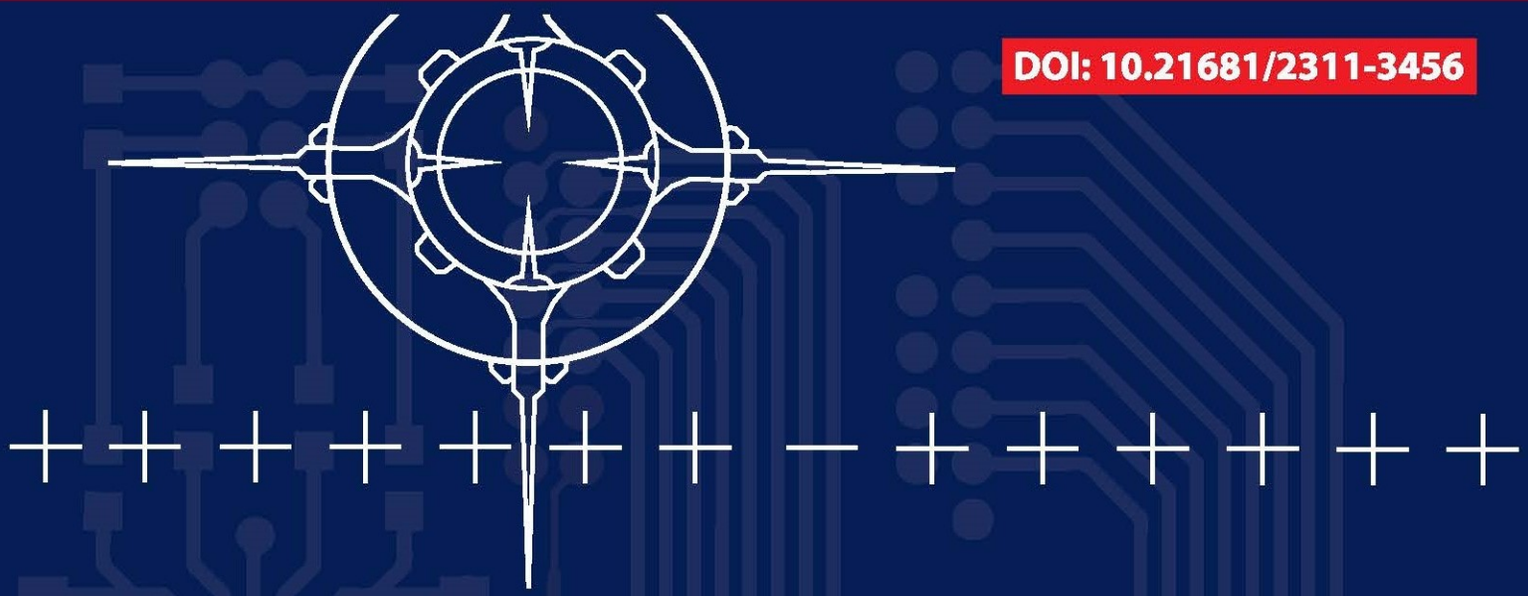
ВОПРОСЫ КИБЕРБЕЗОПАСНОСТИ

№4 2025 (68)

DOI: 10.21681/2311-3456



Методика синтеза квантово-устойчивых блокчейн платформ

Оценка защищенности критической информационной инфраструктуры

Обнаружение фишинговых писем с помощью нейронных сетей

Вышла коллективная монография «Безопасность России. Системная инженерия в проблемах национальной безопасности». Книга вышла при поддержке Совета Безопасности Российской Федерации.

В создании монографии участвовали члены редсовета и редколлегии журнала «Вопросы кибербезопасности»: Гарбук С.В., Марков А.С., Шеремет И.А., Язов Ю.К., а также известные авторы: Климов С.М., Костогрызов А.И., Нистратов А.А., Стрельцов А.А.

Цель настоящей монографии — представление отечественного анализа и демонстрация технологических подходов, методов прогнозирования и системного управления возможностями современной системной инженерии для эффективного применения при решении приоритетных задач обеспечения национальной безопасности. Результаты по всем 12 разделам — небезынтересные, зачастую необычные, скрытые от традиционного взгляда на системы, но принципиальные по существу.

