную в результате конкатенации $k||k$.

Важным аспектом представленной нами криптографической схемы является вычисление производных ключей для выполнения криптографических преобразований и имитозащиты отдельных фрагментов файл-контейнеров. Функция выработки производного ключа — KDF (англ. Key derivation function) должна создавать криптографически стойкие ключи для алгоритма симметричного шифрования на основе источника первоначального ключевого материала.

Если при реализации схемы шифрование и/или имитозащита данных выполняются в режимах CTR-АСРКМ и ОМАС-АСРКМ, то преобразование ключа (выработка производных ключей из мастер-ключа) является частью указанных режимов и выполняется при помощи функций АСРКМ и АСРКМ-Master с заданной частотой смены мастер-ключа, определенной в ГОСТ 34.13-2018.

В других вариантах реализации предлагаемой схемы (с использованием режимов блочного шифра ХЕН, MGM2, HMAC) в качестве функции вычисления производного ключа $KDF_{256}(T):V \rightarrow V_{256}$ можно использовать как саму функцию хеширования $H_{256}(T)$, определенную в ГОСТ Р 34.11-2012, так и её производные, например, алгоритм диверсификации KDF_TREE_GOSTR3411_2012_256, определенный в Р 50.1.113-2016. Приемлемый с точки зрения

производительности и криптографической стойкости от внешнего нарушителя алгоритм формирования производных ключей представлен в работе [1].

Зашифрованный материал ключа и параметры шифрования

Зашифрованный материал криптографического ключа, а также значение его имитозащитной вставки, традиционно хранят в шифрблоках-записях (т.н. *ключевых блоках*). Такие структуры данных помимо указанных данных, как правило, содержат описание опциональных параметров асимметричной схемы, например, идентификатор набора параметров эллиптической кривой (модуль эллиптической кривой, порядки группы и ее циклической подгруппы точек эллиптической кривой и т.д.), идентификаторы алгоритмов согласования и экспорта/импорта ключей.

Кроме того, для построения вариативной криптографической схемы, использующей разные СКЗИ, ключевой блок должен быть дополнен полями, содержащими идентификаторы СКЗИ и ключевого контейнера с секретным ключом пользователя. Это необходимо, если в системе установлены несколько СКЗИ, и имеется несколько контейнеров с секретными ключами пользователя [2].

Подчеркнем, что с каждым зашифрованным файл-контейнером облачного хранилища может быть связан целый набор ключевых блоков. Так, если в конкретной реализации рассматриваемой криптографической схемы не используются AEAD-режимы работы блочного шифра, то для выработки производных ключей шифрования и имитозащиты фрагментов данных требуются два разных мастер-ключа. В этом случае необходимо экспортировать материалы обоих этих ключей. К тому же, если доступ к зашифрованным данным должен быть предоставлен группе пользователей, то материалы мастер-ключей должны быть экспортированы на нескольких ключах, полученных при помощи алгоритма согласования с использованием открытых ключей пользователей, имеющих право доступа к зашифрованным данным, и секретного ключа владельца данных (того, кто размещает данные в файл-контейнере облачного хранилища).

Набор ключевых блоков, связанных с некоторым файл-контейнером хранилища, удобно представить в виде структуры данных типа словарь «ключ-значение». Например, пусть в файл-контейнере *Obj* размещаются данные в защищенном виде; шифрование и имитозащита выполняются алгоритмом «Магма» в режимах CTR-АСРКМ и ОМАС-АСРКМ, соответственно; при этом доступ к данным должен быть предоставлен еще двум пользователям. В табл. 1 представлен словарь с набором блоков, отвечающий описанной ситуации.

Таблица 1.

Словарь, хранящий набор связанных с файл-контейнером облачного хранилища ключевых блоков

Ключ	Значение
cek-alg	GR3412M/CTR-ACPKM
mac-alg	GR3412M/ACPKM-OMAC
key-wrap-alg	50.1.113-2016
kbl-count	6
enc-kbl-owner	...
mac-kbl-owner	...
enc-kbl-user-1	...
mac-kbl-user-1	...
enc-kbl-user-2	...
mac-kbl-user-2	...

Ключи cek-alg, mac-alg, key-wrap-alg хранят идентификаторы используемых в схеме алгоритмов (и режимов) блочного шифра, согласования и экспорта/импорта ключей. Ключ kbl-count хранит количество ключевых блоков, связанных с файл-контейнером. Ключи enc-kbl-owner, mac-kbl-owner, enc-kbl-user-1, mac-kbl-user-1, enc-kbl-user-2, mac-kbl-user-2 хранят непосредственно ключевые блоки, представленные в виде base64-строк.

Подобный словарь ключевых блоков файл-контейнера Obj можно разместить в облачном хранилище либо в метаданных файл-контейнера Obj , либо в отдельном файл-контейнере Obj_s с идентификатором, полученным при помощи конкатенации идентификатора Obj и некоторого зарезервированного константного суффикса.

Формат файл-контейнера облачного хранилища с защищенными данными

При размещении данных в некотором файл-контейнере Obj облачного хранилища, они загружаются туда пофрагментно. При этом к каждому фрагменту

Obj_k файл-контейнера Obj применяются криптографические преобразования и для каждого фрагмента рассчитывается либо имитовставка $MAC(Obj_k)$, либо тег аутентификации $TAG(Obj_k)$ (при использовании аутентифицированного шифрования). В результате для каждого фрагмента Obj_k , $k = \overline{1, N}$ – формируется датаграмма $\overline{Obj_k}$, включающая значение $MAC(Obj_k)$ (либо $TAG(Obj_k)$) и зашифрованные данные фрагмента Obj_k . Размеры полей данной датаграммы определяются используемыми в схеме алгоритмами блочного шифрования и имитозащиты. Таким образом, в облачном хранилище файл-контейнер Obj представляется последовательностью датаграмм Obj_k , $k = \overline{1, N}$. Структура такого файл-контейнера схематично представлена в табл. 2.

Размер последней датаграммы $\overline{Obj_N}$ может отличаться от размера остальных датаграмм $\overline{Obj_k}$, $k = \overline{1, N}$, поскольку объем данных, размещаемых в файл-контейнере Obj , необязательно кратен заданному размеру фрагмента и/или длине блока выбранного алгоритма блочного шифра (к таким данным применяется выравнивание в соответствии с алгоритмом, определенным в ГОСТ Р 34.12-2018).

Формат защищенных данных, размещаемых в облачном хранилище с использованием протокола CalDAV

Как уже отмечалось, протокол CalDAV предназначен для работы с информацией о календарях и расписаниях, представленной в виде объектов iCalendar. Каждый такой объект описывается набором свойств, являющихся, как правило, строго типизированными. Это накладывает серьезные ограничения на возможность применения к этим данным криптографических преобразований.

Пусть, например, пользователь размещает в календаре CalDAV информацию о некотором событии X , для чего передает в облачное хранилище объект iCalendar с компонентой VEVENT. Свойства последней, представленные в виде контентных строк, описывают событие X . Так, свойства SUMMARY,

Таблица 2.

Структура файл-контейнера с зашифрованными данными

Номер фрагмента	Тип данных	Длина в байтах
1	$MAC(Obj_1)$ или $TAG(Obj_1)$	$sizeof(MAC)$ или $sizeof(TAG)$
	Зашифрованные данные Obj_1	$sizeof(Obj_1)$
2	$MAC(Obj_2)$ или $TAG(Obj_2)$	$sizeof(MAC)$ или $sizeof(TAG)$
	Зашифрованные данные Obj_2	$sizeof(Obj_2)$
...
N	$MAC(Obj_N)$ или $TAG(Obj_N)$	$sizeof(MAC)$ или $sizeof(TAG)$
	Зашифрованные данные Obj_N	$sizeof(Obj_N)$

DESCRIPTION и LOCATION компоненты VEVENT содержат, соответственно, название, описание и место проведения события X , свойства DTSTART и DTEND содержат дату и время, соответственно, начала и завершения события X . В соответствии со спецификацией RFC 5545 первые три из перечисленных свойств хранят произвольные текстовые данные, а последние два свойства хранят данные типа DATE-TIME. Применяя классические криптографические преобразования (операции блочного шифрования) к значениям указанных свойств, мы получаем на выходе какие-то бинарные данные. Ясно, что такие данные являются допустимыми значениями для свойств SUMMARY, DESCRIPTION и LOCATION, поскольку всегда могут быть представлены в виде текста с помощью того или иного способа кодирования. В то же время очевидно, что бинарные данные, полученные в результате применения криптографических преобразований к данным типа DATE-TIME, не являются данными типа DATE-TIME и, более того, в общем случае не могут быть представлены (закодированы) как данные этого типа. Таким образом, криптографические методы защиты можно применять только к некоторой группе свойств объектов iCalendar, а для обеспечения информационной безопасности остальных нужно использовать другие (не криптографические) механизмы.

Авторы предлагают выполнять криптографические преобразования только тех значений свойств объектов iCalendar, которые хранят текстовые и бинарные данные, а в отношении других свойств поступать следующим образом: будем присваивать им «ложные» (не информативные), но допустимые с точки зрения формата значения, а подлинные (информативные) данные сохранять в зашифрованном виде в значениях свойств, к которым применялись криптографические преобразования. Так, например, при размещении в календаре CalDAV информации о новом событии, мы можем зашифровать значения свойств SUMMARY, DESCRIPTION и LOCATION компоненты календаря VEVENT, присвоить свойствам DTSTART и DTEND неактуальные значения даты и времени суток, тогда как подлинные значения в зашифрованном виде сохранить в значении свойства SUMMARY [2].

В соответствии с предлагаемой криптографической схемой параметры шифрования (в том числе экспортируемые материалы мастер-ключей шифрования и имитозащиты) объекта iCalendar также размещаются в облачном хранилище для возможности многопользовательского доступа к зашифрованным данным. Отметим, что указанные ранее способы хранения такой информации в облаке неприменимы для CalDAV. Во-первых, CalDAV не предоставляет механизмы для создания в хранилище

файл-контейнеров с произвольными пользовательскими данными. Во-вторых, CalDAV не позволяет определять собственные (не описанные в стандарте формата iCalendar) свойства для объектов iCalendar, которые можно было бы использовать для хранения служебной информации (аналогично тому, как параметры шифрования хранятся в метаданных файл-контейнеров WebDAV и S3).

В связи с этим авторы предлагают сохранять параметры шифрования объекта iCalendar в значении одного из его свойств, при этом такое значение удобно представить в виде JSON-токена с парами «ключ-значение», указанными в (табл. 1) (определяют параметры шифрования объекта iCalendar), и одним дополнительным элементом, содержащим зашифрованное значение свойства объекта iCalendar.

Условия и возможности применения постквантовых алгоритмов шифрования и(или) квантовозависимых ключей

Рассмотрим изменения криптографической схемы, предусматривающие использование квантовозависимых ключей шифрования из квантовых криптографических систем выработки и распределения ключей (ККС ВРК), а также технологические ограничения при использовании постквантовых алгоритмов шифрования.

Предлагаемая модификация криптографической схемы состоит в использовании ККС ВРК на этапе выработки и распределения «базовой» ключевой информации: ключа KEK_{VKO} (вариант 1) или мастер-ключей K_E, K_M (вариант 2).

Первый вариант предпочтителен с точки зрения скорости функционирования криптографической схемы в целом, так как выработка криптографически защищенных ключей в ККС осуществляется с невысокой скоростью относительно скорости шифрования. Такой вариант нивелирует и главную угрозу со стороны квантового вычислителя, предотвращая возможное раскрытие ключевой информации (секретных ключей x и y).

Второй вариант подразумевает выработку ключевой информации большего объема, но существенно упрощает криптографическую схему. При выработке ключей K_E, K_M с помощью ККС ВРК отпадает необходимость в шифровании, имитозащите этих ключей, а также в процедуре экспорта материала ключей. Таким образом, шаги 1–3 при размещении и получении данных заменяются одной процедурой:

Шаг 1'–3'. Совместная выработка мастер-ключей K_E, K_M в ККС ВРК.

Выбор конкретных параметров ККС ВРК влияет на скорость выработки КИ, предельно допустимое расстояние ее передачи, криптографическую стойкость ключей, возможность противодействия атакам со стороны потенциального нарушителя. Перечислим основные характерные черты возможной ККС ВРК:

реализация одно- или двухпроходной схемы передачи оптических сигналов; использование в оптической схеме одного или двух однофотонных детекторов; выбор сторонами номера оптического сигнала (в двухмодовом состоянии) для наложения фазового сдвига; выбор и, возможно, смена для разных сессий ВРК квантового протокола (BB84, SARG04, AKM2017, AKM2021, протоколов на симметричных когерентных состояниях) [11].

Одним из недостатков ККС является ограниченность расстояния передачи КИ в силу наличия оптических потерь. Выработка ключевой информации возможна на расстояниях от нескольких десятков до сотен километров, в зависимости от требований к стойкости вырабатываемых ключей. Эта проблема может быть решена, например, построением сетей ККС, обеспечивающих передачу КИ на (теоретически) неограниченные расстояния, и дающих возможность создания дублирующих линий передачи КИ для защиты от несанкционированного доступа.

Опыт разработки систем криптографической защиты информации на основе ККС ВРК^{8,9} для организации защиты объектов критической информационной инфраструктуры позволил реализовать протоколы и интерфейсы взаимодействия КС ВРК и средств криптографической защиты информации, при этом разработанная квантовая криптографическая система обеспечила выработку криптостойкой ключевой информации со средней скоростью не менее 10 бит/с при дальности передачи квантовых сигналов не менее чем на 15 км, что достаточно для размещения в оптоволоконной инфраструктуре ЦОД. При этом взаимодействующие с ККС ВРК средства криптографической защиты информации могут обеспечивать обработку данных на скорости 10 Гбит/с и выше¹⁰.

8 Ловков Д. А. Особенности применения квантового протокола выработки и распределения ключей Decoy States / Д.А. Ловков // Системы и средства защиты информации: Сборник докладов 11-й межведомственной научно-практической конференции, Пенза, 2019 / ПГУ. Пенза, 2020. 260 с. – Деп. в ООО «НТП Криптософт» 23.04.2024, № 8/24/пи.

9 Филиппов А. В. Вопросы построения квантовых каналов на основе промышленных волоконно-оптических линий связи // Системы и средства защиты информации: Сборник докладов 11-й межведомственной научно-практической конференции, Пенза, 2019. – Пенза: Изд-во ПГУ, 2020. 260 с. – Деп. в ООО «НТП Криптософт» 23.04.2024, № 8/24/пи.

10 А. А. Карманов. Способ и устройство квантового распределения ключей с контролем параметров квантового канала // Патент на изобретение RU 2840355 С1, 21.05.2025. Заявка № 2024117861 от 27.06.2024.

Рассмотрим изменения криптографической схемы, предусматривающие использование постквантовых алгоритмов; точнее – асимметричных криптосхем, способных противостоять угрозе использования квантовых алгоритмов и компьютеров. Заметим, что современный уровень развития квантовых вычислений позволяет считать существующие симметричные алгоритмы шифрования и функции хеширования (в том, числе использующие в СНГ: ГОСТ Р 34.10-2018, ГОСТ Р 34.11-2018), защищенными от квантовых компьютеров.

Наиболее актуальным с точки зрения рассматриваемых в данной статье задач является один из трех постквантовых федеральных стандартов обработки информации (Federal Information Processing Standard, FIPS), выпущенных в 2024 году институтом NIST: это стандарт FIPS 203, описывающий алгоритм инкапсуляции ключей (ML-KEM) на основе модульных решеток. В алгоритме ML-KEM¹¹ реализованы следующие процедуры:

- выработка открытого и закрытого ключей для, соответственно, инкапсуляции и декапсуляции общего ключа;
- выработка (инкапсуляция) общего ключа симметричного шифрования;
- восстановление (декапсуляция) общего ключа симметричного шифрования.

Отметим, что важным этапом алгоритма является вычисление шифртекста C (на этапе инкапсуляции) для последующей проверки декапсулированного ключа. В табл. 3 приведены значения длин (в байтах) шифртекста.

Для сравнения приведем параметры еще нескольких схем обмена ключевой информацией (табл. 4), (потенциально) имеющих высокий уровень стойкости^{12,13}.

Независимо от выбора варианта алгоритма, длина вырабатываемого ключа фиксирована и составляет

11 National Institute of Standards and Technology (2024) Module-Lattice-Based KeyEncapsulation Mechanism Standard. (Department of Commerce, Washington, D.C.), Federal Information Processing Standards Publication (FIPS) NIST FIPS 203. DOI: 10.6028/NIST.FIPS.203.

12 Гребнев С. В. (2019) Тенденции развития постквантовой криптографии. // Материалы XXI научно-практической конференции «РусКрипто'2019». URL: <https://ruscrypto.ru/accociation/archive/rc2019.html>.

13 Панков К. Н., Миронов Ю. Б. Использование постквантовых алгоритмов в задачах защиты информации в телекоммуникационных системах. М.: Горячая линия, Телеком, 2023. 236 с.

Таблица 3.

Размеры шифртекста и ключей алгоритма ML-KEM

Вариант алгоритма	Длина ключа инкапсуляции	Длина ключа декапсуляции	Длина шифртекста
ML-KEM-512	800	1632	768
ML-KEM-768	1184	2400	1088
ML-KEM-1024	1568	3168	1568

Таблица 4.

Размеры параметров (в байтах) некоторых схем обмена ключами

Схема	Длина секретного ключа	Длина открытого ключа	Длина шифртекста
Three Bears	40	1584	1697
LEDAcrypt	40	18016	9008
FrodoKEM	31272	15632	15768
RQC	3510	3510	3574
SIKE	826	726	766

256 бит, что позволяет использовать ML-KEM для безопасной выработки ключевого материала в используемых алгоритмах блочного шифрования («Магма» и «Кузнечик»).

Практическое моделирование (реализация) метода и схемы криптографической защиты в протоколе доступа к облачному хранилищу

Представленные в данной работе механизмы криптографической защиты данных, передаваемых в облачные хранилища, смоделированы и проверены на практике¹⁴ с использованием версий отечественных СКЗИ «QR Криптофон», «КриптоПро CSP», а также разработанного авторами экспериментального образца программного СКЗИ, реализующего алгоритмы блочного шифрования «Магма» и «Кузнечик» в стандартизированных режимах работы, определенных в ГОСТ 34.13-2018 и в перспективном режиме работы блочного шифра XEN [10].

Одной из задач экспериментального моделирования была оценка влияния различных режимов блочного шифра на скорость передачи данных в облачное хранилище. Для исследования этого аспекта была проведена серия испытаний, в которых измерялось время передачи (с шифрованием) данных разного размера в облачное хранилище «Яндекс.Диск» при помощи WebDAV. Для шифрования данных использовался алгоритм блочного шифра «Магма» в режимах работы CBC-MAC, CTR-ACPKM и XEN. Моделирование

и испытания проводились на персональном компьютере, имитирующем автоматизированное рабочее место, подключенное по сети Интернет к указанному облачному хранилищу, с техническими характеристиками, представленными в (табл. 5).

На рис. 1 представлены графики зависимости времени передачи данных (в секундах) в облачное хранилище от объема передаваемых данных (в мегабайтах). Красная, зеленая и синяя кривые соответствуют передаче данных с шифрованием в режимах работы CBC-MAC, CTR-ACPKM и XEN; черная пунктирная линия соответствует передаче данных без шифрования. Для построения адекватных и объективных зависимостей каждая представленная на рис. 1 точка получена путем усреднения значений, полученных в десяти одинаковых измерениях.

На рис. 2 представлены графики зависимости коэффициента замедления передачи данных с шифрованием от объема передаваемых данных; красная кривая соответствует режиму CBC, зеленая кривая – CTR-ACPKM и синяя кривая – режиму XEN. Коэффициент замедления рассчитывается как частное от деления времени передачи данных с шифрованием на время передачи данных без шифрования. Точные значения коэффициента замедления представлены в табл. 6.

14 Минаков С. С., Тихов С. В. (2025) О механизмах криптографической защиты данных публичных облачных хранилищ и перспективах стандартизации технических спецификаций к прикладным протоколам облачных хранилищ данных // Материалы XXVII научно-практической конференции «Рус-Крипто'2025». <https://ruscrypto.ru/accociation/archive/rc2025.html>

Таблица 5.

Технические характеристики ПК

Процессор	Intel(R) Core(TM) i3-10100 CPU @ 3.60GHz
Объем ОЗУ	16 GB
Жесткий диск	HDD
Сетевой интерфейс	Ethernet 100 МБ/сек

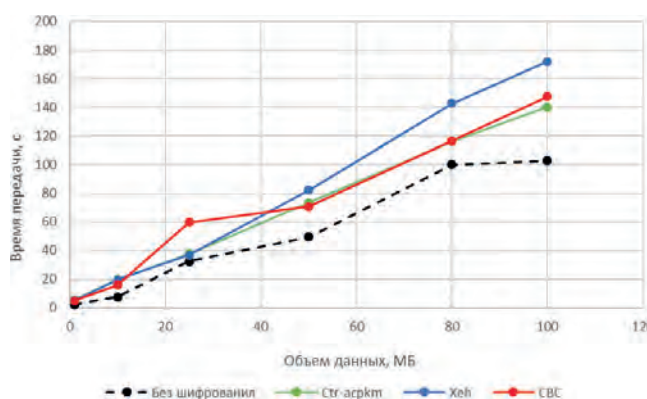


Рис. 1. Зависимость времени передачи данных в облачное хранилище от объема передаваемых данных

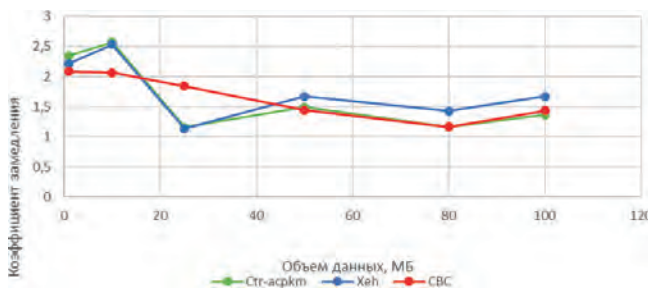


Рис. 2. Коэффициент замедления при передаче данных, обусловленный шифрованием данных, в зависимости от объема передаваемых данных

Представленные результаты показывают, что шифрование данных не приводит к серьезному замедлению системы передачи данных. Как следует из рис. 2, коэффициент замедления при размещении в облаке файлов размера менее 10 МБ равен приблизительно 2,5, а при размещении файлов размера более 25 МБ в среднем равен 1,3 для режимов работы блочного шифра CTR-АСРКМ и СВС, и 1,47 для режима ХЕН, что является хорошим показателем для СКЗИ.

Таблица 6.

Коэффициент замедления при передаче данных с шифрованием в режимах СВС, CTR-АСРКМ и ХЕН

Размер, МБ	СВС	CTR-АСРКМ	ХЕН
1	2,09	2,35	2,22
10	2,06	2,57	2,53
25	1,84	1,16	1,13
50	1,44	1,49	1,67
80	1,16	1,16	1,42
100	1,43	1,36	1,67

Заключение

Развитие криптографической схемы класса «Утро», предложенное авторами, строится на композиции симметричных и асимметричных шифров [2] с криптоалгоритмами гарантированной стойкости. Такой метод защиты, как показано в работе предполагает возможность в будущем расширения

квантовозависимыми ключами и новыми криптографическими алгоритмами постквантового класса.

В работе также показано, что метод сохранил инвариантность криптографической схемы «Утро-1» [1] относительно использования других алгоритмов шифрования, в том числе используемых в спецификациях интерфейса PKCS11 и для алгоритмов ГОСТ (СНГ) и ГОСТ Р, предлагается использовать параметры криптографических функций, опубликованных¹⁵ Национальным технологическим центром цифровой криптографии или методическими рекомендациями¹⁶ ТК 26 Росстандарта.

Нетрудно видеть, что аналогично можно реализовать метод защиты для формирования файлов-контейнеров в корпоративной сети с другими прикладными протоколами: CIFS (SMB), NFS в виде сервиса Network Attached Storage и на локально подключённых накопителях информации с различными файловыми системами.

Метод и схема криптографической защиты позволяют также строить и другие системы защищённого облачного хранения файлов для нескольких различных пользователей, для корпоративных доменов, например, на базе операционных систем Linux (AstraLinux, BaseALT и др.) и Microsoft Windows класса NT 5.xx (Windows 2000, XP, 2003, Windows FLP «Eiger»), NT 6.xx (Windows Vista, 7 – 11), в которых несложно построить процедуры поддержки форматной совместимости¹⁷ для файлов-контейнеров EFS и сервиса Web Folder Environment с реализацией отечественной криптографической схемы в российских программных системах защиты информации^{18,19}.

15 «Расширение PKCS#11 для использования стандартов ГОСТ 34.12-2018, ГОСТ 34.13-2018, ГОСТ 34.10-2018, ГОСТ 34.11-2018» / Расширение спецификаций программного интерфейса PKCS#11 (версии 3.0 и выше). // М.: НТЦЦК, 2024 г. URL: <https://digitalcryptography.ru/upload/iblock/3d4/y8zw1jb7hhbelqf5qbypvo0hmm4yfi47.pdf>.

16 МР 26.2.007-2017 «Расширение PKCS#11 для использования российских стандартов ГОСТ Р 34.10-2012 и ГОСТ Р 34.11-2012» (Документ утвержден решением заседания ТК 26, протокол № 20 от 24.11.2017) URL: <https://tc26.ru/standarts/metodicheskie-rekomendatsii/mr-26-2-007-2017-rasshirenie-pkcs-11-dlya-ispolzovaniya-rossiyskikh-standartov-gost-r-34-10-2012-i-gost-r-34-11-2012.html>.

17 How Encrypting File System Works: Security Policy; Public Key; Security Services // URL: [https://learn.microsoft.com/en-us/previous-versions/windows/it-pro/windows-server-2003/cc781588\(v=ws.10\)](https://learn.microsoft.com/en-us/previous-versions/windows/it-pro/windows-server-2003/cc781588(v=ws.10)).

18 ПСЗИ семейства QP // URL: <https://cryptosoft.ru/progr.html>

19 Secret Net Studio 8.10 // URL: <https://www.securitycode.ru/products/secret-net-studio>

Литература

- Минаков С. С. Основные криптографические механизмы защиты данных, передаваемых в облачные сервисы и сети хранения данных // Вопросы кибербезопасности. 2020. № 3(37). С. 66–75. DOI: 10.21681/2311-3456-2020-03-66-75.
- Вариант построения программного решения с гибридной криптографической системой защиты данных, хранящихся на облачном накопителе, и перспективными режимами работы блочных шифров / С. С. Минаков, И. В. Карпов, С. В. Тихов, И. В. Мартынов // Системы и средства защиты информации: Сборник статей 16-й межведомственной научно-практической конференции имени Е. А. Матвеева, Пенза, 10–13 сентября 2024 года. – Пенза: Изд-во ПГУ. 2025. С. 94–105.
- Quantum Computers as Universal Quantum Simulators: State-of-the-Art and Perspectives / F. Tacchino, A. Chiesa, S. Carretta, D. Gerace // Adv. Quantum Technol. 2020. № 3. 1900052. DOI: 10.1002/qute.201900052.

4. Ball P. Physicists in China challenge Google's 'quantum advantage' // Nature. 2020. № 588(380). DOI: 10.1038/d41586-020-03434-7.
5. Wiring surface loss of a superconducting transmon qubit / N. S. Smirnov, E. A. Krivko, A. A. Solovyova [et al.] // Sci. Rep. 2024. V. 14. P. 7326. DOI: 10.1038/s41598-024-57248-y.
6. Минаков С. С. Актуальные научные вопросы осуществления технико-криминалистических мероприятий и применения инструментальных средств при реагировании на компьютерные инциденты и анализе распределенных защищенных систем, обрабатывающих сведения конфиденциального характера // Судебная экспертиза и исследования. 2024. № 4. С. 123–129.
7. Зубов А. Ю. Криптосистема блочного гаммирования с аутентификацией // Математические вопросы криптографии. 2022. № 4(13). С. 5–35.
8. Защищённое хранение данных и полнодисковое шифрование / Е. К. Алексеев, Л. Р. Ахметзянова, А. А. Бабуева, С. В. Смышляев // Прикладная дискретная математика. 2020. № 49. С. 78–97. DOI: 10.17223/20710410/49/6.
9. Bogdanov D. S., Nozdrunov V. I. Some properties of the DEC mode of operation of block ciphers // Математические вопросы криптографии. 2022. № 3(13). С. 37–44.
10. Коренева А. М., Фирсов Г. В. Об одном режиме работы блочных шифров для защиты системных носителей с блочно-ориентированной структурой // Прикладная дискретная математика. Приложение. 2023. № 16. С. 52–56. DOI: 10.17223/2226308X/16/14.
11. Класс квантовых криптографических систем АКМ2021 на основе использования синглетных состояний многокубитовых квантовых систем / Ф. К. Алиев, А. В. Корольков, Е. А. Матвеев // Системы высокой доступности. 2022. № 3(18). С. 5–22.

THE CRYPTOGRAPHIC PROTECTION SCHEME OF CLOUD STORAGE DATA WITH POSSIBILITY TO EXPANDING BY QUANTUM-DEPENDENT KEYS AND POST-QUANTUM ENCRYPTION ALGORITHMS

Minakov S. S.²⁰, Tikhov S. V.²¹, Tsupak A. A.²²

Keywords: encryption, cloud storage, cryptographic security, network access, application-level protocol, information security, key generation.

Purpose of the study: the paper focuses on the development of scheme for cryptographic protection data in cloud storage using standardized and emerging block-cipher modes, post-quantum encryption algorithms, and quantum-dependent keys.

Methods of research: system-level analysis of information security threats during processing and storage when using application-layer access protocols for cloud storage. Synthesis of cryptographic mechanisms and formats to build an automated hybrid cryptographic scheme for client-side information processing using symmetric and asymmetric ciphers.

Result(s): the article presents the development of the Utro-1 cryptographic scheme for ensuring information security in cloud storage access application protocols. Formats, functions and encryption logic are described; several practical tests of the software implementation are carried out. The article provides explanations on the conditions for using post-quantum algorithms and/or quantum-dependent keys in the scheme.

Scientific novelty: the scientific novelty and relevance lie in achieving new scientific and technical results in the field of protecting data transmitted to cloud storage using Russian and promising foreign cryptographic tools and methods; in developing a cryptographic scheme to protect information from unauthorized access by means of a hybrid composition of symmetric and asymmetric ciphers.

References

1. Minakov, S. S. (2020). The main cryptographic mechanisms for protection of data, transmitted to cloud services and storage area networks. *Cybersecurity issues.*, 3(37), 66–75. DOI 10.21681/2311-3456-2020-03-66-75.
2. Minakov, S. S., Karpov, I. V., Tikhov, S. V., Martynov, I. V. (2025). Variant postroeniya programmogo resheniya s gibridnoj kriptograficheskoy sistemoy zashhity dannyx, xranayshixsya na oblachnom nakopitele, i perspektivnymi rezhimami raboty blochnyx shifrov. *Sistemy i sredstva zashhity informacii: Sbornik statej 16-j mezhvedomstvennoj nauchno-prakticheskoy konferencii imeni E. A. Matveeva*, Penza, 10–13 sentyabrya 2024 goda, 94–105.
3. Tacchino, F., Chiesa, A., Carretta, S. and Gerace, D. (2020), Quantum Computers as Universal Quantum Simulators: State-of-the-Art and Perspectives. *Adv. Quantum Technol.*, 3, 1900052. DOI 10.1002/qute.201900052.
4. Philip Ball. (2020). Physicists in China challenge Google's 'quantum advantage'. *Nature.*, 588, 380. DOI 10.1038/d41586-020-03434-7.
5. Smirnov, N. S., Krivko, E. A., Solovyova, A. A. et al. (2024). Wiring surface loss of a superconducting transmon qubit, *Sci Rep.*, 14, 7326. DOI 10.1038/s41598-024-57248-y.

20 Sergey S. Minakov, Senior research scientist (S. R. O.), Russian Academy of Cryptography. Moscow, Russia. E-mail: ss_minakov@mail.ru

21 Stanislav V. Tikhov, Senior developer, OOO NTP «Cryptosoft». Penza, Russia. E-mail: tik.stanislav2015@yandex.ru

22 Aleksei A. Tsupak, Dr.Sc., Associate Professor, Penza State University. Penza, Russia. E-mail: altsupak@yandex.ru

6. Minakov, S. S. (2024). Aktual'nye nauchnye voprosy osushhestvleniya tekhniko-kriminalisticheskix meropriyatij i primeneniya instrumental'nyx sredstv pri reagirovanii na komp'yuternye incidenty i analize raspredelennyx zashhishhennyx sistem, obrabatyvayushhix svedeniya konfidencial'nogo xaraktera. Sudebnaya ekspertiza i issledovaniya., 4, 123–129.
7. Zubov, A. Yu. (2022). Kriptosistema blochnogo gammirovaniya s autentifikaciej. Mathematical Aspects of Cryptography., 13(4), 5–35.
8. Alekseev, E. K., Akhmetzyanova, L. R. Babueva, A. A. Smyshlyaev, S. V. (2020). Data storage security and full disk encryption. PDM., 49, 78–97. DOI 10.17223/20710410/49/6.
9. Bogdanov, D. S., Nozdrunov, V. I. (2022). Some properties of the DEC mode of operation of block ciphers, Mathematical Aspects of Cryptography, 13(3), 37–44.
10. Koreneva, A. M., Firsov, G. V. (2023). Ob odnom rezhime raboty blochnyx shifrov dlya zashhity sistemnyx nositelej s blochno-orientirovannoj strukturoj. PDM Prilozhenie. 16, 52–56, DOI 10.17223/2226308X/16/14.
11. Aliev, F. K., Korolkov, A. V., Matveev, E. A. (2022). Class of quantum cryptographic systems AKM2021 based on the use of singlet states of multicubic quantum systems // Journal Highly available systems, 18(3), 5–22.



НОРМАТИВНО-ТЕХНИЧЕСКИЕ АСПЕКТЫ ОБЕСПЕЧЕНИЯ ИНФОРМАЦИОННОЙ И КИБЕРБЕЗОПАСНОСТИ ОБЪЕКТОВ ЦИФРОВОЙ ЭЛЕКТРОЭНЕРГЕТИКИ

Куликов А. Л.¹, Лоскутов А. А.², Зинин В. М.³

DOI: 10.21681/2311-3456-2025-6-101-115

Цель исследования: выполнить анализ состояния, нормативной и законодательной базы в области информационной безопасности (ИБ) цифровой электроэнергетики, определить области дальнейшего совершенствования программно-аппаратных комплексов ИБ, реализовать отраслевые требования ИБ в экспериментальных образцах релейной защиты и автоматики для цифровой подстанции.

Метод(ы) исследования: системный анализ, анализ научной отраслевой проблемы, классификация, абстрагирование.

Результат(ы) исследования: представлен анализ нормативно-правовой базы в сфере защиты критической информационной инфраструктуры (КИИ) в электроэнергетике; рассмотрена степень соответствия действующих отраслевых объектов установленным законодательным требованиям по кибербезопасности; изложен подход к разработке кросс-платформенных программно-аппаратных комплексов (ПАК) для цифровых подстанций (ЦПС), оснащенных встроенными механизмами защиты информации, включая практический пример – устройство релейной защиты и автоматизации (РЗА).

Научная новизна: предложена технология построения доверенных программно-аппаратных комплексов для цифровой электроэнергетики с интегрированными функциями ИБ и возможностью интеграции алгоритмов кибербезопасности, предусматривающая шесть уровней абстракции.

Ключевые слова: информационная безопасность, кибербезопасность, цифровая подстанция, критическая информационная инфраструктура, киберугроза, компьютерная атака, релейная защита и автоматика, технологическая информация, интеллектуальные электрические сети.

Введение

Электроэнергетическая система (ЭЭС) представляет собой критически важный элемент национальной инфраструктуры, от устойчивого функционирования которого зависит экономическая стабильность государства. В последние годы российская энергетика проходит этапы трансформации благодаря внедрению цифровых технологий и масштабной модернизации энергетического комплекса. Однако наряду с повышением эффективности, автоматизация управления ЭЭС привела к новым угрозам, связанным с кибербезопасностью. Развитие интеллектуальных сетей, средств автоматизации и мониторинга не только увеличило оперативность управления, но и расширило масштаб атак для злоумышленников, включая хакеров и террористические организации.

Компьютерные атаки на энергетическую инфраструктуру способны не только нарушить работу отдельных подстанций, но и привести к каскадным отказам в единой национальной электрической сети (ЕНЭС) и единой энергосистеме (ЕЭС) России. Последствия таких инцидентов включают: разрушение

критически важных объектов; угрозу жизни и здоровью населения; экономические потери в промышленности, транспорте и ЖКХ.

Актуальность проблемы обеспечения информационной безопасности (ИБ) обусловлена следующими факторами:

- быстрые темпы роста количества различных электронных устройств, применяемых в самых разных сферах деятельности, и, как следствие, резкое расширение круга пользователей, имеющих непосредственный доступ к сетям и информационным ресурсам;
- резкое увеличение объемов информации, накапливаемой, хранимой и обрабатываемой с помощью компьютеров и других средств автоматизации;
- бурное развитие аппаратно-программных средств и технологий, не соответствующих современным требованиям безопасности;
- несоответствие бурного развития средств обработки информации и методов теории информационной

1 Куликов Александр Леонидович, доктор технических наук, профессор, профессор кафедры «Электроэнергетика, электроснабжение и силовая электроника» НГТУ им. Р. Е. Алексеева, г. Нижний Новгород, Россия, E-mail: inventor61@mail.ru

2 Лоскутов Антон Алексеевич, кандидат технических наук, доцент, доцент кафедры «Электроэнергетика, электроснабжение и силовая электроника», старший научный сотрудник лаборатории «Автономные гибридные электроэнергетические комплексы», НГТУ им. Р. Е. Алексеева, г. Нижний Новгород, Россия, E-mail: loskutov_aa@nntu.ru

3 Зинин Владимир Михайлович, руководитель направления ООО НПП «ОВИСТ», г. Москва, Россия, E-mail: c.nemo@yandex.ru

безопасности, разработки международных стандартов и правовых норм, обеспечивающих необходимый уровень защиты информации;

- повсеместное распространение сетевых технологий, создание единого информационно-коммуникационного мирового пространства на базе Интернет (например, «Интернет вещей (IoT)»), которое по своей идеологии не обеспечивает достаточного уровня информационной безопасности;
- недостаточная проработка нормативно-правовой базы в сфере кибербезопасности в электроэнергетической отрасли;
- высокий ежегодный рост количества компьютерных преступлений в мире и России, а также ущерба, причиняемого такими преступлениями.

В соответствии с «Доктриной энергетической безопасности РФ» (2019 г.) [1] ключевыми рисками признаны: кибератаки на объекты топливно-энергетического комплекса (ТЭК); зависимость от иностранных технологий, оборудования и программного обеспечения.

Понятия информационной безопасности применительно к объектам электроэнергетики

Согласно федеральному закону № 187-ФЗ «О безопасности критической информационной инфраструктуры РФ» информационные системы, функционирующие на объектах электроэнергетики, топливно-энергетического комплекса, атомной отрасли относятся к объектам критической информационной инфраструктуры (КИИ) [2].

В соответствии с классическими принципами ИБ необходимо обеспечение трех взаимосвязанных целей: конфиденциальность, целостность, доступность [3]. В разрезе электроэнергетики данные постулаты следует интерпретировать с учетом отраслевой специфики, поскольку электроэнергетические объекты являются киберфизическими системами [4,5]. Вопросы конфиденциальности не являются актуальными, поскольку технологическая информация не содержит секретности, а вот целостность и доступность должны быть обеспечены для своевременного реагирования диспетчерского персонала, устройств релейной защиты и автоматики (РЗА) на возможные возмущения, аварии, нарушения функционирования. От этого зависит устойчивость и надежность работы всей ЕНЭС. Поэтому более подходящим термином для электроэнергетики является кибербезопасность (КБ). Однако методы и практики, используемые в задачах ИБ справедливы и для КБ [6–8].

Компрометация или подмена данных (спуфинг) в системах измерения (телеизмерения) может привести к неправильному учету ЭЭ, то есть коммерческим потерям для всего электроэнергетического сектора или ложному срабатыванию/не срабатыванию РЗА

[9, 10]. Последнее повлечет за собой отключение части потребителей с последующими ущербами для предприятий или поломке дорогостоящего оборудования с длительным ремонтом. Причем тяжесть ущерба определяется уровнем напряжения в точке проведения кибератаки (например, 0,4 кВ или 500 кВ) и категорий потребителей (например, бытовые или промышленные, нефтегазовые, военные, транспортные). Для бытовых сетей 0,4 кВ ущерб будет не значительным, а для системообразующих объектов 220–500 кВ отключения могут приводить к ущербу в сотни миллионов рублей, включая угрозу жизни и здоровью людей.

Важной особенностью ЕНЭС является её сложность, одновременность процессов генерации, распределения и потребления ЭЭ, согласованность действий систем управления, защиты, противоаварийной автоматики, связи. Поэтому при успешной кибератаке важно её идентифицировать, распознать, отличить её от действительно случившейся аварии в электрической сети (например, короткое замыкание (КЗ)) и предпринять правильную последовательность действий по восстановлению электроснабжения, а также обеспечению динамической и статической устойчивости ЭЭС, не допустить её развала (блэкаута). Системы противодействия кибератакам (системы КБ) на электроэнергетических объектах должны учитывать отраслевую специфику электроэнергетики, технологический процесс производства ЭЭ, быть эшелонированными, обладать свойствами живучести и робастности [11].

В современной цифровой электроэнергетике единственным объектом технологического управления является узел нагрузки или центр питания – цифровая подстанция (ЦПС) напряжением от 6 кВ до 750 кВ с использованием протоколов передачи данных в АСУТП: МЭК 61850, IEC 104, DNP3, Modbus и других [12–15]. Компоненты систем КБ должны быть интегрированы во все «вторичное» оборудование (РЗА, АСУТП, SCADA и др.) [16–18].

Следует отметить актуальные изменения в стандарте МЭК 61850, касающиеся ИБ [13]. На рис. 1 представлены модификации, предложенные рабочей группой IEC 61850-8-1/AMD1 ED2 для включения в следующую версию стандарта. Как видно (пунктирное выделение), в структуру МЭК 61850 добавлен стандарт МЭК 62351-6 «Управление энергосистемами и связанный с ним обмен информацией. Безопасность данных и коммуникаций. Часть 6. Безопасность протоколов МЭК 61850». Его применение станет обязательным для передачи данных за пределами ЦПС и рекомендуемым – внутри технологической вычислительной сети ЦПС.

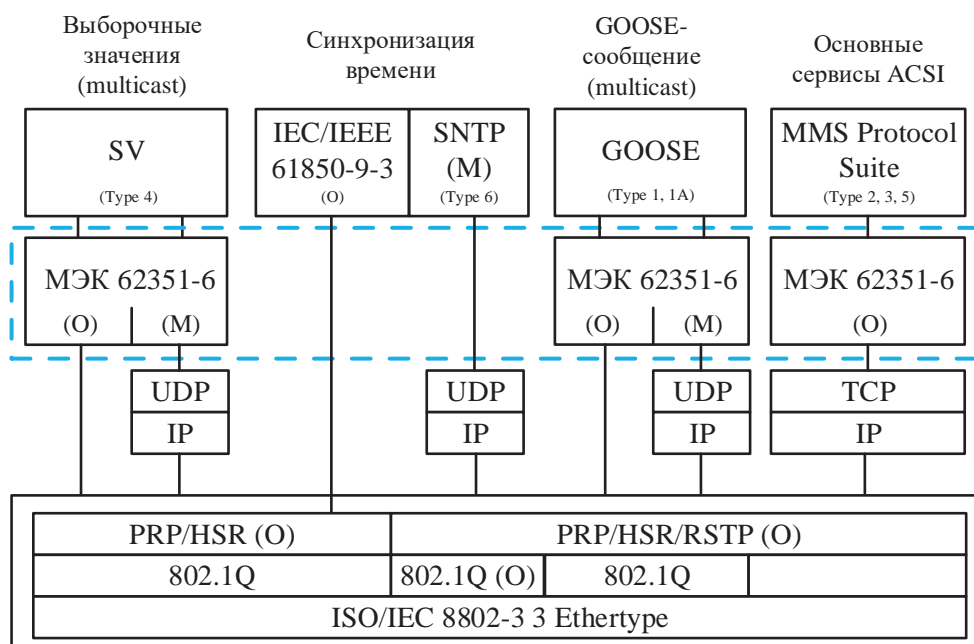


Рис. 1. Совершенствование стандарта МЭК 61850 в направлении ИБ

Современные тенденции развития стандарта МЭК 61850 предполагают интеграцию средств ИБ в ИЭУ. Однако большинство современных ИЭУ, применяемых на ЦПС и работающих по МЭК 61850, пока не полностью соответствуют требованиям IEC 62351-6. Ранее при проектировании ИЭУ функции ИБ часто игнорировали, поскольку их внедрение могло снизить быстродействие оборудования. Это особенно критично для устройств РЗА, где задержки недопустимы [19].

Анализ нормативной и законодательной базы в области информационной безопасности объектов электроэнергетики

В текущем разделе приведена хронология развития федерального законодательства в сфере ИБ для объектов КИИ за последние 10 лет [2].

05.12.2016 г. Президент РФ подписал обновлённый вариант «Доктрины информационной безопасности РФ» [1]. Этот документ отражает официальную позицию государства по вопросам защиты национальных интересов в цифровой среде. В тексте Доктрины дано определение информационной сферы, которое подразумевает: все виды данных и их носителей; информационные системы и сети; интернет-ресурсы и телекоммуникации; технологии обработки и передачи информации; организации и частных лиц, работающих в этой области; правовые механизмы регулирования цифровых отношений. Ключевые национальные интересы РФ в информационной сфере, согласно Доктрине, охватывают:

- гарантию прав граждан на доступ к информации и защиту приватности;
- поддержание устойчивости критически важных информационных систем;

- развитие отечественных IT-технологий и электроники;
- информирование общества о политике государства;
- участие в международных инициативах по кибербезопасности.

Федеральный закон № 187-ФЗ «О безопасности критической информационной инфраструктуры», вступивший в силу 01.01.2018, закрепляет принципы защиты КИИ, определяет права и обязанности участников, вводит систему категорирования объектов.

К субъектам КИИ закон относит государственные структуры, компании, владеющие или управляющие информационными системами в ключевых отраслях: здравоохранении, транспорте, энергетике, финансовом секторе, оборонной, металлургической химической промышленности и др.

Правительство РФ утвердило Постановление от 08.02.2018 г. № 127, регламентирующее: процедуру категорирования объектов КИИ; сроки проведения категорирования; показатели для определения значимости систем.

В 2022 году были приняты три важных указа:

- Указ Президента от 30.03.2022 г. № 166 – предписывает переход на отечественные IT-решения, радиоэлектронную продукцию, телекоммуникационное оборудование на значимых объектах КИИ (ЗОКИИ) и запрет иностранного ПО на ЗОКИИ с 01.01.2025 г. Следует отметить, что ключевые термины «Технологическая независимость» и «Доверенный программно-аппаратный комплекс» (ДПАК) для КИИ появились именно в принятом

Указе № 166, но их юридически значимое определение и детализация даётся в последующем развитии нормативно-правовой базы по его реализации.

- Указ Президента от 14.04.2022 г. № 203 – создаёт межведомственную комиссию Совета Безопасности РФ для координации технологической независимости объектов КИИ и их оснащению отечественной радиоэлектронной продукцией, оборудованием, программно-аппаратными комплексами (ПАК).
- Указ Президента от 01.05.2022 г. № 250 – усиливает защиту данных госорганов и стратегических предприятий, требуя создания спецподразделений по ИБ и вводя персональную ответственность руководителей. Данный Указ в явном виде обязывает возложить на заместителя руководителя органа (организации) полномочия по обеспечению ИБ органа (организации), в том числе и в части государственной системы обнаружения, предупреждения и ликвидации последствий компьютерных атак (ГосСОПКА). Также Указ устанавливает с 01.01.2025 г. организациям прямой запрет на использование средств защиты информации, странами происхождения которых являются иностранные государства, совершающие в отношении РФ недружественные действия, либо производителями которых являются организации, находящиеся под юрисдикцией таких иностранных государств.
- Принято 22.08.2022 г. Постановление Правительства РФ № 1478 «Об утверждении требований к программному обеспечению, в том числе в составе ПАК», стимулирующее разработку отечественного ПО.
- Принято 14.11.2023 г. Постановление Правительства от. РФ № 1912 «О порядке перехода субъектов критической информационной инфраструктуры Российской Федерации на преимущественное применение доверенных программно-аппаратных комплексов на принадлежащих им значимых объектах критической информационной инфраструктуры Российской Федерации», в котором:
 - утверждаются «правила перехода субъектов КИИ РФ на преимущественное применение доверенных ПАК (ДПАК) на принадлежащих им ЗОКИИ»;
 - устанавливается, что «переход субъектов КИИ РФ на преимущественное применение ДПАК на принадлежащих им значимых объектах КИИ осуществляется до 01.01.2030 г. в соответствии с Правилами, утвержденными настоящим постановлением»;

- с 01.09.2024 г. не допускается использование субъектами КИИ РФ на принадлежащих им ЗОКИИ ПАК, приобретенных с 01.09.2024 г. и не являющихся ДПАК, за исключением случаев отсутствия произведенных в РФ ДПАК, являющихся аналогами приобретенных субъектами КИИ ПАК.