За счет рационального использования предложенных идей, моделей, методов и научно обоснованных технических решений системной инженерии ожидается целенаправленное существенное повышение качества, безопасности, эффективности, снижение или удержание на допустимом уровне рисков и/или уменьшение затрат (в т.ч. непроизводительных) на создание и эксплуатацию в России систем в различных областях их приложения.

ВОПРОСЫ КИБЕРБЕЗОПАСНОСТИ

НАУЧНЫЙ РЕЦЕНЗИРУЕМЫЙ ЖУРНАЛ

№4 (68) 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

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 5.08.2025 г.
Общий тираж 120 экз. Цена свободная

Адрес: 107023, Москва, ул. Электрозаводская, 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/**

СОДЕРЖАНИЕ

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

КЛАСТЕРНАЯ МОДЕЛЬ ЗАЩИТЫ РАСПРЕДЕЛЕННОГО РЕЕСТРА

Сундеев П. В.[1]

DOI: 10.21681/2311-3456-2025-4-2-8

Цель исследования: *разработать модель защиты информации для анализа конструктивной безопасности архитектуры распределенного реестра с учетом политики разграничения доступа и квантовой угрозы.*

Методы исследования: *объектно-ориентированный анализ сложных систем, системный анализ, теория модульно-кластерных сетей, теория графов, теория матриц, математическая логика.*

Результат исследования: *разработана расширенная модель защиты информации с полным перекрытием для систем распределенного реестра с учетом влияния квантовой угрозы, которая позволяет оценивать конструктивную защиту, проводить формальный статический или динамический анализ безопасности архитектуры.*

Научная новизна: *на основе методов теории модульно-кластерных сетей разработана расширенная модель защиты информации с полным перекрытием для анализа конструктивной безопасности распределенного реестра за счет кластерной декомпозиции архитектуры и информационных взаимодействий, учета эффективности средств защиты информации. Показаны системные критерии оценки конструктивной защиты.*

Ключевые слова: *модульно-кластерная сеть, квантовая угроза.*

Введение

При разработке систем распределенного реестра (DLTS) необходим анализ конструктивной безопасности архитектуры [1–3]. Особенностью технологии распределенного реестра (DLT) является конструктивная защита на основе криптографии и децентрализации информационного взаимодействия[2]. Сложная топология распределенного информационного взаимодействия, появление эффективных методов взлома криптографии с неидеальной стойкостью и увеличение скорости вычислений создают риск нарушения безопасности транзакций [4–9]. Для анализа безопасности архитектуры DLTS необходима формальная модель защиты, в которой учтены все существенные свойства, применимы критерии для оценки конструктивной защиты с формальным доказательством безопасности архитектуры, имеется возможность учета политики доступа и эффективности средств защиты информации.

Известна теоретическая модель защиты информации с полным перекрытием, которая строится из предположения о необходимости контроля каждого возможного воздействия по схеме «угроза (v) – защита (d) – объект (o)». Для построения модели необходимо определить множества угроз V, средств D и объектов O защиты, а также взаимосвязи между ними. В развитии модели предлагалось ввести в нее множество уязвимостей $\overline{V}$, определяемого подмножеством декартова произведения $T \times O$, и множество барьеров – путей осуществления угроз безопасности, перекрытых средствами защиты, и определяемого декартовым произведением $\overline{V} \times M$.

Практическое применение теоретической модели для распределенных систем со сложной политикой разграничения доступа и динамичной топологией информационного взаимодействия, которая характерна для систем с «открытой» архитектурой, ограничено достоверностью модели защиты из-за проблем с определением элементов указанных множеств и поиском опасных траекторий информационного процесса, а также отсутствием метода оценки их соответствия декларируемой политике доступа. При этом граф состояний системы может иметь большую размерность. Сложность его анализа может соответствовать классу NP-полных задач, поэтому формальная модель защиты должна обеспечивать редукцию графа состояний системы.

Задача моделирования защиты информации

В формальных моделях доступа информационное взаимодействие субъектов и объектов регулируется правилами политики разграничения доступа, которая обеспечивается конструктивно топологией архитектуры системы и средствами защиты, компенсирующими ее уязвимости. Поэтому модель защиты – это, по сути, статическая или динамическая модель собственно информационной системы с включенными в нее элементами, моделирующими источники внешних и внутренних угроз, а также конструктивные и дополнительные элементы защиты информации.