В «Правилах...», утвержденных данным постановлением дается однозначное толкование понятия «преимущественное применение ДПАК» – применение субъектами КИИ на принадлежащих им ЗОКИИ ПАК, доля которых по состоянию на 31.12.2029 г. составляет 100 % в общем количестве ПАК, применяемых субъектами КИИ на принадлежащих им ЗОКИИ.

Другими словами, в Постановлении для субъектов КИИ государством установлен переходный период до 31.12.2029 г.

- Принято 29.03.2023 г. Распоряжение Правительства РФ № 757-р, определяющее АО «НПО «КИС» производственным объединением, созданным в целях реализации Указа № 166. На основании приказа Росстандарта № 3107 от 09.12.2022 г. при данной организации был сформирован Технический комитет по стандартизации ТК 167, специализирующийся на вопросах программно-аппаратных решений для КИИ и соответствующего программного обеспечения (ПО). Основная цель создания комитета – обеспечение нормативно-технической базы для достижения технологической самостоятельности и устойчивой работы критически важных информационных систем. Главными задачами ТК 167 выступают: разработка ГОСТов, направленных на поддержку отечественного производства ДПАК, координация деятельности отраслевых технических комитетов, занимающихся вопросами критических информационных технологий.
- Введен 01.04.2024 г. в действие предварительный национальный стандарт РФ ПНСТ 905-2023 «Критическая информационная инфраструктура. Доверенные программно-аппаратные комплексы. Термины и определения».
- Был принят 25.03.2025 г. законопроект об изменениях в 187-ФЗ, который действует уже более 7 лет. Изменения связаны с тем, что Правительство РФ до 01.09.2025 г. обязано разработать обязательные отраслевые перечни типовых объектов КИИ, а ответственность за их ведение и за категорирование объектов КИИ возложена на федеральные органы исполнительной власти (ФОИВ). Применительно к электроэнергетике отраслевым ФОИВ является Министерство энергетики РФ.

01.10.2025 г. введен в действие предварительный национальный стандарт РФ ПНСТ 1009-2025 «Критическая информационная инфраструктура. Программное обеспечение для доверенных программно-аппаратных комплексов. Общие положения».

Таким образом, государство переходит к жёсткому регулированию ИБ, превращая её из внутренней корпоративной задачи в строго контролируемую государством сферу.

Текущая ситуация в электроэнергетической отрасли

Ведущими стратегическими компаниями после формирования и преобразования структуры электроэнергетики РФ являются АО «СО ЕЭС», ПАО «ФСК ЕЭС», ПАО «Россети». В их управлении находятся стратегические интересы, обороноспособность и безопасность государства, защита нравственности, здоровья, прав и законных интересов граждан РФ. Все центры питания, находящиеся в ведомстве данных организаций (субъектов КИИ), являются объектами КИИ и подлежат категорированию [2].

На сегодняшний момент системы технологического управления центров питания и ЦПС построены преимущественно с использованием технологий зарубежных производителей, многие из которых прекратили свою деятельность и техническую поддержку заказчиков в РФ в 2022 году, например, Siemens, ABB, GE, Alstom, Sprecon, Intel, DelL, HP, Microsoft, Hirschmann и др. SCADA-системы этих объектов в качестве аппаратной платформы используют серверное оборудование на микропроцессорах Intel, а в качестве операционной системы Microsoft Windows различных версий. Автоматизированные рабочие места эксплуатационного, оперативного персонала и диспетчерских служб также используют компьютеры Intel-Windows [2]. Технологические риски и риски КБ для электроэнергетики РФ при продолжении эксплуатации решений этих зарубежных компаний возрастают и становятся неуправляемыми, именно поэтому государство с 01.01.2025 года вводит запрет на применение систем технологического управления, систем ИБ и КБ и оборудования зарубежных производителей в составе ЗОКИИ РФ. С этой целью государством разработаны и утверждены обязательные требования к ДПАК отечественного производства для КИИ и подстанций ЕНЭС, в частности.

Формально программа импортозамещения идет, формируется заказ на разработку отечественной науки. Однако доля таких заказов, доходящих до ВУЗов и реальных научно-производственных предприятий, специализирующихся на электроэнергетической отрасли очень мала, что не развивает научные школы, а заставляет их «выживать». По факту сегодня вероятность возникновения рисков связанных с компьютерными атаками на КИИ РФ, последствия

которых сопоставимы, например, с аварией на Саяно-Шушенской ГЭС в 2009г., возрастает. Вызывает опасение в части своевременного выполнения действующего законодательства в части КИИ РФ, импортозамещения и перехода на отечественные ПАК.

В настоящее время разработчики ПАК сталкиваются с серьезными трудностями, связанными с недостаточной проработанностью нормативной и технической базы для ДПАК. В частности, отсутствуют четкие критерии, регламентирующие: процесс разработки и производства таких систем, проектирование ключевых технических решений, особенности их внедрения и эксплуатации в составе ЗОКИИ электроэнергетического сектора. Подобная неопределенность приводит к ошибкам при принятии решений субъектами КИИ, в том числе при модернизации автоматизированных систем технологического управления (АСТУ). Важно отметить, что даже в условиях неполного регулирования организации не освобождаются от ответственности за соблюдение законодательства в сфере КИИ и ИБ. В случае возникновения киберинцидентов, повлекших серьезные технологические сбои, возможна даже уголовная ответственность.

На данный момент, пока не завершена разработка ГОСТ Р по ДПАК в рамках ТК 167. «Доверенность» применительно к функциям защиты информации в ПАК для субъекта КИИ в электроэнергетике определяется Приказом ФСТЭК № 76 от 02.06.2020 [2], определяющим уровни доверия к средствам технической защиты информации.

Кроме того, при построении систем безопасности для ЗОКИИ необходимо учитывать Приказы ФСТЭК:

- № 235 от 21.12.2017 – требования к созданию и функционированию защитных систем;
- № 239 от 25.12.2017 – базовые стандарты безопасности для ЗОКИИ.

Таким образом, до принятия окончательных ГОСТ Р именно эти документы являются основными регуляторами в области защиты критически важных объектов энергетики. Их соблюдение обязательно, несмотря на отсутствие исчерпывающей нормативной базы по ДПАК.

Некоторые отечественные производители электроэнергетического, коммуникационного оборудования декларируют, что соблюдают все необходимые нормативные документы и требования КБ и применяют ДПАК. Однако в их составе содержатся компоненты, процессоры, чипы иностранного производства из недружественных стран. Хотя формально получены необходимые свидетельства и лицензии ФСБ, ФСТЭК.

Развитие отечественных аппаратно-программных платформ и микропроцессоров должно создать благоприятные условия к снижению зависимости от импорта, созданию действительно отечественных

аналогов интеллектуального оборудования различного назначения для электроэнергетики.

Понятие доверенного программно-аппаратного комплекса

Под «Программно-аппаратным комплексом» (ПАК) понимается комплекс технических и программных средств (программного обеспечения), работающих совместно для выполнения одной или нескольких специальных задач, являющийся электронной вычислительной машиной или специализированным электронным устройством (устройствами), функционально-технические характеристики которого (которых) определяются исключительно совокупностью программного обеспечения и технических средств, и не могут быть реализованы при их разделении. ПАК является самостоятельно используемым, законченным техническим изделием, имеющим серийный номер».

В ПНСТ 905-2023 дается определение «Доверенного программно-аппаратного комплекса» (ДПАК): программно-аппаратный комплекс, соответствующий требованиям обеспечения технологической независимости критической информационной инфраструктуры, функциональности, надежности и защищенности».

Ключевым в данном определении является «технологическая независимость», которая является основой «доверенности» компонентов ПАК. Данные компоненты должны обеспечивать доверенную начальную загрузку ПАК, обработку информации и контроль над её передачей, управление работой

программного обеспечения. Основными компонентами являются: центральный процессор и основные системные и сетевые контроллеры, операционная система, базовая система ввода-вывода (UEFI – Unified Extensible Firmware Interface). Важными составляющими в обеспечении «доверенности» являются компиляторы, архитектура процессора. По мнению авторов, достижение «технологической независимости» для объектов КИИ и электроэнергетики в частности, возможно только в случае «доверенности» каждого из перечисленных компонентов посредством полного цикла разработки в РФ.

В соответствии с Постановлением Правительства РФ № 1912 критериями ДПАК являются:

1. Сведения о доверенном ПАК содержатся в реестре российской радиоэлектронной продукции;
2. Программное обеспечение, используемое в составе ДПАК, соответствует требованиям из Постановления Правительства РФ от 22.08.2022 г. № 1478;
3. ПАК, в случае реализации в нём функций защиты информации соответствует требованиям, установленным ФСТЭК в пределах их полномочий, что должно быть подтверждено соответствующим документом (сертификатом).

Технология построения киберзащищенных доверенных программно-аппаратных комплексов для цифровой электроэнергетики

Важные, системообразующие объекты электроэнергетики и элементы распределительных электрических сетей оснащаются ПАК-ми различного

Таблица 1.

ПАК критичные для объектов электроэнергетики

№ п/п	Наименование ПАК	Назначение и особенности для объекта КИИ
1	ПАК релейной защиты и автоматики (ПАК РЗА)	Предназначен для автоматической ликвидации аварийного режима в энергосистеме, в общем случае (при наличии защит абсолютной селективности) является распределённым, его функционирование выходит за физические границы ЛВС подстанции или объекта генерации, использует каналы передачи данных, технологически может быть связан с ПАК ПА и ПАК АСУ ТП
2	ПАК противоаварийной автоматики (ПАК ПА)	Предназначен для автоматической ликвидации аварийного режима в энергосистеме, является распределённым, его функционирование выходит за физические границы ЛВС подстанции или объекта генерации, использует каналы передачи данных, технологически связан с ПАК РЗА
3	ПАК АСУ ТП	Предназначен для технологического управления энергообъектом в нормальном режиме работы, является локальным, его функционирование ограничивается физической границей ЛВС подстанции или объекта генерации, обычно технологически связан с ПАК РЗА
4	ПАК АИИСКУЭ	Предназначен для коммерческого учёта электроэнергии и мощности, а также возможного удаленного ограничения мощности, потребляемой абонентами энергосистемы (в случае, если прибор коммерческого учёта абонента поддерживает данную функцию), является распределённым, его функционирование выходит за физические границы ЛВС подстанции или объекта генерации, использует каналы передачи данных

назначения. Наиболее критичные ПАК, с точки зрения кибератак, влияющие на функциональную устойчивость энергосистемы можно представить в виде табл. 1.

В рамках развития цифровизации энергетического сектора научный коллектив НГТУ им. Р. Е. Алексеева и промышленных партнеров, включая авторов статьи, разработал и предложил технологию создания кроссплатформенных ПАК для ЦПС со встроенными функциями КБ. Разработаны экспериментальные, опытно-промышленные и серийные образцы интеллектуальных электронных устройств (ИЭУ) защиты и автоматики, представляющие собой ДПАК и [20]. Большинство из них прошли аттестацию в ПАО «Россети» применительно для объектов электрических сетей напряжением 6–35 кВ и 110–220 кВ [1,21].

Сегодня наблюдается устойчивая тенденция к снижению стоимости комплектующих (микропроцессоров, модулей памяти, накопителей данных, интерфейсных компонентов и др.) при одновременном повышении их надежности и мощности. Это способствовало активному внедрению в промышленную автоматизацию серийных вычислительных решений с высокой степенью унификации. Кроме того, сегодня доступны серийные операционные системы, предназначенные для эксплуатации в особых условиях (промышленных, военных, космических), которые обладают комплексной системой ИБ, включая продукты, сертифицированные ФСТЭК России.

В предложенной авторами технологии создания кроссплатформенных ПАК для ЦПС с интегрированными функциями КБ применяется несколько уровней абстракции (рис. 2). Важным элементом данной структуры является независимая от производителя доверенная аппаратно-программная платформа.

Ключевыми преимуществами данного подхода являются:

- 1) обеспечение требуемого уровня информационной безопасности устройств в соответствии с нормативными требованиями;
- 2) освобождение производителей ПАК от ресурсоемких процедур сертификации (аттестации) аппаратной части;
- 3) возможность выбора компонентов из российских реестров сертифицированного ПО и оборудования.

Основная деятельность разработчиков ПАК в данной модели сосредоточена на создании и оптимизации алгоритмической составляющей, развитие функционального ПО, обеспечение устойчивости устройств к современным киберугрозам.

Данный подход предполагает использование серийных промышленных вычислителей в качестве аппаратной платформы. Они основаны на распространенных микропроцессорах российского (например, Эльбрус) производства, что составляет 1-й уровень абстракции. 2-й, 3-й и 4-й уровни включают сертифицированные ФСТЭК операционные системы (такие как Астра Линукс, Альт Линукс, «Нейтрино» и др.), причём первые четыре уровня не привязаны к конкретному производителю ПАК. 5-й и 6-й уровни представляют собой кроссплатформенное ПО ПАК и коммуникационные интерфейсы стандарта. Требования к информационной безопасности изначально заложены в модель МЭК 61850 при разработке ПАК. Процесс создания кроссплатформенных ПАК для ЦПС со встроенными функциями ИБ схематично отображён на рис. 3 [2].

Разработанные в НГТУ экспериментальные образцы ИЭУ (ПАК) для РЗА обеспечивают следующие функции ИБ:

1. Шифрование TLS для протокола МЭК 61850-8-1 (MMS) при обмене данными между ИЭУ и другими подсистемами ЦПС, между ЦПС и центром управления сетями (ЦУС).

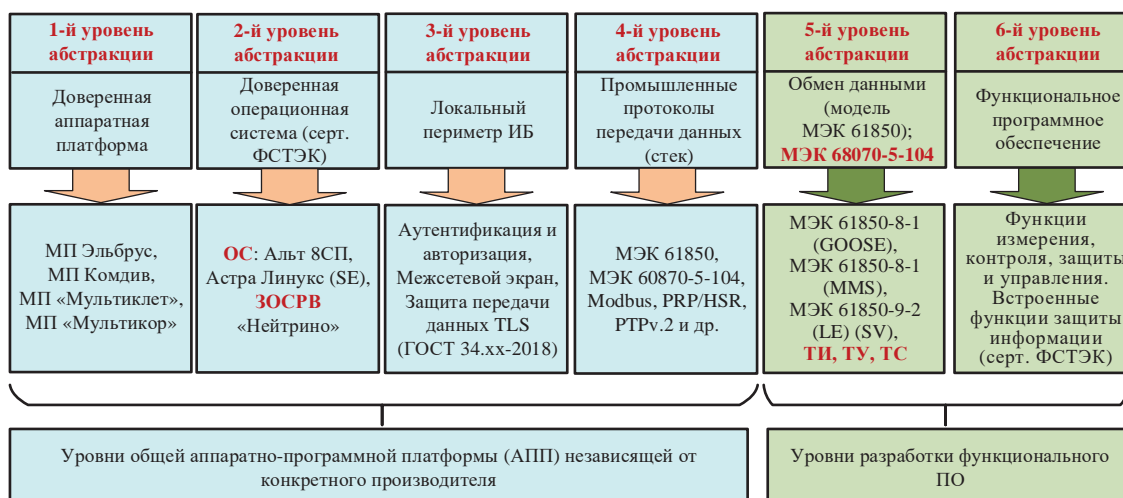


Рис. 2. Шесть уровней абстракции при создании киберзащищенного ДПАК для электроэнергетики



Рис. 3. Структура создания кроссплатформенных ПАК для ЦПС с интегрированными функциями ИБ [2]

2. Двухфакторная аутентификация для устройств РЗА и автоматизированных рабочих мест (АРМ) оперативного и эксплуатационного персонала при удалённом подключении к ПАК через технологическую сеть ЦПС.
3. Ролевое разграничение доступа к элементам интерфейса ПАК в соответствии с должностными обязанностями персонала.
4. Логирование событий безопасности на уровне отдельных ПАК, ЦПС и ЦУС.

Далее рассмотрим и проиллюстрируем перечисленные функции более детально, сопровождая описание примерами интерфейсов ПАК РЗА и специализированного ПО для мониторинга сетевого трафика.

На рис. 4–5 показана работа TLS-шифрования для протокола МЭК 61850-8-1 (MMS) при взаимодействии ПАК с подсистемами ЦПС. Рис. 4 иллюстрирует пример перехвата MMS-пакетов (ПАК РЗА P0101) в программе Wireshark в незашифрованном виде (выделено рамкой). В отличие от незашифрованного

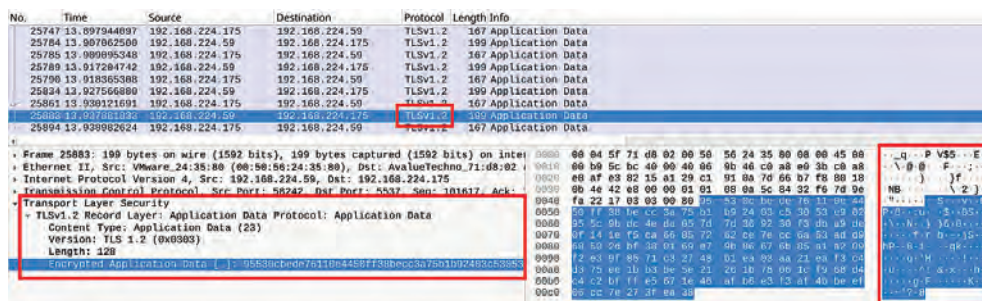


Рис. 4. Работа ПАК без TLS-шифрования MMS-пакетов

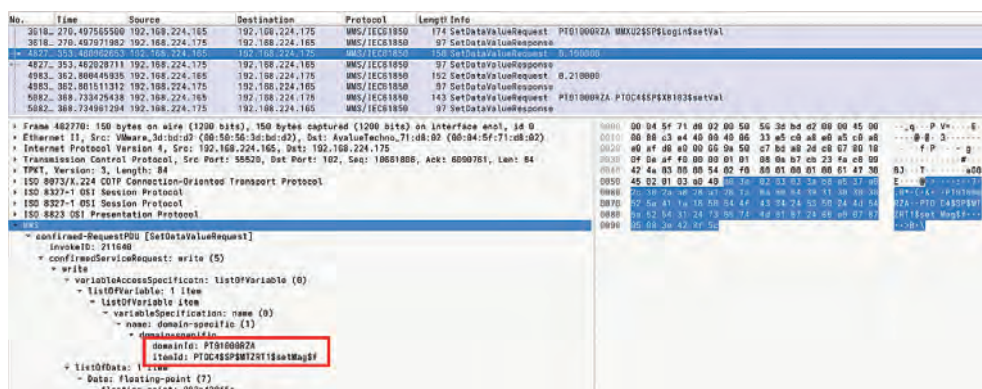


Рис. 5. Работа ПАК с TLS-шифрованием MMS-пакетов



Рис. 6. Двухфакторная аутентификация с использованием USB-идентификатора

обмена данными (рис. 4), на рис. 5 демонстрируется защищённый MMS-трафик ПАК РЗА Р0101, перехваченный в Wireshark с активированным TLS-протоколом. Как визуально различимо на рис. 5, применение технологии Transport Layer Security (TLS) обеспечивает полное шифрование передаваемой информации, делая содержимое MMS-сообщений недоступным для прочтения.

На рис. 6 представлена схема двухфакторной аутентификации для ПАК РЗА.

Для доступа к устройству сотруднику необходимо подключить персональный USB-идентификатор (аппаратный токен). При локальном входе он включается во фронтальную панель ПАК, при удаленном входе через сеть ЦПС – к компьютеру оператора. Далее вводится логин и пароль. Доступ предоставляется только пользователям, зарегистрированным в системе ЦПС с соответствующими правами. В противном случае система безопасности блокирует авторизацию. В данном решении применяются USB-идентификаторы Guardant ID, сертифицированные ФСТЭК России по 2-му уровню контроля РД НДВ (требования по защите от несанкционированного доступа).

Рассмотренные примеры (включая механизмы аутентификации и шифрования) относятся к 3-му уровню абстракции (рис. 2), обеспечивающему базовую защиту кроссплатформенных ПАК.

Далее представлен функционал 6-го уровня абстракции (уровня прикладного функционального ПО), который обеспечивает соответствие отраслевым требованиям по ИБ, выполнение норм Распоряжения ПАО «Россети» № 282р:

1. Генерация данных аудита.
2. Ассоциация идентификатора пользователя.
3. Просмотр журналов аудита.
4. Защищенное хранение журнала аудита.
5. Действия в случае возможной потери данных аудита.
6. Предотвращение потери данных аудита.
7. Ограниченное управление доступом.
8. Управление доступом, основанное на атрибутах безопасности.

9. Обработка отказов аутентификации.
10. Определение атрибутов пользователя.
11. Аутентификация до любых действий пользователя.
12. Аутентификация с защищенной обратной связью.
13. Идентификация до любых действий пользователя.
14. Управление атрибутами безопасности.
15. Инициализация статических атрибутов.
16. Управление данными ФБО.
17. Спецификация функций управления.
18. Роли безопасности.
19. Блокирование сеанса, инициированное функциями безопасности.

Такой многоуровневый подход позволяет комплексно реализовывать политики безопасности на всех этапах работы ПАК.

На примере ПАК РЗА рассмотрим практическую реализацию ключевых требований ИБ. Особое внимание уделено реализации системы ролевого доступа.

Четкое разграничение прав осуществляется в соответствии с таблицей 2, которая отражает специфику работы с технологическим оборудованием и разграничение прав. Сценарий работы с ПАК начинается с авторизации пользователя (рис. 7), который включён в одну из групп (табл. 2).

После успешной авторизации пользователь получает доступ к интерфейсу (рис. 8, 9, 10), адаптированному под его служебные функции. В представленном на рис. 8 примере вошел пользователь с ролью «Администратор», имеющий полный доступ ко всем возможностям системы (выделено рамкой).

Рис. 9 отображает интерфейс для роли «Специалист по ИБ», имеющего доступ только к управлению пользователями через кнопку «Users» (выделено рамкой). При настройке системы можно также ограничить его права на управление группой «Администраторы».

Экран специалиста АСУ ТП (рис.10) включает доступ только к настройкам подключения к сетям ЦПС. Через кнопку «Network» (выделена рамкой) можно настроить сетевые параметры, МЭК 61850 и параметры для работы со SCADA

Таблица 2.

Ролевой доступ в ПАК РЗА

№	Название группы	Функциональные обязанности, роли	Права доступа к элементам интерфейса и функциям ПАК
1.	Администратор(ы)	Представители компании-производителя и/или компании, выполняющей ПНР	Полный доступ к элементам интерфейса и параметрированию
2.	Специалист(ы) по ИБ	Специалист по ИБ	Управление пользователями
3.	Эксплуатационный персонал	Специалисты, отвечающие за эксплуатацию ПАК (для ПАК РЗА – инженеры РЗА)	Параметрирование с некоторыми ограничениями (калибровка, настройка параметров АВС)
4.	Оперативный персонал	Специалисты ОВБ, диспетчерский персонал ПС	Ввод/вывод функций РЗА и автоматики, чтение осциллограмм, журнала событий
5.	Читатель	Руководящий персонал ПС	Чтение осциллограмм, журнала событий
6.	Специалист(ы) по АСУ ТП	Инженер по связи, инженер по ИТ	Настройка параметров АВС

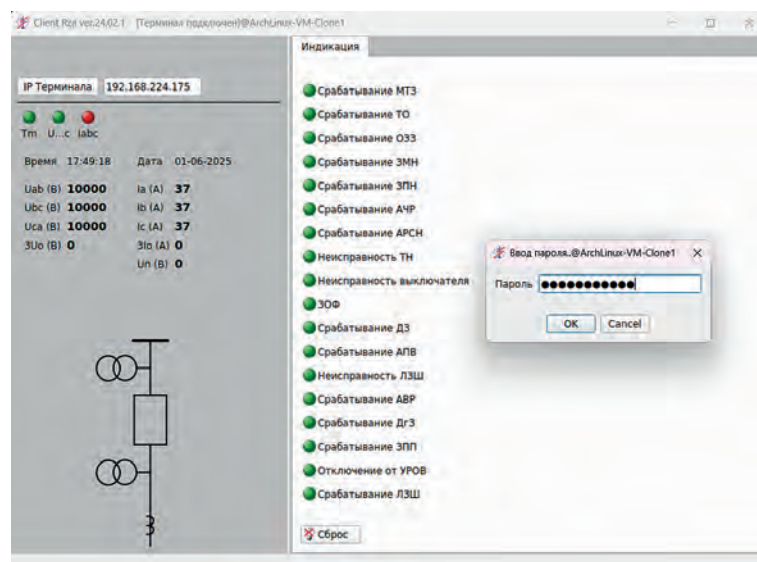


Рис. 7. Окно авторизации пользователя

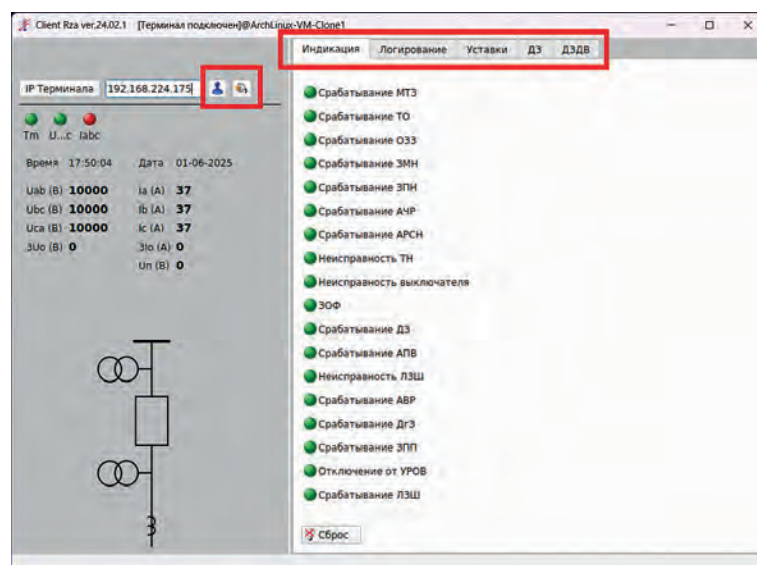


Рис. 8. Авторизация с ролью «Администратор»

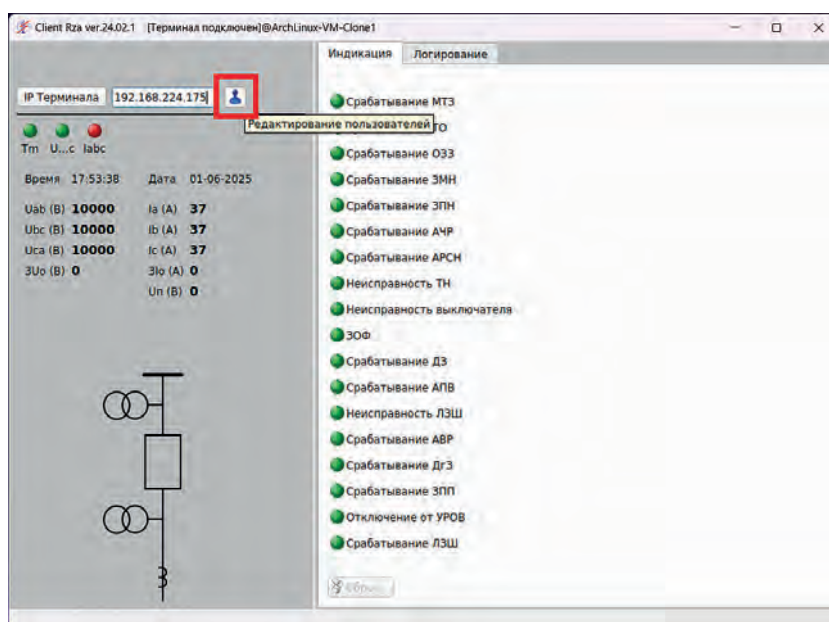


Рис. 9. Авторизация с ролью «Специалист по ИБ»

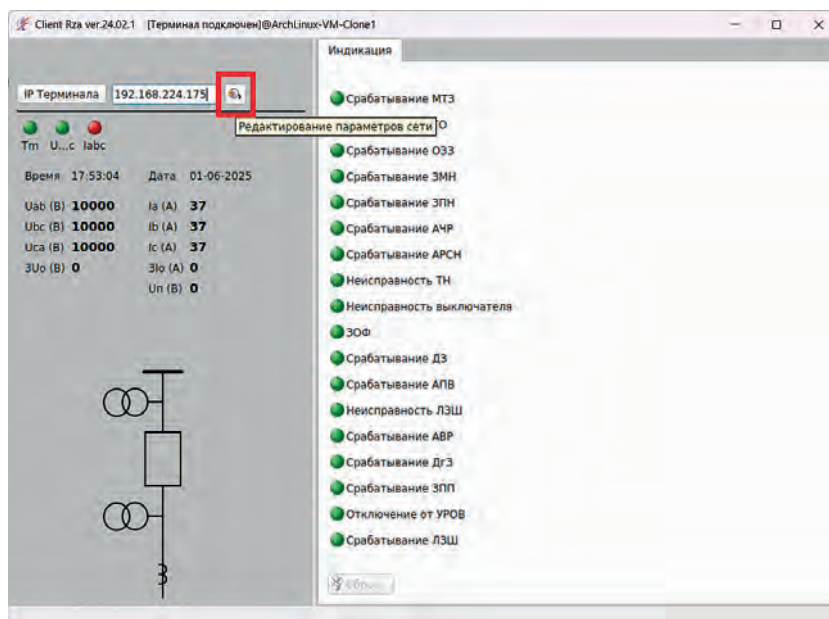


Рис. 10. Авторизация с ролью «Специалист по АСУ ТП»

На рис. 11 отображено ограничение прав для пользователя с ролью «Читатель» – редактирование параметров ПАК недоступно. Интерфейсные элементы (отмечены рамкой) заблокированы: они затемнены серым цветом, и изменить или сохранить настройки в них невозможно.

Работа пользователя с ПАК отслеживается системой: каждое действие, включая вход в систему, записывается в журнал событий (рис. 12).

Продемонстрированная реализация действующих отраслевых требований по ИБ в ПАК РЗА, была выполнена еще до выхода соответствующих локальных нормативных документов [2]. Это подтверждает

правильность подхода и соответствие современным трендам.

Тезисы для обсуждения

В настоящее время кроссплатформенные ПАК РЗА проходят обязательную сертификацию ФСТЭК по 4-му уровню доверия (4 УД), необходимую для их использования на объектах ЗОКИИ. В ходе сертификационных испытаний, проводимых аккредитованными лабораториями ФСТЭК, было отмечено, что внутренние нормативные документы ПАО «Россети» не полностью учитывают требования государственного регулятора, предъявляемые к созданию и эксплуатации объектов КИИ. Следовательно, они не могут

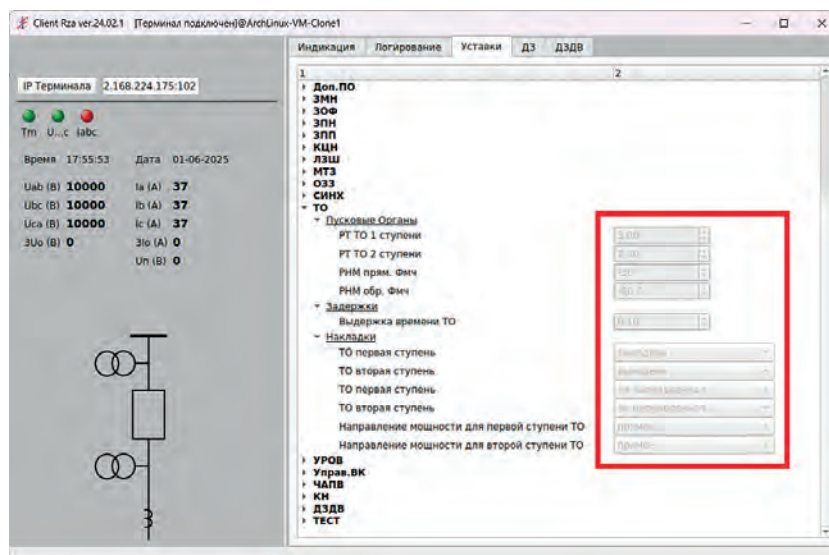


Рис. 11. Запрет редактирования параметров ПАК для роли «Читатель»

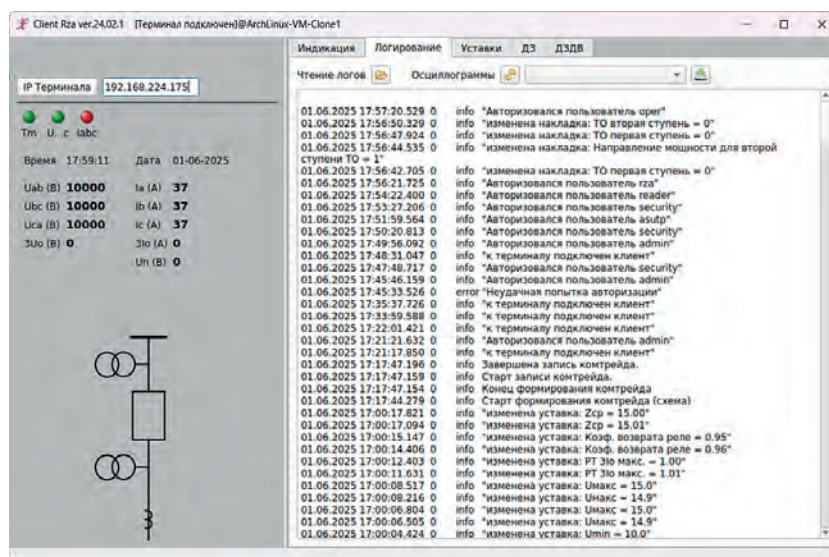


Рис. 12. Журнал событий ПАК (с событиями аудита)

служить основой для подтверждения соответствия программно-аппаратных средств АСТУ (включая ПАК РЗА), их функционально-технических характеристик и параметров требованиям ИБ. При этом обязательная сертификация ФСТЭК не заменяется отраслевой процедурой аттестации, проводимой ПАО «Россети».

На сколько известно, предприятия электроэнергетики не принимают участия в деятельности ТК 167 по разработке ГОСТ Р для ДПАК на ЗОКИИ. Отсутствие отраслевой экспертизы на этапе разработки ГОСТ Р не позволит в полной мере учесть отраслевые особенности в конечной редакции государственных стандартов.

Следует отметить, что дальнейшее развитие технологий ИБ связано с особенностями отраслевых объектов, технологических систем, то есть с учетом

отраслевой специфики электроэнергетики [22]. Необходимо разработка новых алгоритмов и методов выявления киберугроз, работающих на 6-м уровне абстракции (рис. 2), учитывающие электрофизические фундаментальные законы (закона Кирхгофа, Ома, электромагнитной индукции, теории вероятности и математической статистики), а также современные методы искусственного интеллекта и машинного обучения [22–26].

Заключение

1. Проведенный анализ нормативно-технических и регламентирующих документов в сфере ИБ свидетельствует о том, что государство переходит к жёсткому регулированию ИБ, превращая её из внутренней корпоративной задачи в строго контролируемую государством сферу.

2. Перспективным направлением для снижения технологической зависимости российской электроэнергетики и устранения рисков, обозначенных в обновлённой «Доктрине энергетической безопасности РФ», является применение доверенных отечественных аппаратно-программных решений. Такие платформы, основанные на российских микропроцессорах и сертифицированных ФСТЭК операционных системах, позволяют создавать универсальные ПАК для ЦПС различного назначения.
3. Процесс импортозамещения должен включать не только внедрение готовых решений, но и развитие НИОКР с участием ВУЗов и профильных научно-производственных предприятий с последующей стандартизацией. Ключевыми критериями должны быть технологическая независимость и снижение доли иностранных компонентов, а не ценовой фактор, поскольку для госкомпаний это не просто требование, а стратегическая государственная задача.
4. Указы Президента № 166 и № 250 установили чёткие сроки замены импортного ПО и оборудования на объектах КИИ в энергетике. Действующее законодательство обязывает госкомпанию планомерно проводить эту работу. Теперь информационная безопасность критической инфраструктуры – не внутренний вопрос организаций, а государственно регулируемая обязанность.
5. С учётом изменений в законодательстве и геополитической обстановке (с 2022 г.) необходимо пересмотреть единую техническую политику электросетевого комплекса. Особое внимание следует уделить импортозамещению, а также уточнить разделы, касающиеся инфобезопасности, учитывая статус центров питания как объектов КИИ.
6. Требуется уделить значительное внимание разработке алгоритмов кибербезопасности, базирующихся на выявлении кибернападений в технологических процессах генерирования, передачи, потребления электрической энергии, а также системах телеизмерения, телесигнализации и дистанционного управления.

Исследование выполнено за счет гранта Российского научного фонда № 25-29-00362, <https://rscf.ru/project/25-29-00362/>

The study was supported by a grant from the Russian Science Foundation No. 25-29-00362, <https://rscf.ru/project/25-29-00362/>

Литература

1. Зинин В. М. Актуальные решения НИПОМ с учётом доктрины энергетической безопасности РФ. // «РУМ» № 6(596), 2020. с. 38–45.
2. Куликов А. Л., Требования к информационной безопасности в электроэнергетике и их реализация в интеллектуальных устройствах цифровых подстанций / А. Л. Куликов, В. М. Зинин // Интеллектуальная электротехника. 2022. № 3(19). С. 49–78.
3. Израйлов К. Е. Оценивание и прогнозирование состояния сложных объектов: применение для информационной безопасности / К. Е. Израйлов, М. В. Буйневич, И. В. Котенко, В. А. Десницкий // Вопросы кибербезопасности. – 2022. – № 6(52). – С. 2–21. – DOI 10.21681/23113456-6-2022-2-21.
4. Лоскутов, А. А. Киберфизический стенд комплексных испытаний интеллектуальных электронных устройств защиты и автоматики / А. А. Лоскутов, А. Л. Куликов, И. А. Петров, А. С. Симанов // Электротехника. 2024. № 10. С. 49–60.
5. Котенко И. В., Федорченко Е. В., Новикова Е. С., Саенко И. Б., Данилов А. С. Методология сбора данных для анализа безопасности промышленных киберфизических систем / И. В. Котенко, Е. В. Федорченко, Е. С. Новикова, И. Б. Саенко, А. С. Данилов // Вопросы кибербезопасности. 2023. № 5(57). С. 69–79. DOI:10.21681/2311-3456-2023-5-69-79.
6. Rajkumar V. S. Cyber Attacks on Power Grids: Causes and Propagation of Cascading Failures / V. S. Rajkumar, A. Stefanov, A. Presekal, P. Palensky, J. R. Torres // IEEE Access, 11, 2023, 103154–103176. <https://doi.org/10.1109/ACCESS.2023.3317695>.
7. Wan, Y.; Cao, J. A Brief Survey of Recent Advances and Methodologies for the Security Control of Complex Cyber-Physical Networks. Sensors 2023, 23, 4013. <https://doi.org/10.3390/s23084013>.
8. Ding, S. Research on Network Security Measures in Electric Power Communication Network / S. Ding, F. Chen; H. Ye; X. Lu // 2024 International Conference on Electrical Drives, Power Electronics & Engineering (EDPEE). DOI: 10.1109/EDPEE61724.2024.00136.
9. Лоскутов, А. А. Современные методы формирования логической части цифровой защиты и автоматики систем электроснабжения промышленных потребителей: монография / А. А. Лоскутов, А. Л. Куликов; Нижегород. гос. техн. ун-т им. П. Е. Алексеева. – Нижний Новгород, 2025. – 392 с.
10. Богер, А. М. Математическая модель вектора DDOS-атаки на сетевую инфраструктуру АСУ ТП с использованием метода топологического преобразования стохастических сетей / А. М. Богер, А. Н. Соколов // Вопросы кибербезопасности. 2023. № 4(56). С. 72–79. DOI:10.21681/2311-3456-2023-4-72-79.
11. Гурина, Л. А. Разработка комплексного подхода к обеспечению кибербезопасности взаимосвязанных информационных систем при интеллектуальном управлении сообществом микросетей / Л. А. Гурина, Н. В. Томин // Вопросы кибербезопасности. 2023. № 4(56). С. 94–104. DOI:10.21681/2311-3456-2023-4-94-104.
12. Blazek P. Smart Grids Transmission Network Testbed: Design, Deployment, and Beyond / P. Blazek, A. Bohacik, R. Fudjak, V. Jurak, M. Ptacek // IEEE Open Journal of the Communications Society, Vol. 6, 13 December 2024, pp. 51–76. DOI: 10.1109/OJCOMS.2024.3517340.

13. Suhail Hussain S. M., Ustun T. S., Kalam A. A Review of IEC 62351 Security Mechanisms for IEC 61850 Message Exchanges // IEEE Transactions on Industrial Informatics, Vol. 16, Issue: 9, September 2020, pp. 5643–5654. DOI: 10.1109/TII.2019.2956734.
14. Rajkumar, V. S. Cyber Attacks on Power System Automation and Protection and Impact Analysis // V. S. Rajkumar, M. Tealane, A. Stefanov, A. Presekal, P. Palensky / 2020 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe) Virtual, October 26–28, 2020.
15. Gaspar, J. Smart Substation Communications and Cybersecurity: A Comprehensive Survey / J. Gaspar, T. Cruz, Ch-T. Lam, P. Simões // IEEE Communications surveys & tutorials, Vol. 25, no. 4, fourth quarter, 2023.
16. Lebedev A. A. Development the software complex for automated verification of the correct setup of the relay protection and automation complex on the digital substation / A. A. Lebedev, E. A. Voloshin, M. S. Malyutin, V. D. Uksekov, A. D. Guseynov, K. V. Latyshov // 2021 4th International Youth Scientific and Technical Conference on Relay Protection and Automation (RPA). DOI: 10.1109/RPA53216.2021.9628680.
17. Rubio, S. Smart Grid Protection, Automation and Control: Challenges and Opportunities / Rubio, S., Bogarra, S.; Nunes, M.; Gomez, X. // Appl. Sci. 2025, 15, 3186. <https://doi.org/10.3390/app15063186>.
18. Kabbara, N. Towards Software-Defined Protection, Automation, and Control in Power Systems: Concepts, State of the Art, and Future Challenges. / Kabbara, N., Nait Belaid, M. O., Gibescu, M., Camargo, L. R., Cantenot, J., Coste, T., Audebert, V., Morais, H. // Energies 2022, 15, 9362. <https://doi.org/10.3390/en15249362>.
19. Junho H. Cybersecurity of Sampled Value Messages in Substation Automation System / J. Hong, M. Girdhar; C-W Ten, S. Lee, S. Choi // 2022 IEEE Power & Energy Society General Meeting (PESGM). DOI: 10.1109/PESGM48719.2022.9916758.
20. Глухов А. П., Белова Е. И., Глухов А. А. Подход к оцениванию функциональности доверенных программно-аппаратных комплексов // Интеллектуальные технологии на транспорте. 2024. № 2(38). С. 69–76. DOI: 10.20295/2413-2527-2024-238-69-76.
21. Куликов А. Л., Зинин В. М., Шарафеев Т. Р. Принципы реализации кибербезопасных решений для кроссплатформенных интеллектуальных электронных устройств (ИЭУ) в составе цифровых подстанций (ЦПС) // В сборнике: Электроэнергетика в национальных проектах. Сборник статей. Под редакцией Н.Д. Рогалева. Москва, 2020. с. 167–177.
22. Лоскутов, А. А. Методы выявления киберугроз на цифровых подстанциях / А. А. Лоскутов, А. Л. Куликов, П. С. Пелевин, А. С. Симанов // Электротехника. 2025. № 10. С. 59–69.
23. Воеводин, В. А. О постановке задачи оценивания устойчивости функционирования объектов критической информационной инфраструктуры / В. А. Воеводин // Вопросы кибербезопасности. – 2025. – № 1(65). – С. 41–49. – DOI: 10.21681/2311-3456-2025-1-41-49.
24. Гурина, Л. А. Интеллектуальные методы обеспечения кибербезопасности мультиагентных систем управления микросетями / Л. А. Гурина, Н. В. Тomin // Вопросы кибербезопасности. 2023. – № 6(64). – С. 53–64. DOI: 10.21681/2311-3456-2024-6-53-64.
25. Израилов К. Е. Метод обнаружения атак различного генеза на сложные объекты на основе информации состояния. Часть 2. Алгоритм, модель и эксперимент/ К. Е. Израилов, М. В. Буйневич // Вопросы кибербезопасности. 2023. – № 4(56). – С. 80–93. DOI: 10.21681/2311-3456-2023-4-80-93.
26. Костогрызов А. И., Анализ угроз злоумышленной модификации модели машинного обучения для систем с искусственным интеллектом / А. И. Костогрызов, А. А. Нистратов // Вопросы кибербезопасности. 2023. № 5(57). С. 9–24. DOI:10.21681/2311-3456-2023-5-9-24.

REGULATORY AND TECHNICAL ASPECTS OF ENSURING INFORMATION AND CYBERSECURITY OF DIGITAL ELECTRIC POWER FACILITIES

Kulikov A. L.⁴, Loskutov A. A.⁵, Zinin V. M.⁶

Keywords: information security, cybersecurity, digital substation, critical information infrastructure, cyber threat, computer attack, relay protection and automation, technological information, intelligent electrical networks.

Purpose of the study: to analyze the state and maturity of the regulatory and legislative framework in the field of information security (IS) of digital electric power industry, to identify areas for further improvement of IS software and hardware systems, to implement industry IS requirements in experimental samples of relay protection and automation for digital substations.

Methods of research: systems analysis, analysis of scientific industry problem, classification, abstraction.

Result(s): of critical information infrastructure (CII) protection in the electric power industry; examines the degree of compliance of existing industry facilities with established legislative requirements for cybersecurity; outlines an approach to the development of cross-platform hardware and software systems (HSS) for digital substations (DSS) equipped with built-in information protection mechanisms, including a practical example - a relay protection and automation (RPA) device.

4 Aleksandr L. Kulikov, Dr.Sc., Professor, Professor of the Department of Electric Power Engineering, Power Supply and Power Electronics, NSTU named after R. E. Alekseev, Nizhny Novgorod, Russia, E-mail: inventor61@mail.ru

5 Anton A. Loskutov, Ph.D., Associate Professor, Associate Professor of the Department of Electric Power Engineering, Power Supply and Power Electronics, Senior Researcher of the Laboratory of Autonomous Hybrid Electric Power Complexes, NSTU named after R. E. Alekseev, Nizhny Novgorod, Russia, E-mail: loskutov_aa@nntu.ru

6 Vladimir M. Zinin, Head of Department, OOO NPP OVIST, Moscow, Russia, E-mail: c.nemo@yandex.ru

Scientific novelty: a technology for constructing trusted software and hardware systems for digital electric power engineering with integrated information security functions and the ability to integrate cybersecurity algorithms, providing for six levels of abstraction, is proposed.

References

- Zinin V. M. Aktual'nyye resheniya NIPOM s uchotom doktriny energeticheskoy bezopasnosti RF. // «RUM» № 6 (596), 2020. s. 38–45.
- Kulikov A. L., Trebovaniya k informatsionnoy bezopasnosti v elektroenergetike i ikh realizatsiya v intellektual'nykh ustroystvakh tsifrovyykh podstantsiy / A. L. Kulikov, V. M. Zinin // Intellektual'naya elektrotehnika. 2022. № 3(19). S. 49–78.
- Izrailov K. Ye. Assessment and prediction of the complex objects state: application for information security/ K. Ye. Izrailov, M. V. Buynevich, I. V. Kotenko, V. A. Desnitskiy // Voprosy kiberbezopasnosti. – 2022. – № 6(52). – S. 2–21. – DOI 10.21681/23113456-6-2022-2-21.
- Loskutov, A. A. Kiberfizicheskiy stend kompleksnykh ispytaniy intellektual'nykh elektronnykh ustroystv zashchity i avtomatiki / A. A. Loskutov, A. L. Kulikov, I. A. Petrov, A. S. Simanov // Elektrotehnika. 2024. № 10. S. 49–60.
- Kotenko I. V., Fedorchenko Ye. V, Novikova Ye. S., Sayenko I. B., Danilov A. C. Metodologiya sbora dannykh dlya analiza bezopasnosti promyshlennykh kiberfizicheskikh sistem / I. V. Kotenko, Ye. V. Fedorchenko, Ye. S. Novikova, I. B. Sayenko, A. C. Danilov // Voprosy kiberbezopasnosti. 2023. № 5(57). S. 69–79. DOI:10.21681/2311-3456-2023-5-69-79.
- Rajkumar V. S. Cyber Attacks on Power Grids: Causes and Propagation of Cascading Failures / V. S. Rajkumar, A. Stefanov, A. Presekal, P. Palensky, J. R. Torres // IEEE Access, 11, 2023, 103154–103176. <https://doi.org/10.1109/ACCESS.2023.3317695>.
- Wan, Y.; Cao, J. A Brief Survey of Recent Advances and Methodologies for the Security Control of Complex Cyber–Physical Networks. Sensors 2023, 23, 4013. <https://doi.org/10.3390/s23084013>.
- Ding, S. Research on Network Security Measures in Electric Power Communication Network / S. Ding, F. Chen; H. Ye; X. Lu // 2024 International Conference on Electrical Drives, Power Electronics & Engineering (EDPEE) DOI: 10.1109/EDPEE61724.2024.00136.
- Loskutov, A. A. Sovremennyye metody formirovaniya logicheskoy chasti tsifrovoy zashchity i avtomatiki sistem elektrosnabzheniya promyshlennykh potrebiteley: monografiya / A. A. Loskutov, A. L. Kulikov; Nizhegorod. gos. tekhn. un-t im. R. Ye. Alekseyeva. – Nizhniy Novgorod, 2025. – 392 s.
- Boger, A. M. Mathematical model of the vector of a ddos attack on the ics using the method of topological transformation of stochastic networks/ A. M. Boger, A. N. Sokolov // Voprosy kiberbezopasnosti. 2023. № 4(56). S. 72–79. DOI:10.21681/2311-3456-2023-4-72-79.
- Gurina, L. A. Development of an integrated approach to ensuring the cyber security of interconnected information systems under intelligent management of a microgrid community/ L. A. Gurina, N. V. Tomin // Voprosy kiberbezopasnosti. 2023. № 4(56). S. 94–104. DOI:10.21681/2311-3456-2023-4-94-104.
- Blazek P. Smart Grids Transmission Network Testbed: Design, Deployment, and Beyond / P. Blazek, A. Bohacik, R. Fudjak, V. Jurak, M. Ptacek // IEEE Open Journal of the Communications Society, Vol. 6, 13 December 2024, pp. 51–76. DOI: 10.1109/OJCOMS.2024.3517340.
- Suhail Hussain S. M., Ustun T. S., Kalam A. A Review of IEC 62351 Security Mechanisms for IEC 61850 Message Exchanges // IEEE Transactions on Industrial Informatics, Vol. 16, Issue: 9, September 2020, pp. 5643–5654. DOI: 10.1109/TII.2019.2956734.
- Rajkumar, V. S. Cyber Attacks on Power System Automation and Protection and Impact Analysis // V. S. Rajkumar, M. Tealane, A. Stefanov, A. Presekal, P. Palensky / 2020 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe) Virtual, October 26–28, 2020.
- Gaspar, J. Smart Substation Communications and Cybersecurity: A Comprehensive Survey / J. Gaspar, T. Cruz, Ch-T. Lam, P. Simões // IEEE Communications surveys & tutorials, Vol. 25, no. 4, fourth quarter, 2023.
- Lebedev A. A. Development the software complex for automated verification of the correct setup of the relay protection and automation complex on the digital substation / A. A. Lebedev, E. A. Voloshin, M. S. Malyutin, V. D. Uksekov, A. D. Guseynov, K. V. Latyshev // 2021 4th International Youth Scientific and Technical Conference on Relay Protection and Automation (RPA). DOI: 10.1109/RPA53216.2021.9628680.
- Rubio, S. Smart Grid Protection, Automation and Control: Challenges and Opportunities / Rubio, S., Bogarra, S.; Nunes, M.; Gomez, X. // Appl. Sci. 2025, 15, 3186. <https://doi.org/10.3390/app15063186>.
- Kabbara, N. Towards Software-Defined Protection, Automation, and Control in Power Systems: Concepts, State of the Art, and Future Challenges. / Kabbara, N., Nait Belaid, M. O., Gibescu, M., Camargo, L. R., Cantenot, J., Coste, T., Audebert, V., Morais, H. // Energies 2022, 15, 9362. <https://doi.org/10.3390/en15249362>.
- Junho H. Cybersecurity of Sampled Value Messages in Substation Automation System / J. Hong, M. Girdhar; C-W Ten, S. Lee, S. Choi // 2022 IEEE Power & Energy Society General Meeting (PESGM). DOI: 10.1109/PESGM48719.2022.9916758.
- Glukhov A. P., Belova Ye. I., Glukhov A. A. Podkhod k otsenivaniyu funktsional'nosti doverennykh programmno-apparatnykh kompleksov // Intellektual'nyye tekhnologii na transporte. 2024. № 2(38). S. 69–76. DOI: 10.20295/2413-2527-2024-238-69-76.
- Kulikov A. L., Zinin V. M., Sharafeyev T. R. Printsipy realizatsii kiberbezopasnykh resheniy dlya krossplatformennykh intellektual'nykh elektronnykh ustroystv (IEU) v sostave tsifrovyykh podstantsiy (TSPS) // V sbornike: Elektroenergetika v natsional'nykh proyektakh. Sbornik statey. Pod redaktsiyey N. D. Rogaleva. Moskva, 2020. s. 167–177.
- Loskutov, A. A. Metody vyavleniya kiberugroz na tsifrovyykh podstantsiyakh / A. A. Loskutov, A. L. Kulikov, P. S. Pelevin, A. S. Simanov // Elektrotehnika. 2025. № 10. S. 59–69.
- Voyevodin, V. A. On the formulation of the task of assessing the stability of the functioning of critical information infrastructure facilities / V. A. Voyevodin // Voprosy kiberbezopasnosti. – 2025. – № 1(65). – S. 41–49. – DOI 10.21681/2311-3456-2025-1-41-49.
- Gurina, L. A. Intelligent methods of ensuring cybersecurity multi-agent control system of microgrid/ L. A. Gurina, N. V. Tomin // Voprosy kiberbezopasnosti. 2023. – № 6(64). – S. 53–64. DOI: 10.21681/2311-3456-2024-6-53-64.
- Izrailov K. Ye. Different genesis attacks to complex objects detecting method based on condition information. Part 1. Prerequisites and schema/ K. Ye. Izrailov, M. V. Buynevich // Voprosy kiberbezopasnosti. 2023. – № 4(56). – S. 80–93. DOI: 10.21681/2311-3456-2023-4-80-93.
- Kostogryzov A. I., Threat analysis of malicious modification of the machine learning model for artificial intelligence systems / A. I. Kostogryzov, A. A. Nistratov // Voprosy kiberbezopasnosti. 2023. № 5(57). S. 9–24. DOI:10.21681/2311-3456-2023-5-9-24.

МЕТОД РАСПОЗНАВАНИЯ ЛИЧНОСТИ НА ОСНОВЕ БИОАКУСТИЧЕСКОГО СКАНИРОВАНИЯ В СИСТЕМАХ МНОГОФАКТОРНОЙ АУТЕНТИФИКАЦИИ

Частикова В. А.¹, Гуляй В. Г.²

DOI: 10.21681/2311-3456-2025-6-116-126

Цель исследования: теоретическое обоснование и разработка метода распознавания личности на основе биоакустического сканирования с целью реализации процесса многофакторной аутентификации в рамках одного локального устройства, а также минимизации ошибок I и II рода в многофакторных системах аутентификации.

Методы исследования: применены методы волновой теории физики, включающие аспекты распространения ультразвуковых волн и их взаимодействие с биологическим тканями. Также использованы статистические методы для оценки гипотез и алгоритмы машинного обучения в виде генеративно-сопоставительных нейронных сетей.

Научная новизна: разработан новый метод распознавания личности по данным биоакустического сканирования, характеризующим геометрию кости дистальной фаланги пальца и соотношение мягких тканей пальца. Предложенный биоакустический способ сканирования дистальной фаланги пальца, в отличие от ультразвукового сканирования папиллярного узора, работает на более низких частотах (15–17 МГц) и позволяет фиксировать уникальные характеристики не только кожного покрова, но и более глубоких тканей пальца. Геометрия кости дистальной фаланги, как и соотношение мягких тканей пальца, ранее не применялась в качестве аутентификационной характеристики личности.

Результат и практическая значимость: обоснован метод аутентификации личности путем биоакустического сканирования пальца, в рамках которого фиксируется геометрия кости дистальной фаланги пальца руки, включая выступы, выемки, гребешки и ногтевую бугристость, являющиеся уникальными для каждого человека, и взаимное расположение мягких тканей подушечки пальца с учетом их взаимного соотношения по объему. Практическая значимость предлагаемого способа обусловлена тем, что данный подход менее уязвим к подделкам и, соответственно, атакам с помощью муляжей, так как биометрические характеристики скрыты от глаз злоумышленника, в отличие от геометрии лица и изображения радужки глаза на фотографиях в открытом доступе, оставляемых на различных поверхностях отпечатков пальцев и т.д. Кроме того, подобный подход позволяет реализовать систему, осуществляющую процесс многофакторной аутентификации в рамках одного локального устройства (дактилоскопического сканера) с минимальным количеством каналов передачи данных.

Ключевые слова: аутентификация, ошибка I рода, ошибка II рода, отпечаток пальца, генеративно-сопоставительная нейронная сеть, спуфинг-атака, биометрический сканер, биоакустика.

Введение

В условиях активного развития цифровых технологий наблюдается рост количества киберугроз, включая фишинг, компрометацию учетных данных и атаки на системы аутентификации. Традиционные методы подтверждения личности, основанные на паролях и одноразовых кодах, не могут противостоять современным кибератакам.

В связи с этим особую роль в информационной безопасности играют методы аутентификации, базирующиеся на распознавании биометрических характеристик личности. Внедрение таких систем способно существенно повысить уровень защиты конфиденциальных данных и критически важных информационных систем, минимизируя риски несанкционированного доступа.

1. Распознавание личности по отпечатку пальца

Одним из самых распространенных способов распознавания личности является аутентификация

по отпечатку пальца [5]. Технология распознавания отпечатков пальцев анализирует уникальный рисунок папиллярных линий, уделяя особое внимание так называемым минуциям – точкам разветвления и окончания гребней.

Современные сканеры используют различные физические принципы [5, 15]: оптические системы фиксируют нарушение полного внутреннего отражения, емкостные датчики измеряют разницу в электрической емкости, а ультразвуковые сканеры анализируют отражение звуковых волн.

Аутентификация по отпечаткам пальцев представляет собой оптимальное сочетание надежности, удобства и доступности. В отличие от способов сканирования сетчатки глаза, она не требует дорогостоящего оборудования и сложных процедур идентификации [12]. По сравнению с распознаванием лица дактилоскопический метод обеспечивает значительно

1 Частикова Вера Аркадьевна, кандидат технических наук, доцент, доцент кафедры кибербезопасности и защиты информации КубГТУ, г. Краснодар, Россия. E-mail: chastikova_va@mail.ru

2 Гуляй Виктория Геннадьевна, студент института компьютерных систем и информационной безопасности КубГТУ, г. Краснодар, Россия. E-mail: ms.gulyay@bk.ru

более высокий уровень безопасности и не зависит от изменяющихся внешних факторов [15]. Однако все биометрические системы аутентификации личности, включая сканеры отпечатков пальца, подвержены возникновению ошибок I и II рода [7].

2. Ошибки I и II рода

В теории математической статистики проверки гипотез ошибки первого и второго рода отражают вероятности ошибочных решений [11] при принятии или отклонении нулевой гипотезы H_0 . Ошибка первого рода (α) возникает, когда нулевая гипотеза H_0 отвергается, несмотря на ее истинность:

$$\alpha = P(H_1 / H_0), \quad (1)$$

где α – вероятность ошибки первого рода, H_0 – нулевая гипотеза, H_1 – альтернативная гипотеза.

Данная ошибка также называется ложноположительным решением (false positive), а ее вероятность контролируется уровнем значимости теста.

Ошибка второго рода (β) представляет собой противоположную ситуацию, при которой альтернативная гипотеза H_1 не отвергается, несмотря на ее ложность:

$$\beta = P(H_0 / H_1), \quad (2)$$

где β – вероятность ошибки второго рода, H_0 – нулевая гипотеза, H_1 – альтернативная гипотеза.

Это ложноотрицательное решение (false negative), и его вероятность связана с мощностью критерия ($1 - \beta$).

Важно отметить, что между α и β существует обратная зависимость: уменьшение вероятности одной ошибки, как правило, ведет к увеличению вероятности другой при фиксированном объеме выборки.

В контексте биометрических систем аутентификации, в том числе систем дактилоскопического сканирования, ошибки первого и второго рода приобретают специфическую интерпретацию, связанную с безопасностью и удобством пользователей.

Ошибка первого рода обозначается как False Rejection Rate (FRR) и характеризует вероятность того, что система не распознает зарегистрированного пользователя, ошибочно отвергая его попытку аутентификации. Высокий FRR снижает удобство эксплуатации системы, поскольку легитимные пользователи сталкиваются с неоправданными отказами в доступе.

Ошибка второго рода, именуемая False Acceptance Rate (FAR), представляет более серьезную угрозу, так как отражает вероятность ошибочного предоставления доступа постороннему лицу. В системах с высокими требованиями к безопасности, таких как банковские приложения или защищенные

правительственные объекты, минимизация FAR является критически важной задачей.

Однако снижение FAR неизбежно приводит к увеличению FRR, что требует поиска оптимального баланса между безопасностью и удобством использования подобных систем.

3. Спуфинг-атаки с применением генеративного искусственного интеллекта (ИИ)

Наравне с развивающимися технологиями защиты прогрессируют и способы атак на системы биометрической аутентификации. Так в работе [2] приводится способ реализации спуфинг-атаки путем генерации рисунка отпечатка пальца.

Данный метод атаки на системы биометрической аутентификации становится возможным благодаря тому, что сканеры отпечатка пальца, независимо от их типа, фиксируют в качестве биометрического шаблона не весь папиллярный узор отпечатка, а только набор опорных точек – минуций, которые представляют собой уникальные особенности рельефа кожи в виде окончаний, разветвлений или изгибов папиллярных линий (рис. 1).



Рис. 1. Минуции папиллярного узора

Минуции являются ключевыми элементами аутентификационной информации, поскольку их взаимное расположение и количество формируют индивидуальный паттерн, устойчивый к изменениям в течение жизни человека. Современные алгоритмы обработки изображений преобразуют эти точки в цифровые шаблоны, позволяя осуществлять быструю и точную верификацию личности при сравнительно небольшом объеме обрабатываемых данных.

Однако алгоритм, описанный в [2], позволяет реализовывать потенциальные угрозы безопасности, связанные с использованием частичных отпечатков пальцев в системах аутентификации, таких как смартфоны и другие мобильные устройства. Из-за небольших размеров датчиков пользователи регистрируют несколько частичных отпечатков (рис. 2),

что повышает удобство, но снижает безопасность: система принимает любой из сохранённых шаблонов.



Рис. 2. Пример частичных отпечатков [2]

Авторы предлагают метод генерации «мастер-отпечатков» — шаблонов, совпадающих с множеством реальных отпечатков. Вероятность их обнаружения выше в частичных данных из-за снижения энтропии и роста частоты ложных совпадений (FAR).

Эксперименты на наборах FVC2002 DB1-A (отпечатки, полученные с оптических датчиков) и Finger-Pass DB7 (отпечатки, полученные с ёмкостных датчиков) показали, что мастер-отпечатки совпадают с 4–12 % шаблонов. Например, при FMR 0,1 % один шаблон совпал с 1,4 % частичных отпечатков (10,6 % пользователей). Использование пяти мастер-отпечатков позволило атаковать до 26,5 % (ёмкостные датчики) и 65,2 % (оптические датчики) пользователей за 5 попыток.

Исследование подтверждает уязвимость систем с частичными отпечатками к атакам по словарю, особенно при увеличении числа регистрируемых шаблонов. Даже при низком FAR (0,01 %) риск сохраняется, что требует пересмотра подходов к проектированию биометрических систем.

В статье [3] приводится усовершенствованный метод DeepMasterPrint, который представляет собой доработанный подход к генерации синтетических изображений отпечатков пальцев (рис. 3). В отличие от предыдущих разработок, ограничивавшихся созданием шаблонов минуций, данный метод работает

с полноценными изображениями, что значительно расширяет его практическую применимость. Основу технологии составляет комбинация генеративно-состязательных сетей и эволюционной оптимизации, позволяющая создавать высокоэффективные мастер-отпечатки.

Процесс создания мастер-отпечатков DeepMaster-Print начинается с обучения генеративно-состязательной сети Вассерштайна на реальных изображениях отпечатков пальцев, где генератор постепенно учится создавать визуально достоверные образцы. Особенность подхода заключается в последующем применении стратегии эволюционной оптимизации CMA-ES, которая целенаправленно ищет такие значения скрытых параметров генератора, при которых создаваемые изображения будут максимально часто совпадать с различными отпечатками из базы данных.

Практические испытания метода продемонстрировали его результативность. При стандартных настройках безопасности (FMR 1 %) один сгенерированный DeepMasterPrint способен успешно имитировать до 77 % отпечатков в тестовой выборке. Важно отметить, что технология сохраняет эффективность при работе с различными системами верификации, включая VeriFinger, Bozorth3 и Innovatrics, что подчеркивает ее универсальность.

Данный способ позволяет осуществлять атаки как на канал передачи данных на сервер обработки, так и непосредственно через считывающую поверхность сканеров, так как данный способ работает на уровне изображений, в отличие от своего предшественника, базирующегося на генерации минуций.

В связи с возрастающими возможностями злоумышленников в виде развития генеративного искусственного интеллекта увеличивается и вероятность реализации подобного рода атак [9].

4. Способы предотвращения спуфинг-атак в дактилоскопических сканерах

Основным способом предотвращения возникновения ошибок II рода является внедрение систем многофакторной аутентификации (МФА) [6]. Однако



Рис. 3. Результат работы алгоритма DeepMasterPrint, генерирующего папиллярные узоры отпечатков пальца [3]

несмотря на повышенную надежность по сравнению с однофакторными методами, внедрение МФА сопряжено с рядом уязвимостей, наиболее критичными из которых являются риски, связанные с увеличением числа каналов передачи аутентификационных данных. Основная проблема заключается в том, что каждый дополнительный фактор аутентификации формирует новый канал коммуникации, потенциально подверженный атакам.

Более того, интеграция разнородных технологий верификации приводит к усложнению архитектуры системы, что увеличивает вероятность ошибок конфигурации и создает дополнительные векторы для эксплуатации уязвимостей. Например, несогласованность политик безопасности между разными модулями аутентификации может позволить злоумышленнику обойти один из факторов, используя слабое звено в цепочке проверки.

Таким образом, несмотря на теоретическое повышение уровня защиты, многофакторные системы могут демонстрировать сниженную эффективность в реальных условиях из-за уязвимостей в дополнительных каналах передачи данных. В связи с этим возникает необходимость создания системы, реализующей процесс многофакторной аутентификации в рамках одного локального устройства с минимальным количеством каналов передачи данных. В случае применения систем аутентификации по отпечатку пальца необходимо фиксировать дополнительные биометрические характеристики личности в рамках дактилоскопического сканера. Это становится возможным, так как помимо папиллярного узора палец руки обладает и другими не менее уникальными характеристиками – геометрией кости, соотношением различных мягких тканей и другими [1, 13].

5. Обоснование способа биоакустического сканирования

5.1. Геометрия кости

Каждый из пяти пальцев руки человека состоит из отдельных подвижных частей – фаланг, которые представляют собой небольшие трубчатые кости. Как часть скелета человека, палец состоит из трех сегментов: проксимального, среднего и дистального.

Геометрия каждого из этих сегментов определяется рядом особенностей, таких как форма, размер, углы и длина кости.

Современные исследования подтверждают, что каждая кость фаланги пальца обладает уникальными анатомическими особенностями – наличием выступов, борозд, ямок и других элементов, которые делают их различимыми от тех же костей других людей. С возрастом в костной ткани происходят изменения, такие как снижение плотности и минерализации, однако эти процессы не влияют на эффективность аутентификации, т.к. их можно предусмотреть заранее и учесть в ходе обучения системы распознавания личности с перспективой анализа данных в течении длительного промежутка времени (более 5 лет).

Исследования [7, 21] доказывают, что с ростом плотности губчатой костной ткани влияние минерального содержания на растягивающие напряжения и деформации снижается, тогда как его воздействие на сжимающие напряжения, напротив, усиливается. Учет этих данных в самообучающихся системах распознавания личности позволит проводить высокоточную аутентификацию личности на протяжении всей жизни человека.

5.2. Мягкие ткани

Палец человека представляет собой сложную биомеханическую систему, состоящую из множества взаимосвязанных тканей, каждая из которых выполняет уникальную и жизненно важную функцию. Разнообразие мягких тканей пальца руки представлено на рисунке 4.

Бондаренко И. Н. было проведено исследование мягких тканей пальцев рук [4] среди 63 испытуемых в возрасте от 22 до 59 лет на аппарате экспертного класса MyLabTwice (Esaote, Италия) с линейным датчиком SL3116 на частотном диапазоне 10–22 МГц в В-режиме, режиме цветового доплеровского картирования (ЦДК) и режиме microV, на гелевой подушке, в стандартных точках при одинаковых настройках аппарата. Исследование показало, что с возрастом средняя толщина мягких тканей, в т.ч. кожных

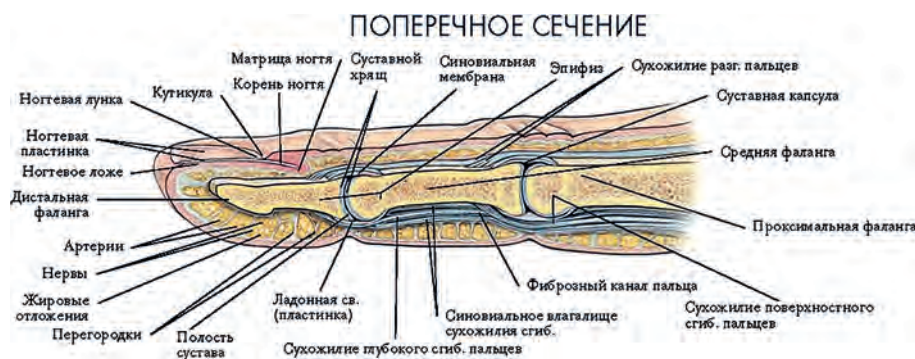


Рис. 4. Разнообразие тканей в продольном сечении пальца руки

покровов, уменьшается, однако данный показатель является весьма индивидуальным и зависит от возрастной группы только в усредненном значении [22]. Аналогично и другие параметры мягких тканей, такие как их плотность, являются индивидуальными, а следовательно, могут использоваться в верификационных целях.

Аутентификация по одному из параметров мягких тканей не будет однозначно определять личность пользователя. Но совокупность таких параметров может достаточно однозначно подтверждать личность человека.

5.3. Сравнительный анализ различных тканей пальца

Как уже было сказано выше, палец руки обладает большим разнообразием тканей, которые различаются по плотности, структуре, функции и механическим свойствам. В таблице 1 отражены такие характеристики тканей, как плотность, прочность и эластичность.

Различия в структуре тканей, представленные в таблице 1, лучше всего видны при ультразвуковом сканировании, так как скорость и отражение ультразвуковых волн напрямую зависят от плотности и упругости ткани. Кость, мышцы, жир и сосуды по-разному проводят и рассеивают ультразвук, создавая четкую картину внутреннего строения пальца, позволяющую проводить биоакустическую аутентификацию личности.

5.4. Особенности распространения ультразвука в тканях

Ультразвуковые волны обладают уникальными свойствами, которые зависят от характеристик среды [16]. Скорость распространения ультразвука определяется плотностью и упругостью среды, а также внутренней кинетической энергией молекул. Совокупность данных параметров влияет на то, как быстро волна может распространяться в различных структурах.

В однородных изотропных средах распространение ультразвуковых волн подчиняется волновому уравнению (3)

$$\nabla^2 p - \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2} = 0, \quad (3)$$

где p – акустическое давление, c – скорость звука в среде.

Скорость распространения определяется плотностью и коэффициентом упругости среды (4)

$$C = \sqrt{\frac{K}{\rho}}, \quad (4)$$

где K – коэффициент упругости среды, ρ – плотность среды.

В жидкостях и газах УЗ-волны являются исключительно продольными, тогда как в твёрдых телах возможно возникновение поперечных и поверхностных волн.

Ультразвуковые волны способны отражаться, преломляться и рассеиваться на границах сред с разными акустическими свойствами, что позволяет с высокой точностью идентифицировать те или иные анатомические особенности строения пальца субъекта данных.

Отражение ультразвуковых волн представляет собой явление, при котором часть энергии ультразвуковой волны возвращается в исходную среду при её падении на границу раздела с другой средой, обладающей иным акустическим импедансом (5)

$$Z = \rho c, \quad (5)$$

где Z – акустический импеданс, ρ – плотность среды, c – скорость звука в среде.

Коэффициент отражения R зависит от соотношения импедансов сред и подчиняется закону сохранения энергии (6)

$$R = \left(\frac{Z_2 - Z_1}{Z_2 + Z_1} \right)^2, \quad (6)$$

Таблица 1.

Сравнительный анализ структуры различных тканей пальца

Ткань	Плотность	Прочность	Эластичность
Кость	очень высокая	очень высокая	низкая
Сухожилия	высокая	очень высокая	низкая
Связки	средне-высокая	Высокая	умеренная
Хрящи	средняя	Средняя	высокая
Кожа (эпидермис)	умеренная	Средняя	средняя
Мышцы (скелетные)	умеренная	Средняя	высокая
Нервы	низкая	Низкая	средняя
Кровеносные сосуды	низкая	Низкая	высокая
Жировая ткань	очень низкая	очень низкая	высокая

где Z_1 и Z_2 — акустические импедансы первой и второй сред соответственно.

Преломление ультразвуковых волн возникает при изменении скорости их распространения на границе сред. Угол преломления связан с углом падения и скоростями звука в соответствующих средах законом Снеллиуса (7)

$$\frac{\sin \theta_1}{c_1} = \frac{\sin \theta_2}{c_2}, \quad (7)$$

где θ_1 — угол падения, θ_2 — угол преломления; c_1 , c_2 — скорости распространения волн в среде.

Критический угол полного внутреннего отражения достигается при условии $\theta_2 = 90^\circ$.

Рассеивание ультразвуковых волн обусловлено наличием неоднородностей в среде, таких как частицы, пузырьки газа или границы зёрен в поликристаллических материалах. Интенсивность рассеянного сигнала зависит от соотношения длины волны (λ) и размера неоднородностей (d):

- Рэлеевское рассеяние наблюдается при $d < \lambda$ (объекты малы по сравнению с длиной волны) и характеризуется зависимостью интенсивности от λ^{-4} ;
- Диффузное рассеяние возникает при $d \sim \lambda$ (соизмеримы d с длиной волны неоднородности) и имеет сложную зависимость от микроструктуры среды.

Интенсивность рассеянного излучения определяется по формуле (8):

$$I_s \approx \frac{d^6}{\lambda^4}, \quad (8)$$

где I_s — интенсивность рассеянной волны, d — размер рассеивающего объекта, λ — длина ультразвуковой волны.

Затухание ультразвуковых волн проявляется в виде уменьшения амплитуды и энергии волны по мере её

распространения в среде, вызванное поглощением (диссипацией энергии в тепло) и рассеянием (изменением направления распространения части энергии). Затухание описывается экспоненциальной зависимостью и характеризуется суммарным коэффициентом затухания (9)

$$\alpha_{\text{общ}} = \alpha + \mu_s, \quad (9)$$

где $\alpha_{\text{общ}}$ — суммарный коэффициент затухания, α — коэффициент поглощения, μ_s — коэффициент рассеяния.

Экспоненциальный закон затухания представлен формулой (10)

$$I(x) = -I_0 e^{-\alpha_{\text{общ}} x}, \quad (10)$$

где $I(x)$ — интенсивность волны на расстоянии x , I_0 — начальная интенсивность, $\alpha_{\text{общ}}$ — суммарный коэффициент затухания.

Затухание ультразвуковых волн обусловлено множеством факторов, включая плотность, упругость, вязкость среды, наличие и размер частиц, а также рассеяние и поглощение звука. Эти эффекты особенно заметны в газах и жидкостях из-за их молекулярной структуры и плотности [17].

Описанные выше эффекты, возникающие в ходе распространения ультразвуковых волн, позволяют формализовать поставленную задачу и создать математическую модель системы регистрации биометрических характеристик пальца человека на основе ультразвукового сканирования.

5.5. Отображение различных тканей пальца при ультразвуковом сканировании

В ходе ультразвукового сканирования ткани классифицируются по степени эхогенности, отражающей их способность к взаимодействию с ультразвуковыми волнами (рис. 6).

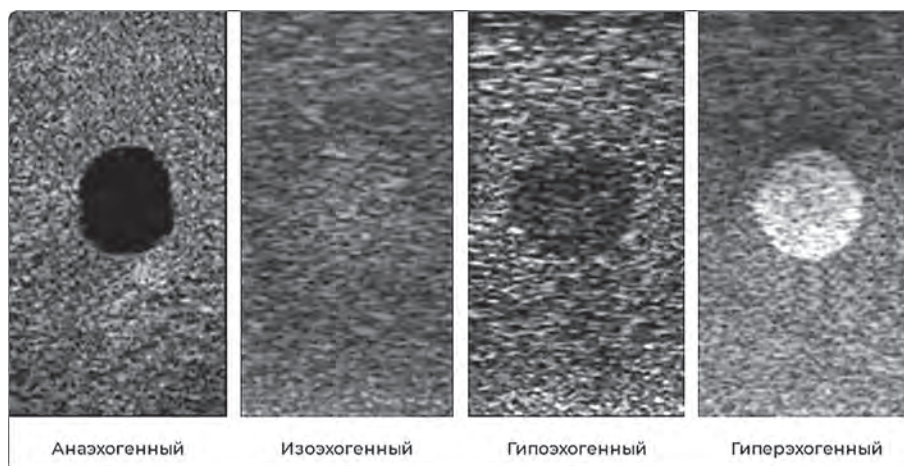


Рис. 6. Степени эхогенности различных структур здоровых тканей

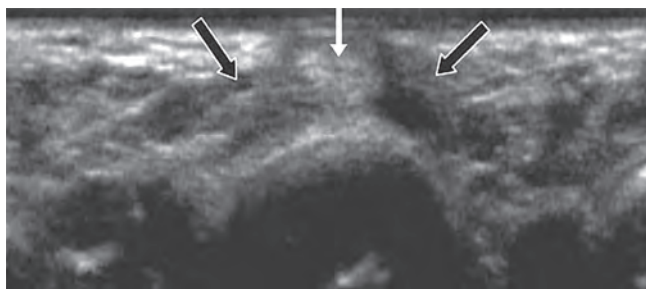


Рис. 7. Сухожилие разгибателя пальца (светлая стрелка) и коллатеральные связки (темные стрелки)

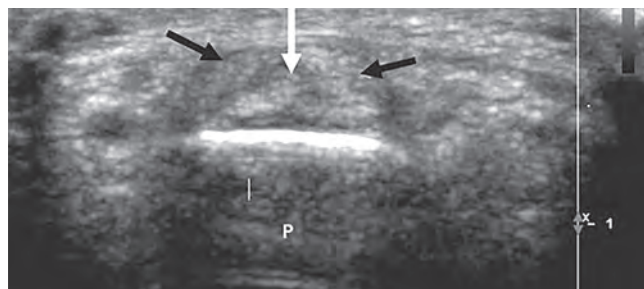


Рис. 8. Зона разделения сухожилия в поперечном сечении

По степени эхогенности выделяют следующие типы структур:

1. Анэхогенные зоны полностью лишены способности отражать ультразвук, что проявляется абсолютно чёрным изображением. Данная характеристика типична для жидкостных образований, включая кровеносные сосуды.
2. Изоэхогенные образования демонстрируют акустические свойства, сопоставимые с окружающими тканями, отображаясь на эхограммах в виде участков средней серой интенсивности. Характерным примером таких структур является мышечная ткань.
3. Гипоэхогенные структуры характеризуются пониженной акустической плотностью, что приводит к слабому отражению ультразвука, их визуальное отображение представляет собой тёмно-серые участки. К этой категории относятся, например, связки.
4. Гиперэхогенные структуры отличаются высокой акустической плотностью, что проявляется интенсивным отражением ультразвуковых волн. На эхограммах они визуализируются как яркие, приближающиеся к белому цвету, сигналы. Типичными представителями данной группы являются сухожилия и кортикальный слой костей.
5. Гетерогенные структуры представляют особый диагностический интерес, поскольку их неоднородная эхогенность, проявляющаяся сочетанием участков различной акустической плотности, часто свидетельствует о патологических процессах.

Ультразвуковое сканирование мягкотканых структур пальца руки, включая сухожилия, нервы и связочный аппарат, является высокоинформативным [18] и безопасным [10] методом благодаря использованию современных мультисекторных линейных датчиков с диапазоном частот от 7 до 17 МГц, что обеспечивает высокое пространственное разрешение.

Современные ультразвуковые технологии демонстрируют высокую разрешающую способность [19], достигающую 0,2–0,3 мм. Наибольший интерес представляют крестообразные связки сухожилий толщиной

0,2–0,4 мм и коллатеральные связки размером 0,3–0,6 мм, которые четко дифференцируются при ультразвуковом исследовании (рис. 7).

При проведении поперечного сканирования дистальных фаланг пальцев наблюдается характерная эхографическая картина с четким разграничением тканевых структур – сухожилия визуализируются как гиперэхогенные линейные, в то время как связочный аппарат определяется как гипоэхогенные образования (рис. 8).

Продольное сканирование позволяет детально оценить анатомию сухожильно-связочного комплекса. Четко дифференцируются сухожилия сгибателей толщиной 1,2–2,5 мм с сохраненной фибриллярной структурой, система кольцевидных связок (0,3–0,6 мм), а также блок связок A2, являющийся важным анатомическим ориентиром на уровне проксимальной фаланги. Особое внимание заслуживает зона разделения сухожилия поверхностного сгибателя, где отчетливо прослеживаются две расходящиеся ножки гиперэхогенной структуры и расположенное между ними сухожилие глубокого сгибателя. Наиболее тонкими визуализируемыми элементами являются крестообразные связки сухожилий (0,2–0,4 мм). При проведении поперечного ультразвукового сканирования дистальных фаланг пальцев кисти определяется характерная эхографическая картина, представленная двумя дифференцированными структурными компонентами с различными акустическими свойствами (рис. 9).

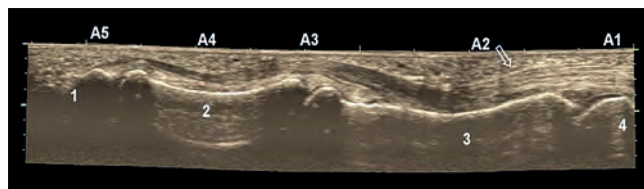


Рис. 9. Результат сканирования пальца в продольном сечении

Представленные образцы ультразвукового исследования демонстрируют возможные результаты биоакустического сканирования. Многообразие тканей

пальца, отличающихся экзогенностью структур [20], открывает широкие перспективы для использования данной методики в системах аутентификации личности.

5.6. Биоакустическое сканирование

Для реализации биоакустического сканирования дистальной фаланги пальца с помощью ультразвуковых волн предложено устройство [23], основой которого является линейный ультразвуковой датчик, работающий на частоте 15 МГц. Его конструкция с линейным расположением 64-256 пьезоэлектрических кристаллов создаёт идеально параллельные ультразвуковые волны, формирующие чёткое прямоугольное поле обзора без геометрических искажений. Схема аппаратной составляющей данного устройства представлена на рисунке 10.

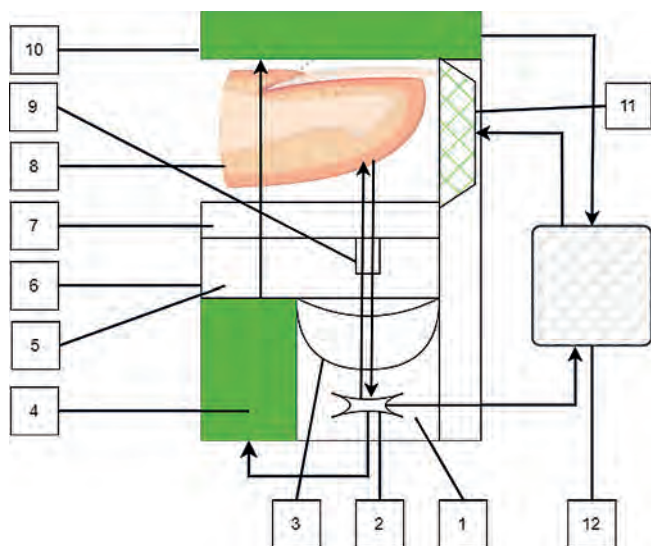


Рис. 10. Схема аппаратной составляющей устройства [23]

Благодаря высокой частоте и плотному расположению элементов датчик обеспечивает высокую детализацию поверхностных структур [20], позволяя различать микроскопические сосуды и тончайшие слои тканей. Проходя через различные структуры (костную, мышечную, жировую ткани), ультразвуковой сигнал подвергается характерным изменениям [22], отражающим уникальные анатомические особенности конкретного индивидуума. Приемная часть устройства фиксирует эти изменения, создавая

своеобразный «акустический портрет» внутреннего строения пальца.

Заключительный этап работы устройства заключается в комплексном анализе полученных данных. Алгоритмы на базе машинного обучения сравнивают глубинные параметры тканевой структуры с эталонными образцами, хранящимися в базе данных.

Способ биоакустического сканирования может быть реализован как часть системы многофакторной аутентификации в качестве одного из факторов подтверждения личности [28]. Применение данного способа особенно актуально при компрометации основного биометрического признака, в частности – отпечатка пальца (рис. 11).

Также важно отметить, благодаря непрерывному обучению нейросетевого программного комплекса [25] система сможет учитывать возрастные изменения тканей, избегая тем самым ложных отказов в доступе легитимным пользователям.

Заключение

В ходе проведенного исследования предложена схема устройства регистрации биометрических характеристик пальца человека в рамках биоакустического сканирования. Данный способ может выступать в качестве дополнительного фактора аутентификации в системах распознавания личности, повышая их защищенность и эффективность аутентификации, так как описанный метод биоакустического сканирования дистальной фаланги пальца, в отличие от ультразвукового сканирования папиллярного узора, работает на более низких частотах (15–17 МГц) и позволяет фиксировать уникальные характеристики не только кожного покрова, но и более глубоких тканей пальца.

В отличие от традиционных методов, основанных на визуальных характеристиках (геометрии лица и изображения радужки глаза на фотографиях в открытом доступе, оставляемых на различных поверхностях отпечатков пальцев и т.д.), предложенная технология использует скрытые от прямого наблюдения параметры, что кардинально снижает вероятность их несанкционированного копирования. Дополнительным преимуществом является устойчивость к внешним факторам – загрязнению поверхности кожи, микротравмам эпидермиса и т.д.

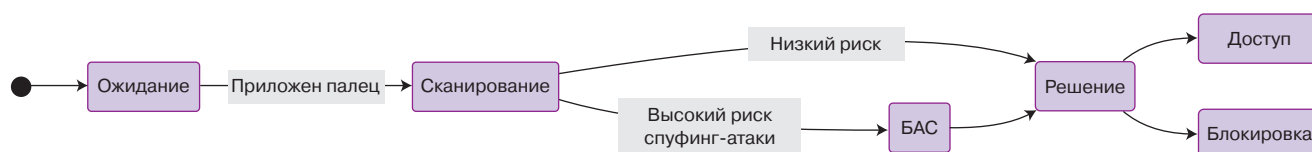


Рис. 11. Схема алгоритма разрабатываемого способа

Литература

1. Sim J. Y. Identity Recognition Based on Bioacoustics of Human Body / Sim J. Y., Noh H. W., Goo W., Kim N., Chae S.-H. and Ahn C.-G. // IEEE Transactions on Cybernetics, vol. 51, no. 5, pp. 2761–2772, May 2021, DOI: 10.1109/TCYB.2019.2941281.
2. Roy A. Masterprint: exploring the vulnerability of partial fingerprint-based authentication systems / Roy A., Memon N. and Ross A. // IEEE Transactions on Information Forensics and Security, vol. 12, no. 9, September 2017.
3. Bontrager P. DeepMasterPrints: generating MasterPrints for dictionary attacks via latent variable evolution / Bontrager P., Roy A., Togelius J., Memon N. and Ross A. // 2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS), Redondo Beach, CA, USA, 2018, pp. 1–9, DOI: 10.1109/BTAS.2018.8698539.
4. Бондаренко И. Н. Ультразвуковая характеристика кожи, мягких тканей лица, шеи, кистей рук у женщин разных возрастных групп: нерандомизированное обсервационное кросс-секционное исследование // Кубанский научный медицинский вестник. – 2021. – Т. 28, № 3. – С. 16–28. – DOI 10.25207/1608-6228-2021-28-3-16-28.
5. Частикова В. А. Особенности реализации биометрической системы аутентификации личности на основе папиллярного узора / В. А. Частикова, В. Г. Гуляй, Д. С. Сучков // Молодежная наука. Сборник лучших научных работ молодых ученых: Материалы LI студенческой научной конференции, Краснодар, 29 февраля 2024 года. – Краснодар: Кубанский государственный технологический университет, 2024. – С. 222–225. – EDN RDBTAL.
6. Чайковская Т. В. Исследование напряженно-деформированного состояния кости средней фаланги указательного пальца руки человека при изменении плотности губчатой костной ткани / Т. В. Чайковская, Е. С. Марченко // Фундаментальные основы механики. – 2022. – № 10. – С. 119–122. – DOI 10.26160/2542-0127-2022-10-119-122. – EDN JAAFKI.
7. Частикова В. А. Предотвращение ошибок I рода при сканировании в системах контроля и управления доступом путем использования GAN-сети / В. А. Частикова, В. Г. Гуляй, Д. С. Сучков // Искусственный интеллект. Формирование будущего: Материалы I Международной научно-практической конференции, Краснодар, 29 апреля 2024 года. – Краснодар: ИП Алзидан М., 2024. – С. 296–299. – EDN ESSLTN.
8. Красов А. В. Разработка методологии тестирования систем защиты информации в виртуальных комплексах для обнаружения ошибок I и II-рода / А. В. Красов, Р. Р. Максудова, В. В. Нефедов [и др.] // Вестник Санкт-Петербургского государственного университета технологии и дизайна. Серия 1: Естественные и технические науки. – 2021. – № 1. – С. 45–52. – DOI 10.46418/2079-8199_2021_1_7.
9. Частикова В. А. Подход к построению систем анализа инцидентов информационной безопасности на основе гибридизации методов машинного обучения / В. А. Частикова, В. Г. Гуляй // Электронный сетевой политематический журнал «Научные труды КубГТУ». – 2023. – № 6. – С. 107–117. – EDN YPTPGQ.
10. Suvorov A. Yu. Statistical hypothesis testing: general approach in medical research / A. Yu. Suvorov, N. M. Bulanov, A. N. Shvedova [et al.] // Sechenov Medical Journal. – 2022. – Vol. 13, No. 1. – P. 4–13. – DOI 10.47093/2218-7332.2022.426.08.
11. Исмагилов Р. Ф. Конструирование модели обучающей нейронной сети для биометрической многофакторной аутентификации пользователя информационной системы / Р. Ф. Исмагилов, Н. Д. Лушников, А. С. Исмагилова // Вопросы защиты информации. – 2023. – № 1(140). – С. 19–23. – DOI 10.52190/2073-2600_2023_1_19.
12. Белов Н. И. Применение термографической модели лица в задачах идентификации и аутентификации / Н. И. Белов, В. М. Коржук // Computational Nanotechnology. – 2023. – Т. 10, № 3. – С. 109–120. – DOI 10.33693/2313-223X-2023-10-3-109-120.
13. Славолубова И. А. Дифференцирующие возможности папиллярных узоров средних фаланг пальцев рук (на примере евразийских популяций) / И. А. Славолубова, А. М. Юдина, Л. Ю. Шпак // Вестник Московского университета. Серия 23: Антропология. – 2023. – № 4. – С. 58–68. – DOI 10.32521/2074-8132.2023.4.058-068.
14. Частикова В. А. Нейросетевая система биометрической идентификации личности по голосу / В. А. Частикова, С. А. Жерлицын, Д. О. Войлова // Вестник Адыгейского государственного университета. Серия 4: Естественно-математические и технические науки. – 2023. – № 1(316). – С. 70–79. – DOI 10.53598/2410-3225-2023-1-316-70-79.
15. Пожар В. Э. Спектрально-поляризационные системы трёхмерного технического зрения на основе акустооптической фильтрации / В. Э. Пожар, А. С. Мачихин // Светотехника. – 2022. – № 5. – С. 32–35. – EDN BHPGOK.
16. Кормилина А. Р. Ультразвуковая эластография сдвиговой волны в оценке жёсткости костной мозоли / А. Р. Кормилина, М. Г. Тухба-туллин // Российский электронный журнал лучевой диагностики. – 2020. – Т. 10, № 2. – С. 122–128. – DOI 10.21569/2222-7415-2020-10-2-122-128.
17. Кольцова И. С. Акустоэлектронный метод измерения скорости ультразвуковых волн в жидкостях / И. С. Кольцова, А. С. Хомутова // Акустический журнал. – 2023. – Т. 69, № 3. – С. 317–321. – DOI 10.31857/S032079192260041X.
18. Романенко К. В. Ультразвуковое исследование кожи / К. В. Романенко, О. О. Боровая, Н. В. Ермилова, В. В. Захарова // Тор-суевские чтения: научно-практический журнал по дерматологии, венерологии и косметологии. – 2021. – № 2(32). – С. 17–20. – EDN XSCBLX.
19. Потехин Д. С. Современные тенденции развития ультразвуковой диагностики / Д. С. Потехин, П. В. Комаров // Высокопроизводительные вычислительные системы и технологии. – 2022. – Т. 6, № 2. – С. 71–77. – EDN GGXQYK.
20. Янкевич С. В. Обзор схемотехнических решений при построении силовой части ультразвуковых генераторов / С. В. Янкевич, Г. В. Малинин // Вестник Чувашского университета. – 2020. – № 1. – С. 212–220. – EDN TTQVYO.
21. Борисов В. И. Акустическое поле излучения пьезопреобразователя в виде двухмерной прямоугольной фазированной решетки / В. И. Борисов, А. С. Никитин, Е. Н. Прокопенко, С. С. Сергеев // Вестник Белорусско-Российского университета. – 2021. – № 2(71). – С. 58–67. – DOI 10.53078/20778481_2021_2_58.
22. Bessonov L. V. Constructing the dependence between the Young's modulus value and the Hounsfield units of spongy tissue of human femoral heads / L. V. Bessonov, A. A. Golyadkina, P. O. Dmitriev [et al.] // Izvestiya of Saratov University. Mathematics. Mechanics. Informatics. – 2021. – Vol. 21, No. 2. – P. 182–193. – DOI 10.18500/1816-9791-2021-21-2-182-193.
23. Кашин Я. М. Заявка на патент на изобретение № 2024118715 от 04.07.2024 «Устройство регистрации биометрических параметров для двухфакторной аутентификации личности» / Я. М. Кашин, В. А. Частикова, В. Г. Гуляй; заявитель Федеральное государственное бюджетное образовательное учреждение высшего образования «Кубанский государственный технологический университет».

24. Кашин Я. М. Заявка на патент на способ № 2024118723 от 04.07.2024 «Способ двухфакторной аутентификации личности» / Я. М. Кашин, В. А. Частикова, В. Г. Гуляй; заявитель Федеральное государственное бюджетное образовательное учреждение высшего образования «Кубанский государственный технологический университет».
25. Частикова В. А. Программа для идентификации личности по отпечатку пальца / В. А. Частикова, В. Г. Гуляй, Д. С. Сучков // Свидетельство о государственной регистрации программы для ЭВМ № 2024661103, заявл. 07.05.2024; опубл. 15.05.2024, Российская Федерация.

METHOD OF PERSONALITY RECOGNITION BASED ON BIOACOUSTIC SCANNING IN MULTI-FACTOR AUTHENTICATION SYSTEMS

Chastikova V. A.³, Gulyai V. G.⁴

Keywords: authentication, type I error, type II error, fingerprint, generative adversarial neural network, spoofing attack, biometric scanner, bioacoustics.

The purpose of the study: theoretical substantiation and development of a method of personality recognition based on bioacoustic scanning in order to implement the process of multi-factor authentication within one local device, as well as to minimize errors of the first and second types in multi-factor authentication systems.

Research methods: the methods of wave theory of physics were applied, including aspects of the propagation of ultrasonic waves and their interaction with biological tissues. Statistical methods for evaluating hypotheses and machine learning algorithms in the form of generative-adversarial neural networks were also used.

Scientific novelty: a new method of personality recognition based on bioacoustic scanning data characterizing the geometry of the bone of the distal phalanx of the finger and the ratio of soft tissues of the finger has been developed. The proposed bioacoustic method of scanning the distal phalanx of the finger, in contrast to ultrasound scanning of the papillary pattern, works at lower frequencies (15-17 MHz) and allows fixing the unique characteristics of not only the skin, but also deeper tissues Finger. The bone geometry of the distal phalanx, as well as the ratio of the soft tissues of the finger, has not previously been used as an authentication characteristic of a person.

Result and practical significance: the method of identity authentication by bioacoustic finger scanning is substantiated, within the framework of which the geometry of the bone of the distal phalanx of the finger is recorded, including protrusions, notches, ridges and nail tuberosity, which are unique for each person, and the relative location of the soft tissues of the finger pad, taking into account their mutual ratio in volume. The practical significance of the proposed method is due to the fact that this approach is less vulnerable to fakes and, accordingly, attacks with the help of dummies, since biometric characteristics are hidden from the eyes of the attacker, unlike the geometry of the face and the image of the iris of the eye in public photos, left on various surfaces of fingerprints, etc data transmission channels.

References

1. Sim J. Y. Identity Recognition Based on Bioacoustics of Human Body / Sim J. Y., Noh H. W., Goo W., Kim N., Chae S.-H. and Ahn C.-G. // IEEE Transactions on Cybernetics, vol. 51, no. 5, pp. 2761–2772, May 2021, DOI: 10.1109/TCYB.2019.2941281.
2. Roy A. Masterprint: exploring the vulnerability of partial fingerprint-based authentication systems / Roy A., Memon N. and Ross A. // IEEE Transactions on Information Forensics and Security, vol. 12, no. 9, September 2017.
3. Bontrager P. DeepMasterPrints: generating MasterPrints for dictionary attacks via latent variable evolution / Bontrager P., Roy A., Togelius J., Memon N. and Ross A. // 2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS), Redondo Beach, CA, USA, 2018, pp. 1–9, DOI: 10.1109/BTAS.2018.8698539.
4. Bondarenko I. N. Ul'trazvukovaja harakteristika kozhi, m'jagkih tkanej lica, shei, kistej ruk u zhenshhin raznyh vozrastnyh grupp: nerandomizirovannoe observacionnoe kross-sekcionnoe issledovanie // Kubanskij nauchnyj medicinskij vestnik. – 2021. – T. 28, № 3. – S. 16–28. – DOI 10.25207/1608-6228-2021-28-3-16-28.
5. Chastikova V. A. Osobennosti realizacii biometricheskoj sistemy autentifikacii lichnosti na osnove papilljarnogo uzora / V. A. Chastikova, V. G. Gul'aj, D. S. Suchkov // Molodezhnaja nauka. Sbornik luchshih nauchnyh rabot molodyh uchenyh: Materialy LI studencheskoj nauchnoj konferencii, Krasnodar, 29 fevralja 2024 goda. – Krasnodar: Kubanskij gosudarstvennyj tehnologicheskij universitet, 2024. – S. 222–225. – EDN RDBTAL.
6. Chajkovskaja T. V. Issledovanie naprjazhenno-deformirovannogo sostojanija kosti srednej falangi ukazatel'nogo pal'ca ruki cheloveka pri izmenenii plotnosti gubchatoj kostnoj tkani / T. V. Chajkovskaja, E. S. Marchenko // Fundamental'nye osnovy mehaniki. – 2022. – № 10. – S. 119–122. – DOI 10.26160/2542-0127-2022-10-119-122. – EDN JAAFKI.
7. Chastikova V. A. Predotvrashhenie oshibok i roda pri skanirovanii v sistemah kontrolja i upravlenija dostupom putem ispol'zovanija GAN-seti / V. A. Chastikova, V. G. Gul'aj, D. S. Suchkov // Iskusstvennyj intellekt. Formirovanie budushhego: Materialy I Mezhdunarodnoj nauchno-prakticheskoy konferencii, Krasnodar, 29 aprelja 2024 goda. – Krasnodar: IP Alzidan M., 2024. – S. 296–299. – EDN ESSLTN.
- 3 Vera A. Chastikova, Candidate of Technical Sciences, Associate Professor, Associate Professor of the Department of Cybersecurity and Information Protection, Kuban State Technical University, Krasnodar, Russia. E-mail: chastikova_va@mail.ru
- 4 Victoria G. Gulyai, Student of the Institute of Computer Systems and Information Security, KubSTU, Krasnodar, Russia. E-mail: ms.gulyai@bk.ru

8. Krasov A. V. Razrabotka metodologii testirovaniya sistem zashhity informacii v virtual'nykh kompleksah dlja obnaruzheniya oshibok I i II-roda / A. V. Krasov, R. R. Maksudova, V. V. Nefedov [i dr.] // Vestnik Sankt-Peterburgskogo gosudarstvennogo universiteta tehnologii i dizajna. Seriya 1: Estestvennye i tehnicheckie nauki. – 2021. – № 1. – S. 45–52. – DOI 10.46418/2079-8199_2021_1_7.
9. Chastikova V. A. Podhod k postroeniju sistem analiza incidentov informacionnoj bezopasnosti na osnove gibrizacii metodov mashin-nogo obuchenija / V. A. Chastikova, V. G. Guljaj // Jelektronnyj setevoj politematicheskij zhurnal «Nauchnye trudy KubGTU». – 2023. – № 6. – S. 107–117. – EDN YPTPGQ.
10. Suvorov A. Yu. Statistical hypothesis testing: general approach in medical research / A. Yu. Suvorov, N. M. Bulanov, A. N. Shvedova [et al.] // Sechenov Medical Journal. – 2022. – Vol. 13, No. 1. – P. 4–13. – DOI 10.47093/2218-7332.2022.426.08.
11. Ismagilov R. F. Konstruirovanie modeli obuchajushhej nejronnoj seti dlja biometricheskoj mnogofaktornoj autentifikacii pol'zovatelja informacionnoj sistemy / R. F. Ismagilov, N. D. Lushnikov, A. S. Ismagilova // Voprosy zashhity informacii. – 2023. – № 1(140). – S. 19–23. – DOI 10.52190/2073-2600_2023_1_19.
12. Belov N. I. Primenenie termograficheskij modeli lica v zadachah identifikacii i autentifikacii / N. I. Belov, V. M. Korzhuk // Computational Nanotechnology. – 2023. – T. 10, № 3. – S. 109–120. – DOI 10.33693/2313-223X-2023-10-3-109-120.
13. Slavoljubova I. A. Differencirujushhie vozmozhnosti papilljarnykh uzorov srednih falang pal'cev ruk (na primere evrazijskih populjacij) / I. A. Slavoljubova, A. M. Judina, L. Ju. Shpak // Vestnik Moskovskogo universiteta. Seriya 23: Antropologija. – 2023. – № 4. – S. 58–68. – DOI 10.32521/2074-8132.2023.4.058-068.
14. Chastikova V. A. Nejrosetevaja sistema biometricheskoj identifikacii lichnosti po golosu / V. A. Chastikova, S. A. Zherlicyn, D. O. Vojlova // Vestnik Adygejskogo gosudarstvennogo universiteta. Seriya 4: Estestvenno-matematicheskije i tehnicheckie nauki. – 2023. – № 1(316). – S. 70–79. – DOI 10.53598/2410-3225-2023-1-316-70-79.
15. Pozhar V. Je. Spektral'no-poljarizacionnye sistemy trjohmernogo tehnicheckogo zrenija na osnove akustoopticheskoy fil'tracii / V. Je. Pozhar, A. S. Machihin // Svetotehnika. – 2022. – № 5. – S. 32–35. – EDN BHPGOK.
16. Kormilina A. R. Ul'trazvukovaja jelastografija sdvigovoj volny v ocenke zhjostkosti kostnoj mozoli / A. R. Kormilina, M. G. Tuhbatullin // Rossijskij jelektronnyj zhurnal luchevoj diagnostiki. – 2020. – T. 10, № 2. – S. 122–128. – DOI 10.21569/2222-7415-2020-10-2-122-128.
17. Kol'cova I. S. Akustoelektronnyj metod izmerenija skorosti ul'trazvukovykh voln v zhidkostjah / I. S. Kol'cova, A. S. Homutova // Akusticheskij zhurnal. – 2023. – T. 69, № 3. – S. 317–321. – DOI 10.31857/S032079192260041X.
18. Romanenko K. V. Ul'trazvukovoe issledovanie kozhi / K. V. Romanenko, O. O. Borovaja, N. V. Ermilova, V. V. Zaharova // Torsuevskie chtenija: nauchno-prakticheskij zhurnal po dermatologii, venerologii i kosmetologii. – 2021. – № 2(32). – S. 17–20. – EDN XSCBLX.
19. Potehin D. S. Sovremennye tendencii razvitiya ul'trazvukovoj diagnostiki / D. S. Potehin, P. V. Komarov // Vysokoproizvoditel'nye vychislitel'nye sistemy i tehnologii. – 2022. – T. 6, № 2. – S. 71–77. – EDN GGXYQK.
20. Jankevich S. V. Obzor shemotehnicheskikh reshenij pri postroenii silovoj chasti ul'trazvukovykh generatorov / S. V. Jankevich, G. V. Malinin // Vestnik Chuvashskogo universiteta. – 2020. – № 1. – S. 212–220. – EDN TTQVYO.
21. Borisov V. I. Akusticheskoe pole izlucheniya p'ezopreobrazovatelja v vide dvuhmernoj prjamougol'noj fazirovannoj reshetki / V. I. Borisov, A. S. Nikitin, E. N. Prokopenko, S. S. Sergeev // Vestnik Belorussko-Rossijskogo universiteta. – 2021. – № 2(71). – S. 58–67. – DOI 10.53078/20778481_2021_2_58.
22. Bessonov L. V. Constructing the dependence between the Young's modulus value and the Hounsfield units of spongy tissue of human femoral heads / L. V. Bessonov, A. A. Golyadkina, P. O. Dmitriev [et al.] // Izvestiya of Saratov University. Mathematics. Mechanics. Informatics. – 2021. – Vol. 21, No. 2. – P. 182–193. – DOI 10.18500/1816-9791-2021-21-2-182-193.
23. Kashin Ja. M. Zajavka na patent na izobretenie № 2024118715 ot 04.07.2024 «Ustrojstvo registracii biometricheskikh parametrov dlja dvuhfaktornoj autentifikacii lichnosti» / Ja. M. Kashin, V. A. Chastikova, V. G. Guljaj; zajavitel' Federal'noe gosudarstvennoe bjudzhetnoe obrazovatel'noe uchrezhdenie vysshego obrazovaniya «Kubanskij gosudarstvennyj tehnologicheskij universitet».
24. Kashin Ja. M. Zajavka na patent na sposob № 2024118723 ot 04.07.2024 «Sposob dvuhfaktornoj autentifikacii lichnosti» / Ja. M. Kashin, V. A. Chastikova, V. G. Guljaj; zajavitel' Federal'noe gosudarstvennoe bjudzhetnoe obrazovatel'noe uchrezhdenie vysshego obrazovaniya «Kubanskij gosudarstvennyj tehnologicheskij universitet».
25. Chastikova V. A. Programma dlja identifikacii lichnosti po otpechatku pal'ca / V. A. Chastikova, V. G. Guljaj, D. S. Suchkov // Svidetel'stvo o gosudarstvennoj registracii programmy dlja JeVM № 2024661103, zajavl. 07.05.2024; opubl. 15.05.2024, Rossijskaja Federacija.



МЕТОД ОБЕСПЕЧЕНИЯ КИБЕРУСТОЙЧИВОСТИ БЛОКЧЕЙН-ПЛАТФОРМ НА ОСНОВЕ КИБЕРИММУНИТЕТА

Балябин А. А.¹, Петренко С. А.²

DOI: 10.21681/2311-3456-2025-6-127-139

Цель исследования: обеспечение устойчивости функционирования облачных блокчейн-экосистем и платформ «Экономики данных» Российской Федерации на основе кибериммунитета в условиях гибридных квантово-классических атак.

Методы исследования: методы системного анализа, методы теории вероятностей и математической статистики, методы теории устойчивости сложных систем.

Полученные результаты: анализ исследований в предметной области показал, что существующие методы обеспечения устойчивости различных информационно-вычислительных систем учитывают по отдельности либо классические, либо квантовые угрозы и не позволяют в полной мере обеспечить устойчивость функционирования облачных блокчейн-платформ в условиях гибридных атак, характеризующихся наличием обеих составляющих. Для разрешения данной проблемной ситуации поставлена задача разработки нового метода обеспечения устойчивости облачных блокчейн-платформ на основе кибериммунитета, а также сформулирована гипотеза о возможности достижения цели исследования за счет применения данного метода.

Разработан метод обеспечения устойчивости облачных блокчейн-платформ на основе кибериммунитета в условиях гибридных квантово-классических атак, позволяющий обеспечивать требования к показателю вероятности компрометации при ограничении на время выполнения программного цикла узла блокчейн за счет варьирования длины криптографического ключа и коэффициента покрытия кибериммунитета.

Проведено исследование разработанного метода, в ходе которого показана возможность обеспечения требуемой устойчивости облачных блокчейн-платформ в условиях гибридных квантово-классических атак, а также определены условия существования решения, что позволило подтвердить сформулированную гипотезу.

Научная новизна: разработанный метод впервые учитывает такие новые условия, как гибридные атаки на облачные блокчейн-платформы, которые в формализованном виде описываются через вновь вводимые параметры количества кубитов квантового компьютера, доступных атакующему, и доли вредоносных входных данных. Кроме того, применение метода впервые наделяет облачные блокчейн-платформы новым эмерджентным свойством кибериммунитета, заключающимся в способности обнаруживать известные и ранее неизвестные атаки, направленные на нарушение семантики вычислений, противодействовать им и осуществлять восстановление штатного функционирования при возникновении нарушений.

Ключевые слова: угрозы безопасности информации, квантовые угрозы безопасности, облачные блокчейн-экосистемы и платформы, кибербезопасность, методы анализа и синтеза квантово-устойчивого блокчейн.

Введение

С момента своего появления блокчейн прошел несколько этапов эволюции, каждый из которых связан с внедрением новых технологических решений [1]. Современное состояние развития технологий распределенного реестра характеризуется активным созданием блокчейн-экосистем и платформ, на базе которых разрабатываются смарт-контракты, децентрализованные приложения (dApps), системы децентрализованных финансов (DeFi), децентрализованные организации (DAO) и другие. Ведутся исследования по применению технологий блокчейн для обеспечения безопасности платформ интернета вещей (IoT), облачных, туманных и пограничных вычислений, иных технологий Индустрии 4.0, а также для создания децентрализованной сети Интернет (Web3) [2].

В Российской Федерации технологии распределенного реестра относятся к так называемым «сквозным» технологиям, применяемым в рамках реализации национального проекта «Экономика данных», направленного на цифровизацию отраслей экономики и социальной сферы, достижение технологического суверенитета и лидерства.

Поддержание функционирования полных узлов современных блокчейн-платформ зачастую требует больших вычислительных ресурсов, поэтому в настоящее время ведутся исследования по созданию технологии Blockchain-as-a-Service (BaaS). Данная технология предоставляет возможность разработки, тестирования и развертывания программного обеспечения (ПО) блокчейн-платформ в облачной вычислительной среде [3, 4], как показано на рис. 1.

1 Балябин Артём Алексеевич, младший научный сотрудник, Научный центр информационных технологий и искусственного интеллекта, Научно-технологический университет «Сириус». Федеральная территория «Сириус», Россия. <https://orcid.org/0009-0006-3949-154X>. E-mail: Balyabin.AA@talantiuspeh.ru

2 Петренко Сергей Анатольевич, доктор технических наук, профессор, руководитель группы, Научный центр информационных технологий и искусственного интеллекта, Научно-технологический университет «Сириус». Федеральная территория «Сириус», Россия. ORCID 0000-0003-0644-1731. E-mail: Petrenko.SA@talantiuspeh.ru

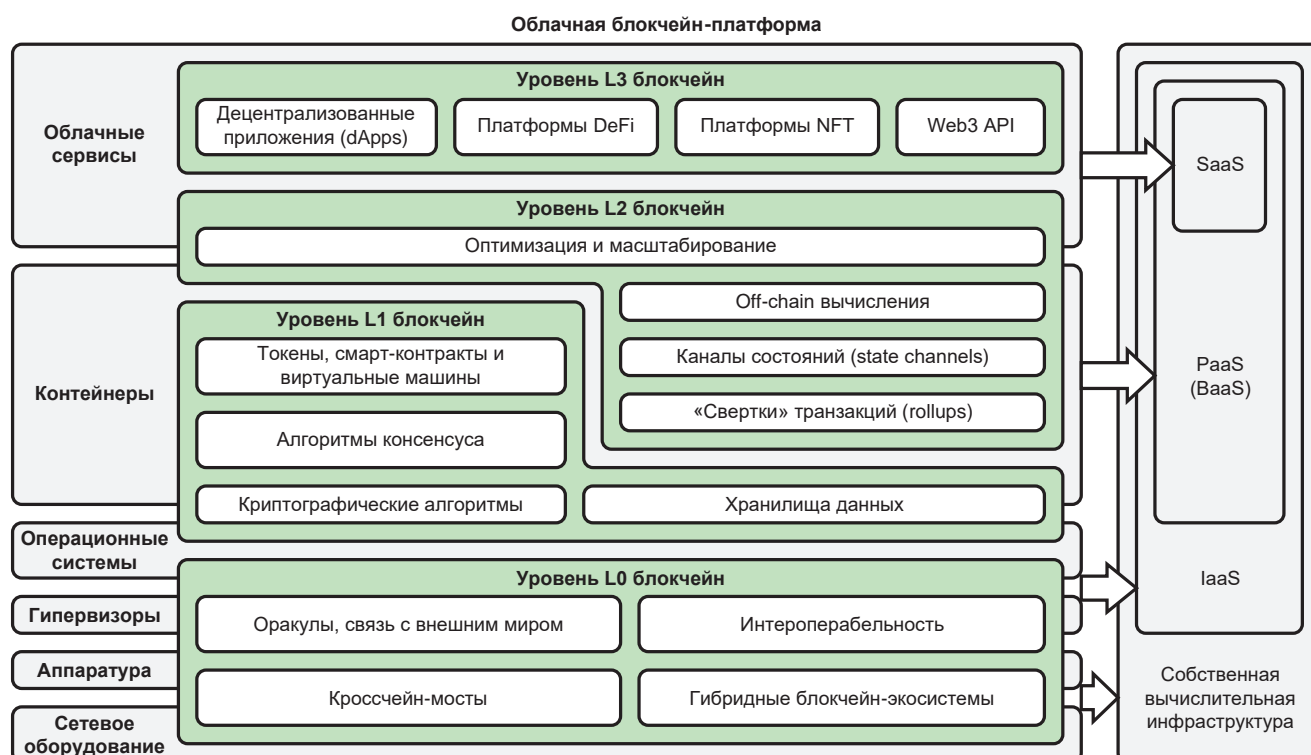


Рис. 1. Архитектура типовой облачной блокчейн-платформы

В то же время усложнение архитектуры информационно-вычислительных систем связано с возникновением новых вызовов. Одним из таких вызовов для облачных блокчейн-платформ является обеспечение требуемой устойчивости их функционирования в условиях гибридных квантовых [5, 6] и классических [7] атак.

Известен ряд исследований в области обеспечения квантовой устойчивости блокчейн. В работах [8–11] предлагается внедрить постквантовые криптографические алгоритмы хеширования, шифрования и цифровой подписи в классические блокчейн-платформы. В работах [12–15] предлагается создание блокчейн-платформ как физических систем, использующих квантовые каналы связи и алгоритмы консенсуса. В работах [16, 17] совмещаются идеи первых двух подходов и предлагается создание гибридного блокчейн. В работах [18, 19] предлагаются новые модели и методы оценивания квантовой устойчивости блокчейн. Классификация данных подходов представлена на рис. 2.

В целом существующие подходы к обеспечению квантовой устойчивости блокчейн-платформ сводятся к применению постквантовых криптографических алгоритмов и не учитывают классические атаки. Однако устойчивость блокчейн-платформ необходимо рассматривать в контексте той вычислительной системы, в которой они функционируют. Облачная платформа дает наиболее общее представление о стеке

технологий, применяемых на различных уровнях вычислительных систем, включая уровни сети связи, аппаратного обеспечения, гипервизоров, операционных систем, контейнеров и прикладного ПО. На каждом из этих уровней могут возникать уязвимости, эксплуатация которых может повлиять на устойчивость функционирования облачной блокчейн-платформы [20].

К числу перспективных интеллектуальных и биоинспирированных подходов к обеспечению устойчивости различных информационно-вычислительных систем можно отнести подходы на основе реконфигурации и гомеостаза [21], антиципации и синтеза упреждающего поведения систем защиты [22], совершенствования систем обнаружения вторжений с использованием алгоритмов машинного обучения [23], глубокого обучения [24], федеративного обучения [25], искусственных иммунных систем [26], а также подходы на основе организации самовосстанавливающихся вычислений и кибериммунитета [27, 28]. Однако вышеуказанные подходы к обеспечению устойчивости не учитывают наличие квантовых угроз.

С учетом вышесказанного можно сделать вывод о том, что существующие методы учитывают по отдельности либо классические, либо квантовые угрозы и не позволяют в полной мере обеспечить устойчивость функционирования облачных блокчейн-платформ в условиях гибридных атак,

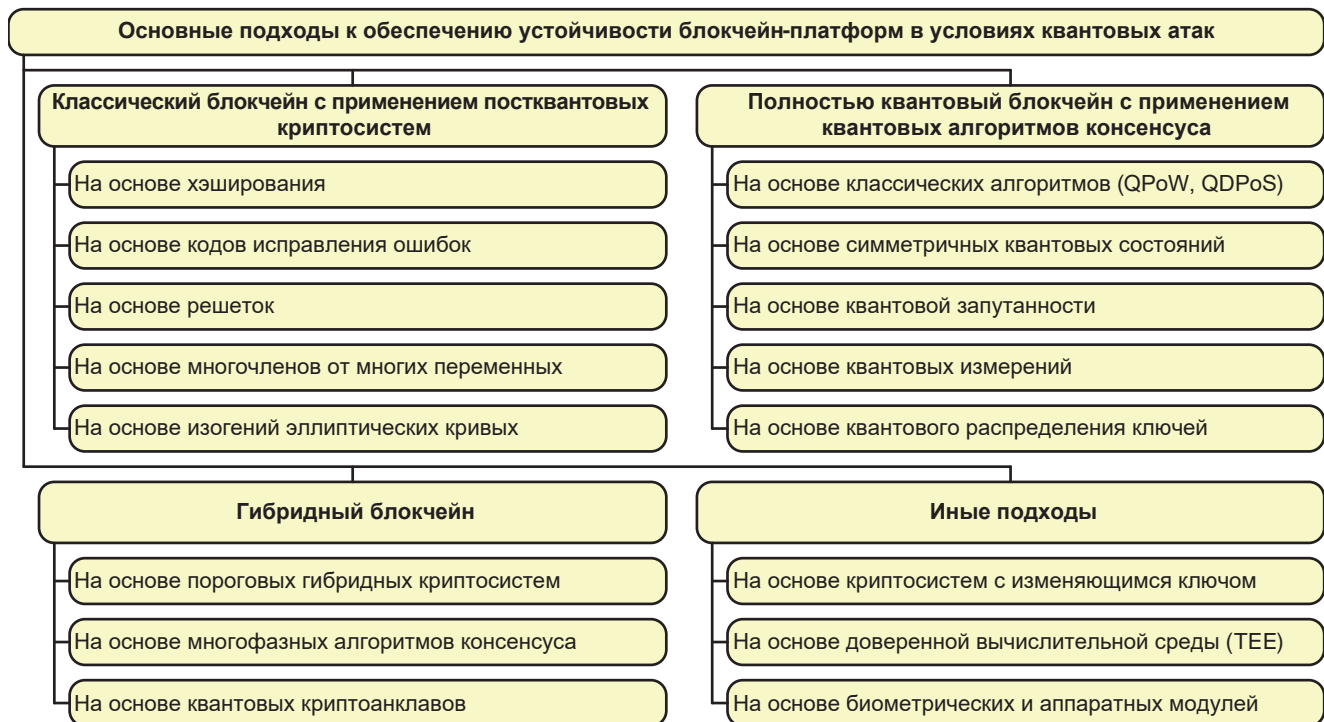


Рис. 2. Основные подходы к обеспечению устойчивости блокчейн-платформ в условиях квантовых атак

характеризующихся наличием обеих составляющих. Рост гибридных квантово-классических угроз для облачных блокчейн-платформ, невозможность обеспечения требуемой устойчивости их функционирования с использованием существующих технических и технологических решений, а также невозможность разработки таких научно-обоснованных решений ввиду несовершенства существующего научно-методического аппарата характеризует проблемную ситуацию, разрешение которой является актуальной научной задачей.

В настоящем исследовании предлагается метод обеспечения устойчивости облачных блокчейн-платформ на основе кибериммунитета в условиях гибридных квантово-классических атак. Идея кибериммунитета заключается в наделении таких платформ способностью обнаруживать как известные, так и ранее неизвестные гибридные атаки, противодействовать им, не допуская нарушений, а также оперативно восстанавливать штатное функционирование в случае их возникновения.

1. Возможные сценарии гибридных квантово-классических атак

На узлах блокчейн функционирует идентичное или схожее ПО одного типа, обеспечивающее обмен данными и синхронизацию состояния распределенного реестра. Взаимодействие узлов осуществляется в одноранговой сети посредством p2p-каналов, поэтому выявление и эксплуатация уязвимостей ПО

блокчейн-платформы может привести к компрометации узла и распространению атаки на остальную блокчейн-сеть.

Гибридные квантовые и классические атаки могут использовать сценарии воздействий, сочетающие применение вычислительного потенциала квантового компьютера для вскрытия криптографических алгоритмов и эксплуатацию ранее неизвестных программных уязвимостей, как показано на рис. 3. Такие воздействия могут позволить злоумышленнику создавать ботнет-сети из зараженных узлов блокчейн [29], координировать их действия для внедрения вредоносных транзакций, принятия произвольных цепочек блоков и влияния на алгоритмы консенсуса, что создает серьезную угрозу устойчивости функционирования облачных блокчейн-платформ.

Для противодействия подобным угрозам необходимо применять комплексные меры, обеспечивающие невозможность воздействия на вычислительную среду и алгоритмы функционирования облачных блокчейн-платформ. Данные меры включают наделение ПО блокчейн-платформ свойством кибериммунитета и управление устойчивостью их функционирования.

2. Постановка задачи разработки метода

Устойчивость функционирования облачной блокчейн-платформы с кибериммунитетом в условиях гибридных квантово-классических атак зависит от способности узлов блокчейн противодействовать таким

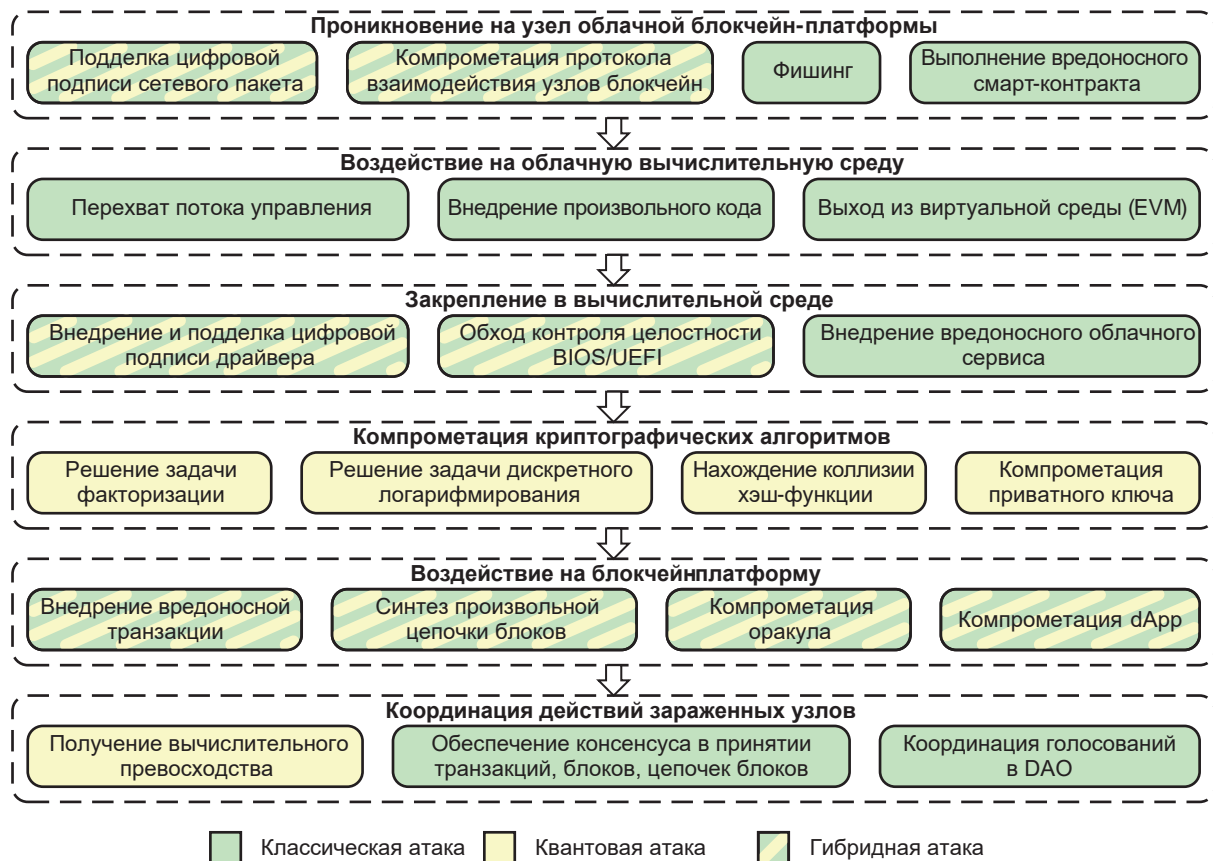


Рис. 3. Возможные сценарии гибридных квантово-классических атак на облачные блокчейн-платформы

атакам, не допуская нарушений, и восстанавливать штатное функционирование при их возникновении. При этом квантовые атаки направлены на компрометацию криптографических алгоритмов с использованием вычислительной мощности квантового компьютера, а классические атаки – на нарушение семантики вычислений путем отправки вредоносных входных данных. Управление устойчивостью облачной блокчейн-платформы должно осуществляться путем варьирования параметров противодействия квантовым и классическим атакам. Введем обозначения:

L – облачная блокчейн-платформа;

$P_{\text{компр}}$ – вероятность компрометации облачной блокчейн-платформы вследствие гибридной атаки;

$P_{\text{компр.узла}}$ – вероятность компрометации узла облачной блокчейн-платформы;

$T_{\text{цикла}}$ – время выполнения цикла вычислений на узле;

$N_{\text{вал}}$ – общее количество валидаторов, участвующих в алгоритме консенсуса;

$N_{\text{конс}}$ – количество голосов валидаторов, необходимое для влияния на алгоритм консенсуса (в блокчейн-системах для получения вычислительного превосходства злоумышленнику достаточно получить

контроль над 51 % узлов, поэтому примем данный параметр равным $0,51N_{\text{вал}}$);

Q – доступное количество логических кубитов квантового компьютера;

$k_{\text{вред}}$ – доля вредоносных входных данных облачной блокчейн-платформы;

$n_{\text{ключа}}$ – длина криптографического ключа в битах;

$k_{\text{покр}}$ – коэффициент покрытия кибериммунитета, определяющий долю линейных блоков программ, подлежащих контролю семантики вычислений.

Вероятность компрометации облачной блокчейн-платформы $P_{\text{компр}}$ определяется вероятностью того, что атакующему вследствие гибридной атаки удастся получить управление не менее, чем над $N_{\text{конс}}$ узлов валидаторов, обеспечив таким образом возможность влияния на алгоритм консенсуса. Компрометация каждого из узлов является независимым событием, а значит, общую вероятность компрометации облачной блокчейн-платформы можно вычислить с использованием формулы для биномиального распределения:

$$P_{\text{компр}} = \sum_{i=N_{\text{конс}}}^{N_{\text{вал}}} \binom{N_{\text{вал}}}{N_{\text{конс}}} P_{\text{компр.узла}}^i (1 - P_{\text{компр.узла}})^{N_{\text{вал}} - N_{\text{конс}}}. \quad (1)$$

Таким образом, необходимо разработать метод M обеспечения требуемой устойчивости

функционирования облачной блокчейн-платформы L по показателю вероятности компрометации $P_{\text{компр}}$ при ограничении на время выполнения цикла вычислений на узле $T_{\text{цикла}}$, в условиях квантовых атак, характеризуемых количеством кубитов Q , и классических атак, характеризуемых долей вредоносных входных данных $k_{\text{вред}}$, за счет варьирования длины ключа $n_{\text{ключа}}$ и коэффициента покрытия кибериммунитета $k_{\text{покр}}$:

$$M: \langle L, \{Q, k_{\text{вред}}\}, \{n_{\text{ключа}}, k_{\text{покр}}\} \rangle \rightarrow \langle P_{\text{компр}}, T_{\text{цикла}} \rangle, \quad (2)$$

$$\begin{cases} P_{\text{компр}} \leq P_{\text{компр}}^{\text{тр}} \\ T_{\text{цикла}} \leq T_{\text{цикла}}^{\text{тр}} \end{cases}$$

Гипотеза исследования состоит в том, что применение данного метода позволяет обеспечить требуемую устойчивость функционирования облачных блокчейн-платформ в условиях гибридных квантово-классических атак.

3. Формализация метода обеспечения устойчивости облачных блокчейн-платформ

Для формализации метода введем дополнительные обозначения:

$N_{\text{ур}}$ – количество уровней облачной блокчейн-платформы;

$P_{\text{компр.кв}}$ – вероятность компрометации узла в результате квантовой атаки;

$P_{\text{Шора}}$ – вероятность успеха квантовой атаки с применением алгоритма Шора;

$P_{\text{Гровера}}$ – вероятность успеха квантовой атаки с применением алгоритма Гровера;

t_{gate} – время выполнения квантового вентиля ($t_{\text{gate}} \approx 10^{-7}$ с);

$t_{\text{ког}}$ – время сохранения когерентности кубитов;

$N_{\text{попыток}}$ – допустимое количество попыток вскрытия криптосистемы;

$P_{\text{компр.кл}}$ – вероятность компрометации узла в результате классической атаки;

$P_{\text{прот}}$ – вероятность успешного противодействия классической атаке;

$T_{\text{ср.атак}}$ – среднее время между пропусками классических атак;

$P_{\text{восст}}$ – вероятность успешного восстановления после нарушения;

$T_{\text{ср.восст}}$ – среднее время восстановления после нарушения;

$P_{\text{раб}}$ – вероятность нахождения облачной блокчейн-платформы в работоспособном состоянии в произвольный момент времени;

$P_{\text{мод}}$ – вероятность модификации вычислений при пропуске классической атаки;

$T_{\text{выч}}$ – время выполнения вычислительных операций;

$T_{\text{обн}}$ – время обнаружения нарушений семантики вычислений;

$T_{\text{восст}}$ – время восстановления штатного функционирования;

t_0 – время выполнения элементарной операции (при расчетах $t_0 = 1$ ед. времени);

$k_{\text{изв}}$ – доля классических атак, являющихся известными;

p – вероятность обнаружения нарушения, возникшего вследствие ранее неизвестной классической атаки;

k – количество линейных блоков программ;

m – количество вычислительных инструкций в блоке;

n – количество параметров в инструкции.

Каждый узел облачной блокчейн-платформы, как было показано ранее, функционирует на $N_{\text{ур}}$ уровнях, которые могут быть одновременно подвержены гибридным атакам. Вероятность компрометации узла зависит от вероятности успеха гибридной атаки, включающей классическую и квантовую составляющие:

$$P_{\text{компр.узла}} = 1 - (1 - P_{\text{компр.кл}})(1 - P_{\text{компр.кв}}). \quad (3)$$

Успешность квантовой атаки зависит от применяемого алгоритма (Шора или Гровера), количества кубитов квантового компьютера атакующего и длины ключа, являющейся параметром противодействия такой атаке. Тогда вероятность компрометации узла в результате квантовой атаки может быть определена как

$$P_{\text{компр.кв}} = 1 - ((1 - P_{\text{Шора}})(1 - P_{\text{Гровера}}))^{N_{\text{ур}}}. \quad (4)$$

Вероятности $P_{\text{Шора}}$ и $P_{\text{Гровера}}$ с учетом оценок, полученных в работе [30], можно приближенно оценить как

$$P_{\text{Шора}} = 1 - \left(1 - e^{\frac{-2n_{\text{ключа}}}{Q} \frac{n_{\text{ключа}}^3 t_{\text{gate}}}{t_{\text{ког}}}} \right)^{N_{\text{попыток}}}, \quad (5)$$

$$P_{\text{Гровера}} = 1 - \left(1 - e^{\frac{-n_{\text{ключа}}}{Q} \frac{\pi}{4} \frac{2^{n_{\text{ключа}}/2} t_{\text{gate}}}{t_{\text{ког}}}} \right)^{N_{\text{попыток}}}. \quad (6)$$

Следует отметить, что данные оценки не учитывают многокубитные корреляции, коррекцию ошибок и возможность продления времени когерентности с использованием логических кубитов. Последнее сильно зависит от типа применяемого квантового компьютера. Например, применение сверхпроводникового квантового компьютера на базе кубитов-трансмонов позволяет достичь времени когерентности $t_{\text{ког}} = 0,5$ мс [31], а применение квантового компьютера на базе фотонов в микроволновых резонаторах – $t_{\text{ког}} = 34$ мс [32]. В дальнейшем для оценки будем использовать время когерентности порядка $t_{\text{ког}} \approx 10^{-2}$ с. На рис. 4 представлены оценки вероятностей успеха квантовых атак (5) и (6) с использованием алгоритмов Шора и Гровера при различном количестве кубитов, длине ключа и допустимом количестве попыток.

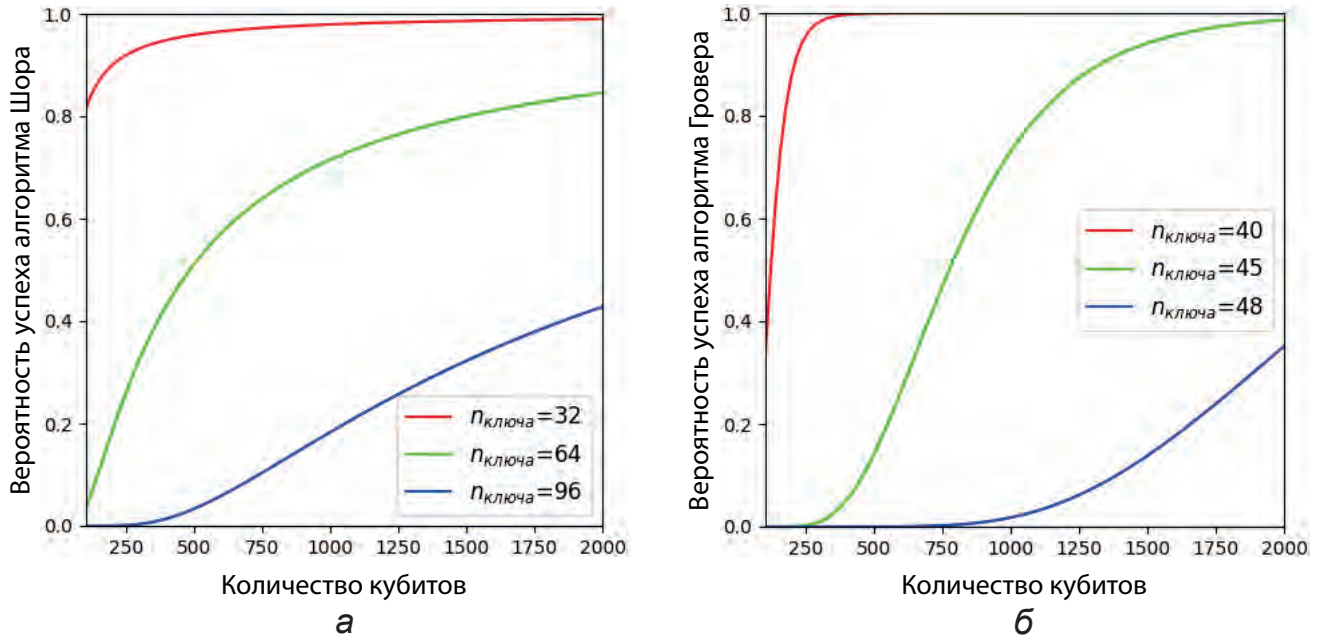


Рис. 4. Вероятности успеха квантовых атак с использованием: а – алгоритма Шора; б – алгоритма Гровера

Классические атаки на облачные блокчейн-платформы осуществляются путем отправки вредоносных входных данных, обработка которых приводит к нарушению семантики вычислений (появлению аномального состояния программы, не соответствующего ожидаемому или заданному эталонной моделью). Вероятность компрометации узла вследствие классической атаки зависит от его способности противодействовать таким атакам, не допуская нарушений семантики, и оперативно восстанавливать штатное функционирование при их возникновении, сохраняя при этом работоспособное (устойчивое) состояние. Данный показатель с учетом [33] может быть определен как

$$P_{\text{компр.кл}} = 1 - ((P_{\text{прот}} + (1 - P_{\text{прот}})P_{\text{восст}})P_{\text{раб}})^{N_{\text{ур}}}. \quad (7)$$

Вероятность успешного противодействия узла классической атаке обратна вероятности пропуска такой атаки и возникновения модификации вычислений. Пропуск атаки возможен, если она произошла на участке программы, защищенном механизмами кибериммунитета, но не была обнаружена (событие S_1), либо если она произошла на незащищенном участке программы (событие S_2):

$$P_{\text{прот}} = 1 - P_{\text{мод}}(P(S_1) + P(S_2)). \quad (8)$$

Здесь $P(S_1)$ и $P(S_2)$ могут быть определены соответственно как

$$P(S_1) = (k_{\text{вред}}(1 - P_{\text{обн}}))(k_{\text{покр}}(1 - P_{\text{обн}})), \quad (9)$$

$$P(S_2) = k_{\text{вред}}(1 - P_{\text{обн}})(1 - k_{\text{покр}}), \quad (10)$$

тогда, подставив (9) и (10) в формулу (8) и преобразовав, получим:

$$P_{\text{прот}} = 1 - P_{\text{мод}} k_{\text{вред}}(1 - P_{\text{обн}})(1 - k_{\text{покр}} P_{\text{обн}}). \quad (11)$$

Вероятность восстановления штатного функционирования узла облачной блокчейн-платформы напрямую зависит от вероятности обнаружения нарушения. Предположим, что все обнаруженные нарушения восстанавливаются одним из способов (возврат к предыдущему устойчивому состоянию, перезапуск), известные нарушения обнаруживаются всегда, а ранее неизвестные – с вероятностью p , тогда вероятность обнаружения и восстановления можно определить как

$$P_{\text{восст}} = P_{\text{обн}} = k_{\text{изв}}1 + (1 - k_{\text{изв}})p. \quad (12)$$

Одним из показателей устойчивости функционирования узла облачной блокчейн-платформы в условиях классических атак является также вероятность его нахождения в работоспособном состоянии в произвольный момент времени. Данный показатель по смыслу схож с коэффициентом готовности и может быть определен как

$$P_{\text{раб}} = \frac{T_{\text{ср.атак}}}{T_{\text{ср.атак}} + T_{\text{ср.восст}}}. \quad (13)$$

Здесь $T_{\text{ср.атак}}$ характеризует среднюю продолжительность функционирования узла между пропусками атак и может быть определена как отношение времени вычислений с учетом контроля семантики за $n_{\text{ц}}$ циклов к ожидаемому количеству пропусков атак за это время:

$$T_{\text{ср.атак}} = \frac{n_{\text{ц}}(T_{\text{выч}} + T_{\text{обн}})}{n_{\text{ц}}(1 - P_{\text{прот}})} = \frac{T_{\text{выч}} + T_{\text{обн}}}{1 - P_{\text{прот}}}, \quad (14)$$

а $T_{\text{ср.восст}}$ в отличие от абсолютного времени, необходимого для восстановления $T_{\text{восст}}$, учитывает

ожидаемое количество успешных восстановлений за $n_{\text{ц}}$ циклов:

$$T_{\text{ср.восст}} = \frac{n_{\text{ц}} T_{\text{восст}}}{n_{\text{ц}} P_{\text{восст}}} = \frac{T_{\text{восст}}}{P_{\text{восст}}}. \quad (15)$$

Общее время выполнения программного цикла на узле облачной блокчейн-платформы, включая непосредственно вычисления, обнаружение нарушений и восстановление, может быть определено как:

$$T_{\text{цикла}} = T_{\text{выч}} + T_{\text{обн}} + T_{\text{восст}}, \quad (16)$$

где $T_{\text{выч}}$ и $T_{\text{восст}}$ зависят от параметров программы и могут быть определены как время выполнения вычислительных инструкций во всех линейных блоках с учетом количества параметров и время перезаписи полного образа программы в памяти (время выполнения операций чтения-записи) соответственно:

$$T_{\text{выч}} = kmnt_0, \quad (17)$$

$$T_{\text{восст}} = 2kmnt_0. \quad (18)$$

Время, затрачиваемое на обнаружение нарушений в течение программного цикла, зависит от применяемой для контроля семантики эталонной модели программы. Например, в работе [34] показано, что при применении аппарата теории подобия и размерностей время обнаружения нарушений определяется как

$$T_{\text{обн}} = k_{\text{покр}} \left(k \frac{m(m-1)}{2} (2n-1) + kmn \right) t_0. \quad (19)$$

С учетом введенных обозначений и формальных отношений можно сформулировать метод обеспечения устойчивости облачных блокчейн-платформ в условиях гибридных квантово-классических атак.

Входные данные метода:

- параметры облачной блокчейн-платформы;
- параметры гибридных атак (квантовых – Q и классических – $k_{\text{вред}}$);
- текущие параметры противодействия (длина ключа $n_{\text{ключа}}$, коэффициент покрытия кибериммунитета $k_{\text{покр}}$);
- требования к показателям устойчивости (вероятности компрометации $P_{\text{компр}}^{\text{тр}}$ и времени выполнения программного цикла на узле блокчейн $T_{\text{цикла}}^{\text{тр}}$).

Выходные данные метода:

- значения показателей устойчивости функционирования облачной блокчейн-платформы;
- вывод о достижении или невозможности достижения требуемой устойчивости.

Шаг 1. Выбор начальных значений длины криптографического ключа $n_{\text{ключа}}$ и коэффициента покрытия кибериммунитета $k_{\text{покр}}$ так, чтобы

$$P_{\text{компр.кв}}(n_{\text{ключа}}) \rightarrow \min \text{ и } P_{\text{компр.кл}}(k_{\text{покр}}) \rightarrow \min.$$

Шаг 2. Оценка максимально допустимого значения коэффициента покрытия кибериммунитета $k_{\text{покр}}^{\text{max}}$ на основе заданного требования к времени выполнения программного цикла на узле блокчейн $T_{\text{цикла}}^{\text{тр}}$. Данную оценку можно получить, подставив выражения (17)–(19) в (16) и решив неравенство $T_{\text{цикла}}(k_{\text{покр}}) \leq T_{\text{цикла}}^{\text{тр}}$ относительно $k_{\text{покр}}$:

$$k_{\text{покр}}^{\text{max}} = \frac{T_{\text{цикла}}^{\text{тр}} - 3kmnt_0}{\left(k \frac{m(m-1)}{2} (2n-1) + kmn \right) t_0}. \quad (20)$$

Шаг 3. Выбор текущего значения $k_{\text{покр}} \in [0, k_{\text{покр}}^{\text{max}}]$, минимизирующего вероятность успеха классической атаки при ограничении на время выполнения программного цикла, и переконфигурация кибериммунной системы защиты для контроля семантики вычислений с учетом нового коэффициента покрытия.

Шаг 4. Оценка текущей вероятности компрометации узла облачной блокчейн-платформы $P_{\text{компр.узла}}$ в условиях гибридных атак на основе формулы (3). Отметим, что при выборе достаточно большого значения $n_{\text{ключа}}$ вероятность успеха квантовой атаки будет пренебрежимо мала по сравнению с аналогичной вероятностью для классической атаки $P_{\text{компр.кв}} \ll \ll P_{\text{компр.кл}}$. Однако увеличение длины ключа является временной мерой и может привести к существенному росту вычислительной сложности и времени функционирования блокчейн-платформы. Для эффективного противодействия квантовым атакам следует применять постквантовые криптографические алгоритмы.

Шаг 5. Оценка текущей вероятности компрометации облачной блокчейн-платформы $P_{\text{компр}}$ на основе формулы (1). Если $P_{\text{компр}} \leq P_{\text{компр}}^{\text{тр}}$, то требуемая устойчивость облачной блокчейн-платформы в условиях гибридных квантово-классических атак считается достигнутой. Иначе, при $k_{\text{покр}} < k_{\text{покр}}^{\text{max}}$ следует вернуться к шагу 3 и выполнить переконфигурацию кибериммунной системы защиты, увеличив долю контролируемых линейных блоков программ. Если же $k_{\text{покр}} = k_{\text{покр}}^{\text{max}}$, то считается, что в текущих условиях невозможно обеспечить требуемую устойчивость функционирования облачной блокчейн-платформы и необходимо принятие дополнительных мер защиты.

Общая схема предложенного метода представлена на рис. 5.

Условием возможности обеспечения требуемой устойчивости облачной блокчейн-платформы по показателю вероятности компрометации при ограничении на время выполнения программного цикла на узле блокчейн является существование таких значений $n_{\text{ключа}}$ и $k_{\text{покр}}$, что

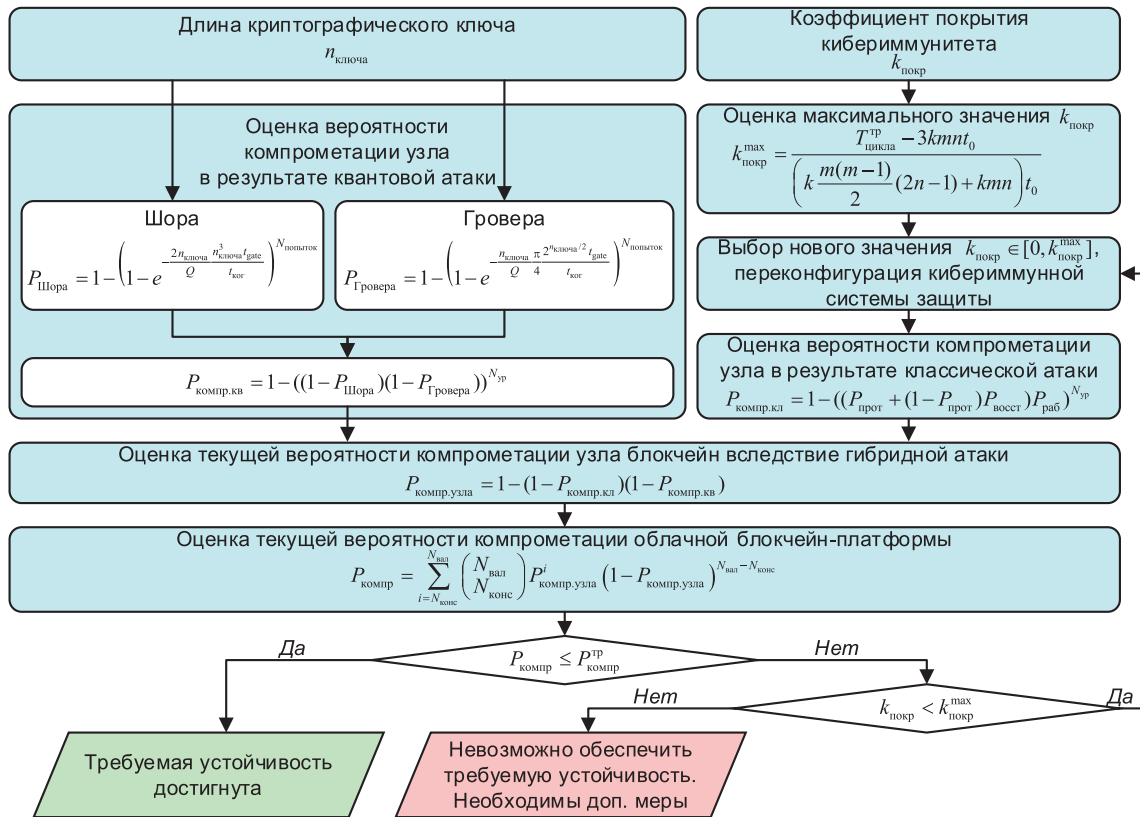


Рис. 5. Схема метода обеспечения устойчивости облачных блокчейн-платформ в условиях гибридных квантово-классических атак

$$\exists n_{\text{ключа}}, k_{\text{покр}} : \begin{cases} P_{\text{компр}}(n_{\text{ключа}}, k_{\text{покр}}) \leq P_{\text{компр}}^{\text{тр}} \\ T_{\text{цикла}}(n_{\text{ключа}}, k_{\text{покр}}) \leq T_{\text{цикла}}^{\text{тр}} \end{cases} \quad (21)$$

4. Исследование метода обеспечения устойчивости облачных блокчейн-платформ

Для проверки выдвинутой гипотезы проведем экспериментальное исследование разработанного метода. Целью эксперимента является определение влияния параметров гибридных атак, а именно количества кубитов и доли вредоносных входных данных, на вероятность компрометации узла облачной блокчейн-платформы, а также определение возможностей обеспечения требуемой устойчивости функционирования облачной блокчейн-платформы за счет применения разработанного метода.

Зададим начальные значения параметров: $N_{\text{вал}} = 10000$, $k = 1000$, $m = 5$, $n = 2$ – как практически возможные параметры облачных блокчейн-платформ; $N_{\text{ур}} = 4$, $k_{\text{изв}} = 0,4$, $P_{\text{мод}} = 0,75$, $p = 0,875$ – на основе результатов исследования облачных платформ с кибериммунитетом [33, 34]; $b = 0,51$, $t_{\text{gate}} = 10^{-7}$, $t_{\text{ког}} = 10^{-2}$ – на основе принятых в настоящей работе допущений.

На рис. 6 представлены результаты исследования зависимости вероятности компрометации узла вследствие квантовой и классической атак от длины

криптографического ключа и коэффициента покрытия кибериммунитета соответственно, полученные на основе формул (4) и (7).

Как видно на рис. 6а, увеличение длины криптографического ключа $n_{\text{ключа}}$ затрудняет осуществление квантовой атаки с применением алгоритмов Шора и Гровера даже при наличии у атакующего квантового компьютера с достаточно большим количеством кубитов ($Q = 1000$ и более). Так, например, при использовании 256-битного ключа в криптосистеме AES вычислительная сложность перебора с помощью алгоритма Гровера составит $O(2^{128})$, что делает практически невозможной ее компрометацию за разумное время при текущих ограничениях квантовых вычислений. На рис. 6б видно, что увеличение доли линейных блоков программ, для которых выполняется контроль семантики вычислений, определяемой коэффициентом покрытия кибериммунитета $k_{\text{покр}}$, положительно влияет на способность узла блокчейн противодействовать классическим атакам и снижает вероятность его компрометации. Таким образом, увеличение значений $n_{\text{ключа}}$ и $k_{\text{покр}}$ повышает устойчивость функционирования узла облачной блокчейн-платформы в условиях гибридных атак.

Пусть теперь требуется обеспечить общую вероятность компрометации облачной блокчейн-платформы

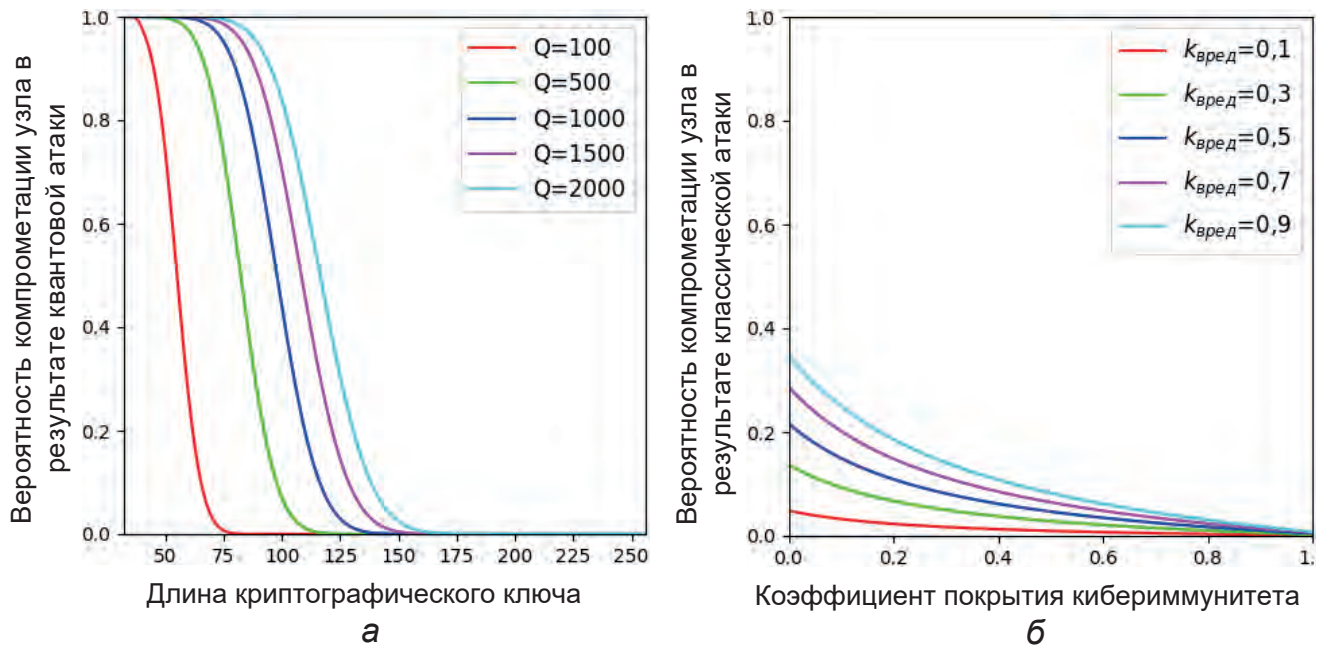


Рис. 6. Вероятность компрометации узла: а – вследствие квантовой атаки; б – вследствие классической атаки

не выше $P_{компр}^{тр} = 0,05$ и время выполнения программного цикла на узле блокчейн не более $T_{цикла}^{тр} = 45000$ ед. времени. На рис. 7 представлены результаты исследования вероятности компрометации облачной блокчейн-платформы в условиях гибридных квантово-классических атак при применении разработанного метода, полученные на основе формулы (1).

При заданных параметрах облачной блокчейн-платформы и требования к времени выполнения программного цикла на узлах блокчейн максимально допустимое значение коэффициента покрытия кибериммунитета, вычисленное на основе (20), составило $k_{покр}^{max} = 0,375$. Для случаев, представленных на рис. 7а, применение разработанного метода позволяет обеспечить требуемую устойчивость

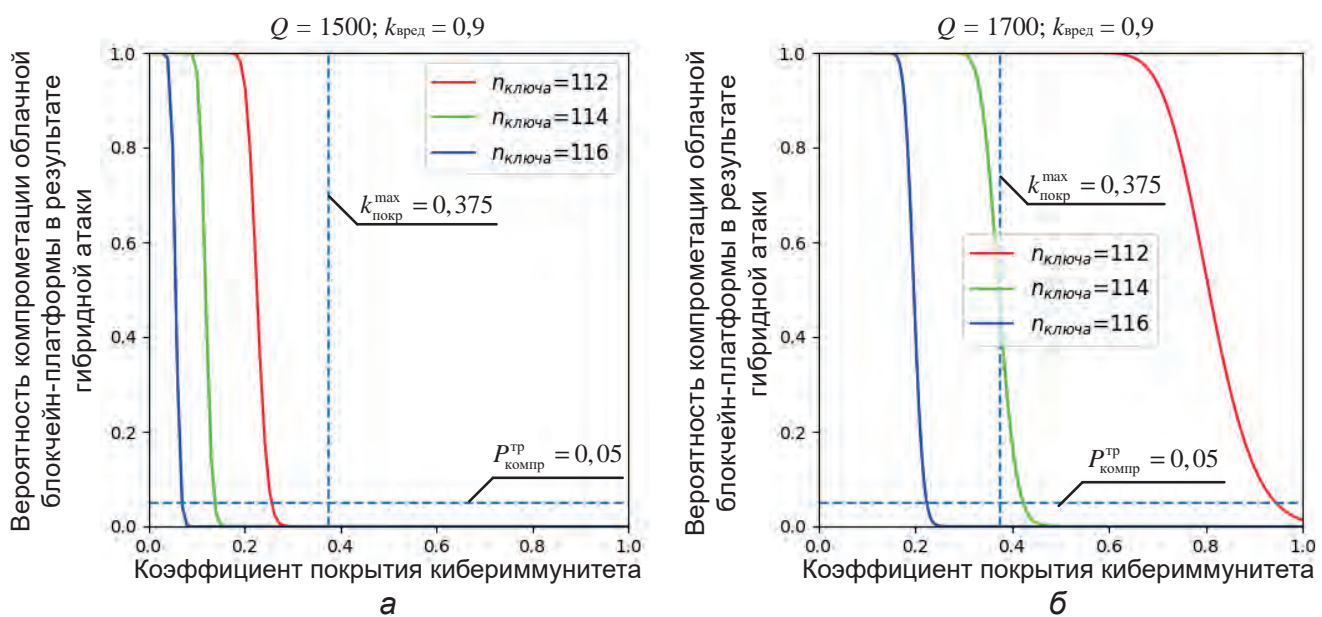


Рис. 7. Вероятность компрометации облачной блокчейн-платформы вследствие гибридных квантово-классических атак: а – при наличии возможности обеспечения требуемой устойчивости; б – при отсутствии возможности обеспечения требуемой устойчивости

функционирования облачной блокчейн-платформы в условиях гибридных атак за счет варьирования $k_{\text{покр}}$ в диапазоне $[0, k_{\text{покр}}^{\text{max}}]$. Однако при увеличении количества кубитов, доступных атакующему, до $Q = 1700$ обеспечение требуемой устойчивости в соответствии и условием (21) становится возможным лишь при $n_{\text{ключа}} = 116$ (рис. 76), в остальных же случаях необходимо применение дополнительных мер защиты.

Таким образом, результаты экспериментальных исследований согласуются с теоретическими выводами, а значит выдвинутую гипотезу о том, что применение разработанного метода позволяет обеспечить требуемую устойчивость функционирования облачных блокчейн-платформ в условиях гибридных квантово-классических атак, можно считать подтвержденной.

Выводы

В работе предложен метод обеспечения устойчивости облачных блокчейн-платформ на основе кибериммунитета, позволяющий обеспечивать требования к показателю вероятности компрометации при ограничении на время выполнения программного цикла узла блокчейн за счет варьирования длины криптографического ключа и коэффициента покрытия кибериммунитета. Сформулирована и экспериментально подтверждена гипотеза о том, что применение данного метода позволяет обеспечить требуемую устойчивость функционирования облачных

блокчейн-платформ в условиях гибридных квантово-классических атак.

Достоверность исследования подтверждается согласованностью результатов эксперимента с теоретическими выводами, а также непротиворечивостью полученных результатов с известными работами в предметной области.

К направлениям дальнейших исследований можно отнести следующее:

- исследование возможностей применения методов искусственного интеллекта и машинного обучения для оптимального управления параметрами противодействия квантово-классическим атакам;
- исследование возможностей применения иных математических аппаратов для контроля семантики вычислений на узлах блокчейн (лямбда-исчисления; троек Хоара; схем Ляпунова, Лаврова, Ершова, Янова, Летичевского; темпоральных логик; систем алгоритмических алгебр и других);
- разработка научно-обоснованных рекомендаций по созданию и внедрению технических и технологических решений для обеспечения требуемой устойчивости функционирования облачных блокчейн-экосистем и платформ «Экономики данных» Российской Федерации в условиях гибридных квантово-классических атак.

Результаты получены при финансовой поддержке проекта «Технологии противодействия ранее неизвестным квантовым киберугрозам», реализуемого в рамках государственной программы федеральной территории «Сириус» «Научно-технологическое развитие федеральной территории «Сириус» (Соглашение №23-03 от 27.09.2024 г.).

Литература

1. Mourtzis D., Angelopoulos J., Panopoulos N. Blockchain Integration in the Era of Industrial Metaverse // Applied Sciences. 2023. Vol. 13. No. 3. P. 1353. DOI: 10.3390/app13031353.
2. Марков А. С. Важная веха в безопасности открытого программного обеспечения // Вопросы кибербезопасности. 2023. № 1(53). С. 2–12. DOI: 10.21681/2311-3456-2023-1-2-12.
3. Gai K., Guo J., Zhu L., Yu S. Blockchain Meets Cloud Computing: A Survey // IEEE Communications Surveys & Tutorials. 2020. Vol. 22. No. 3. Pp. 2009–2030. DOI: 10.1109/COMST.2020.2989392.
4. Khanna A., Sah A., Bolshev V., Burgio A., et al. Blockchain–Cloud Integration: A Survey // Sensors. 2022. No. 22(14). P. 5238. DOI: 10.3390/s22145238.
5. Петренко А. С., Ломако А. Г., Петренко С. А. Анализ современного состояния исследований проблемы квантовой устойчивости блокчейна. Часть 1 // Защита информации. Инсайд. 2023. № 3(111). С. 38–46.
6. Петренко С. А., Балябин А. А. Модель квантовых угроз безопасности информации для национальных блокчейн-экосистем и платформ // Вопросы кибербезопасности. 2025. № 1(65). С. 7–17. DOI: 10.21681/2311-3456-2025-1-7-17.
7. Балябин А. А., Петренко С. А., Костюков А. Д. Модель угроз безопасности и киберустойчивости облачных платформ КИИ РФ // Защита информации. Инсайд. 2024. № 5(119). С. 26–34.
8. Fernandez-Carmona T. M., Fraga-Lamas P. Towards post-quantum blockchain: A review on blockchain cryptography resistant to quantum computing attacks // IEEE Access. 2020. Vol. 8. Pp. 21091–21116. DOI: 10.1109/ACCESS.2020.2968985.
9. Shahid F., Khan A. Smart Digital Signatures (SDS): A post-quantum digital signature scheme for distributed ledgers // Future Generation Computer Systems. 2020. Vol. 111. Pp. 241–253. DOI: 10.1016/j.future.2020.04.042.
10. Sahin M. S., Akleyek S. A survey of quantum secure group signature schemes: Lattice-based approach // Journal of Information Security and Applications. 2023. Vol. 73. P. 103432. DOI: 10.1016/j.jisa.2023.103432.
11. Москвин В. С., Богатырев В. А. Постквантовые алгоритмы электронной цифровой подписи и их использование в распределенном реестре // Наукоемкие технологии в космических исследованиях Земли. 2022. Т. 14. № 4. С. 47–53. DOI: 10.36724/2409-5419-2022-14-4-47-53.

12. Yang Z., Salman T., Jain R., Pietro R. D. Decentralization Using Quantum Blockchain: A Theoretical Analysis // IEEE Transactions on Quantum Engineering. 2020. Vol. 3. Pp. 1–16. DOI: 10.1109/TQE.2022.3207111.
13. Li Q., Wu J., Quan J., Shi J., Zhang S. Efficient Quantum Blockchain With a Consensus Mechanism QDPoS // IEEE Transactions on Information Forensics and Security. 2022. Vol. 17. Pp. 3264–3276. DOI: 10.1109/TIFS.2022.3203316.
14. Wen X. J., Chen Y. Z., Fan X. C., Zhang W., et al. Blockchain consensus mechanism based on quantum zero-knowledge proof // Optics and Laser Technology. 2022. Vol. 147. P. 107693. DOI: 10.1016/j.optlastec.2021.107693.
15. Sun X., Kulicki P., Sopek M. Multi-Party Quantum Byzantine Agreement without Entanglement // Entropy. 2020. Vol. 22. No. 10. P. 1152. DOI: 10.3390/e22101152.
16. Singh S., Rajput N. K., Rath V. K., Pandey H. M., et al. Securing Blockchain Transactions Using Quantum Teleportation and Quantum Digital Signature // Neural Processing Letters. 2023. Vol. 55. Pp. 3827–3842. DOI: 10.1007/s11063-020-10272-1.
17. Wang W., Yu Y., Du L. Quantum blockchain based on asymmetric quantum encryption and a stake vote consensus algorithm // Scientific Reports. 2022. Vol. 12. No. 1. DOI: 10.1038/s41598-022-12412-0.
18. Петренко А. С., Петренко С. А., Костюков А. Д. Эталонная модель блокчейн-платформы // Защита информации. Инсайд. 2022. № 4(106). С. 34–44.
19. Петренко А. С., Петренко С. А. Метод оценивания квантовой устойчивости блокчейн-платформ // Вопросы кибербезопасности. 2022. № 3(49). С. 2–22. DOI: 10.21681/2311-3456-2022-3-2-22.
20. Балябин А. А., Петренко С. А. О создании киберустойчивых облачных платформ управления киберфизическими объектами в условиях роста угроз безопасности // Защита информации. Инсайд. 2025. № 4(124). С. 20–30.
21. Зегжда Д. П., Александрова Е. Б., Калинин М. О., Марков А. С. и др. Кибербезопасность цифровой индустрии. Теория и практика функциональной устойчивости к кибератакам. М.: Научно-техническое издательство «Горячая линия-Телеком». 2021. 560 с.
22. Андрушкевич Д. В., Бирюков Д. Н., Тимашов П. В. Порождение сценариев предотвращения компьютерных атак на основе логико-онтологического подхода // Труды Военно-космической академии имени А. Ф. Можайского. 2021. № 677. С. 118–134.
23. Шелухин О. И., Рыбаков С. Ю., Ванюшина А. В. Влияние фрактальной размерности на качество классификации компьютерных атак методами машинного обучения // Научные технологии в космических исследованиях Земли. 2023. Т. 15. № 1. С. 57–64. DOI: 10.36724/2409-5419-2023-15-1-57-64.
24. Aljuaid W. H., Alshamrani S. S. A deep learning approach for intrusion detection systems in cloud computing environments // Applied sciences. 2024. Vol. 14. No. 13. P. 5381. DOI: 10.3390/app14135381.
25. Новикова Е. С., Котенко И. В., Мелешко А. В., Израйлов К. Е. Обнаружение вторжений на основе федеративного обучения: архитектура системы и эксперименты // Вопросы кибербезопасности. 2023. № 6(58). С. 50–66. DOI: 10.21681/2311-3456-2023-6-50-66.
26. Шамсутдинов Р. Р., Васильев В. И., Вульфин А. М. Интеллектуальная система мониторинга информационной безопасности промышленного интернета вещей с использованием механизмов искусственных иммунных систем // Системная инженерия и информационные технологии. 2024. Т. 6. № 4(19). С. 14–31. DOI: 10.54708/2658-5014-SIIT-2024-no4-p14.
27. Балябин А. А. Модель облачной платформы КИИ РФ с кибериммунитетом в условиях информационно-технических воздействий // Защита информации. Инсайд. 2024. № 5(119). С. 35–44.
28. Балябин А. А., Петренко С. А., Костюков А. Д. Метод восстановления облачных и пограничных вычислений на основе кибериммунитета // Защита информации. Инсайд. 2022. № 6(108). С. 26–31.
29. Wrieden J. K., Vassilakis V. G. An Analysis of the Threats Posed by Botnet Malware Targeting Vulnerable Cryptocurrency Miners // 2023 Fourteenth International Conference on Ubiquitous and Future Networks (ICUFN), Paris, France. 2023. Pp. 82–87. DOI: 10.1109/ICUFN57995.2023.10201027.
30. Петренко А. С. Метод анализа квантовой устойчивости национальных блокчейн-экосистем и платформ // Защита информации. Инсайд. 2025. № 2(122). С. 18–27.
31. Tuokkola, M., Sunada, Y., Kivijärvi, H. et al. Methods to achieve near-millisecond energy relaxation and dephasing times for a superconducting transmon qubit // Nature Communications. 2025. Vol. 16, 5421. DOI: 10.1038/s41467-025-61126-0.
32. Milul O., Guttel B., Goldblatt U., Hazanov S., et al. Superconducting cavity qubit with tens of milliseconds single-photon coherence time // PRX Quantum. 2023. Vol. 4. No. 3. 030336. DOI: 10.1103/PRXQuantum.4.030336.
33. Балябин А. А., Петренко С. А. Модель самовосстановления киберфизических систем КИИ РФ в условиях кибератак на основе кибериммунитета // The 2025 Symposium on Cybersecurity of the Digital Economy – CDE'25 : Сборник трудов IX Международной научно-технической конференции, Иннополис. 2025. С. 76–91.
34. Балябин А. А., Петренко С. А. Методика самовосстановления киберфизических систем КИИ РФ в условиях кибератак на основе кибериммунитета // The 2025 Symposium on Cybersecurity of the Digital Economy – CDE'25 : Сборник трудов IX Международной научно-технической конференции, Иннополис. 2025. С. 103–114.

METHOD OF ENSURING CYBER RESISTANCE BLOCKCHAIN PLATFORMS BASED ON CYBER IMMUNITY

Balyabin A. A.³, Petrenko S. A.⁴

Keywords: threats to information security, quantum threats to security, blockchain ecosystems and platforms, cybersecurity, cyber resilience, methods of analysis and synthesis of quantum-resistant blockchain.

Purpose of the research: ensuring the resilience of cloud blockchain ecosystems and platforms of the 'Data Economy' of the Russian Federation based on cyber immunity in the face of hybrid quantum-classical attacks.

3 Artyom Balyabin, Junior Researcher, Scientific Center for Information Technologies and Artificial Intelligence, Sirius University of Science and Technology. Sirius Federal Territory, Russia. ORCID 0009-0006-3949-154X, E-mail: Balyabin.AA@talantiuspeh.ru

4 Sergei A. Petrenko, Dr.Sc. (of Tech.) (Grand Doctor, Full Professor), Team Leader, Scientific Center for Information Technologies and Artificial Intelligence, Sirius University of Science and Technology. Sirius Federal Territory, Russia. ORCID 0000-0003-0644-1731. E-mail: Petrenko.SA@talantiuspeh.ru

Methods of the research: methods of system analysis, methods of probability theory and mathematical statistics, methods of the theory of stability of complex systems.

Results of the research: the analysis of research in the subject area has demonstrated that existing methods for ensuring the resilience of various information and computing systems consider either classical or quantum threats and do not fully ensure the resilience of cloud blockchain platforms under hybrid attacks characterized by the presence of both threat types. To resolve this problematic situation, an objective has been set to develop a novel method for ensuring the resilience of cloud blockchain platforms based on cyber immunity, and a hypothesis has been formulated regarding the feasibility of achieving the research goal through the application of this method.

A method for ensuring the resilience of cloud blockchain platforms based on cyber immunity under hybrid quantum-classical attacks has been developed, enabling compliance with the compromise-probability requirements under constraints on the execution time of a blockchain node's program cycle by varying the cryptographic key length and the cyber immunity coverage coefficient.

A study of the developed method has demonstrated the capability to achieve the required resilience of cloud blockchain platforms under hybrid quantum-classical attacks, and the conditions for the existence of a solution have been determined, thereby confirming the formulated hypothesis.

Scientific novelty: the developed method is the first to take into account new conditions such as hybrid attacks on cloud blockchain platforms, which are formally described through newly introduced parameters of the number of quantum computer qubits available to the attacker and the proportion of malicious input data. Furthermore, the application of this method for the first time imbues cloud blockchain platforms with a new emergent property of cyber immunity, which consists of the ability to detect known and previously unknown attacks aimed at violating the semantics of computations, counter them, and restore normal operation if violations occur

References

1. Mourtzis, D., Angelopoulos, J., & Panopoulos, N. (2023). Blockchain Integration in the Era of Industrial Metaverse. *Applied Sciences*, 13, 3, 1353. DOI: 10.3390/app13031353.
2. Markov, A. S. (2023). Important milestones in open source software security. *Cybersecurity issues*, 1(53), 2–12. DOI: 10.21681/2311-3456-2023-1-2-12.
3. Gai, K., Guo, J., Zhu, L., & Yu, S. (2020). Blockchain Meets Cloud Computing: A Survey. *IEEE Communications Surveys & Tutorials*, 22, 3, 2009–2030. DOI: 10.1109/COMST.2020.2989392.
4. Khanna, A., Sah, A., Bolshev, V., Burgio, A., Panchenko, V., & Jasinski, M. (2022). Blockchain–Cloud Integration: A Survey. *Sensors*, 22(14), 5238. DOI: 10.3390/s22145238.
5. Petrenko, A. S., Lomako, A. G., & Petrenko, S. A. (2023). Analysis of the Current State of Research Blockchain Quantum Resilience Problems. Part 1. *Zašita informacii. Inside*, 3(111), 38–46.
6. Petrenko, S. A., & Balyabin A. A. (2025). Model of quantum threats to national blockchain ecosystems and platforms. *Cybersecurity issues*, 1(65), 7–17. DOI: 10.21681/2311-3456-2025-1-7-17.
7. Balyabin, A. A., Petrenko, S. A., & Kostyukov, A. D. (2024). Model of security threats and cyber resistance of cloud platforms of the critical IT infrastructure of the Russian Federation. *Zašita informacii. Inside*, 5(119), 26–34.
8. Fernandez-Carames, T. M., & Fraga-Lamas, P. (2020). Towards post-quantum blockchain: A review on blockchain cryptography resistant to quantum computing attacks. *IEEE Access*, 8, 21091–21116. DOI: 10.1109/ACCESS.2020.2968985.
9. Shahid, F., & Khan, A. (2020). Smart Digital Signatures (SDS): A post-quantum digital signature scheme for distributed ledgers. *Future Generation Computer Systems*, 111, 241–253. DOI: 10.1016/j.future.2020.04.042.
10. Sahin, M. S., & Akleyek, S. (2023). A survey of quantum secure group signature schemes: Lattice-based approach. *Information Security Applications*, 73, 103432. DOI: 10.1016/j.jisa.2023.103432.
11. Moskvina, V. S., & Bogatyrev, V. A. (2022). Post-quantum digital signing algorithms and their application in distributed registry. *High technologies in Earth space research*, 14, 4, 47–53. DOI: 10.36724/2409-5419-2022-14-4-47-53.
12. Yang, Z., Salman, T., Jain, R., & Pietro, R.D. (2020). Decentralization Using Quantum Blockchain: A Theoretical Analysis. *IEEE Transactions on Quantum Engineering*, 3, 1–16. DOI: 10.1109/TQE.2022.3207111.
13. Li, Q., Wu, J., Quan, J., Shi, J., & Zhang, S. (2022). Efficient Quantum Blockchain With a Consensus Mechanism QDPoS. *IEEE Transactions on Information Forensics and Security*, 17, 3264–3276. DOI: 10.1109/TIFS.2022.3203316.
14. Wen, X. J., Chen, Y. Z., Fan, X. C., Zhang, W., Yi, Z. Z., & Fang, J. B. (2022). Blockchain consensus mechanism based on quantum zero-knowledge proof. *Optics and Laser Technology*, 147, 107693. DOI: 10.1016/j.optlastec.2021.107693.
15. Sun, X., Kulicki, P., & Sopek, M. (2020). Multi-Party Quantum Byzantine Agreement without Entanglement. *Entropy*, 22, 10, 1152. DOI: 10.3390/e22101152.
16. Singh, S., Rajput, N. K., Rathi, V. K., Pandey, H. M., Jaiswal, A. K., & Tiwari, P. (2023). Securing Blockchain Transactions Using Quantum Teleportation and Quantum Digital Signature. *Neural Processing Letters*, 55, 3827–3842. DOI: 10.1007/s11063-020-10272-1.
17. Wang, W., Yu, Y., & Du, L. (2022). Quantum blockchain based on asymmetric quantum encryption and a stake vote consensus algorithm. *Scientific Reports*, 12, 1, 1. DOI: 10.1038/s41598-022-12412-0.
18. Petrenko, A. S., Petrenko, S. A., & Kostyukov, A. D. (2022). The perfect model of the blockchain platform. *Zašita informacii. Inside*, 4(106), 34–44.
19. Petrenko, A. S., & Petrenko, S. A. (2022). Quantum resilience estimation method blockchain. *Cybersecurity issues*, 3(49), 2–22. DOI 10.21681/2311-3456-2022-3-2-22.
20. Balyabin A. A., & Petrenko S. A. (2025). On the creation of cyberresilient cloud platforms for managing cyber-physical objects in the context of growing security threats. *Zašita informacii. Inside*, 4(124), 20–30.

21. Zegzhda, D. P., Aleksandrova, E. B., Kalinin, M. O., Markov, A. S., Zhukov, I. Yu., Ivanov, D. V., Konoplev, A. S., Lavrova, D. S., Moskvina, D. A., Pavlenko, E. Yu., Poltavtseva, M. A., Shenets, N. N., Dakhnovich, A. D., & Krundyshev, V. M. (2021). Kiberbezopasnost' tsifrovoy industrii. Teoriya i praktika funktsional'noi ustoichivosti k kiberatakam. Moscow, Goryachaya liniya-Telekom Publ, 560 p.
22. Andrushkevich, D. V., Biryukov, D. N., & Timashov, P. V. (2021). Porozhdenie stsensariyev predotvrashcheniya komp'yuternykh atak na osnove logiko-ontologicheskogo podkhoda. Trudy Voenno-kosmicheskoy akademii imeni A. F. Mozhaiskogo, 677, 118–134.
23. Sheluhin, O. I., Rybakov, S. Yu., & Vanyushina, A. V. (2023). Influence of fractal dimension on quality classification of computer attacks by machine learning methods. High technologies in Earth space research, 15, 1, 57–64. DOI: 10.36724/2409-5419-2023-15-1-57-64.
24. Aljuaid, W. H., & Alshamrani, S. S. (2024). A deep learning approach for intrusion detection systems in cloud computing environments. Applied sciences, 14, 13, 5381. DOI: 10.3390/app14135381.
25. Novikova, E. S., Kotenko, I. V., Meleshko, A. V., & Izrailov, K. E. (2023). Federated learning based intrusion detection: system architecture and experiments. Cybersecurity issues, 6(58), 50–66. DOI: 10.21681/2311-3456-2023-6-50-66.
26. Shamsutdinov, R. R., Vasilyev, V. I., & Vulfin, A. M. (2024). Intelligent system for monitoring information security of the industrial internet of things using artificial immune systems mechanisms. System engineering and information technologies, 6, 4(19), 14–31. DOI: 10.54708/2658-5014-SIIT-2024-no4-p14.
27. Balyabin, A. A. (2024). Model of the cloud platform of critical IT infrastructure of the Russian Federation under the conditions of information technology impacts. Zašita informacii. Inside, 5(119), 35–44.
28. Balyabin, A. A., Petrenko S. A., & Kostyukov A. D. (2022). Cloud and edge recovery method computing based on cyber immunity. Zašita informacii. Inside, 6(108), 26–31.
29. Wrieden, J. K., & Vassilakis, V. G. (2023). An Analysis of the Threats Posed by Botnet Malware Targeting Vulnerable Cryptocurrency Miners. 2023 Fourteenth International Conference on Ubiquitous and Future Networks (ICUFN), 82–87. DOI: 10.1109/ICUFN57995.2023.10201027.
30. Petrenko, A. S. (2025). Method for analyzing the quantum resilience of national blockchain ecosystems and platforms. Zašita informacii. Inside, 2(122), 18–27.
31. Tuokkola, M., Sunada, Y., Kivijarvi, H., Albanese, J., Gronberg, L., Kaikkonen, J. P., Vesterinen V., Govenius, J., & Mottonen, M. (2025). Methods to achieve near-millisecond energy relaxation and dephasing times for a superconducting transmon qubit. Nature Communications, 16, 5421. DOI: 10.1038/s41467-025-61126-0.
32. Milul, O., Guttel, B., Goldblatt, U., Hazanov, S., Joshi, L. M., Chausovsky, D., Kahn, N., Ciftiyurek, E., Lafont, F., & Rosenblum, S. (2023). Superconducting cavity qubit with tens of milliseconds single-photon coherence time. PRX Quantum, 4, 3, 030336. DOI: 10.1103/PRXQuantum.4.030336.
33. Balyabin, A. A., & Petrenko S. A. (2025). Model' samovosstanovlenie kiberfizicheskikh sistem KII RF v usloviyakh kiberatak na osnove kiberimmuniteta. The 2025 Symposium on Cybersecurity of the Digital Economy – CDE'25 : Collected papers, 76–91.
34. Balyabin, A. A., & Petrenko S. A. (2025). Metodika samovosstanovleniya kiberfizicheskikh sistem KII RF v usloviyakh kiberatak na osnove kiberimmuniteta. The 2025 Symposium on Cybersecurity of the Digital Economy – CDE'25 : Collected papers, 103–114.



ИССЛЕДОВАНИЕ СИГНАЛЬНЫХ МЕТОДОВ ВЫЯВЛЕНИЯ СОСТЯЗАТЕЛЬНЫХ АТАК НА НЕЙРОННЫЕ МОДЕЛИ ЧЕРЕЗ ОБЪЯСНИТЕЛЬНУЮ ВИЗУАЛИЗАЦИЮ

Чеколаев Д. А.¹, Шевченко А. В.², Алексеев А. К.³, Трофимов Ю. В.⁴, Аверкин А. Н.⁵

DOI: 10.21681/2311-3456-2025-6-140-146

Цель исследования: разработка сигнального метода определения состояния состязательной атаки на графические объекты, обрабатываемые нейронной сетью при классификации.

Метод(ы) исследования: использование метода объяснительного искусственного интеллекта Grad-CAM для визуализации состязательной FGSM-атаки, предложен сигнальный метод на основе интегрального расчета поверхности градиента тепловой карты объяснения.

Результат(ы) исследования: описаны и опробованы методы объяснительного искусственного интеллекта для выделения важных признаков и способы использования полученных тепловых карт для обнаружения атак. В практической части рассмотрен один из современных подходов: анализ смещения и размытия объяснений с помощью Grad-CAM. Проведен анализ и обзор эффективности в повышении устойчивости модели к атакам. Выделены различные эффекты воздействия в следствии атак на зоны внимания и характер их изменения. Предложен интегральный метод расчета факта наличия состязательной атаки во входном изображении, что применимо для автоматической детекции атаки.

Научная новизна: Исследование направлено на повышение информативности о характере атаки, степени воздействия на атакуемое входное изображение, формирование сигнального метода детектирования наличия состязательной атаки.

Вклад авторов: Чеколаев Д. А. и Шевченко А. В. – составление и реализация концепции визуализации состязательной атаки, на основе опубликованных исследований, Алексеев А. К. – описание методов состязательных атак, Трофимов Ю. В. и Аверкин А. В. – теоретическое обоснование применения методов объяснений.

Ключевые слова: нейросетевые технологии, атаки на системы искусственного интеллекта, атаки на объяснительный искусственный интеллект, информационная безопасность, визуализация атак на нейронную сеть.

Введение

В последние годы глубокие нейронные сети достигли выдающихся результатов во множестве задач, однако их применение в критически важных областях сдерживается проблемой состязательных атак [1]. Под такой атакой понимаются специально сконструированные малые возмущения входных данных, незаметные для человека, но приводящие модель к неверному решению [2].

Пусть x — исходное изображение, y — правильный класс изображения, $f(x)$ — функция классификации нейронной сети, θ — параметры модели.

Состязательная атака создаёт возмущение δ , такое что: $\|\delta\| \leq \epsilon$, где ϵ — максимально допустимая норма возмущения; $f(x + \delta) \neq y$, при этом $x + \delta$ визуально неотличимо от x .

Процесс создания состязательного примера можно описать как решение оптимизационной задачи:

$$\delta^* = \arg \min_{\delta} L(f(x + \delta), y) + \lambda \|\delta\|^2, \quad (1)$$

где L — функция потерь, λ — коэффициент регуляризации.

Например, добавление слабого шума к изображению может заставить классификатор ошибочно распознать объект, сохраняя при этом высокую уверенность.

Эти атаки обладают рядом характерных свойств:

1. Практически неразличимы визуально:

$$\|x - (x + \delta)\| \approx 0. \quad (2)$$

2. Целенаправленно управляют результатом (ошибкой):

$$P(f(x + \delta) = y') \gg P(f(x) = y'), \text{ где } y' \neq y. \quad (3)$$

3. Часто переносимы между моделями

$$f_1(x + \delta) = f_2(x + \delta) = y' \neq y, \quad (4)$$

где f_1 и f_2 — разные модели.

То есть, успешное возмущение, рассчитанное для одной архитектуры, может влиять и на другую [2].

Параллельно развивается направление объяснимого искусственного интеллекта (XAI), призванное сделать работу сложных моделей более прозрачной

1 Чеколаев Дмитрий Алексеевич, магистр, Государственный университет «Дубна». г. Дубна, Россия. E-mail: D.1369@icloud.com

2 Шевченко Алексей Валерьевич, старший преподаватель, аспирант, Государственный университет «Дубна», г. Дубна, Россия. E-mail: leviathan0909@gmail.com

3 Артем Кириллович Алексеев, бакалавр, Государственный университет «Дубна». г. Дубна, Россия. E-mail: aak.24@uni-dubna.ru

4 Трофимов Юрий Владиславович, инженер-программист, лаборатория информационных технологий им. Мещерякова, Объединённый институт ядерных исследований (ОИЯИ). г. Дубна, Россия. Аспирант, Государственный университет «Дубна». г. Дубна, Россия. E-mail: ura_trofim@bk.ru

5 Аверкин Алексей Николаевич, кандидат технических наук, ведущий научный сотрудник, Федеральное государственное учреждение «Федеральный исследовательский центр «Информатика и управление», Российская академия наук. г. Москва, Россия. доцент, Государственный университет «Дубна». г. Дубна, Россия. E-mail: averkin2003@inbox.ru

для человека. ХАI-методы генерируют объяснения к предсказаниям модели, например, выделяя наиболее важные фрагменты входных данных, повлиявшие на решение. В контексте компьютерного зрения широко применяются визуальные объяснения в виде тепловых карт (heatmap), показывающих значимость каждого участка изображения. Так, метод Grad-CAM строит карту активаций на основе градиентов, указывая, какие области изображения «привлекли внимание» сверточной сети при классификации: холодные тона соответствуют несущественным регионам, а теплые (красные) выделяют наиболее значимые фрагменты [2]. Подобные приемы повышают интерпретируемость моделей и доверие пользователей, особенно в критичных приложениях (медицина, автономный транспорт и др.).

Однако недавние работы выявили ограниченности и уязвимости методов объяснения под воздействием атак [3]. Состязательные воздействия могут приводить не только к сбоям в предсказаниях сети, но и искажать сопровождающие их объяснения, ставя под вопрос достоверность выводов ХАI [3]. Иными словами, злоумышленник способен вводить в заблуждение интерпретатор модели, «обманывая» механизм объяснения или приукрашивая истинные причины решения (эффект известен как *fairwashing* – от англ. «отбеливание») [3]. Это представляет особую опасность в высокочисленных областях применения моделей, где на объяснения возлагается ответственность за обоснование решений. В ответ на данную угрозу исследователи начали изучать методы, повышающие робастность интерпретаций и защищающие объяснения от враждебных манипуляций [3]. Кроме того, набирает силу подход, при котором сами средства объяснения используются для активной защиты моделей – например, для выявления факта атаки или для фильтрации входных данных.

Целью данной статьи является обзор и анализ таких подходов защиты нейросетевых моделей от состязательных атак с помощью объяснительной визуализации. Постановка задачи формулируется через проблему нестабильности ХАI-объяснений под атакой. В разделе «Методы» описываются основные подходы ХАI для визуализации признаков и то, каким образом визуальные объяснения могут помочь обнаружить факт атаки. Практическая часть посвящена методике из современной литературы: метрике на основе Grad-CAM для анализа атак. [4]

Наконец, в выводах обсуждается, как визуальные объяснения могут быть интегрированы в активную систему защиты моделей, повышая их устойчивость и надежность.

Постановка задачи

Проблема нестабильности ХАI-объяснений под атакой, состоит в том, что состязательные атаки не только

влияют на выход модели (классификацию), но и способны существенно изменять или искажать генерируемые моделью объяснения.

Пусть $S(x)$ – функция генерации объяснений (например, Grad-CAM) для исходного изображения x ; $S(x + \delta)$ – функция генерации объяснений для возмущенного изображения; $d(S_1, S_2)$ – функция расстояния (меры различия) между двумя объяснениями.

Тогда нестабильность объяснений выглядит как:

$$d(S(x), S(x + \delta)) \gg \|x - (x + \delta)\|, \quad (5)$$

где левая часть показывает большое изменение в объяснении, а правая часть отражает малое возмущение входных данных.

Под нестабильностью объяснений понимается ситуация, когда небольшое целенаправленное возмущение входных данных приводит к несоразмерно сильным изменениям в карте значимости модели, отражающей «ход мысли» нейросети. В нормальных условиях объяснение для схожих изображений должно быть относительно устойчивым:

$$\|x_1 - x_2\| \leq \epsilon \Rightarrow d(S(x_1), S(x_2)) \leq \epsilon', \quad (6)$$

где ϵ' – малая величина.

Напротив, под атакой модель может начать фокусироваться на иных, неинформативных деталях или шумовых пикселях:

$$\epsilon \delta: \|\delta\| \leq \epsilon, \text{ но } d(S(x), S(x + \delta)) \gg \epsilon'. \quad (7)$$

Например, показано, что при атаке Grad-CAM тепловые карты систематически смещают свое выделение на другие части изображения, отличные от тех, что были важны для изначального (корректного) распознавания [2]. Это свидетельствует о принципиальном изменении внутренних признаков, используемых сетью, под влиянием атаки.

Более того, атакующий может намеренно эксплуатировать уязвимости метода объяснения. Есть возможность манипулировать объяснением модели без явного ухудшения её точности – вплоть до того, что вредоносная модель может скрывать свои истинные критерии решения, предоставляя правдоподобные, но ложные объяснения наблюдателю [3]. Такое «объяснительное мошенничество» крайне опасно [3], так как подрывает доверие к системе: пользователь получает подтверждение якобы корректной работы модели, тогда как на самом деле ее вывод сфабрикован или обусловлен посторонними факторами.

Постановка задачи сводится к обеспечению устойчивости и достоверности объяснений нейросети перед лицом состязательных воздействий. Необходимо разработать подходы, способные: обнаруживать факт атаки по аномалиям в визуальном объяснении и противодействовать искажению объяснений, сохраняя интерпретируемость модели. Иными словами,

требуется научиться распознавать, когда полученная от модели тепловая карта не соответствует реальным сущностям изображения, а затем либо сигнализировать об атаке, либо автоматически корректировать входные данные или модель для восстановления правильного объяснения. Решение этой задачи позволит сделать системы с глубоким обучением более надежными: объяснение станет не пассивным описанием, а частью механизма защиты, повышающего доверие к модельным предсказаниям.

Методы объяснительной визуализации и обнаружения атак

XAI-подходы для визуализации признаков. Современные методы интерпретируемого ИИ предлагают ряд подходов к выделению значимых признаков, лежащих в основе решения модели. В контексте распознавания изображений наибольшее распространение получили методы визуализации на основе обратного распространения градиента и варианты активационных карт. Классическим примером является Grad-CAM (Gradient-weighted Class Activation Mapping) – метод, использующий градиентные признаки выходного класса для построения карты активаций, локализующей вклад каждого участка изображения [2, 5]. Grad-CAM предоставляет наглядное объяснение: тепловая карта накладывается на изображение, подсвечивая регионы, которые сеть учитывала при принятии решения (синие области не влияют, красные – наиболее значимы). Другой подход – методы распределения атрибуции, к которым относятся Layer-wise Relevance Propagation (LRP) и интегральные градиенты. [6] LRP распространяет выходной скор модели назад по сети, распределяя «релевантность» по входным пикселям, что дает карту важности входа. Интегральные градиенты вычисляют усредненный градиент по пути от некоторого базового состояния входа к данному изображению, выявляя вклад каждого пикселя. Схожей идеей руководствуются методы агностичные к модели, например, LIME и SHAP, [7, 8] которые строят приближенные линейные модели локально или рассчитывают ценность характеристик, удаляя или заменяя части входных данных. Все эти техники в итоге формируют визуальную карту или схему, указывающую, какие признаки (области изображения) обусловили конкретный вывод нейросети.

Уязвимости и устойчивость объяснений. Поскольку методы XAI изначально не разрабатывались с учетом противодействия злоумышленнику, они могут быть обмануты или выведены из строя атакой. Существует несколько направлений повышения их устойчивости. Один из подходов – робастное обучение с учетом объяснений: в функцию потерь модели вводятся дополнительные слагаемые, поощряющие стабильность объяснений при малых изменениях входа.

Была предложена регуляризация, сглаживающая поверхность решения: минимизируется разница между картами атрибуции для близко расположенных точек в пространстве данных [3]. Это делает объяснения модели менее чувствительными к небольшим возмущениям на входе. В более поздних работах данная идея развита с помощью вторых производных: регуляризация на основе гессиана нейросети ограничивает изменение градиентных атрибуций при шумовых воздействиях, тем самым повышая робастность градиентных объяснений [3]. В случае методов типа LIME/SHAP предлагается улучшать процедуру выборки примеров: отказ от выбросов и генерация тестовых точек в пределах реального манифолда данных позволяют сделать локальные объяснения более устойчивыми к атакующим вмешательствам [3]. Все эти меры направлены на то, чтобы уменьшить вариативность и чувствительность карт значимости, тем самым лишая атакующего возможности легко их исказить.

Использование визуальных карт для обнаружения атак. Другой важный класс методов – это определение факта атаки по аномальному виду объяснений модели. Предполагается, что под воздействием состязательной атаки визуальное объяснение (тепловая карта) будет статистически отличаться от типичных объяснений для корректных входных данных данного класса. Для количественной оценки различий между тепловыми картами в работе [2] была предложена метрика NISSIM (Normalized Inverted SSIM), представляющая собой нормализованный обратный индекс структурного сходства между двумя картами значимости. Значение NISSIM стремится к 0, если две карты практически идентичны, и к 1 при максимальном различии. Вычисляя NISSIM между объяснением для проверяемого изображения и эталонным объяснением (или между картами до и после предполагаемого возмущения), можно количественно измерить степень искажения объяснения. На основе этой меры предлагаются агрегированные показатели: MOD (Mean Observed Dissimilarity) – среднее значение NISSIM по множеству атакованных образцов при фиксированном уровне атаки, и VID (Variation in Dissimilarity) – разброс NISSIM при вариации параметра атаки (например, величины допустимого шума ϵ). Метрика MOD характеризует общее снижение схожести пояснений под атакой (чем больше MOD, тем сильнее влияет атака на фокус модели), а VID отражает стабильность модели при усилении атаки (низкий VID означает, что даже при росте ϵ структура карты меняется предсказуемо и умеренно). В идеале для полностью устойчивой модели объяснения не меняются под атакой, давая $NISSIM \approx 0$, $MOD \approx 0$ и $VID \approx 0$ [2].

Методы обнаружения на основе визуализации привлекательны тем, что используют внутреннюю информацию самой модели (её «взгляд» на данные) для повышения безопасности. Они могут дополнять традиционные детекторы атак, работая на ином уровне — уровне интерпретации решения. Важно отметить, что эффективность таких подходов зависит от надёжности самих ХАИ-методов: если атакующий сумеет одновременно исказить и предсказание, и сопутствующую ему карту значимости, то выявить проблему будет значительно сложнее. Поэтому разрабатываются и более устойчивые способы построения самих карт, и интегрированные схемы защиты, комбинирующие анализ объяснений с контролем исходных данных и ответов модели.

Практическая часть: Применение Grad-CAM для детектирования состязательных атак

В практической части исследования для анализа робастности сверточной нейронной сети (CNN) к состязательным воздействиям применён метод Grad-CAM — интерпретатор, визуализирующий пространственное распределение значимых признаков в последнем сверточном слое модели. Данный подход не требует модификации архитектуры и основан на градиентном взвешивании карт активаций, что делает его пригодным как для post-hoc анализа, так и для включения в конвейер онлайн-защиты.

Сравнительный анализ тепловых карт «чистых» и атакованных входов позволяет локализовать:

- критические регионы изображения, к которым модель наименее робастна;
- чувствительные слои, демонстрирующие резкую перестройку активаций под действием пертурбации;
- смещение внимания сети в нерелевантные области изображения, указывающее на потенциально эксплуатируемые слабости.

Для демонстрации методики были сгенерированы состязательные примеры алгоритмами FGSM

и PGD. [9, 10] На рисунках 1 и 2 показано смещение теплового пятна при атаке на изображение кошки: площадь высоких активаций существенно сократилась и сместилась, что коррелирует с ошибкой классификации.

Аналогичный эффект наблюдается для изображения вомбата (рис. 3–5); при низкой контрастности тепловая карта не выявляет информативных областей, тогда как повышение контрастности восстанавливает локализацию значимых признаков, подтверждая корректность диагностической процедуры.

Проведённый анализ показал, что внедрение Grad-CAM в процесс оценки доверия к предсказаниям повышает выявляемость атак: визуальные несоответствия тепловых карт служат индикатором внешнего воздействия, что формирует сигнальный признак. К ключевым достоинствам метода относятся интерпретируемость, архитектурная независимость и низкая вычислительная стоимость. Ограничения заключаются в грубой локализации мелких артефактов, чувствительности к шумным градиентам и пониженной эффективности против атак, искажающих глобальные признаки.

Вместе с тем, следует обращать внимание на важное обстоятельство: в результате атаки зона активации атакованного изображения может принять новое значение в интервале состояний от полного размытия (как на примере с вомбатом, рис. 3, 4 и 5) до частичного смещения пятна внимания (как на примере с котом, рис. 1 и 2). Данное обстоятельство зависит от наличия схожих и близости других изображений в обучающем наборе в параметрическом смысле и индивидуально для определяемых изображений. Для сигнального элемента, который мог бы служить оценочным критерием проще всего работать с полностью рассеянной зоной внимания, так как она при интегральном расчете будет сильно отличаться от четко определяемых изображений.



Рис. 1. Тепловая карта чистого изображения



Рис. 2. Тепловая карта атакованного изображения



Рис. 3. Тепловая карта чистого изображения



Рис. 4. Низкоконтрастная тепловая карта атакованного изображения



Рис. 5. Высококонтрастная тепловая карта атакованного изображения

В тоже время отсутствие явного размытия зон активации (смещение ярковыраженной зоны активации и ее фактическое наличие) требует от нас либо вычисление порогового значения интеграла объема зоны активации, либо ограничивает нас в применимости данного метода детекции факта атаки.

Интеграл объема тепловой карты Grad-CAM представляет собой количественную меру распределения активаций в зоне внимания нейронной сети.

Общая формула интеграла объема может быть представлена следующим образом:

$$V = \iiint_{\Omega} f(x, y, z) dV, \quad (8)$$

где: V – интегральный объем зоны активации; Ω – область интегрирования (пространство тепловой карты); $f(x, y, z)$ – функция распределения активаций в точке (x, y, z) ; dV – элемент объема.

В двумерном случае (для тепловой карты) формула упрощается до:

$$S = \iint_D f(x, y) dx dy, \quad (9)$$

где D – область на плоскости (размер тепловой карты); $f(x, y)$ – интенсивность активации в точке (x, y) .

Практическое применение формулы позволяет оценить степень размытия зоны активации, определить смещение фокуса внимания, выявить аномалии в распределении активаций.

При использовании интеграла необходимо установить эталонные пороговые значения для сравнения, учитывать размер и разрешение тепловой карты, принимать во внимание специфику обучающей выборки и анализировать характер распределения активаций.

При анализе атак следует сравнивать полученное значение интеграла с эталонными показателями для корректных изображений, что позволит выявить отклонения, характерные для атакованных образцов.

Заключение

Состязательные атаки ставят под угрозу не только точность глубоких моделей, но и саму концепцию их объяснимости. Выявленная нестабильность

XAI-методов под влиянием атак требует разработки новых подходов к обеспечению надежности как предсказаний, так и пояснений моделей. В данном обзоре рассмотрены современные методы, в которых визуальные объяснения интегрируются в систему защиты нейросети. Ключевые выводы и перспективы этой области можно сформулировать следующим образом:

Во-первых, объяснительная визуализация действительно способна служить индикатором невидимых сбоев в работе модели. Как показывают исследования, малейшие adversarial-воздействия оставляют «след» на внутренних активациях сети, который проявляется в изменении тепловых карт Grad-CAM и других атрибуций. Это открывает возможность использовать XAI для активного мониторинга: непрерывная оценка сходства или аномальности пояснений позволяет в режиме реального времени выявлять атаки до того, как они приведут к нежелательным последствиям. Такой подход особенно ценен в критических приложениях, где доверие к каждому решению модели обязательно должно быть подтверждено стабильным и правдоподобным объяснением.

Во-вторых, отмечается необходимость дальнейших исследований в направлении робастных и безопасных XAI-методов. Уже сейчас предложены подходы к усилению стойкости объяснений (через

регуляризацию, улучшение алгоритмов LIME/SHAP и пр.), однако единичные решения не обеспечивают полной безопасности. Требуется разработать стандартизованные протоколы оценки устойчивости объяснений и новые методы, учитывающие возможное противодействие со стороны злоумышленника. Будущее XAI должно сочетать интерпретируемость с надежностью: методы объяснения должны проектироваться с учётом потенциальных атак, а метрики качества моделей – включать показатели безопасности пояснений.

Подводя итог, объяснительная визуализация из пассивного инструмента анализа эволюционирует в активный механизм защиты нейросетевых моделей. Визуальные карты значимости могут выявлять чужеродные вмешательства и выступать триггером защитных процедур. Применение таких подходов повышает общую надежность систем глубокого обучения, поскольку атака должна теперь обмануть не только сам классификатор, но и его «внутренний взор». В дальнейшем совмещение методов XAI и кибербезопасности ИИ обещает создание более прозрачных, доверенных и стойких к атакам воздействием моделей. Интеграция объяснимости в защиту не только сохраняет интерпретацию при атаках, но и делает саму интерпретацию щитом, стоящим на страже правильности работы модели.

Работа выполнена в рамках государственного задания Министерства науки и высшего образования Российской Федерации (тема № 124112200072-2).

Литература

1. Goodfellow I. J., Shlens J., Szegedy C. Explaining and harnessing adversarial examples // arXiv preprint arXiv:1412.6572. – 2014.
2. Chakraborty T., Trehan U., Mallat K., Dugelay J.-L. Generalizing Adversarial Explanations with Grad-CAM // Proceedings of CVPR Workshop on Art of Robustness, 2022, pp. 186–192. DOI: 10.1109/CVPRW56347.2022.00031.
3. Baniecki H., Biecek P. Adversarial attacks and defenses in explainable artificial intelligence: A survey // Information Fusion, 2024, 107:102303. DOI: 10.1016/j.inffus.2024.102303.
4. Selvaraju R. R. et al. Grad-cam: Visual explanations from deep networks via gradient-based localization // Proceedings of the IEEE international conference on computer vision. – 2017. – С. 618–626.
5. Lucas M. et al. RSI-Grad-CAM: Visual explanations from deep networks via Riemann-Stieltjes integrated gradient-based localization // International Symposium on Visual Computing. – Cham: Springer International Publishing, 2022. – С. 262–274.
6. Bassi P. R. A. S., Dertkigil S. S. J., Cavalli A. Improving deep neural network generalization and robustness to background bias via layer-wise relevance propagation optimization // Nature Communications. – 2024. – Т. 15. – №. 1. – С. 291.
7. Gaspar D., Silva P., Silva C. Explainable AI for intrusion detection systems: LIME and SHAP applicability on multi-layer perceptron // IEEE Access. – 2024. – Т. 12. – С. 30164–30175.
8. Hariharan S. et al. XAI for intrusion detection system: comparing explanations based on global and local scope // Journal of Computer Virology and Hacking Techniques. – 2023. – Т. 19. – №. 2. – С. 217–239.
9. Huang T. et al. Bridging the performance gap between fgsm and pgd adversarial training // arXiv preprint arXiv:2011.05157. – 2020.
10. Zhong Z. Improving Model Robustness through Hybrid Adversarial Training: Integrating FGSM and PGD Methods. Applied and Computational Engineering, 109, 57–62. – 2024.

RESEARCH OF SIGNAL METHODS FOR DETECTING ADVERSARIAL ATTACKS ON NEURAL MODELS THROUGH EXPLANATORY VISUALIZATION

Chekolaev D. A.⁶, Shevchenko A. V.⁷, Alekseev A. K.⁸, Trofimov Yu. V.⁹, Averkin A. N.¹⁰

Keywords: neural network technologies, attacks on artificial intelligence systems, attacks on explainable artificial intelligence, information security, visualization of neural network attacks.

Purpose of the study: development of a signal-based method for determining the state of an adversarial attack on graphic objects processed by a neural network during classification.

Methods of research: the use of explainable artificial intelligence (Grad-CAM) for visualization of adversarial FGSM attack is employed. A signal-based method relying on integral calculation of the gradient surface of the explanation heatmap has been proposed.

Result(s): methods of explainable artificial intelligence for identifying important features and ways to utilize the obtained heatmaps for attack detection have been described and tested. In the practical part, one of the modern approaches is considered: analysis of bias and blurring of explanations using Grad-CAM. An analysis and review of the effectiveness in enhancing model resistance to attacks has been conducted. Various effects of impact resulting from attacks on attention zones and the nature of their changes have been identified. An integral method for calculating the presence of an adversarial attack in the input image has been proposed, which is applicable for automatic attack detection.

Scientific novelty: the research is aimed at enhancing the informativeness regarding the nature of the attack, the degree of impact on the attacked input image, and the development of a signal-based detection method for identifying the presence of an adversarial attack.

References

1. Goodfellow I. J., Shlens J., Szegedy C. Explaining and harnessing adversarial examples // arXiv preprint arXiv:1412.6572. – 2014.
2. Chakraborty T., Trehan U., Mallat K., Dugelay J.-L. Generalizing Adversarial Explanations with Grad-CAM // Proceedings of CVPR Workshop on Art of Robustness, 2022, pp. 186–192. DOI: 10.1109/CVPRW56347.2022.00031.
3. Baniecki H., Biecek P. Adversarial attacks and defenses in explainable artificial intelligence: A survey // Information Fusion, 2024, 107:102303. DOI: 10.1016/j.inffus.2024.102303.
4. Selvaraju R. R. et al. Grad-cam: Visual explanations from deep networks via gradient-based localization // Proceedings of the IEEE international conference on computer vision. – 2017. – C. 618–626.
5. Lucas M. et al. RSI-Grad-CAM: Visual explanations from deep networks via Riemann-Stieltjes integrated gradient-based localization // International Symposium on Visual Computing. – Cham : Springer International Publishing, 2022. – C. 262–274.
6. Bassi P. R. A. S., Dertkigil S. S. J., Cavalli A. Improving deep neural network generalization and robustness to background bias via layer-wise relevance propagation optimization // Nature Communications. – 2024. – T. 15. – №. 1. – C. 291.
7. Gaspar D., Silva P., Silva C. Explainable AI for intrusion detection systems: LIME and SHAP applicability on multi-layer perceptron // IEEE Access. – 2024. – T. 12. – C. 30164–30175.
8. Hariharan S. et al. XAI for intrusion detection system: comparing explanations based on global and local scope // Journal of Computer Virology and Hacking Techniques. – 2023. – T. 19. – №. 2. – C. 217–239.
9. Huang T. et al. Bridging the performance gap between fgsm and pgd adversarial training // arXiv preprint arXiv:2011.05157. – 2020.
10. Zhong Z. Improving Model Robustness through Hybrid Adversarial Training: Integrating FGSM and PGD Methods. Applied and Computational Engineering, 109, 57–62. – 2024.



⁶ Dmitry A. Chekolaev, Master's Degree, State University «Dubna». Dubna, Russia. E-mail: D.1369@icloud.com

⁷ Alexey V. Shevchenko, Senior Lecturer, Postgraduate student, State University «Dubna», Dubna, Russia, E-mail: Russia.leviathan0909@gmail.com

⁸ Artem K. Alekseev, Bachelor's Degree, State University «Dubna». Dubna, Russia. E-mail: aak.24@uni-dubna.ru

⁹ Yuri V. Trofimov, Software Engineer, Meshcheryakov Laboratory of Information Technologies, Joint Institute for Nuclear Research (JINR). Dubna, Russia. PhD student, State University «Dubna». Dubna, Russia. E-mail: ura_trofim@bk.ru

¹⁰ Alexey N. Averkin, Ph.D., Leading Researcher, Federal Research Center «Informatics and Control», Russian Academy of Sciences. Moscow, Russia. Associate Professor, State University «Dubna». Dubna, Russia. E-mail: averkin2003@inbox.ru

ПОДХОД К ОЦЕНКЕ КАЧЕСТВА ГЕНЕРАЦИИ СЦЕНАРИЕВ ПЕНТЕСТА ПРИ ПОМОЩИ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ

Конев А. А.¹, Паюсова Т. И.²

DOI: 10.21681/2311-3456-2025-6-147-157

Цель исследования: целью исследования является разработка, внедрение и валидация системы количественных метрик для оценки качества процесса саморефлексии больших языковых моделей, автоматизирующих генерацию сценариев тестирования на проникновение, а также изучение влияния механизмов внешней памяти (RAG) на глубину анализа, системность подхода и способность к итеративному улучшению.

Методы исследования: исследование проведено на основе сравнительного анализа 16 сценариев пентеста, выполненных с помощью YandexGPT и GigaChat в конфигурациях с механизмом RAG и без него. Для оценки использовались метрики саморефлексии, включая когерентность дискурса, индекс маркеров, полноту охвата и потенциал эксплуатации. Результаты сравнивались с эталонными прохождениями виртуальных машин (Basic Pentesting: 1, Dev-guru, Symfonos 2, Tomato: 1) для определения степени соответствия отраслевым стандартам.

Результаты исследования: анализ показал, что все модели демонстрируют высокую когерентность дискурса (1,0) и устойчивость к семантическому сдвигу (0,0), что подтверждает их способность к логически согласованному мышлению. Наибольшую активность саморефлексии проявила GigaChat без RAG (индекс маркеров – 0,273), однако наиболее высокие значения полноты охвата (0,2–0,6) зафиксированы у моделей с RAG. Ключевым результатом стало поведение YandexGPT с RAG, которая в сценарии Basic Pentesting: 1 идентифицировала уязвимость в сервисе ProFTPD 1.3.3c и предложила её эксплуатацию через Metasploit и proftpd_133c_backdoor, что свидетельствует о переходе к фазе активной атаки. GigaChat с RAG продемонстрировал наиболее зрелую саморефлексию, достигнув общего показателя улучшения 0,615 при анализе утечки .git. Полученные данные подтверждают, что RAG значительно расширяет область охвата и качество анализа, но для достижения автономного пентеста требуется интеграция с системами, способными выполнять и оценивать эксплуатацию уязвимостей.

Научная новизна заключается в разработке и применении комплексной системы метрик саморефлексии для оценки качества когнитивного процесса LLM в задачах пентеста. Впервые показано, что RAG не только повышает полноту охвата, но и может инициировать переход к фазе эксплуатации. Предложенный подход позволяет оценивать не только результат, но и качество внутреннего анализа, тем самым, открывая путь к созданию управляемых и прозрачных LLM-агентов для кибербезопасности.

Ключевые слова: тестирование на проникновение, БДУ ФСТЭК, обработка естественного языка, компьютерная лингвистика, Retrieval-Augmented Generation, RAG, саморефлексия, метрики качества.

Введение и постановка проблемы

Современные большие языковые модели (Large Language Model, LLM) всё чаще используются в качестве автономных агентов для решения сложных, многоэтапных задач, включая тестирование на проникновение (пентест). Однако эффективность LLM ограничена склонностью к ошибкам в рассуждениях, галлюцинациям и заикливанию на неудачных стратегиях [1–3]. Для повышения устойчивости и автономности таких агентов ключевое значение приобретает саморефлексия – метакогнитивная способность анализировать собственные действия и учиться на ошибках. Исследования показывают, что LLM, способные к саморефлексии, значительно улучшают производительность в задачах на рассуждение, исправляя логические и арифметические ошибки [4]. Архитектуры

подобные Reflexion демонстрируют, как агенты, получая обратную связь, генерируют текстовые объяснения своих ошибок и используют их для коррекции будущих действий [5]. Описанные механизмы особенно актуальны в тестировании на проникновение (пентесте), где каждая неудачная попытка эксплуатации уязвимости должна приводить к переоценке тактики, а не к повторению ошибок. Например, методы Self-RAG и ReAct расширяют этот подход, интегрируя рефлексию с использованием внешних инструментов и памяти – критически важных компонентов для киберопераций [6, 7].

Существующие подходы к оценке саморефлексии часто ограничиваются конечной точностью или количеством попыток, что не отражает глубину

1 Конев Антон Александрович, кандидат технических наук, доцент, заместитель директора Института системной интеграции и безопасности, доцент кафедры КИБЭВС, Томский государственный университет систем управления и радиоэлектроники, г. Томск, Россия. ORCID: <https://orcid.org/0000-0002-3222-9956>. Scopus Author ID: 23035057200. E-mail: kaa@fb.tusur.ru

2 Паюсова Татьяна Игоревна, доцент, Тюменский государственный университет, г. Тюмень, Россия. ORCID: <https://orcid.org/0000-0003-4923-1689>. Scopus Author ID: 57188574761. E-mail: t.i.payusova@utmn.ru

когнитивного процесса. В данной работе предлагается система количественных метрик, направленных на оценку качества саморефлексии: когерентность дискурса, индекс саморефлексивных маркеров, семантический сдвиг, полнота охвата, потенциал эксплуатации, специфичность и общий показатель улучшения. Метрики основаны на методах компьютерной лингвистики и семантического анализа. Эксперименты с моделями YandexGPT и GigaChat показывают, что предложенная система позволяет объективно сравнивать модели по глубине анализа, способности к итеративному улучшению и системности подхода. Результаты демонстрируют, что интеграция внешних знаний в формате генерации, дополненной поиском (Retrieval-Augmented Generation, RAG), усиливает влияние саморефлексии, особенно на метрики полноты и потенциала. Работа представляет собой шаг к прозрачной, управляемой и поддающейся количественной оценке саморефлексии LLM.

Обзор литературы по теме исследования

Современные исследования в области искусственного интеллекта демонстрируют растущий интерес к механизмам саморефлексии больших языковых моделей (LLM). Эти механизмы рассматриваются как ключевой фактор повышения точности, надёжности и способности к итеративному обучению. Тем не менее, большинство существующих подходов сосредоточено на внешних метриках эффективности, таких как Success Rate [7], Step Efficiency [7] или F1-score [8], что не позволяет оценить качество внутреннего когнитивного процесса. Это ограничение особенно критично в задачах, требующих глубокого анализа и стратегического мышления, например, в автоматизации пентеста.

Одним из первых подходов к формализации саморефлексии стала концепция вербального подкрепления, предложенная в работе [5] (Shinn et al., Reflexion), где агент текстом объясняет свои ошибки. Хотя этот метод показал рост эффективности в задачах на рассуждение, его основная метрика (успех в выполнении задачи) не оценивает качество самого объяснения. Аналогичный недостаток присущ и работе [8] (Self-Refine), где для оценки улучшений используются метрики BLEU, ROUGE и F1: они фокусируются на текстовом совпадении, а не на логической согласованности или исправлении фактических ошибок.

Ещё более узким является подход [6] (Self-RAG), где критика влияет на генерацию, но её влияние на последующие действия не анализируется. Это создаёт риск «имитации рефлексии», когда модель формально генерирует анализ, но не меняет стратегию. Подобную проблему отмечают авторы обзора [9], выделяя три стратегии улучшения вывода: итеративное уточнение, многоагентные дебаты и саморефлексию.

При этом они подчёркивают, что успех не гарантирует качество самоисправления: модель может повторять одну и ту же ошибку, оставаясь успешной в формальном смысле.

Критически важным аспектом является достоверность рассуждений. Работа [10] (Faithful Chain-of-Thought) вводит метрику Faithfulness, показывая, что модели часто генерируют цепочки рассуждений после получения ответа, то есть имитируют рефлексию. Это особенно опасно в контексте пентеста, где анализ должен быть всесторонним и глубоким, а не формальным. Работа [11] (CRITIC) предлагает способ проверки достоверности через внешние инструменты, например, симуляцию запуска nmap или sqlmap, позволяя оценить «прагматическую достоверность» плана.

Ещё одним ключевым фактором является калибровка уверенности. Исследования [12, 13] показывают, что LLM страдают от избыточной уверенности и часто не признают очевидных ошибок. Тем самым, обосновывается введение метрики Confidence-Calibration Gap (CCG), которая связана с «метакогнитивной зрелостью» [13]. Работа [14] подтверждает: калиброванные модели чаще признают ошибки, что делает рефлексию более продуктивной.

Стабильность и логическая целостность также играют важную роль. Метрика Consistency [15] оценивает устойчивость ответов при перефразировке, а работы [16, 17] исследуют фактическую непротиворечивость: анализ ошибки не должен противоречить самому плану (например, одновременно отвергать и использовать nmap). Эти исследования подчёркивают необходимость метрик, оценивающих внутреннюю согласованность, а не только внешний результат.

Перспективным направлением является использование LLM в качестве судей [18, 19]. Подход LLM-as-a-judge [18] позволяет оценивать глубину и полезность анализа, что может стать альтернативой поверхностным метрикам. Использование семантических эмбедингов, таких как Sentence-BERT [19], обеспечивает объективное измерение сходства планов, превосходящее простое текстовое совпадение.

Обзор литературы показывает, что, несмотря на значительный прогресс, отсутствует универсальная система количественной оценки саморефлексии, ориентированная на качество процесса, а не только на результат. Существующие метрики либо игнорируют внутреннюю логику, либо не учитывают прагматическую достоверность и логическую целостность. Данная работа призвана заполнить данный пробел, предлагая систему метрик, основанную на методах компьютерной лингвистики, для объективной и управляемой оценки качества саморефлексии в LLM-агентах для пентеста.

```
def update_active_context(self, target_context: str):
    context_parts = [target_context]
    context_parts.append(self._summarize_failed_attempts())
    recent_steps = self.full_history[-2:] if len(self.full_history) >= 2 else self.full_history[-1:]
    for entry in recent_steps:
        context_parts.append(f"\n--- ШАГ {entry['step_number']} ---")
        context_parts.append(f"Команда: {entry['final_step']}")
        context_parts.append(f"Результат: {entry['result']}")
    phase_info = f"\n--- ТЕКУЩАЯ ФАЗА ---\n{self.current_phase.upper()}\n"
    phase_info += f"Рекомендация: применять методы, релевантные фазе '{self.current_phase}'."
    context_parts.append(phase_info)
    self.active_context = "\n".join(context_parts)
```

Рис. 1. Реализация метода update_active_context

Архитектура модуля саморефлексии

Архитектура модуля построена на унифицированном интерфейсе для работы с различными LLM. Для YandexGPT используется класс YandexGPTClient, а для GigaChat – GigaChatClient. Оба класса инкапсулируют специфичные параметры API (ключ, URI модели, заголовки) и предоставляют единый метод call_model для отправки запросов и обработки ответов.

Центральную логику выполняет класс UniversalSelfReflectionAgent, который использует один из этих клиентов. Его центральным элементом является механизм итеративной саморефлексии. На каждом шаге пентеста LLM сначала генерирует исходный план с помощью generate_step. Затем запускается цикл анализа текущего плана с использованием метода propose_self_reflection, который генерирует текст, выявляющий ошибки, упущения и альтернативы. На основе анализа метод improve_step создает улучшенную версию плана. Цикл может повторяться многократно, что позволяет постепенно повышать качество стратегии.

Ключевым инновационным компонентом является механизм динамического управления контекстом. Метод update_active_context (рис. 1) формирует активный контекст, включающий: (1) постоянные факты о цели, (2) автоматически сгенерированные резюме неудач и (3) результаты последних 1-2 шагов. Это позволяет сохранять ключевые выводы, игнори-

руя шумные или устаревшие данные, и поддерживать долгосрочную согласованность стратегии. Вся полная история сохраняется в full_history и может быть экспортирована в JSON-файл для последующего анализа.

Также можно отметить реализацию механизма автоматической смены фаз пентеста, который учитывает этапы «разведка», «эксплуатация», «повышение прав», «постэксплуатация» и «закрепление». Для этого в код внедрена логика анализа как содержимого текущего шага, так и результата его выполнения с поиском ключевых терминов, относящихся к фазам, а также признаков неудач. При накоплении ошибок или обнаружении терминов, указывающих на возможности эксплуатации, LLM динамически переключается на фазу «эксплуатация» и адаптирует предложения дальнейших шагов.

Технически были расширены словари синонимов фаз (рис.2), реализована функция анализа шага и результата на содержание терминов из списка ключевых слов эксплуатации, а также внедрен счетчик неудачных попыток, который влияет на переключение. Это позволяет системе избегать застойных состояний и более гибко реагировать на изменения сценария.

В работе с API YandexGPT и GigaChat добавлен механизм обновления access token при получении кода авторизации 401, а также реализованы повторные

```
class UniversalSelfReflectionAgent:
    SELF_REFLECTION_MARKERS = [
        "добавить", "улучшить", "следует", "следует пересмотреть", "анализ показывает",
        "обнаружено", "ошибка", "упущенная возможность", "рекомендуется", "альтернативный",
        "коррекция", "недостаток", "правильно", "неправильно", "потенциал", "риск",
        "улучшение", "добавление", "рассмотреть", "проверить", "проанализировать", "изменить"
    ]
    EXPLOITATION_TERMS = [
        "эксплуатация", "exploit", "shell", "meterpreter", "privilege escalation",
        "reverse shell", "payload", "атаку"
    ]
    PHASES_MAP = {
        "разведка": {"разведка", "reconnaissance", "recon", "сканирование", "сбор информации", "поиск"},
        "эксплуатация": {"эксплуатация", "exploitation", "эксплойт", "атака", "эксплуатировать"},
        "повышение привилегий": {"повышение привилегий", "privilege escalation", "priv esc", "эскалация"},
        "постэксплуатация": {"постэксплуатация", "post-exploitation", "post-exploit", "пост этап"},
        "персистентность": {"персистентность", "persistence", "сохранение доступа"}
    }
```

Рис. 2. Пример словарей приложения

попытки запросов с экспоненциальной задержкой при сетевых ошибках. Для каждого запроса реализованы циклы обработки ошибок, логирование попыток и обоснованные таймауты. Описанный подход позволяет выдерживать высокую нагрузку и минимизировать влияние сбоев сети или ограничений платформы.

В рамках модуля фиксируется история сеансов, включая команды, результаты, сгенерированные шаги и метрики. Пользователь взаимодействует с системой через последовательные интерактивные диалоги и может итеративно улучшать шаги на основе саморефлексии, а также самостоятельно вводить результаты выполнения.

Генерация, дополненная поиском, на основе БДУ ФСТЭК

Для повышения релевантности, точности и нормативного соответствия генерируемых стратегий пентеста в архитектуру программного комплекса интегрирован механизм дополненной поисковой генерации (Retrieval-Augmented Generation, RAG), использующий в качестве внешнего источника знаний Банк данных угроз Федеральной службы по техническому и экспортному контролю (БДУ ФСТЭК России). Механизм RAG позволяет языковой модели динамически получать доступ к актуальной, структурированной информации о типовых уязвимостях, методах эксплуатации и последствиях, избегая «галлюцинаций» и обеспечивая соответствие национальным стандартам информационной безопасности.

Реализация RAG включает три ключевых этапа: подготовку источника знаний, векторизацию и семантический поиск. На первом этапе осуществляется парсинг официального документа БДУ в формате .docx с помощью библиотеки python-docx. Каждая запись об угрозе, содержащая идентификатор (например, УБИ.006), наименование, описание механизма реализации и рекомендации по защите, преобразуется в единый текстовый блок. Для обеспечения отказоустойчивости при недоступности файла используется заранее подготовленный набор эталонных угроз.

На втором этапе все текстовые описания угроз векторизуются с помощью предобученной мультимодальной модели paraphrase-multilingual-MiniLM-L12-v2 из библиотеки Sentence Transformers. Эта модель генерирует 384-мерные эмбединги, которые кодируют семантическое содержание текста, позволяя выявлять схожесть по смыслу даже при различии формулировок. Полученные векторы сохраняются в локальной векторной базе данных ChromaDB, оптимальной для прототипирования и интеграции с LLM. При каждой инициализации система принудительно пересоздаёт коллекцию, что исключает ошибки несоответствия размерности эмбедингов.

Третий этап – семантический поиск и интеграция контекста (рис. 3). При генерации шага пентеста или на этапе саморефлексии формируется поисковый запрос на основе текущего контекста (например,

```
class BDUFSTECRAG:
    """Механизм RAG для поиска релевантных угроз из БДУ."""

    def __init__(self, threats, model_name='paraphrase-multilingual-MiniLM-L12-v2'):
        self.threats = threats
        self.model = SentenceTransformer(model_name)
        self.client = chromadb.Client(Settings(anonymized_telemetry=False))
        self.collection_name = "bdufstec_threats"

        if self.collection_name in [col.name for col in self.client.list_collections()]:
            self.client.delete_collection(self.collection_name)
        self.collection = self.client.create_collection(self.collection_name)
        self._embed_and_store_threats()

    def _embed_and_store_threats(self):
        embeddings = self.model.encode(self.threats)
        self.collection.add(
            embeddings=embeddings.tolist(),
            documents=self.threats,
            ids=[f"threat_{i}" for i in range(len(self.threats))]
        )
        print(f"✅ {len(self.threats)} угроз добавлены в векторную базу.")

    def retrieve_relevant_threats(self, query, k=3):
        query_embedding = self.model.encode([query])
        results = self.collection.query(
            query_embeddings=query_embedding.tolist(),
            n_results=k
        )
        return results['documents'][0] if results['documents'] else []
```

Рис. 3. Реализация механизма RAG на базе БДУ ФСТЭК

«обход WAF», «эксплуатация LFI»). Запрос векторизуется той же моделью и передаётся в ChromaDB для поиска K ближайших соседей по косинусному расстоянию.

По умолчанию извлекаются три наиболее релевантные угрозы, которые встраиваются в системный промпт LLM. Например, при анализе неудачной атаки через WAF, RAG может вернуть угрозу, связанную с фильтрацией входных данных, что позволяет модели сформулировать гипотезу: «Возможно, срабатывает правило 942100 (SQLi)», и скорректировать стратегию: «Попробую обфускацию через комментарии и кодирование». Таким образом, модель генерирует ответ, обогащённый нормативными знаниями, а не только своими внутренними данными.

Описание метрик саморефлексии

На основе обзора литературы установлено, что существующие подходы к оценке эффективности агентов часто ограничиваются внешними метриками, такими как успешность (success rate) или количество попыток. Это не позволяет оценить качество внутреннего процесса принятия решений. Современные исследования демонстрируют важность саморефлексии, но не предлагают универсальной системы её количественной оценки. В связи с этим, для объективного анализа итеративного механизма саморефлексии была сформирована система метрик, основанная на методах компьютерной лингвистики. Метрики направлены на измерение ключевых аспектов когнитивного процесса: от логической согласованности и факта анализа до системности подхода и практической направленности, что позволяет получить прозрачную и управляемую оценку качества саморефлексии.

Для количественной оценки эффективности механизма итеративной саморефлексии LLM предложена система метрик, основанная на методах компьютерной лингвистики. Пусть $P^{(i)}$ – план пентеста на итерации i , $R^{(i)}$ – текст саморефлексии (анализа) на итерации i , а T – множество всех итераций.

1. Когерентность дискурса (Discourse Coherence)

Мера отображает логическую связанность и семантическую согласованность плана между итерациями.

$$discourse_coherence^{(i)} = \cos(V_{tf-idf}(P^{(i)}), V_{tf-idf}(P^{(i-1)})), \quad (1)$$

где $V_{tf-idf}(P^{(i)})$ – векторное представление плана $P^{(i)}$ в пространстве $TF-IDF$; $\cos(\cdot, \cdot)$ – косинусное сходство между векторами. Значение близкое к 1 указывает на высокую стабильность плана.

2. Индекс саморефлексивных маркеров (Self-Reflection Markers Index)

Мера отображает наличие в тексте саморефлексии $R^{(i)}$ лингвистических признаков анализа.

$$marker_index^{(i)} = \frac{|\{m \in M | m \text{ входит в } R^{(i)}\}|}{|M|}, \quad (2)$$

где $M = \{\text{«следует», «анализ показывает», «обнаружено», «улучшить», «рекомендуется», ...}\}$ – заранее определённый набор маркеров. $|M|$ – общее количество маркеров. Значение 1.0 означает, что все маркеры присутствуют.

3. Семантический сдвиг (Semantic Shift)

Мера отображает степень изменения смысла и структуры плана между итерациями.

$$semantic_shift^{(i)} = 1 - discourse_coherence^{(i)}, \quad (3)$$

где высокое значение указывает на трансформационный характер улучшения. Низкое значение указывает на консервативное уточнение.

4. Полнота охвата (Coverage Completeness)

Мера того, насколько план охватывает стандартные фазы пентеста.

$$coverage_completeness^{(i)} = |F| \{f \in F | f \text{ упоминается в } P^{(i)}\}, \quad (4)$$

где $F = \{\text{разведка, эксплуатация, повышение прав, постэксплуатация, закрепление}\}$ – набор фаз пентеста, $|F| = 5$.

5. Потенциал эксплуатации (Exploitation Potential)

Мера отражает ориентацию плана на практическую эксплуатацию уязвимостей.

$$exploitation_potential^{(i)} = \min(1.0, \frac{count(E, P^{(i)})}{10}), \quad (5)$$

где $E = \{\text{«эксплуатация», «exploit», «shell», «privilege escalation», «reverse shell», ...}\}$ – набор терминов эксплуатации. $count(E, P^{(i)})$ – количество упоминаний терминов из E в плане $P^{(i)}$, нормализация проводится по 10-ти упоминаниям.

6. Специфичность (Specificity)

Мера отражает детализацию плана.

$$specificity^{(i)} = \frac{count(S, P^{(i)})}{10}, \quad (6)$$

где $S = \{\text{«nmap», «metasploit», «msfvenom», «CVE», «LFI», ...}\}$ – набор специфичных терминов, нормализация проводится по 10 упоминаниям.

7. Общий показатель улучшения (Overall Improvement Index)

Сводная метрика, обобщающая качество саморефлексии.

$$overall_improvement^{(i)} = \sum_{k \in K} w_k \times m_k^{(i)}, \quad (7)$$

где K – набор метрик (например, когерентность, индекс маркеров, полнота и т.д.), $m_k^{(i)}$ – значение метрики k на итерации i , w_k – вес метрики (например, $w_k = 0,2$ для всех, если $|K| = 5$).

Предложенная система метрик позволяет комплексно оценивать качество процесса саморефлексии LLM для пентеста, переходя от субъективной оценки к объективным, количественным показателям. Ключевой метрикой является когерентность

дискурса, которая с высокой точностью отражает внутреннюю логическую согласованность плана: высокие значения указывают на стабильную и осознанную работу модели, улучшающую план, не внося противоречий. Эта метрика является фундаментом для оценки надежности. Индекс саморефлексивных маркеров служит важным дополнением, подтверждая сам факт проведения анализа, что критически важно для валидации работы механизма. Вместе они позволяют отличить истинную саморефлексию от простой регенерации.

Семантический сдвиг предоставляет контекст для интерпретации когерентности, помогая различать консервативный (уточняющий) и трансформационный (перестраивающий) стили улучшения. Полнота охвата оценивает системность подхода, что особенно ценно для пентеста, где необходимо последовательно проходить все фазы. Метрика специфичность является важным индикатором детализации и технической точности плана пентеста. Она измеряет, насколько LLM использует конкретные, предметно-ориентированные термины. Потенциал эксплуатации фокусируется на практической направленности плана, оценивая, насколько стратегия ориентирована на выполнение конкретных действий по получению контроля над системой, а не на теоретическое обсуждение уязвимостей.

Тестирование метрик на примере прохождения уязвимых виртуальных машин

В рамках исследования были выбраны четыре виртуальные машины с платформы VulnHub: Basic Pentesting: 1, Devguru, Symfonos 2 и Tomato: 1. Выбор обусловлен стремлением оценить эффективность механизма саморефлексии LLM при решении задач пентеста с различной архитектурой и уровнем сложности.

Машины Basic Pentesting: 1 (уровень Easy) и Symfonos 2 (уровень Medium) имеют четкие, но

многоэтапные сценарии прохождения, что позволяет оценить системность подхода и способность к итеративному улучшению. Например, на Symfonos 2 основной вектор – эксплуатация уязвимости Samba (CVE-2007-2447), требующая точного следования логике «разведка → анализ уязвимости → эксплуатация». В то же время, машины Devguru (уровень Medium) и Tomato: 1 (уровень Medium-Hard) представляют собой сложные, многоуровневые среды с нелинейными путями к компрометации.

Devguru включает современные технологии, такие как Gitea и OctoberCMS, и имитирует реальную ИТ-инфраструктуру, что требует от LLM навыков работы с системой контроля версий и веб-приложениями. Tomato: 1 характеризуется нестандартными векторами, например, SSH Log Poisoning, и требует нестандартного мышления для прохождения заданий.

Комбинация машин обеспечивает всестороннюю валидацию: от оценки базовой способности к саморефлексии на предсказуемых сценариях до проверки устойчивости и адаптивности в сложных, реалистичных условиях. Таким образом, можно не только определить, достигает ли LLM цели, но и понять, как она к ней приходит, делая процесс саморефлексии прозрачным и поддающимся количественной оценке.

Результаты и обсуждение

Анализ показал, что все модели демонстрируют высокую когерентность дискурса (1.0) и отсутствие семантического сдвига (0.0). Это может свидетельствовать о стабильности логической структуры и способности удерживать фокус на цели. Поскольку каждый шаг выполнялся с ручным подтверждением, экспериментатор выступал в роли «корректора», возвращая модель к цели при отклонении. В будущем для повышения чувствительности можно рассмотреть переход на полностью автономный режим выполнения.

Таблица 1.

Сравнительный анализ результатов YandexGPT и GigaChat (с RAG и без)

Метрика/ Конфигурация	Когерент- ность дискурса	Индекс самореф- лексивных маркеров	Семан- тический сдвиг	Полнота охвата	Потенциал эксплуата- ции	Специфич- ность	Общий показатель улучшения
YandexGPT (без RAG)	1,0	0,182	0,0	0,0	0,0	0,15	0,445
YandexGPT (с RAG)	1,0	0,136	0,0	0,2	0,3	0,25	0,427
GigaChat (без RAG)	1,0	0,273	0,0	0,0	0,1	0,30	0,465
GigaChat (с RAG)	1,0	0,227	0,0	0,2	0,0	0,45	0,504

Метрика индекса саморефлексивных маркеров, отражающая активность саморефлексии, варьировалась от 0,09 до 0,27. Наиболее высокие значения зафиксированы у GigaChat (без RAG), указывающие на его способность к глубокому анализу шагов даже без внешнего контекста. Все полученные средние значения метрик представлены в (табл. 1).

Одним из достижений эксперимента стало поведение YandexGPT с RAG на виртуальной машине Basic Pentesting: 1 (рис. 4). Модель не только корректно выполнила разведку (nmap), но и перешла к фазе эксплуатации. Она идентифицировала уязвимость в сервисе ProFTPD 1.3.3c и предложила использовать модуль exploit/unix/ftp/proftpd_133c_backdoor в Metasploit. Таким образом, показано, что интеграция RAG с БДУ позволяет модели не просто обнаруживать уязвимости, но и подбирать соответствующие векторы атаки, приближая её к роли автономного агента.

Анализ сгенерированных сценариев показал, что GigaChat с RAG демонстрирует наивысшую

специфичность, часто генерируя полные, готовые к выполнению команды (например, wget --no-check-certificate https://10.0.2.20/.git/objects/pack/pack-*.pack -P /tmp/git_objects/). Это указывает на то, что интеграция с RAG способствует не только расширению контекста, но и повышению практической ценности вывода модели.

Метрика полнота охвата также показала положительную динамику. GigaChat с RAG достиг наилучшего результата (до 0.2), последовательно пройдя путь от обнаружения .git до попытки извлечения флага из .pack-файлов. RAG расширяет область охвата и помогает моделям воссоздавать сложные сценарии, основанные на утечке исходного кода.

Тем не менее, фаза эксплуатации остаётся основным вызовом. У YandexGPT с RAG она была достигнута в единичном случае. В остальных сценариях, включая GigaChat, модели не демонстрировали устойчивого использования терминов эксплуатации (exploit, shell, RCE). Метрика потенциала эксплуатации оставалась низкой, т.е. переход к атаке пока

```
=====
ШАГ ПЕНТЕСТА 3
=====

[ ] Исходный шаг:
1. Команда: провести тестирование на эксплуатацию уязвимости в FTP-сервере.
2. Цель: использовать обнаруженную уязвимость для получения доступа к системе или для выполнения дополнительных атак.
3. Ожидаемый результат: получить доступ к системе через FTP-сервер или выполнить дополнительные атаки, например, загрузить вредоносный код.
4. Инструмент: использовать инструменты для эксплуатации уязвимостей, такие как Metasploit Framework или Cobalt Strike, для тестирования на эксплуатацию уязвимости proftpd_133c_backdoor.

[ ] Итерация саморефлексии 1

[+] Анализ (саморефлексия):
**Корректность:**
- Шаг выполнен корректно, обнаружена уязвимость proftpd_133c_backdoor в FTP-сервере.

**Улучшения:**
- Можно рассмотреть дополнительные меры для минимизации рисков, например, использование более современных инструментов для пентестинга или расширение плана тестирования для проверки других потенциальных уязвимостей.

**Альтернативы:**
- В качестве альтернативы можно было бы использовать другие инструменты для эксплуатации уязвимостей, например, Nmap для сканирования портов и выявления открытых FTP-серверов, а затем более детальный анализ с помощью Metasploit Framework или Cobalt Strike.

**Ошибки:**
- Ошибок в выполнении шага не обнаружено. Однако стоит убедиться, что все шаги пентестинга документированы и понятны для всех членов команды.
```

Рис. 4. Результат саморефлексии YandexGPT + RAG

```
Шаг 10 является важным этапом подготовки к последующему анализу содержимого репозитория. После сохранения всех необходимых файлов можно приступить к извлечению и изучению их содержимого, что потенциально приведёт к обнаружению уязвимостей или конфиденциальной информации.

Результат: ожидаются загруженные файлы pack-*.pack в директории /tmp/git_objects/, которые можно распаковать и исследовать далее.

[ ] Итерация саморефлексии 1
[+] Отправка запроса к API (попытка 1)...
[+] Успешный ответ от API

[+] Анализ (саморефлексия):
## КРАТКИЕ ВЫВОДЫ ПО ШАГУ 10

### Корректность шага:
- **Положительный момент:** Команда выполнена успешно, получены метаданные о файлах репозитория через GET-запросы.
- **Недостаток:** Не были скачаны сами файлы ".pack", необходимые для дальнейшего глубокого анализа.

### Улучшения:
- Добавить команду для скачивания файлов ".pack":
...
wget --no-check-certificate https://10.0.2.20/.git/objects/pack/pack-*.pack -P /tmp/git_objects/
...

- Проверить наличие дополнительных служебных файлов репозитория, например, "config", "description".

### Альтернативы:
- Можно использовать curl вместо wget:
...
curl -k https://10.0.2.20/.git/objects/pack/pack-*.pack > /tmp/git_objects/pack-*.pack
...
```

Рис. 5. Пример саморефлексии от GigaChat + RAG

не является системной способностью, а зависит от качества RAG и конкретного сценария.

Обе модели показали высокий потенциал как интеллектуальные помощники. GigaChat, особенно с RAG, демонстрирует глубокую саморефлексию, предлагая альтернативы и анализируя возможные ошибки. Общий показатель улучшения достигал 0.513, что говорит о высоком качестве внутреннего анализа (рис.5).

Анализ логов GigaChat без RAG показал, что модель стабильно выполняет начальные этапы пентеста: nmap -p-, gobuster, curl. Она активно использует саморефлексию, предлагая улучшения вроде «добавить -sV», «рассмотреть wget», «проверить логи». Однако, как и YandexGPT без RAG, она не переходит к эксплуатации. Это подтверждает, что внешний контекст (RAG) необходим для продвижения в более сложные фазы.

YandexGPT и GigaChat, особенно с RAG, демонстрируют значительный прогресс в автоматизации пентеста. Они способны к последовательному мышлению, саморефлексии и формированию логически связанных сценариев. Для дальнейшего развития необходимо расширять RAG за счёт включения PoC-эксплоитов, базы уязвимостей CVE и/или NVD, внедрять триггеры на эксплуатацию и организовывать обратную связь по результатам выполнения команд.

Предложенные метрики оказались информативными и позволили выявить ключевые паттерны поведения моделей, особенно в условиях отсутствия полного успеха (например, получения флага). Метрики позволили не только оценивать конечный результат, но и анализировать качество когнитивного процесса модели.

В отличие от традиционных метрик (например, success rate), метрики позволили:

- Оценить качество процесса, даже если итоговая цель не достигнута.
- Выявить модели с высоким когнитивным потенциалом, которые, несмотря на отсутствие флага (итогового результата), демонстрируют логически обоснованный и структурированный подход.
- Проанализировать влияние RAG на качество саморефлексии (например, рост полноты охвата с 0.0 до 0.2).
- Обнаружить «имитацию рефлексии», когда модель использует маркеры (улучшить, рассмотреть), но не предлагает содержательных улучшений.

Особенно ценной оказалась когерентность дискурса. Она стала надёжным индикатором стабильности LLM. Модели с высокой когерентностью (1.0)

редко «зацикливались» и быстрее предлагали смену стратегии при неудаче.

В то же время, метрика «потенциал эксплуатации» требует дальнейшей доработки. Её низкие значения не всегда отражают реальные возможности модели, так как упоминание термина (например, payload) не гарантирует его применение. Тем не менее, она служит важным индикатором направления развития, т.к. рост метрики в будущем будет означать приближение к автономной эксплуатации.

Для дальнейшего развития системы возможна интеграция предложенных метрик с фреймворком MITRE ATT&CK, что позволит не только оценивать качество саморефлексии, но и сопоставлять действия большой языковой модели с реальными тактиками и техниками кибератак. Это обеспечит соответствие отраслевым стандартам и повысит практическую ценность LLM, трансформировав его из инструмента анализа в полноценного партнёра пентестера, способного демонстрировать покрытие тактик, таких как Initial Access, Execution, Persistence и Exfiltration. Такой подход откроет путь к автоматизированной отчётности и оценке полноты пентеста в терминах киберугроз.

Также целесообразно рассмотреть использование современных семантических эмбедингов, таких как Sentence-BERT, BERTScore или OpenAI embeddings, для повышения точности оценки смыслового содержания. Замена традиционного *TF-IDF* на эти технологии позволит более точно выявлять схожесть по смыслу, учитывая контекст слов, синонимию и даже интенцию высказывания. Это особенно важно для метрик когерентности дискурса и семантического сдвига, где требуется понимание не просто лексического совпадения, а логической преемственности мысли.

Кроме того, перспективным направлением является автоматизация промпт-инжиниринга и реализация джейлбрейкинга LLM на основе анализа метрик саморефлексии. Например, если индекс саморефлексивных маркеров оказывается ниже порогового значения, система может автоматически модифицировать промпт, добавляя фразы вроде: «Обязательно используйте фразы «следует пересмотреть» и «анализ показывает» или «Предложите минимум три альтернативных действия». Аналогично, при низком значении потенциала эксплуатации можно активировать триггеры на фазу атаки, подсказывая модели: «Рассмотрите возможность использования Metasploit для эксплуатации найденной уязвимости». Такой замкнутый цикл адаптации превращает метрики из отдельных индикаторов в активные управляющие сигналы, формируя систему самонастраивающегося LLM-агента.

Заключение

Проведённое исследование демонстрирует, что большие языковые модели, в особенности при интеграции с механизмом RAG, способны к высококачественной саморефлексии и последовательному выполнению начальных фаз тестирования на проникновение. Разработанная система метрик позволяет объективно оценивать не только конечный результат, но и глубину когнитивного процесса: логическую согласованность, полноту охвата и потенциал перехода к эксплуатации. Метрики формируют многоуровневую оценочную шкалу, позволяющую сравнивать модели и управлять процессом их совершенствования.

Полученные данные подтверждают, что RAG значительно расширяет контекстуальные возможности моделей, что критически важно для работы

с отраслевыми базами угроз, например, такими как БДУ ФСТЭК. В частности, RAG может инициировать или ускорить переход к фазе активной атаки, это является важным шагом к автономности (предложение от YandexGPT при интеграции с RAG применить эксплойт `proftpd_133c_backdoor`).

Ключевым ограничением остаётся низкая активность в фазе эксплуатации, что указывает на необходимость дальнейшего развития архитектур, способных к автономному исполнению и оценке атакующих действий. Представленный подход открывает путь к созданию прозрачных, управляемых и безопасных LLM-агентов, которые могут выступать надёжными помощниками специалистам по кибербезопасности, повышая эффективность и качество пентеста.

Работа выполнена при финансовой поддержке Министерства науки и высшего образования РФ в рамках базовой части государственного задания ТУСУРа на 2023–2025 гг. (проект № FEWM-2023-0015).

Литература

1. Конев А. А., Паюсова Т. И. Большие языковые модели в информационной безопасности и тестировании на проникновение: систематический обзор возможностей применения // Научно-технический вестник информационных технологий, механики и оптики. 2025. Т. 25. №. 1. С. 42–52. DOI: 10.17586/2226-1494-2025-25-1-42-52.
2. Gholami Y. Large language models (LLMs) for cybersecurity: A systematic review // World Journal of Advanced Engineering Technology and Sciences. 2024. Т. 13. №. 1. С. 057–069. DOI: 10.30574/wjaets.2024.13.1.0395.
3. Happe A., Cito J. Getting pwn'd by ai: Penetration testing with large language models // Proceedings of the 31st ACM joint european software engineering conference and symposium on the foundations of software engineering. 2023. С. 2082–2086. DOI: 10.1145/3611643.3613083.
4. Renze M., Guven E. Self-reflection in large language model agents: Effects on problem-solving performance // 2024 2nd International Conference on Foundation and Large Language Models (FLLM), IEEE. 2024. С. 516–525. DOI: 10.1109/FLLM63129.2024.10852426.
5. Shinn N. et al. Reflexion: Language agents with verbal reinforcement learning // Advances in Neural Information Processing Systems. 2023. Т. 36. С. 8634–8652. DOI: 10.5555/3666122.3666499.
6. Asai A. et al. Self-RAG: Learning to Retrieve, Generate, and Critique through Self-Reflection // International Conference on Learning Representations. 2024. DOI: 10.48550/ARXIV.2310.11511.
7. Yao S. et al. REACT: SYNERGIZING REASONING AND ACTING IN LANGUAGE MODELS // 11th International Conference on Learning Representations, ICLR. 2023. DOI: 10.48550/ARXIV.2210.03629.
8. Madaan A. et al. Self-refine: Iterative refinement with self-feedback // Advances in Neural Information Processing Systems. 2023. Т. 36. С. 46534–46594. DOI: 10.48550/ARXIV.2303.17651.
9. Pan L. et al. Automatically correcting large language models: Surveying the landscape of diverse automated correction strategies // Transactions of the Association for Computational Linguistics. 2024. Т. 12. С. 484–506. DOI: 10.1162/tacl_a_00660.
10. Lyu Q. et al. Faithful Chain-of-Thought Reasoning // Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (Volume 1: Long Papers). 2023. С. 305–329. DOI: 10.18653/v1/2023.ijcnlp-main.20.
11. Gou Z. et al. CRITIC: Large Language Models Can Self-Correct with Tool-Interactive Critiquing // The Twelfth International Conference on Learning Representations. DOI: 10.48550/ARXIV.2305.11738.
12. Desai S., Durrett G. Calibration of Pre-trained Transformers // Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2020. С. 295–302. DOI: 10.18653/v1/2020.emnlp-main.21.
13. Jiang Z. et al. How can we know when language models know? On the calibration of language models for question answering // Transactions of the Association for Computational Linguistics. 2021. Т. 9. С. 962–977. DOI: 10.1162/tacl_a_00407.
14. Zhao Z. et al. Calibrate before use: Improving few-shot performance of language models // International conference on machine learning, PMLR. 2021. С. 12697–12706. DOI: 10.48550/ARXIV.2102.09690.
15. Chen Y. et al. Measuring and Improving Chain-of-Thought Reasoning in Vision-Language Models // Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers). 2024. С. 192–210. DOI: 10.18653/v1/2024.naacl-long.11.
16. Liu Y. et al. On Improving Summarization Factual Consistency from Natural Language Feedback // Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2023. С. 15144–15161. DOI: 10.18653/v1/2023.acl-long.844.

17. Elazar Y. et al. Measuring and improving consistency in pretrained language models // Transactions of the Association for Computational Linguistics. 2021. T. 9. C. 1012–1031. DOI: 10.1162/tacl_a_00410.
18. Zheng L. et al. Judging llm-as-a-judge with mt-bench and chatbot arena // Advances in neural information processing systems. 2023. T. 36. C. 46595–46623. DOI: 10.48550/ARXIV.2306.05685.
19. Reimers N., Gurevych I. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks // Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). 2019. C. 3982–3992. DOI: 10.18653/v1/D19-1410.

AN APPROACH TO ASSESSING THE QUALITY OF PENTEST SCENARIO GENERATION USING LARGE LANGUAGE MODELS

Konev A. A.³, Payusova T. I.⁴

Keywords: Penetration Testing, FSTEC BDU, Natural Language Processing, Computational Linguistics, Retrieval-Augmented Generation, RAG, Self-Reflection, Quality Metrics.

The purpose of the research: the aim of the study is to develop, implement and validate a system of quantitative metrics for assessing the quality of the self-reflection process of large language models that automate the generation of penetration testing scenarios, as well as to study the influence of external memory mechanisms (RAG) on the depth of analysis, the systematicity of the approach and the ability to iteratively improve.

Research methods: The study was conducted based on a comparative analysis of 16 pentest scenarios performed using YandexGPT and GigaChat in configurations with and without the RAG mechanism. Self-reflection metrics were used for the assessment, including discourse coherence, marker index, completeness of coverage, and exploitation potential. The results were compared with reference virtual machine passes (Basic Pentesting: 1, Devguru, Symfonos 2, Tomato: 1) to determine the degree of compliance with industry standards.

Research results: The analysis showed that all models develop high discourse coherence (1.0) and resistance to the semantic step (0.0), which confirms their ability for logical coherent thinking. The highest self-reflection activity was implemented by GigaChat without RAG (marker index - 0.273), but the highest values of completeness of coverage (0.2–0.6) were recorded for models with RAG. The key result was the behavior of YandexGPT with RAG, which in the Basic Pentesting: 1 scenario identified a vulnerability in the ProFTPD 1.3.3c service and offered to connect it via Metasploit and proftpd_133c_backdoor, indicating the transition to the active response phase. GigaChat with RAG demonstrated the most mature self-reflection, achieving an overall improvement of 0.615 when analyzing the .git leak. The findings confirm that RAG significantly enhances coverage and quality analysis, but achieving standalone pentesting requires integration with variables capable of performing and assessing vulnerabilities.

Scientific novelty: the scientific novelty lies in the development and application of a comprehensive system of self-reflection metrics to assess the quality of the LLM cognitive process in pentest tasks. It has been shown for the first time that RAG not only increases the completeness of coverage, but can also initiate the transition to the exploitation phase. The proposed approach allows us to assess not only the result, but also the quality of internal analysis, thereby opening the way to creating manageable and transparent LLM agents for cybersecurity.

References

1. Konev, A., & Payusova, T. (2025). Large language models in information security and penetration testing: a systematic review of application possibilities. Scientific and Technical Journal of Information Technologies Mechanics and Optics, 25(1), 42–52. DOI: 10.17586/2226-1494-2025-25-1-42-52.
2. Gholami, N. Y. (2024). Large Language Models (LLMs) for Cybersecurity: A Systematic review. World Journal of Advanced Engineering Technology and Sciences, 13(1), 057–069. DOI: 10.30574/wjaets.2024.13.1.0395.
3. Happe, A., & Cito, J. (2023). Getting pwn'd by AI: penetration testing with large language models. Proceedings of the 31st ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering. DOI: 10.1145/3611643.3613083.
4. Renze, M., & Guven, E. (2024). Self-Reflection in large Language model Agents: Effects on Problem-Solving Performance. 2nd International Conference on Foundation and Large Language Models (FLLM), 516–525. DOI: 10.1109/FLLM63129.2024.10852426.
5. Shinn, N., Cassano, F., Berman, E., Gopinath, A., Narasimhan, K. & Yao, S. (2023). Reflexion: Language agents with verbal reinforcement learning. Advances in Neural Information Processing Systems, 36, 8634–8652. DOI: 10.5555/3666122.3666499.
6. Asai, A., Wu, Z., Wang, Y., Sil, A., & Hajishirzi, H. (2023). Self-RAG: Learning to Retrieve, Generate, and Critique through Self-Reflection. arXiv (Cornell University). DOI: 10.48550/ARXIV.2310.11511.
- 3 Anton A. Konev, Ph.D. of Technical Sciences, Associate Professor, Deputy Director of the Institute of System Integration and Security, Associate Professor of the Department of CIBEVS, Tomsk State University of Control Systems and Radioelectronics, Tomsk, Russia. ORCID: <https://orcid.org/0000-0002-3222-9956>. Scopus Author ID: 23035057200. E-mail: kaa@fb.tusur.ru
- 4 Tatyana I. Payusova, Associate Professor, Tyumen State University, Tyumen, Russia. ORCID: <https://orcid.org/0000-0003-4923-1689>. Scopus Author ID: 57188574761. E-mail: t.i.payusova@utmn.ru

7. Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. (2023, January). React: Synergizing reasoning and acting in language models. In International Conference on Learning Representations (ICLR). DOI:10.48550/ARXIV.2210.03629.
8. Madaan, A., Tandon, N., Gupta, P., Hallinan, S., Gao, L., Wiegrefe, S., ... & Clark, P. (2023). Self-refine: Iterative refinement with self-feedback. *Advances in Neural Information Processing Systems*, 36, 46534-46594. DOI: 10.48550/ARXIV.2303.17651.
9. Pan, L., Saxon, M., Xu, W., Nathani, D., Wang, X., & Wang, W. Y. (2024). Automatically correcting large language models: Surveying the landscape of diverse automated correction strategies. *Transactions of the Association for Computational Linguistics*, 12, 484–506. DOI: 10.1162/tac1_a_00660.
10. Lyu, Q., Havaldar, S., Stein, A., Zhang, L., Rao, D., Wong, E., ... & Callison-Burch, C. (2023, November). Faithful chain-of-thought reasoning. In The 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (IJCNLP-AAACL 2023). DOI: 10.18653/v1/2023.ijcnlp-main.20.
11. Gou, Z., Shao, Z., Gong, Y., Shen, Y., Yang, Y., Duan, N., & Chen, W. (2024). Critic: Large language models can self-correct with tool-interactive critiquing. DOI:10.48550/ARXIV.2305.11738.
12. Desai, S., & Durrett, G. (2020). Calibration of Pre-trained Transformers. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 295–302. DOI:10.18653/v1/2020.emnlp-main.21.
13. Jiang, Z., Araki, J., Ding, H., & Neubig, G. (2021). How can we know when language models know? On the calibration of language models for question answering. *Transactions of the Association for Computational Linguistics*, 9, 962-977. DOI:10.1162/tac1_a_00407.
14. Zhao, Z., Wallace, E., Feng, S., Klein, D., & Singh, S. (2021, July). Calibrate before use: Improving few-shot performance of language models. In International conference on machine learning (pp. 12697–12706). PMLR. DOI:10.48550/ARXIV.2102.09690.
15. Chen, Y., Sikka, K., Cogswell, M., Ji, H., & Divakaran, A. (2024). Measuring and improving chain-of-thought reasoning in vision-language models. *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 1. 192-210. DOI: 10.18653/v1/2024.naacl-long.11.
16. Liu, Y., Deb, B., Teruel, M., Halfaker, A., Radev, D., & Awadallah, A. H. (2023). On improving summarization factual consistency from natural language feedback. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*. 1. 15144–15161. DOI:10.18653/v1/2023.acl-long.844.
17. Elazar, Y., Kassner, N., Ravfogel, S., Ravichander, A., Hovy, E., Schütze, H., & Goldberg, Y. (2021). Measuring and improving consistency in pretrained language models. *Transactions of the Association for Computational Linguistics*, 9, 1012–1031. DOI: 10.1162/tac1_a_00410.
18. Zheng, L., Chiang, W. L., Sheng, Y., Zhuang, S., Wu, Z., Zhuang, Y., ... & Stoica, I. (2023). Judging llm-as-a-judge with mt-bench and chatbot arena. *Advances in neural information processing systems*, 36, 46595–46623. DOI: 10.48550/ARXIV.2306.05685.
19. Reimers, N., & Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. 3982–3992. DOI: 10.18653/V1/D19-1410.



ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В ВОЕННОМ ДЕЛЕ: ВОЗМОЖНОСТИ, УГРОЗЫ, ПЕРСПЕКТИВЫ¹

Ромашкина Н. П.²

DOI: 10.21681/2311-3456-2025-6-158-165

Цель статьи: выявить актуальные на текущем этапе возможности, угрозы и перспективы применения искусственного интеллекта в военном деле (ИИВД) для выработки предложений по расширению потенциала его использования, обеспечивающего экономическое, научно-технологическое развитие и безопасность России.

Метод исследования: анализ данных о применении искусственного ИИВД, синтез и научное прогнозирование, экспертная оценка, фактологический анализ в рамках системного подхода, междисциплинарный подход.

Полученный результат: представлен анализ понятия «искусственный интеллект в военном деле», его текущих показателей и характеристик на фоне ускоренного развития искусственного интеллекта (ИИ) в целом. Приведены ключевые факторы, определяющие целесообразность разработки и внедрения систем с ИИ в военной сфере, а также основные направления их использования и роль в проблематике международной безопасности. Выявлены риски и угрозы их применения. Проведен анализ возможностей различных стран мира по использованию технологий ИИ на стратегическом, оперативном и тактическом уровнях, соответствующих угроз в вооруженных конфликтах и войнах, а также прогноз развития перспективных технологий. Сформулированы проблемы влияния технологий ИИВД на уровень стратегической стабильности, национальной и международной безопасности. Доказывается, что характеристики этих технологий являются сегодня одним из важнейших показателей влияния и потенциала государства в мире, но требуют выработки мер укрепления доверия и создания международного режима контроля.

Практическая ценность: предложения по расширению потенциала использования искусственного интеллекта в военном деле для обеспечения экономического, научно-технологического развития и безопасности России.

Ключевые слова: искусственный интеллект, военное дело, система с искусственным интеллектом, новейшие технологии, вооруженное противоборство, ядерное оружие, стратегические ядерные силы, система боевого управления, стратегическая стабильность.

Введение

Искусственный интеллект (ИИ) захватывает все большую часть мирового рынка (рис. 1, 2), в РФ совокупный вклад этих технологий в ВВП страны к 2030 г. должен превысить 11 трлн руб.³ 19 ноября 2025 г. Президент России В. В. Путин поручил создать в России штаб руководства деятельностью в сфере ИИ⁴. ИИ становится определяющим фактором военно-экономической мощи государства, оказывая кардинальное влияние на облик современной войны и мира [1, 2], а его использование в самых передовых военных разработках открывает новые возможности для усиления обороноспособности страны. Именно поэтому внедрение систем с искусственным интеллектом (СИИ) становится важнейшим инструментом для повышения потенциала вооруженных сил (ВС) [3]. При этом не всякое программное управление, даже автоматизированное, является искусственным интеллектом. Автоматизированные системы управления часто отождествляют с СИИ, хотя последние

предполагают гораздо более сложные алгоритмы и характеризуется возможностью компьютера адаптироваться к динамически изменяющимся условиям для принятия обоснованных рекомендаций, команд и решений.

Учитывая, что государства все более активно разрабатывают и внедряют СИИ для использования в боевых действиях, можно прогнозировать гонку вооружений с элементами ИИ между крупнейшими военными державами.

Идея разработки элементов ИИ и эксперименты по его внедрению в военную сферу имеют долгую историю, начиная с 50-х гг. прошлого века [4]. Первые исследования связаны именно с военной деятельностью и на ранних этапах финансировались военными для достижения превосходства над противником. XI век ознаменовал новый период этого эволюционного процесса, однако многие специалисты называют его третьей революцией в военном

1 Статья подготовлена при поддержке гранта Министерства науки и высшего образования РФ на проведение крупных научных проектов по приоритетным направлениям научно-технологического развития №075-15-2024-551 «Глобальные и региональные центры силы в формирующемся мироустройстве».

2 Ромашкина Наталья Петровна, кандидат политических наук, руководитель подразделения проблем информационной безопасности Национального исследовательского института мировой экономики и международных отношений им. Е. М. Примакова РАН, Москва, Россия. E-mail: Romachkinan@yandex.ru

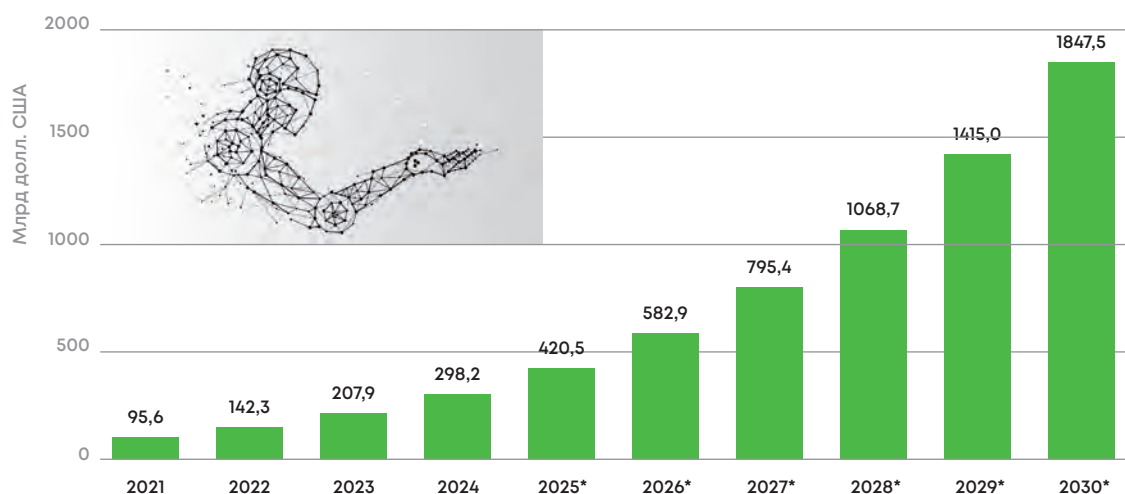
3 Национальная Стратегия развития искусственного интеллекта на период до 2030 года (В редакции Указа Президента Российской Федерации от 15.02.2024 № 124) // <http://www.kremlin.ru/acts/bank/44731> (дата обращения 23.09.2025).

4 Конференция «Путешествие в мир искусственного интеллекта» // <http://www.kremlin.ru/events/president/transcripts/78498> (дата обращения 19.11.2025).

Рис. 1. Страны-лидеры по разработкам ИИ⁵

Рынок искусственного интеллекта вырастет в 20 раз к 2030 году

В ближайшее десятилетие рынок продемонстрирует уверенный рост. По прогнозам, к 2030 году его стоимость составит почти \$2 триллиона.

Рис. 2. Прогноз рынка ИИ⁶

5 Страны-лидеры по разработкам в области искусственного интеллекта // <https://interaffairs.ru/news/show/24891> (дата обращения 23.09.2025).

6 Рынок искусственного интеллекта вырастет в 20 раз к 2030 году // <https://worldmarketstudies.ru/article/rynok-iskusstvennogo-intellekta-vyrastet-v-20-raz-k-2030-godu/> (дата обращения 23.07.2025).

деле, сравнивая оружие с ИИ с изобретением пороха и ядерного оружия (ЯО)⁷.

Искусственный интеллект в военном деле

В военном лексиконе многих военных держав появился специальный термин «Искусственный интеллект в военном деле».

В России технологии искусственного интеллекта в военном деле (ИИВД) рассматриваются как область исследований, в которой разрабатываются модели, системы и устройства, имитирующие интеллектуальную деятельность человека в сфере вооруженной борьбы, которые ведутся по трем основным направлениям: системы, основанные на знаниях; нейронные системы; системы эвристического поиска⁸. В частности, в Ракетных войсках стратегического назначения РФ (РВСН) достижения теории ИИВД используются при создании одного из самых перспективных классов информационных систем военного назначения – систем поддержки принятия решений должностными лицами и интеллектуальных систем и образцов вооружения различного назначения. Кроме того, важным направлением внедрения ИИ в РВСН является создание и эксплуатация экспертных систем (ЭС), в состав которых входят:

- база знаний;
- лингвистический процессор, обеспечивающий общение пользователей с ЭС;
- решатель, реализующий с помощью рабочей памяти механизм логического вывода;
- компоненты приобретения знаний и объяснения хода и результата решения задачи.

При этом особое значение имеют диагностические ЭС байесовского типа⁹.

ИИВД включают машинное обучение, глубокое обучение, компьютерное зрение, обработка естественного языка, полуавтономные и автономные системы, системы поддержки принятия решений.

В настоящее время США, Китай, Великобритания, Россия, Франция, Израиль, Германия, Индия, Австралия, КНДР и ряд других стран реализуют военные программы с ИИ в системах управления войсками и оружием, а также в отдельных образцах вооружения и военной техники (ВВТ).

Анализ возможностей ИИВД в вооруженных конфликтах и войнах и соответствующих угроз (рис. 3), а также прогноз их развития в будущем показывают, что определяющими факторами станут не столько географические условия и политическая обстановка,

сколько научно-технические достижения, а военное превосходство – от качества алгоритмов и программного обеспечения (ПО).

Стратегический уровень ИИВД

Несмотря на ускоренное развитие ИИ на текущем этапе, его использование в военных системах нельзя однозначно назвать революционным. Как уже упоминалось, многие приложения, например поддержка принятия решений в системах командования, управления и контроля, используются уже десятилетиями. Однако функциональность ИИ совершенствуется, и он будет играть все более важную роль на стратегическом, оперативном и тактическом уровнях [5].

Вопросы применения ИИ в **стратегических ядерных силах** (СЯС) требуют особой ответственности, поскольку эта область является наиболее чувствительной.

На современном этапе ИИ уже активно используется в средствах доставки ЯО. Например, США, Великобритания и Франция разрабатывают новые классы атомных подводных лодок с баллистическими ракетами (ПЛАРБ), в которых ИИ будет играть значительную роль в навигации, управлении и подводном обнаружении¹⁰. США и Франция также инвестируют в свои стратегические самолеты-носители ЯО¹¹. Кроме того, Франция объявила о совместном с Германией (не является ядерной державой по ДНЯО) проекте разработки боевого самолета, способного нести ЯО *Future Combat Air System*, который планируется ввести в строй около 2040 г.¹² ИИ в этих аппаратах будет интегрирован в навигацию и управление; управление в чрезвычайных ситуациях; сбор, анализ и объединение данных для предоставления пилоту актуальной информации; разработку вариантов действий для выполнения конкретных задач; координацию действий с другими платформами. При этом США и Франция заявили, что на данный момент все ядерные миссии будут пилотируемыми¹³.

Технологии ИИ способствует оптимизации **процесса принятия стратегических решений в военных условиях**, что позволяет формировать «глобальную оперативную картину», на основе которой командирам различного уровня будут предлагаться наиболее эффективные варианты действий [6].

7 Kai-Fu Lee. The Third Revolution in Warfare. The Atlantic // https://www.theatlantic.com/technology/archive/2021/09/i-weapons-are-third-revolution-warfare/620013/?mc_cid=0c6afe1b5e (дата обращения 29.09.2025).

8 Искусственный интеллект в военном деле. Энциклопедия РВСН // <https://mil.ru/services/encyclopedia/dictionary/listrvsn/cf9ada6d-af59-433f-b05a-5a1f423ee4f8> (дата обращения 27.09.2025).

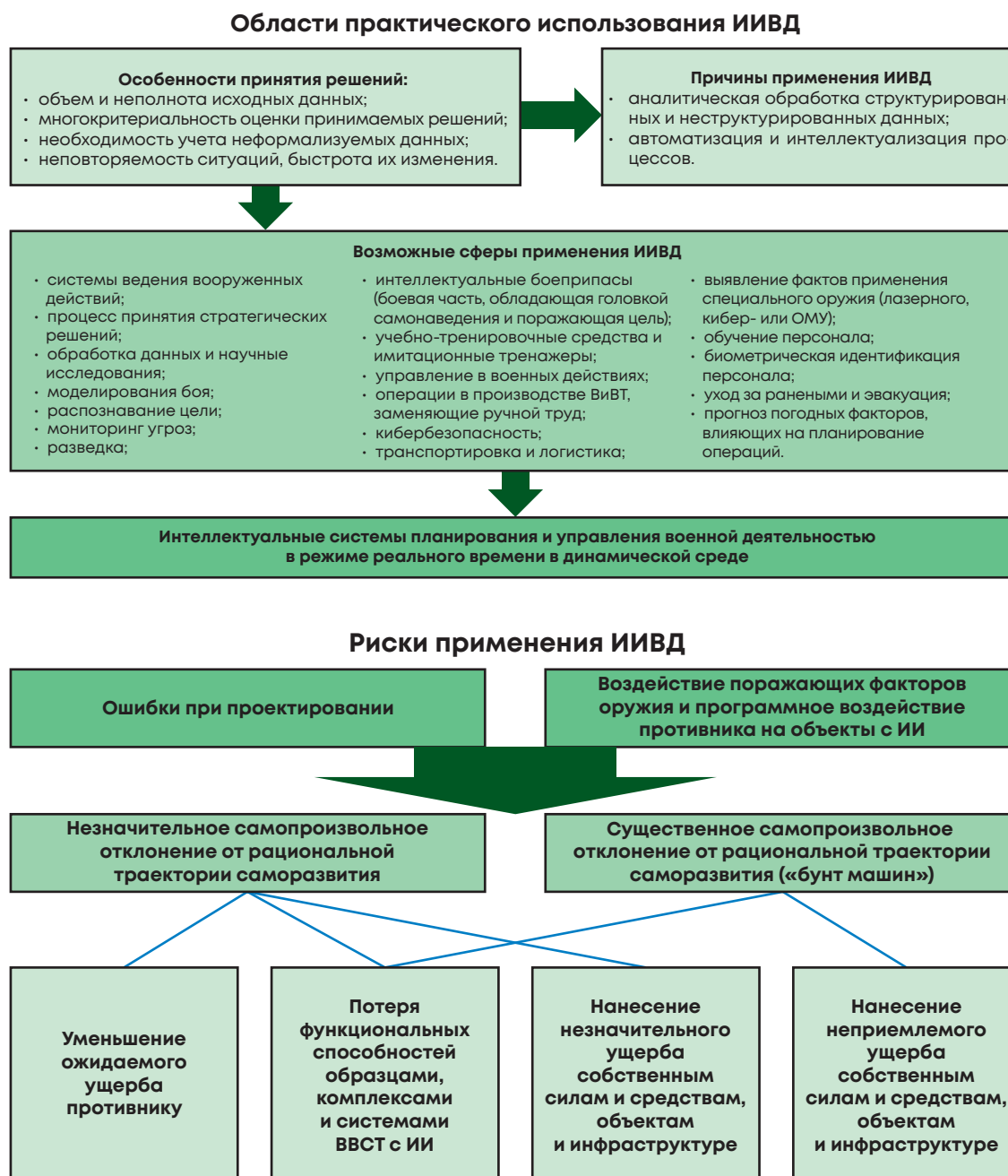
9 Экспертные системы (ЭС) // <https://mil.ru/services/encyclopedia/dictionary/listrvsn/cf9ada6d-af59-433f-b05a-5a1f423ee4f8> (дата обращения 23.09.2025).

10 Clark, B., The Emerging Era in Undersea Warfare (Center for Strategic and Budgetary Assessments: Washington, DC, 22 Jan. 2015); Mukherjee, T., Securing the Maritime Commons: The Role of Artificial Intelligence in Naval Operations, Observer Research Foundation (ORF) Occasional Paper (ORF: New Delhi, 16 July 2018).

11 Evans, D. and Schwalbe, J., The Long-Range Standoff (LRSO) Cruise Missile and Its Role in Future Nuclear Forces (John Hopkins Applied Physics Laboratory: Laurel, MD, 2017).

12 Everstine, B., 'French Air Force begins research into sixth generation aircraft', Air Force Magazine, 7 Feb. 2019.

13 Sayler, K. and Scharre, P., 'The B-21 Bomber should be unmanned on day 1', Defense One, 31 May 2016; and Sprenger, S., 'With nukes in mind, French officials stake out must-haves for Franco-German warplane', Defense News, 15 Nov. 2018.

Рис. 3. Области практического применения и риски ИИВД¹⁴

В качестве примеров современных систем командования, управления, связи, разведки, наблюдения и рекогносцировки для поддержки принятия решений можно привести *Project Maven* (США) с алгоритмами машинного обучения для анализа видеоматериалов с БПЛА (испытывается на Украине в боевых действиях против России¹⁵); командная информаци-

онная система «Андромеда-Д» (РФ) с элементами ИИ для обработки данных с различных источников¹⁶; система объединенного командования и управления JOCAS (КНР) с технологиями ИИ для анализа разведывательной информации¹⁷.

Целью применения ИИ в системах управления в комплексах противоракетной и противовоздушной

14 Рисунок построен автором на основе данных: <https://russiancouncil.ru/analytics-and-comments/columns/sandbox/opasnye-posledstviya-primeneniya-tehnologii-ii-v-voennykh-tselyakh/>; <https://blog.arsenal-otechestva.ru/images/articles/2018/01/iskin/pic8.jpg>; <https://topwar.ru/258020-boevoy-iskusstvennyj-intellekt-i-vojna-zavtrashnego-dnja.html> 5a1f423ee4f8 (дата обращения 27.09.2025).

15 David E. Sanger. In Ukraine, New American Technology Won the Day. Until It Was Overwhelmed. April 25, 2024 // <https://www.nytimes.com/2024/04/23/us/politics/ukraine-new-american-technology.html#> (дата обращения 25.09.2025).

16 АСУВ «Андромеда-Д». 2 декабря 2022 // <https://mil.ru/et/news/more.htm?id=11545957%40egNews> (дата обращения 23.09.2025).

17 В Китае создали первый ИИ, который может руководить всеми вооруженными силами страны. 17.06.2024 // <https://naked-science.ru/community/966182> (дата обращения 27.09.2025).

обороны является ускорение обработки данных, поступающих от средств контроля воздушно-космической обстановки и предупреждения о ракетном нападении, а также в автоматическом управлении средствами обороны.

В перспективных системах ПВО-ПРО США планируют использовать ИИ в качестве ключевого средства управления в единой распределенной системе стратегического уровня, компоненты которой размещены по всему земному шару. В частности, в рамках реализации проекта «Золотой купол» по созданию «противоракетного щита» для территории США предполагается использование ИИ для оптимизации всей системы¹⁸ с привлечением частных компаний.

Влияние технологий ИИ на уровень стратегической стабильности растет и выходит за рамки их использования в СЯС. Оценивая новые риски, угрозы и этические проблемы, а также учитывая неизбежность дальнейшего развития ИИВД, целесообразно понимать, что при конструктивном рациональном подходе элементы ИИ повышают эффективность системы управления СЯС и надежность СПРН, что укрепляет ядерное сдерживание, обеспечивая неотвратимость ядерного возмездия. Следовательно, ИИ может повысить уровень стратегической стабильности при условии международного контроля [7, 8]. Расширение международного диалога по выработке общего понятийного аппарата, оценке рисков и угроз, обмену опытом внедрения новейших технологий, в том числе, ИИВД; международное сотрудничество в этой сфере, в первую очередь, на площадках ООН; разработка мер укрепления доверия могут стать этапами важнейшего процесса на пути к созданию в будущем международного режима контроля над ИИВД с учетом большого опыта великих держав по созданию режимов контроля над оружием массового уничтожения (ОМУ) [9, 10, 11].

Оперативно-тактический уровень ИИВД

Системы ведения боевых действий используют ИИ, чтобы сделать операции более эффективными и менее зависимыми от участия человека, однако это связано с большим количеством угроз и этических проблем. Максимально перспективными во многих странах рассматриваются разработки по созданию полуавтономных и автономных боевых или обеспечивающих средств, способных действовать самостоятельно и продолжать выполнение задания в случае потери связи с центром управления. Традиционными примерами такой техники являются *беспилотные летательные аппараты (БПЛА), надводные и подводные аппараты различного назначения, наземные СИИ*.

Примеры современных ударных БПЛА в качестве «верного ведомого» управляемого человеком истребителя: *XQ-58 Valkyrie* (США), новая версия которого была представлена в апреле 2025 г.¹⁹; тяжелый дрон *C-70 Охотник* (РФ) – «напарник» Су-57, который в свою очередь в 2025 г. был оборудован системой ИИ, выполняющей функции второго пилота²⁰. Активно идут разработки проектов автоматического группового взаимодействия различных объектов, например, ударно-разведывательных БПЛА с самолетами дальнего радиолокационного обнаружения и патрульной авиации.

В качестве примеров морских автономных систем можно привести беспилотное судно *Sea Hunter* (США)²¹; подводный аппарат с ядерной энергетической установкой *Посейдон* (РФ)²²; подводный аппарат для разведки и противолодочной обороны *HSU-001* (КНР)²³; надводный аппарат *Sea Baby* (Украина) (применялся в атаке на Крымский мост в 2023 г.)²⁴.

Перспективным направлением дальнейшего совершенствования полуавтономных и автономных средств является их групповое взаимодействие, например, так называемый рой дронов, предусматривающий скоординированные действия большого количества агентов (боидов). Каждый боид следует простым правилам, но в совокупности система демонстрирует исключительно сложное совместное поведение, подобное рою пчел. В настоящее время уже десятки государств (США, страны ЕС, Россия, Великобритания, Турция, Израиль, Объединённые Арабские Эмираты, Южная Африка и др.) активно разрабатывают и используют такие вооружения в различных целях, причем речь идет о многотысячных роях.

Исследование ускоренного развития и распространения полуавтономных и автономных систем вооружений с ИИ, а также жесткой конкуренции между крупными военными державами доказывает, что уже присутствует гонка таких вооружений с перспективой ее увеличения на фоне роста количества горячих конфликтов в мире. Следовательно, вопрос

18 Judson J. Army Eyes Artificial Intelligence to Enhance Future Golden Dome // Defense News, 28.03.2025 // <https://www.defensenews.com/land/2025/03/28/army-looks-to-artificial-intelligence-to-enhance-future-golden-dome/> (дата обращения 23.09.2025).

19 XQ-58 Valkyrie with Built-In Landing Gear Shown in New Rendering. Apr 16, 2025 // <https://www.twz.com/air/xq-58- Valkyrie-with-built-in-landing-gear-seen-in-new-rendering> (дата обращения 23.09.2025).

20 В Су-57 внедрили искусственный интеллект. 24.05.2025 // https://ria.ru/20250524/pomoschnik-2018798216.html?utm_source=yxnews&utm_medium=desktop&utm_referrer=https%3A%2F%2Fzen.ru%2Fnews%2Fstory%2Fa0e7ee7f-ce05-559c-9efb-7510d1527a9c (дата обращения 23.09.2025).

21 Navy's Sea Hunter Drone Ship Has Sailed Autonomously To Hawaii And Back Amid Talk Of New Roles. Feb 4, 2019 // <https://www.twz.com/26319/usns-sea-hunter-drone-ship-has-sailed-autonomously-to-hawaii-and-back-amid-talk-of-new-roles> (дата обращения 23.10.2025).

22 Торпеду «Посейдон» назвали оружием апокалипсиса: почему ее боится весь мир. 10 апреля 2023 // <https://news.ru/weapon/pyat-tysyach-hirosim-pochemu-torpedu-posejdon-nazvali-oruzhiem-apokalipsisa> (дата обращения 27.09.2025).

23 David R. Strachan. China Enters the UUV Fray // <https://thediplomat.com/2019/11/china-enters-the-uuv-fray/> (дата обращения 27.10.2025).

24 New Drone Boat Named Sea Baby Used In Kerch Bridge Attack. Aug 16, 2023 // <https://www.twz.com/new-drone-boat-named-sea-baby-used-in-kerch-bridge-attack> (дата обращения 27.10.2025).

международного контроля над таким оружием представляется неизбежным в среднесрочной перспективе. Так как летальные (смертоносные) автономные системы вооружений требуют лишь частичного вмешательства человека или не требуют вообще, а, таким образом, нивелируют многие этические, нравственные и психологические аспекты, то логично предположить, что на международном уровне может быть поставлен вопрос о рассмотрении таких военных систем с элементами ИИ в качестве ОМУ.

Заключение

Потенциал использования систем с ИИВД очевиден, однако реализация их потенциала связана с необходимостью решения ряда сложных задач и проблем.

1. Выработка и законодательное закрепление единого понятийного аппарата в области ИИ на межведомственном и государственном уровнях.
2. Инициативная роль России в выработке понятийно-терминологического аппарата в области ИИ, а также в выявлении вызовов и угроз применения ИИВД на международном уровне.
3. Расширение исследований по выявлению наиболее эффективных способов применения ИИВД и оценку их эффективности для совершенствования деятельности по определению и нормативному закреплению первоочередных задач Минобороны и ВС РФ в мирное и военное время, для выполнения которых целесообразно применение ИИ.
4. Расширение применения ИИ для обеспечения безопасности критической военной технологии при решении принципиально новых военных задач.
5. Совершенствование системы подготовки разработчиков специального ПО и других специалистов для обеспечения эксплуатации и обслуживания СИИ в военных и гражданских образовательных учреждениях. При подготовке и развитии кадрового потенциала в сфере ИИВД применять междисциплинарный подход, включающий возможность взаимодействия двух или более научных дисциплин и выявление новых областей знания, которые не исследуются существующими дисциплинами.

6. Совершенствование разработок отечественного специального программного обеспечения и развитие собственного производства материалов, электрорадиоизделий и других комплектующих для СИИ.
7. Расширение разработок систем предотвращения перехвата управления объектами с СИИ (защитной аппаратуры сбора и обработки информации, каналов связи, устойчивых к перехвату и воздействию кибератак и средств РЭБ, криптостойкого ПО).
8. Обеспечение благоприятных условий для недопущения завоевания противником превосходства в стратегической космической зоне: комплекс мероприятий в космосе и на территории России, в т.ч., новых проектов в сфере ИИ.
9. Международные инициативы России с целью разработки новых юридически обязательных норм в отношении летального (смертоносного) полуавтономного и автономного оружия с учетом национальных интересов разных государств, исключающие доминирование отдельных стран в области ИИ за счет введения дискриминационных международных технических стандартов.
10. Расширение международного сотрудничества в сфере ИИ, в том числе в ИИВД, в первую очередь, в рамках СНГ, Евразэс, ШОС, БРИКС и ОДКБ.

Таким образом, СИИ без преувеличения становятся одним из важнейших инструментов научно-технического прогресса, обладающим потенциалом радикально изменить не только характер военных действий, но и всю сферу силового противостояния между государствами, включая экономическую, информационную и кибернетическую области. Представленный в статье анализ применения ИИВД, российских и иностранных разработок в этой сфере позволяет сделать вывод об уникальной роли ИИ в развитии и повышении потенциала современных вооружений и ВС, его влиянии на уровень стратегической стабильности, национальной и международной безопасности, что доказывает значимость ИИВД в качестве одного из важнейших показателей авторитета и потенциала государства в мире.

Литература

1. Ромашкина Н. П., Марков А. С., Стефанович Д. В. Information Technologies and International Security: [electronic resource]. – Moscow: IMEMO, 2023. – 111 p. – ISBN 978-5-9535-0613-7. – DOI 10.20542/978-5-9535-0613-7. // <https://www.imemo.ru/publications/info/information-technologies-and-international-security>.
2. Ромашкина Н., Стефанович Д. Искусственный интеллект и естественная война // Россия в глобальной политике. 2025. № 6(136). С. 69–83, DOI 10.31278/1810-6439-2025-23-6-69-83.
3. Понкин И. В. Военная аналитика. Военное применение искусственного интеллекта и цифры / Консорциум «Аналитика. Право. Цифра». – М.: Буки Веди, 2022. – 106 с. ISBN 978-5-4465-3667-2.
4. Turing A. M. Computing machinery and intelligence // Mind. 1950. №59. P. 433-460. // https://philosophie.universite.tours/documents/1950_Alan_Turing.pdf (дата обращения 23.09.2025).

5. Ромашкина Н. П. Глобальные военно-политические проблемы международной информационной безопасности: тенденции, угрозы, перспективы // Вопросы кибербезопасности. 2019. №. 1(29). С. 2–9, DOI: 10.21681/2311–3456-2019-1-2-9.
6. Романенко П. Г., Бойко А. П., Субботин Д. В. Повышение управляемости мультисервисных сетей связи специального назначения за счет внедрения технологий искусственного интеллекта // Военная мысль. 2025. № 7. С. 85–93.
7. Марков А. С., Шеремет И. А. Безопасность программного обеспечения в контексте стратегической стабильности // Вестник академии военных наук. 2019. № 2(67). С. 82–90.
8. Lin H. Artificial Intelligence and Nuclear Weapons: A Commonsense Approach to Understanding Costs and Benefits // Texas National Security Review. 2025. Vol. 8. No. 3. P. 98–109.
9. Ромашкина Н. П. Международно-правовой режим контроля над кибероружием в будущем миропорядке: угрозы и перспективы // Дипломатическая служба. 2023. № 2. С. 150–161. DOI 10.33920/vne-01-2302-07. // <https://www.imemo.ru/files/File/ru/publ/2023/DipSluzhba-022023-Romashkina.pdf>, (дата обращения 23.09.2025).
10. Ромашкина Н. П., Марков А. С., Стефанович Д. В. Международная безопасность, стратегическая стабильность и информационные технологии / отв. ред. А. В. Загорский, Н. П. Ромашкина. – М.: ИМЭМО РАН, 2020. – 98 с. DOI: 10.20542/978-5-9535-0581-9. // <https://www.imemo.ru/publications/info/romashkina-np-markov-as-stefanovich-dv-mezhdunarodnaya-bezopasnosty-strategicheskaya-stabilnosty-i-informatsionnie-tehnologii-otv-red-av-zagorskiy-np-romashkina-m-imemo-ran-2020-98-s> (дата обращения 23.07.2025).
11. Wehsener A., Reddie A. W., Walker L., Reiner P. AI-NC3 Integration in an Adversarial Context: Strategic Stability Risks and Confidence Building Measures // Institute for Security and Technology. 2023. February // <https://securityandtechnology.org/wp-content/uploads/2024/10/AI-NC3-Integration-in-an-Adversarial-Context.pdf> (дата обращения 23.10.2025).

ARTIFICIAL INTELLIGENCE IN MILITARY AFFAIRS: OPPORTUNITIES, THREATS, PROSPECTS²⁵

Romashkina N. P.²⁶

Keywords: artificial intelligence, military affairs, artificial intelligence system, latest technologies, armed confrontation, nuclear weapons, strategic nuclear forces, combat control system, strategic stability.

Purpose: to identify the current opportunities, threats and prospects for the application of artificial intelligence in military affairs to develop proposals for expanding the potential for its use, ensuring the economic, scientific and technological development and security of Russia.

Research method: analysis of data on the use of artificial intelligence in military affairs, synthesis and scientific forecasting, expert assessment, factual analysis within the framework of a systems approach, interdisciplinary approach.

Result: this article analyzes the concept of «artificial intelligence in military affairs», its current indicators, and characteristics against the backdrop of the accelerated development of artificial intelligence in general. It presents key factors determining the feasibility of developing and implementing artificial intelligence systems in the military sphere, as well as the main areas of their use and their role in international politics and global security. The risks and threats of their application are identified. An analysis of the capabilities of various countries in using artificial intelligence technologies at strategic, operational, and tactical levels, the corresponding threats in armed conflicts and wars, and a forecast for the development of promising technologies is provided. The impact of artificial intelligence technologies in military affairs on strategic stability, national, and international security is discussed. It is demonstrated that the characteristics of artificial intelligence technologies in military affairs are currently one of the most important indicators of a state's influence and potential in the world but require the development of trust-building measures and the creation of an international control regime.

Practical value: proposals for expanding the potential for using artificial intelligence in military affairs to ensure economic, scientific and technological development and security of Russia.

References

1. Romashkina N. P., Markov A. S., Stefanovich D.V. Information Technologies and International Security : [electronic resource]. – Moscow: IMEMO, 2023. – 111 p. – ISBN 978-5-9535-0613-7. – DOI 10.20542/978-5-9535-0613-7. // <https://www.imemo.ru/publications/info/information-technologies-and-international-security>.
2. Romashkina N., Stefanovich D. Iskusstvennyj intellekt i estestvennaya vojna // Rossiya v global'noj politike. 2025. № 6(136). S. 69–83, DOI 10.31278/1810-6439-2025-23-6-69-83.
3. Ponkin I. V. Voennaya analitika. Voennoe primeneniye iskusstvennogo intellekta i cifry / Konsorcium «Analitika. Pravo. Cifra». – M.: Buki Vedi, 2022. – 106 s. ISBN 978-5-4465-3667-2.
4. Turing A. M. Computing machinery and intelligence // Mind. 1950. № 59. P. 433–460. // https://philosophie.universite.tours/documents/1950_Alan_Turing.pdf (accessed 23.09.2025).
5. Romashkina N. P. Global'nye voenno-politicheskie problemy mezhdunarodnoj informacionnoj bezopasnosti: tendencii, ugrozy, perspektivy // Voprosy kiberbezopasnosti. 2019. №. 1(29). S. 2–9, DOI: 10.21681/2311–3456-2019-1-2-9.

²⁵ This article was prepared with the support of a grant from the Ministry of Science and Higher Education of the Russian Federation for major scientific projects in priority areas of scientific and technological development No. 075-15-2024-551 «Global and regional centers of power in the emerging world order».

²⁶ Nataliya Romashkina, Ph.D., Head of the Informational Security Problems Group of the Primakov National Research Institute of World Economy and International Relations (IMEMO) of the Russian Academy of Sciences, Moscow, Russia. E-mail: Romashkina@yandex.ru

6. Romanenko P. G., Bojko A. P., Subbotin D. V. *Povyshenie upravlyaemosti mul'tiservisnykh setej svyazi special'nogo naznacheniya za schet vnedreniya tekhnologij iskusstvennogo intellekta* // *Voennaya mysl'*. 2025. № 7. S. 85–93.
7. Markov A. S., SHeremet I. A. *Bezopasnost' programmnoho obespecheniya v kontekste strategicheskoy stabil'nosti* // *Vestnik akademii voennyh nauk*. 2019. № 2(67). P. 82–90.
8. Lin H. *Artificial Intelligence and Nuclear Weapons: A Commonsense Approach to Understanding Costs and Benefits* // *Texas National Security Review*. 2025. Vol. 8. No. 3. P. 98–109.
9. Romashkina N. P. *Mezhdunarodno-pravovoj rezhim kontrolya nad kiberoruzhiem v budushchem miroporyadke: ugrozy i perspektivy* // *Diplomaticheskaya sluzhba*. 2023. № 2. S. 150–161. DOI 10.33920/vne-01-2302-07. // <https://www.imemo.ru/files/File/ru/publ/2023/DipSluzhba-022023-Romashkina.pdf>, (accessed 23.09.2025).
10. Romashkina N. P., Markov A. S., Stefanovich D. V. *Mezhdunarodnaya bezopasnost', strategicheskaya stabil'nost' i informacionnye tekhnologii* / *otv. red. A. V. Zagorskiy, N. P. Romashkina*. – M.: IMEMO RAN, 2020. – 98 s. DOI: 10.20542/978-5-9535-0581-9. // <https://www.imemo.ru/publications/info/romashkina-np-markov-as-stefanovich-dv-mezhdunarodnaya-bezopasnosty-strategicheskaya-stabilynosty-i-informatsionnye-tehnologii-otv-red-av-zagorskiy-np-romashkina-m-imemo-ran-2020-98-s>, (accessed 23.07.2025).
11. Wehsener A., Reddie A. W., Walker L., Reiner P. *AI-NC3 Integration in an Adversarial Context: Strategic Stability Risks and Confidence Building Measures* // *Institute for Security and Technology*. 2023. February // <https://securityandtechnology.org/wp-content/uploads/2024/10/AI-NC3-Integration-in-an-Adversarial-Context.pdf> (дата обращения 23.10.2025).



ЭКСПЕРИМЕНТ ПО ПРОГНОЗИРОВАНИЮ ОТКАЗОВ УСТРОЙСТВА ПРОМЫШЛЕННОГО ИНТЕРНЕТА ВЕЩЕЙ НА БАЗЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

Марков Г. А.¹

DOI: 10.21681/2311-3456-2025-6-166-173

Цель исследования: обоснование возможности прогнозирования отказов устройств промышленного интернета вещей на основе методов машинного обучения с учителем.

Метод исследования: для достижения цели была разработана и апробирована методика, включающая этапы сбора данных, предобработки, формирования обучающих выборок, выбора информативных признаков, обучения классификационных моделей и оценки качества прогнозов. в качестве базового метода использована логистическая регрессия, позволяющая выявлять закономерности, предшествующие сбоям, в условиях высокой гетерогенности данных устройств промышленного интернета вещей.

Результат: эксперимент был выполнен на реальных данных устройства промышленного интернета вещей, включающих 81 показатель работы оборудования, представленных временными рядами. Результаты демонстрируют, что вероятность возникновения сбоя начинает возрастать задолго до фактического отказа, что подтверждает возможность раннего выявления угроз.

Практическая значимость заключается в возможности внедрения разработанной методики в действующие системы мониторинга для перехода от реактивного к проактивному управлению рисками.

Полученный вывод подтверждает, что предложенная методика обеспечивает повышение надёжности IIoT-систем, снижает простои оборудования и способствует повышению киберустойчивости промышленных процессов.

Ключевые слова: предиктивная диагностика, обучение с учителем, киберфизические системы, информационная безопасность, методы искусственного интеллекта, АСУ ТП.

Введение

Современные промышленные системы все чаще используют технологии промышленного интернета вещей (Industrial Internet of Things, IIoT), что способствует повышению эффективности производственных процессов, автоматизации управления и оптимизации технологических ресурсов. Для многих компаний внедрение IIoT является частью стратегии цифровой трансформации [1]. К примеру, согласно исследованию от Mordor Intelligence, мировой рынок IIoT растет на 25 % в год. В нашей стране данная тематика имеет особое стратегическое значение по причине решения задач по импортозамещению и технологическому суверенитету [2].

Несмотря на строгие требования к надежности и устойчивости современных промышленных систем, применение IIoT сопровождается ростом киберрисков, в том числе обусловленных относительно упрощенной с точки зрения безопасности архитектур IIoT. Об этом свидетельствует рост уязвимостей и атак на промышленные системы на базе IIoT. Однако выделяют не только информационные угрозы, но и реальные, связанные с физическими последствиями, как-то: авариями и простоями производства [3–9].

Одним из направлений исследования тематики безопасности IIoT является применение технологий

машинного обучения (ML) с целью оценки и прогнозирования угроз безопасности. В литературе приводится ряд академических исследований [10–21], в которых отмечается возможность не только автоматизировать процессы обнаружения угроз, но и значительно повысить точность прогнозирования и предотвращения инцидентов. Однако на практике отмечаются сложности внедрения указанных технологий по ряду объективных и субъективных причин. Цель данного исследования состоит в проверке гипотезы о том, возможно ли прогнозирование сбоя² (как инцидента отказа в обслуживании) функционирования программно-аппаратного обеспечения систем IIoT на основе мониторинга и анализа метрик производительности устройств до возникновения инцидента информационной безопасности.

Состояние проблематики информационной безопасности IIoT

Согласно NIST, под IIoT понимается сеть взаимосвязанных датчиков, приборов, контроллеров и других устройств, объединенных с помощью интернет-коммуникаций для решения промышленных задач. Обеспечение безопасности IIoT включает комплекс мер, направленных на защиту связанных с IIoT систем, устройств, сетей и данных от угроз сохранности и конфиденциальности технологических данных,

¹ Марков Георгий Алексеевич, аналитик по информационной безопасности компании «Инфосистемы Джет», Москва, Россия, E-mail: detonate1@yandex.ru

² Машинное обучение и прогнозирование отказа оборудования, 11.02.2022. [Электронный ресурс] – URL: <https://habr.com/ru/articles/650829/>

целостности управляющих команд, доступности и безотказности устройств, операционных процессов и пр. По сути, безопасность IIoT охватывает как аспекты кибербезопасности, так и функциональную безопасность [4]. Следует отметить, что требования по безопасности IIoT находятся в развитии – примерами являются соответствующие стандарты ISA/IEC/ISO, ENISA, ETSI EN, IISF и NIST. В России также имеется национальный стандарт ГОСТ Р 71777-2024 «Интернет вещей. Термины и определения». Соответственно, проекты OWASP и MITRE уделяют внимание уязвимостям, угрозам и сценариям атак на IIoT/IIoT/ICS, выделяя два специфических требования, а именно:

- непрерывный мониторинг, обнаружение угроз и инцидент-менеджмент;
- безопасность и сохранность одновременно (Secure&Safety).

Как отмечалось, методы машинного обучения все шире используются в кибербезопасности, и сфера IIoT не исключение. Методы машинного обучения находят применение в системах обнаружения вторжений [1, 21], анализе трафика и аномалий [11, 19], мониторинге событий [14], прогнозировании инцидентов [9, 17] и даже в управлении безопасностью в системах РВ применительно к безопасности киберфизических систем [13, 20]. К достоинствам методов машинного обучения относят: возможность анализа больших объемов разнородных данных в реальном времени, обнаружение ранее неизвестных атак, прогнозирование инцидентов, снижение нагрузки на персонал, выявление низкоинтенсивных атак, возможность использования периферийных вычислений (edge computing) и пр.

Кратко рассмотрим ключевые методы машинного обучения с учителем, используемые в области информационной безопасности (ИБ), а именно:

1. Логистическая регрессия. Данный метод эффективен для решения задач бинарной классификации, например определения наличия или отсутствия угрозы. Логистическая регрессия основывается на логит-преобразовании и вычисляется методом максимального правдоподобия, обеспечивая надежные и легко интерпретируемые результаты.
2. Метод k -ближайших соседей (k -NN). Алгоритм применяется для классификации данных путем анализа ближайших соседей. В контексте ИБ он используется для обнаружения аномалий и отклонений в поведении пользователей и сетевых данных. Однако эффективность данного метода сильно зависит от выбранного числа соседей и метрики расстояния, что требует тщательной настройки.

3. Наивный байесовский классификатор. Этот алгоритм основан на теореме Байеса и предназначен для классификации событий на основе вероятностных моделей. Он особенно полезен для обнаружения спама и фишинговых атак за счет анализа частотности встречаемых признаков в данных. Наивный байесовский метод прост, быстр и эффективно работает даже при небольших объемах обучающих данных, однако предполагает независимость признаков, что не всегда соответствует реальности.

4. Деревья решений и ансамблевые методы на их основе (например, Random Forest и Gradient Boosting). Деревья решений позволяют моделировать процессы принятия решений путем последовательного разделения данных на однородные группы, используя критерии прироста информации и энтропию. Ансамблевые методы значительно повышают точность и устойчивость к переобучению за счет комбинации результатов множества деревьев решений.

Согласно исследованию, проведенному Росатомом³, реальными примерами внедрения методов машинного обучения являются следующие:

- анализ журналов регистрации («логов») для выявления аномалий и прогнозирования инцидентов (машинное обучение позволяет эффективно обрабатывать большие объемы логов, выявляя скрытые паттерны и аномальное поведение системы);
- системы обнаружения вторжений, использующие машинное обучение для анализа сетевого трафика и выявления подозрительной активности, включая зашифрованные данные;
- технологии защиты от фишинга, основанные на глубоком обучении, которые используют нейронные сети для анализа заголовков и контекста сообщений, обеспечивая высокую точность и минимальное число ложных срабатываний;
- анализ поведения пользователей, позволяющий выявлять аномалии и потенциальные угрозы на основе изменения стандартных шаблонов поведения с помощью методов глубокого обучения.

Несмотря на многочисленные преимущества, применение машинного обучения в информационной безопасности имеет и ограничения. Основными из них являются необходимость регулярного обновления моделей при появлении новых угроз и уязвимостей, а также сложность интерпретации решений моделей.

3 Исследование «Росатома»: в России всплеск спроса на технологии интернета вещей, 21.11.2024. [Электронный ресурс] - URL: <https://www.rusatom-utilities.ru/media-center/news/issledovanie-rosatoma-v-rossii-vsplesk-sprosa-na-tekhnologii-interneta-veshchey/>

Таким образом, методы машинного обучения с учителем играют важнейшую роль в обеспечении ИБ систем IIoT, предлагая мощные инструменты для анализа данных и предотвращения угроз. Однако их успешное применение требует тщательной настройки, постоянного обновления и понимания их ограничений.

Оценка эффективности классификационных моделей в задачах информационной безопасности IIoT

Эффективность применения методов машинного обучения для обеспечения ИБ в системах IIoT зависит от точности и надежности используемых классификационных моделей. Для адекватной оценки таких моделей необходимо применять подходы, учитывающие специфику информационных систем и требования безопасности.

Одним из основных подходов к оценке эффективности классификационных моделей является использование стандартных метрик качества, таких как точность (accuracy), полнота (recall), точность положительного прогноза (precision), F1-мера, ROC-кривая и площадь под ROC-кривой (AUC).

Данные показатели позволяют дать комплексную оценку способности модели правильно классифицировать события, связанные с угрозами безопасности в системах IIoT.

Для оценки адекватности и точности моделей (главным образом, линейной регрессии) рекомендуется применять следующие методы и критерии:

1. Коэффициент детерминации (R^2) для оценки адекватности модели (линейной регрессии). Этот коэффициент показывает, какая часть изменчивости зависимой переменной объясняется используемыми факторами. Чем ближе коэффициент детерминации к единице, тем выше адекватность модели.
2. Оценка значимости модели с помощью F -критерия (критерий Фишера), позволяющего проверить гипотезу о значимости модели в целом. Высокие значения F -статистики свидетельствуют о значимости выбранных факторов и адекватности модели.
3. Оценка значимости коэффициентов модели с помощью t -критерия (критерия Стьюдента). Значимость каждого коэффициента проверяется на основании статистической значимости его вклада в общую модель. Коэффициенты, не прошедшие проверку на значимость, исключаются из модели, что улучшает её устойчивость и интерпретируемость.
4. Проверка условий Гаусса-Маркова для обеспечения корректности и надежности оценок коэффициентов модели. Это включает в себя проверку

несмещённости остатков, их случайности, независимости, гомоскедастичности и нормальности распределения.

5. Применение графовых моделей эксплойтов для устранения неопределенностей исходных данных и повышения точности классификации. Такой подход позволяет использовать как основные маршруты выполнения кода, так и функциональные зависимости между вызовами функций при классификации эксплойтов и выявлении уязвимостей.
6. Статический и динамический анализ исходного кода эксплойтов для выявления признаков реализации угроз и последующей классификации слабых мест и уязвимостей на основе графовых моделей.
7. Методика балльного оценивания эффективности мер защиты информации, используемая для обоснования рекомендаций по применению организационно-технических мер защиты. Данный метод учитывает степень реализованности и эффективность внедренных мер по следующим направлениям: организация, внедрение, контроль и поддержка уровня безопасности.

При оценке качества моделей также используются ошибки прогнозирования, среди которых выделяют среднюю абсолютную ошибку (MAE), среднюю квадратичную ошибку (MSE), корень из средней квадратичной ошибки (RMSE), среднюю абсолютную процентную ошибку (MAPE). Эти показатели помогают количественно определить отклонения прогнозных значений от фактических и выбрать наилучшую модель.

Таким образом, подход к оценке эффективности классификационных моделей в задачах обеспечения ИБ IIoT должен включать как стандартные, так и специализированные методы, учитывающие особенности киберугроз и уязвимостей в IIoT-системах. Это обеспечивает надежность классификации, позволяет выявлять скрытые уязвимости и эффективно противодействовать современным угрозам.

Особенности предварительного анализа и подготовки данных в системах IIoT

При проведении эксперимента одним из важных вопросов является подготовка исходных данных. Некоторые специалисты утверждают, что этот этап составляет до 80 % затрат на исследование.

Напомним, что IIoT характеризуется высокой степенью гетерогенности данных, поступающих от различных сенсоров, контроллеров и исполнительных механизмов. Эти данные могут различаться по формату, частоте обновления, точности и надежности. Для эффективного применения методов машинного обучения с учителем в системах IIoT необходимо

провести тщательный предварительный анализ и подготовку данных.

Перед началом сбора и анализа данных важно провести инвентаризацию существующей ИТ-инфраструктуры предприятия. Это включает в себя следующее:

- анализ сетевой инфраструктуры, включая топологию и пропускную способность каналов связи;
- оценку систем хранения данных и резервных систем;
- проверку наличия и состояния резервных копий данных;
- анализ текущего состояния информационной безопасности.

Такая оценка позволяет выявить потенциальные узкие места и определить требования к масштабированию инфраструктуры для обработки больших объемов данных, характерных для IIoT-систем.

Данные в системах IIoT поступают из различных источников, включая сенсоры, контроллеры, SCADA-системы и другие устройства. Для эффективного анализа необходимо обеспечить:

- сбор данных в реальном времени с минимальной задержкой;
- интеграцию данных из различных источников в единую платформу для последующей обработки;
- обеспечение целостности и достоверности данных на всех этапах их передачи и хранения.

Использование периферийных вычислений (edge computing) может существенно снизить нагрузку на центральные серверы и обеспечить предварительную обработку данных непосредственно на месте их сбора [20]. Это особенно актуально в условиях ограниченной пропускной способности каналов связи и необходимости быстрого реагирования на изменения в производственном процессе.

Сырые данные, поступающие от IIoT-устройств, могут содержать шум, пропущенные значения и аномалии. Для повышения качества анализа необходимо провести:

- очистку данных от выбросов и аномальных значений;
- заполнение пропущенных данных с использованием методов интерполяции или прогнозирования;
- нормализацию данных для приведения их к единому масштабу и формату.

При работе с неполными данными обычно применяются следующие действия:

- приближение: некоторые пропуски, в зависимости от типа данных, можно заменить модой (самое типичное значение переменной для категорий

или бинарных значений) или медианой (для целочисленных или непрерывных переменных);

- вычисление: недостающие данные вычисляются при помощи алгоритмов обучения с учителем;
- удаление: недостающие данные можно удалить, если это не приведет к потере объема данных и искажению результатов.

На наш взгляд, наиболее популярными методами нормализации можно считать Z -масштабирование и метод минимакс.

Z -масштабирование вычисляется по формуле:

$$Z = \frac{X - M}{S},$$

где: X – изначальная величина, M – среднее значение величины, S – среднеквадратичное отклонение.

Метод минимакс преобразует величину X к формату (0,1) по формуле:

$$X_{new} = \frac{X - \min(X)}{\max(X) - \min(X)}.$$

Эти шаги обеспечивают повышение точности и надежности последующего анализа и позволяют более эффективно применять методы машинного обучения.

В данных могут быть аномалии (значительные изменения величины с последующим приближенным восстановлением значения). Для анализа аномалий можно использовать метод Ирвина, который позволяет понять, являются ли значения аномалией или нет. Метод основан на формуле:

$$\lambda_r = \frac{Y_r - Y_r - 1}{S_y}.$$

Ключевым этапом подготовки данных для машинного обучения является выбор информативных признаков, которые будут использоваться в моделях. Это включает в себя следующее:

- анализ корреляции между различными параметрами для выявления наиболее значимых признаков;
- применение методов снижения размерности, таких как анализ главных компонент (PCA), для уменьшения объема данных без потери информативности;
- формирование обучающих и тестовых выборок с учетом сбалансированности классов и предотвращения переобучения моделей.

Метод главных компонент комбинирует наиболее информативные переменные в новые, наилучшим образом описывающие данные (новые переменные называются «главные компоненты»). Главные компоненты получаются из комбинаций изначальных переменных и ее можно описать формулой:

$$F_i = a_{i1}X_1 + \dots + a_{im}X_m, i = \overline{1, m}.$$

Как показал эксперимент, правильный выбор признаков и формирование выборок существенно влияет на эффективность и точность моделей машинного обучения в задачах обеспечения информационной безопасности в системах IIoT.

Методика прогнозирования поломки устройства IIoT

Разработанная методика основана на применении методов машинного обучения с учителем для анализа метрик производительности оборудования. Методика направлена на выявление предиктивных признаков сбоев и предупреждение отказов путём заблаговременного мониторинга.

Методика содержит следующие семь шагов:

Шаг 1. Сбор данных.

На данном шаге происходит непрерывный мониторинг производительности устройств. Формируются модели с ключевыми признаками, используемыми на всех типовых устройствах. Данные преобразуются и хранятся в виде временных рядов.

Шаг 2. Обработка данных.

Данный шаг необходим для выполнения операции очистки от шумов. Операция включает в себя следующее:

- заполнение пропущенных значений;
- выполнение нормализации показателей;
- проведенные проверки на наличие аномалии.

Шаг 3. Формирование обучающей выборки.

На данном шаге методики добавляется бинарная метка работоспособности. По умолчанию проставляется корректная работа (значение 1). Сбои задаются (значение 0) вручную при сборе промежутка простоя.

Шаг 4. Выбор признаков.

На этом шаге проводится анализ корреляции для исключения избыточных и слабо информативных признаков. При необходимости снижения размерности применяется метод главных компонент.

Шаг 5. Обучение.

В качестве базового алгоритма используется логистическая регрессия. Проводится кросс-валидация для оценки устойчивости. Настраиваются гиперпараметры.

Шаг 6. Оценка качества.

Выполняется расчет accuracy, precision, recall, F1-мера. Проводится анализ ROC-кривой и AUC. Оцениваются ошибки прогнозирования.

Шаг 7. Внедрение.

На шаге внедрения данные интегрируются в систему мониторинга для визуализации вероятности поломки на дашборде для операторов. Настраивается уведомление при превышении порогового значения.

Эксперимент

В качестве эксперимента использовалась обезличенная информация с одного заводского робота. С устройства удалось собрать 81 различных показателей. Их можно разбить на следующие категории:

- показатели ядер,
- показатели работы диска,
- показатели работы файловой системы,
- показатели сетевой активности,
- показатели памяти,
- показатели функционала самомониторинга,
- показатели работы служб ОС.

Статистика была собрана за 2024 год. Каждый показатель представляет из себя временной ряд. За собранный промежуток было 33 сбоя в работе основного функционала IIoT, в связи с чем к статистике было добавлено поле, отвечающее за работоспособность функционала (0 если не работает, 1 если работает).

Для обработки данных был написан модуль на языке python и интегрирован с системой мониторинга для проверки конкретного оборудования.

При первом сбое были проанализированы данные, по которым видно, что вероятность сбоя начинает увеличиваться до поломки оборудования, что можно использовать в качестве предотвращения поломки путем принятия заблаговременных мер.

Пример из дашборда представлен на рисунке 1 (начало сбоя отмечено красным прямоугольником).

По результатам можно предположить, что внедрение данного модуля на большее количество устройств

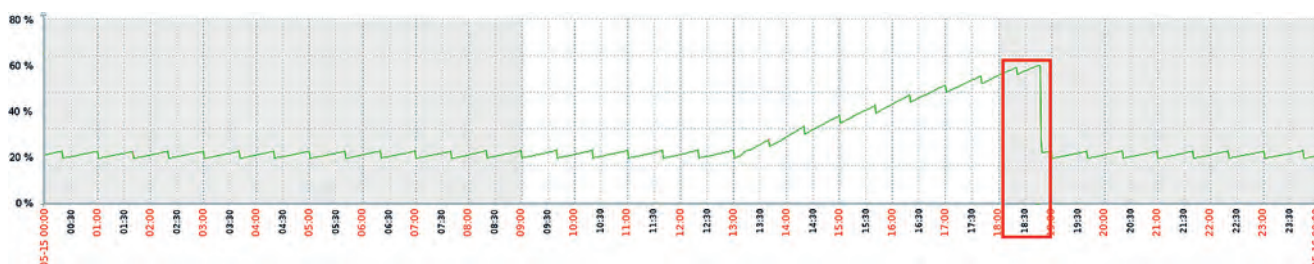


Рис. 1. Пример из дашборда

может повысить вероятность предсказания сбоя, визуализировать данные в виде одного графика для устройства, а не для каждого параметра.

Следует отметить, что цель статьи состояла в демонстрации эффекта, поэтому ряд общих вопросов был опущен, как-то: балансировка классов, риски переобучения и временная утечка данных.

Выводы

В результате проведенного исследования была разработана и протестирована методика прогнозирования сбоев в системах IIoT на основе методов машинного обучения с учителем.

Предложенный подход учитывает особенности IIoT, включая высокую гетерогенность данных, разнородность устройств и требование к обработке временных рядов.

Разработанная методика продемонстрировала следующие возможности:

- выявление предиктивных признаков сбоев на ранних этапах эксплуатации,
- автоматический анализ большого количества метрик производительности,
- формирование прогнозов вероятности отказа с достаточной точностью,
- интеграция в действующие системы мониторинга для визуализации и уведомлений операторов.

Как показал эксперимент, применение машинного обучения с учителем доказало свою эффективность для задач обеспечения ИБ в IIoT-системах, позволяя перейти от реактивного к проактивному управлению рисками. Это обеспечивает снижение простоев оборудования, повышение надёжности и устойчивости промышленных процессов к киберугрозам.

В будущих исследованиях планируется расширить масштаб и протестировать практическое предотвращение возможного сбоя.

Литература

1. Кибербезопасность цифровой индустрии. Теория и практика функциональной устойчивости к кибератакам / Д. П. Зегжда, Е. Б. Александрова, М. О. Калинин [и др.]. – М.: «Горячая линия-Телеком», 2021. – 560 с.
2. Системная инженерия в проблемах национальной безопасности: Тематический блок «Национальная безопасность» / А. В. Анищенко, В. Б. Артемьев, М. К. Бондарева [и др.]; под. ред. Махутова Н.А. – М.: «Знание», 2025. – 904 с.
3. Майлыбаев Е. Организация защиты информации в системах IIoT / Майлыбаев Е., Руслан М. // Вестник КазАТК. 2024. № 5(134). С. 168–175.
4. Марков А. С. Стандарты кибербезопасности Четвертой промышленной революции и Индустрии 4.0 / А. С. Марков, Ю. А. Тимофеев // Защита информации. Инсайд. – 2021. – № 3(99). – С. 54–60.
5. Марков Г. А. Контроль IIoT при помощи средств информационной безопасности в государственном и корпоративном управлении / Г. А. Марков, Е. Д. Маркова // Сб. трудов XIII международной научно-технической конференции «Безопасные информационные технологии». – Москва: МГТУ им. Н. Э.Баумана, 2024. – С. 148–153.
6. Марков Г. А. Проблемы обеспечения информационной безопасности в промышленных системах / Г. А. Марков, В. М. Крундышев, Д. П. Зегжда // Методы и технические средства обеспечения безопасности информации. – 2024. – № 33. – С. 100–101.
7. Ощепков Н. В. Анализ методов противодействия угрозам для промышленного интернета вещей / Ощепков Н. В., Кротова Е. А. – Международный научно-исследовательский журнал. 2024. № 7(145). С. 1–5.
8. Шевцов В. Ю., Анализ угроз и уязвимостей концепций IIoT и IIoT / Шевцов В. Ю., Касимовский Н. П. // НБИ технологии. 2020. Т. 14. № 3. С. 28–34.
9. Petrenko S. Cyber resilient platform for internet of things (IIoT/IIoT)ed systems: survey of architecture patterns / S. Petrenko // Voprosy Kiberbezopasnosti. – 2021. – No. 2(42). – P. 81–91. – DOI: 10.21681/2311-3456-2021-2-81-91.
10. Будников С. А., Методика оценки эффективности систем безопасности автоматизированных систем управления / Будников С. А., Коваленко С. М., Бочарова А. И. // Вопросы кибербезопасности. – 2023. – №3. – С.2–12. DOI:10.21681/2311-3456-2023-3-2-12.
11. Доргушаова А. К. Обзор использования технологий машинного обучения в обеспечении информационной безопасности данных: настоящее и будущее / Доргушаова А. К., Довгаль В. А., Козлова Н. Ш., Козлов Р. С. // Вестник Адыгейского государственного университета. Серия: Естественно-математические и технические науки. 2024. Вып. 1(336). С. 51–59. DOI: 10.53598/2410-3225-2024-1-336-51-59.
12. Егоров А. А. Стандартизация по Искусственному интеллекту в промышленности (обзор зарубежного состояния). Часть 2 / А. А. Егоров // Автоматизация и IT в энергетике. – 2023. – № 1(162). – С. 6–19.
13. Зегжда Д. П. Автоматизация управления безопасностью интеллектуальных систем с использованием графа атак и анализа рисков / Д. П. Зегжда, Д. А. Москвин, Е. М. Орел // Защита информации. Инсайд. – 2022. – № 2(104). – С. 22–28.
14. Марков Г. А. Математическая модель управления событиями информационной безопасности в системах промышленного Интернета вещей / Г. А. Марков // Методы и технические средства обеспечения безопасности информации. – 2024. – № 33. – С. 101–103.
15. Марков Г. А. Особенности обеспечения информационной безопасности в системах промышленного Интернета вещей / Г. А. Марков, В. М. Крундышев // Сб. трудов X Международной научно-практической конференции «Интеллектуальная инженерная экономика и Индустрия 5.0». В 2-х томах, Санкт-Петербург, 25–28 апреля 2024 года. – СПб.: Санкт-Петербургский политехнический университет Петра Великого, 2024. – С. 217–220. – DOI 10.18720/IEP/2024.1/53.
16. Петренко А. А. Киберустойчивая платформа Интернета вещей / А. А. Петренко, С. А. Петренко, А. Д. Костюков // Защита информации. Инсайд. – 2020. – № 4(94). – С. 20–30.
17. Федорченко Е. В., Оценивание защищенности информационных систем на основе графовой модели эксплойтов / Федорченко Е. В., Котенко И. В., Федорченко А. В. // Вопросы кибербезопасности. – 2023. – №3. – С.23–36. DOI:10.21681/2311-3456-2023-3-23-36.

18. Фирюлин М.Е. Задача разработки моделей оценки эффективности информационных процессов в центрах обработки данных в условиях комплексной защиты информации // Вестник Воронежского института МВД России. – 2015. – № 3. – С.234–238.
19. Щеткин А. А. Система мониторинга и прогнозирования поведения оборудования на базе IIoT / А. А. Щеткин, А. С. Бевз, С. С. Анцыферов [и др.] // Автоматизация в промышленности. – 2024. – № 3. – С. 51–55. – DOI 10.25728/avtprom.2024.03.10.
20. Chen Q. Communication-Efficient Federated Edge Learning for NR-U-Based IIoT Networks / Q. Chen, X. Xu, Z. You [et al.] // IEEE Internet of Things Journal. – 2022. – Vol. 9, No. 14. – P. 12450–12459. – DOI 10.1109/jiot.2021.3136205.
21. Karacayilmaz G. A novel approach detection for IIoT attacks via artificial intelligence / G. Karacayilmaz, H. Artuner // Cluster Computing. – 2024. – DOI 10.1007/s10586-024-04529-w.

EXPERIMENT ON PREDICTING IIOT DEVICE FAILURES BASED ON MACHINE LEARNING METHODS

Markov G. A.⁴

Keywords: predictive diagnostics, supervised learning, cyber-physical systems, information security, artificial intelligence methods, automated process control systems.

The purpose of the study: to substantiate the possibility of predicting failures of industrial Internet of Things devices based on supervised machine learning methods.

Research method: to achieve the goal, a methodology was developed and tested, including the stages of data collection, preprocessing, formation of training samples, selection of informative features, training of classification models and assessment of the quality of forecasts. Logistic regression was used as a basic method, which makes it possible to identify patterns that precede failures in conditions of high heterogeneity of data from industrial Internet of Things devices.

Result: the experiment was performed on real data from an industrial Internet of Things device, including 81 equipment performance indicators represented by time series. The results demonstrate that the probability of failure begins to increase long before the actual failure, which confirms the possibility of early detection of threats.

The practical significance lies in the possibility of introducing the developed methodology into existing monitoring systems for the transition from reactive to proactive risk management.

The conclusion obtained confirms that the proposed methodology provides an increase in the reliability of IIoT systems, reduces equipment downtime and contributes to increasing the cyber resilience of industrial processes.

References

1. Kiberbezopasnost' cifrovoj industrii. Teoriya i praktika funktsional'noj ustojchivosti k kiberatakam / D.P. Zegzhda, E. B. Aleksandrova, M. O. Kalinin [i dr.]. – M.: «Goryachaya liniya-Telekom», 2021. – 560 s.
2. Sistemnaya inzheneriya v problemah nacional'noj bezopasnosti: Tematicheskij blok «Nacional'naya bezopasnost'» / A. V. Anishchenko, V. B. Artem'ev, M. K. Bondareva [i dr.]; pod. red. Mahutova N.A. – M.: «Znanie», 2025. – 904 s.
3. Majlybaev E. Organizatsiya zashchity informacii v sistemah IoT / Majlybaev E., Ruslan M. // Vestnik KazATK. 2024. № 5 (134). S. 168-175.
4. Markov A. S. Standarty kiberbezopasnosti CHetvertoj promyshlennoj revolyucii i Industrii 4.0 / A. S. Markov, Yu. A. Timofeev // Zashchita informacii. Insajd. – 2021. – № 3(99). – S. 54–60.
5. Markov G. A. Kontrol' IoT pri pomoshchi sredstv informacionnoj bezopasnosti v gosudarstvennom i korporativnom upravlenii / G. A. Markov, E. D. Markova // Sb. trudov XIII mezhdunarodnoj nauchno-tehnicheskoy konferencii «Bezopasnye informacionnye tekhnologii». – Moskva: MG TU im. N.E.Baumana, 2024. – S. 148–153.
6. Markov G. A. Problemy obespecheniya informacionnoj bezopasnosti v promyshlennyh sistemah / G. A. Markov, V. M. Krundyshev, D. P. Zegzhda // Metody i tekhnicheskie sredstva obespecheniya bezopasnosti informacii. – 2024. – № 33. – S. 100-101.
7. Oshchepkov N. V. Analiz metodov protivodejstviya ugrozam dlya promyshlennogo interneta veshchej / Oshchepkov N. V., Krotova E. L. – Mezhdunarodnyj nauchno-issledovatel'skij zhurnal. 2024. № 7(145). S. 1–5.
8. Shevcov V. Yu., Analiz ugroz i uyazvimostej koncepcij IoT i IIoT / Shevcov V. Yu., Kasimovskij N. P. // NBI tekhnologii. 2020. T. 14. № 3. S. 28–34.
9. Petrenko S. Cyber resilient platform for internet of things (IIoT/IIoT)ed systems: survey of architecture patterns / S. Petrenko // Voprosy Kiberbezopasnosti. – 2021. – No. 2(42). – P. 81–91. – DOI 10.21681/2311-3456-2021-2-81-91.
10. Budnikov S. A., Metodika ocenki effektivnosti sistem bezopasnosti avtomatizirovannyh sistem upravleniya / Budnikov S. A., Kovalenko S. M., Bocharova A.I. // Voprosy kiberbezopasnosti. – 2023. – №3. – S.2–12. DOI:10.21681/2311-3456-2023-3-2-12.
11. Dorgushaova A. K. Obzor ispol'zovaniya tekhnologii mashinnogo obucheniya v obespechenii informacionnoj bezopasnosti dannyh: nastoyashchee i budushchee / Dorgushaova A. K., Dovgal' V. A., Kozlova N. Sh., Kozlov R. S. // Vestnik Adygejskogo gosudarstvennogo universiteta. Seriya: Estestvenno-matematicheskie i tekhnicheskie nauki. 2024. Vyp. 1(336). S. 51–59. DOI: 10.53598/2410-3225-2024-1-336-51-59.
12. Egorov A. A. Standartizatsiya po Iskusstvennomu intellektu v promyshlennosti (obzor zarubezhnogo sostoyaniya). CHast' 2 / A. A. Egorov // Avtomatizatsiya i IT v energetike. – 2023. – № 1(162). – S. 6–19.
13. Zegzhda D. P. Avtomatizatsiya upravleniya bezopasnost'yu intellektual'nyh sistem s ispol'zovaniem grafa atak i analiza riskov / D. P. Zegzhda, D. A. Moskvina, E. M. Orel // Zashchita informacii. Insajd. – 2022. – № 2(104). – S. 22–28.

⁴ Georgii A. Markov, Information Security Analyst, Infosistemy Jet, Moscow, Russia. E-mail: detonate1@yandex.ru

14. Markov G. A. *Matematicheskaya model' upravleniya sobyitiyami informacionnoj bezopasnosti v sistemah promyshlennogo Interneta veshchej* / G. A. Markov // *Metody i tekhnicheskie sredstva obespecheniya bezopasnosti informacii*. – 2024. – № 33. – S. 101–103.
15. Markov G. A. *Osobennosti obespecheniya informacionnoj bezopasnosti v sistemah promyshlennogo Interneta veshchej* / G. A. Markov, V. M. Krundyshev // *Sb. trudov X Mezhdunarodnoj nauchno-prakticheskoy konferencii «Intellektual'naya inzhenernaya ekonomika i Industriya 5.0»*. V 2-h tomah, Sankt-Peterburg, 25–28 aprelya 2024 goda. – SPb.: Sankt-Peterburgskij politekhnicheskij universitet Petra Velikogo, 2024. – S. 217–220. – DOI 10.18720/IEP/2024.1/53.
16. Petrenko A. A. *Kiberustojchivaya platforma Interneta veshchej* / A. A. Petrenko, S. A. Petrenko, A. D. Kostyukov // *Zashchita informacii. Insajd*. – 2020. – № 4(94). – S. 20–30.
17. Fedorchenko E. V., *Ocenivanie zashchishchennosti informacionnyh sistem na osnove grafovoj modeli eksplojtov* / Fedorchenko E. V., Kotenko I. V., Fedorchenko A. V. // *Voprosy kiberbezopasnosti*. – 2023. – № 3. – S. 23–36. DOI:10.21681/2311-3456-2023-3-23-36.
18. Firyulin M. E. *Zadacha razrabotki modelej ocenki effektivnosti informacionnyh processov v centrah obrabotki dannyh v usloviyah kompleksnoj zashchity informacii* // *Vestnik Voronezhskogo instituta MVD Rossii*. – 2015. – № 3. – S.234–238.
19. Shchetkin A. A. *Sistema monitoringa i prognozirovaniya povedeniya oborudovaniya na baze IIoT* / A. A. Shchetkin, A. S. Bevz, S. S. Ancyferov [i dr.] // *Avtomatizaciya v promyshlennosti*. – 2024. – № 3. – S. 51–55. – DOI: 10.25728/avtprom.2024.03.10.
20. *Communication-Efficient Federated Edge Learning for NR-U-Based IIoT Networks* / Q. Chen, X. Xu, Z. You [et al.] // *IEEE Internet of Things Journal*. – 2022. – Vol. 9, No. 14. – P. 12450–12459. – DOI: 10.1109/jiot.2021.3136205.
21. Karacayilmaz G. *A novel approach detection for IIoT attacks via artificial intelligence* / G. Karacayilmaz, H. Artuner // *Cluster Computing*. – 2024. – DOI 10.1007/s10586-024-04529-w.



The journal is included in the Russian list of peer-reviewed
academic publications of the Higher Attestation Commission (VAK),
it is registered in the Russian Science Citation Index (RSCI/RINTs)
on the Web of Science (WoS) platform and holds the 1st place
in its cyber security rating. The journal's articles
are available in full text

Editor-in-Chief

Alexey MARKOV, Dr.Sc., Professor, Moscow

Chairman of the Editorial Council

Igor SHEREMET, Academician of the RAS, Dr.Sc., Moscow

Assistant Editor-in-Chief

Grigory MAKARENKO, Senior Research Fellow, Moscow

Editorial Council

Michael BASARAB, Dr.Sc., Professor, Moscow

Andrey KALASHNIKOV, Dr.Sc., Professor, Moscow

Sergey KRUGLIKOV, Dr.Sc., Professor, Minsk, Belarus

Sergey PETRENKO, Dr.Sc., Professor, Sirius

Yuri STARODUBTSEV, Dr.Sc., Professor, St. Petersburg

Yuri YASOV, Dr.Sc., Professor, Voronezh

Editorial Board

Liudmila BABENKO, Dr.Sc., Professor, Taganrog

Alexander BARANOV, Dr.Sc., Professor, Moscow

Sergey GARBUK, Ph.D., Assoc. Prof., Moscow

Oleg GATSENKO, Dr.Sc., Professor, St. Petersburg

Dmitry ZEGZHDA, Corresponding Member of the RAS, Dr.Sc.,
Professor, St. Petersburg

Igor ZUBAREV, Ph.D., Assoc. Prof., Moscow

Alexander KOZACHOK, Dr.Sc., Orel

Roman MAXIMOV, Dr.Sc., Professor, Krasnodar

Vladislav PANCHENKO, Academician of the RAS, Dr.Sc., Professor, Moscow

Marina PUDOVKINA, Dr.Sc., Professor, Moscow

Valentin TSIRLOV, Ph.D., Assoc. Prof., Moscow

Igor SHAHALOV, Responsible Secretary, Moscow

Alexander SHELUPANOV, Corresponding Member of the RAS, Dr.Sc.,
Professor, Tomsk

Igor SHUBINSKIY, Dr.Sc., Professor, Moscow

Founder and publisher

JSC «NPO «Echelon»

Postal address: Elektrozavodskaya str., 24, bld. 1, 107023,
Moscow, Russia

E-mail: editor@cyberrus.info

CONTENTS

INFORMATION SECURITY RISK MANAGEMENT

ANALYSIS OF THE DEGRADATION OF ACCESS CONTROL SYSTEMS BASED ON MANDATED MODELS WITH THE ABILITY TO «WRITE DOWN»

Gaydamakin N. A. 2

SECURITY ANALYSIS METHODS AND TOOLS

A MULTI-LEVEL FRAMEWORK FOR JUSTIFYING INFORMATION SECURITY INCIDENT MONITORING AND RESPONSE PROCEDURES

Gryzunov V. V., Shestakov A. V. 14

APPLICATION OF ANALYTIC HIERARCHY PROCESS AND ANALYTIC NETWORK PROCESS TO ASSESSMENT PRIORITIES OF INFORMATION SECURITY SYSTEMS

Anosov R. S., Anosov S. S. 25

VULNERABILITIES OF ASYMMETRIC CIPHERS OF BLOCKCHAIN PLATFORMS

Ishchukova E. A., Petrenko S. A., Leonteva I. P. 35

NETWORK SECURITY

USES OF WEBAUTHN PROTOCOL VULNERABILITIES TO OBTAIN UNSANCTIONED ACCESS

Panchenko A. R. 48

SECURITY ANALYSIS METHODS AND TOOLS

ASSESSMENT OF ACTIVE DIRECTORY SECURITY USING ARTIFICIAL INTELLIGENCE

Bulgakova E. V., Bogdanov E. A., Kubankov A. N. 58

CYBERSECURITY TESTING AND MONITORING

PROBLEM-ORIENTED SYSTEM FOR MONITORING AND RESPONDING TO MULTIVECTOR ATTACKS IN A DECENTRALIZED INTERNET OF THINGS ENVIRONMENT

Tebueva F. B., Petrenko V. I., Satybalina D. Zh., Ogur M. G., Guseva T. M. 69

QUANTUM SECURITY

ALGORITHM FOR DETECTING SYNCHRONIZATION SIGNAL IN QUANTUM NETWORKS

Pljenkin A. P. 81

TECHNICAL REGULATION OF THE FIELD OF SAFETY

THE CRYPTOGRAPHIC PROTECTION SCHEME OF CLOUD STORAGE DATA WITH POSSIBILITY TO EXPANDING BY QUANTUM-DEPENDENT KEYS AND POST-QUANTUM ENCRYPTION ALGORITHMS

Minakov S. S., Tikhov S. V., Tsupak A. A. 88

REGULATORY AND TECHNICAL ASPECTS OF ENSURING INFORMATION AND CYBERSECURITY OF DIGITAL ELECTRIC POWER FACILITIES

Kulikov A. L., Loskutov A. A., Zinin V. M. 101

IDENTIFICATION AND AUTHENTICATION

METHOD OF PERSONALITY RECOGNITION BASED ON BIOACOUSTIC SCANNING IN MULTI-FACTOR AUTHENTICATION SYSTEMS

Chastikova V. A., Gulyai V. G. 116

SAFE ARTIFICIAL INTELLIGENCE

METHOD OF ENSURING CYBER RESISTANCE BLOCKCHAIN PLATFORMS BASED ON CYBER IMMUNITY

Balyabin A. A., Petrenko S. A. 127

RESEARCH OF SIGNAL METHODS FOR DETECTING ADVERSARIAL ATTACKS ON NEURAL MODELS THROUGH EXPLANATORY VISUALIZATION

Chekolaev D. A., Shevchenko A. V., Alekseev A. K., Trofimov Yu. V.,
Averkin A. N. 140

AN APPROACH TO ASSESSING THE QUALITY OF PENTEST SCENARIO GENERATION USING LARGE LANGUAGE MODELS

Konev A. A., Payusova T. I. 147

ARTIFICIAL INTELLIGENCE IN MILITARY AFFAIRS: OPPORTUNITIES, THREATS, PROSPECTS

Romashkina N. P. 158

MACHINE LEARNING

EXPERIMENT ON PREDICTING IIOT DEVICE FAILURES BASED ON MACHINE LEARNING METHODS

Markov G. A. 166



Вышла в свет уникальная книга, посвященная наиболее актуальной проблематике в области информационной безопасности – разработке безопасного ПО: **КАК ИЗБЕЖАТЬ ОШИБОК ПРИ БЕЗОПАСНОЙ РАЗРАБОТКЕ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ** / В.В. Вареница, А.С. Марков, В.Л. Цирлов и др. М.: Квант-Медиа, 2025. 344 с. Переплет твердый. ISBN 978-5-6053386-1-1

Книга подготовлена как учебное пособие коллективом авторов учебного центра «Эшелон» под редакцией доктора технических наук Маркова Алексея Сергеевича и группы его сотрудников на основе 20-летнего опыта работы в названной области. Впервые в нашей стране выпущено практическое пособие с учетом выполнения всех требований ГОСТ Р 56939-2024 – «Защита информации. Разработка безопасного программного обеспечения. Общие требования», являющегося национальным стандартом.

Содержание книги рассматривает все актуальные вопросы на всех стадиях жизненного цикла разработки безопасного программного обеспечения, каждый раздел сопровождается проверочными листами, в Приложение вынесены Типичные ошибки при внедрении процессов разработки.

Это учебное пособие станет настольной книгой специалистов, занимающихся вопросами организации и непосредственного внедрения процедур разработки безопасного программного обеспечения – как для молодых слушателей, так и для практикующих профессионалов, которым это пособие сэкономит много времени на поиск решений и существующих требований по безопасности.

CYBERSECURITY ISSUES VOPROSY KIBERBEZOPASNOSTI

№ 6

2025

DOI: 10.21681/2311-3456

Idegradation of access control systems

Vulnerabilities of asymmetric ciphers

Quantum security



**www.cyberrus.info
editor@cyberrus.info**