1 Сундеев Павел Викторович, доктор технических наук, ведущий инженер-исследователь Научного центра информационных технологий и искусственного интеллекта АНОО ВО «Университет «Сириус», Федеральная территория «Сириус», Россия. E-mail: sundeev.pv@talantiuspeh.ru

2 Recommendation ITU-T X.1410 (03/2023), Distributed ledger technology (DLT) security. Security architecture of data sharing management based on the distributed ledger technology.

Пусть задана система распределённого реестра W и ее информационное окружение V. Физические и логические модули M_N^W и $M_{N^*}^V$, которые имеют выделенное функциональное значение при реализации информационного взаимодействия (средства защиты, субъекты и объекты доступа, способные реализовать информационные примитивы), являются элементами множеств W и V. Вместе они составляют множество вершин графа $G(M_{N+N^*},R_M)$, где $N+N^*$ – число вершин графа, и R_M – множество дуг, которые обозначают информационные взаимосвязи между модулями.

Для анализа безопасной архитектуры DLTS формальная модель защиты информации должна включать и обеспечивать:

- разделение субъектов и объектов доступа M_N^W и $M_{N^*}^V$ на непересекающиеся подмножества, для которых установлены разные политики доступа;
- средства ограничения или управления доступом (защиты информации) $D(M_D^W \subset M_N^W)$, которые обеспечивают разделение множества вершин графа G на непересекающиеся подмножества;
- информационные связи R_M между средствами защиты, субъектами и объектами, которые реализуют информационный процесс и существенны для управления доступом к объектам защиты;
- критерии для оценки конструктивной защиты архитектуры DLTS;
- редукцию порождающего графа состояний системы W и ее окружения V для снижения размерности области поиска опасных состояний;
- формальную верификацию безопасности архитектуры DLTS.

Кластерная модель защиты

В расширенной кластерной модели защиты информации с полным перекрытием проблема точности формальной модели решается определением субъектов и объектов доступа в качестве функциональных информационных модулей и декомпозицией информационных взаимодействий между ними на физические (F), синтаксические (L) и семантические (S) отношения с учетом их функциональных свойств, существенных для защиты информации, на основе методов теории модульно-кластерных сетей [11] с последующей оценкой безопасности статичной или динамичной топологии системы и эффективности средств защиты. Граф $G(M_{N+N^*},R_M)$ преобразуется в мультиграф $G^{FLS}(M_{N+N^*},R_M^{FLS})$. В мультиграфе G^{FLS} к множеству дуг R_M^{FLS} относятся только кратные дуги вида $\{r_{ij}^F \cup r_{ij}^L \cup r_{ij}^S\} \subseteq R_M^{FLS}$. Состав дуг мультиграфа определяется наличием входных и выходных интерфейсов модулей, обеспечивающих реализацию информационных примитивов через FLS-отношения, которые имеют иерархическую зависимость вида $r^F \longrightarrow r^L \longrightarrow r^S$. Состав вершин и дуг мультиграфа может меняться при наличии условий для информационного взаимодействия модулей по правилу «если взаимодействие возможно, то оно реализуется». Некратные FLS-дуги включаются в мультиграф при анализе угроз безопасности информации.

Все субъекты и объекты распределяются по кластерам K^W в соответствии с правами доступа, установленными политикой разграничения доступа, с учетом топологии системы. Внешние источники угроз из множества $M_{N^*}^V$ выделяются в отдельные кластеры K^V. Одному кластеру могут принадлежать только «доверенные» субъекты и объекты с одинаковым уровнем доступа. Субъекты доступа из других кластеров рассматриваются как потенциальные источники угроз. Распределение вершин $M_{N^*}^V$ и M_N^W по кластерам, которые являются источниками внешних и внутренних угроз, позволяет использовать модель защиты в качестве модели угроз. В DLTS, которые предназначены для взаимодействия недоверенных субъектов в конкурентной среде, каждый субъект может быть выделен в отдельный кластер. Одинаковые правила доступа для объектов и субъектов позволяют редуцировать граф состояний системы и свести сложность задачи поиска опасных состояний к разрешимости за полиномиальное время.

На рис. 1 представлена расширенная кластерная модель защиты информации с полным перекрытием. В качестве объекта защиты рассматривается кластер K_2^W системы W. К множеству угроз V для кластера K_2^W отнесены внешние нарушители из кластера K_3^V и внутренние субъекты из кластера K_1^W системы W, для которых установлена иная политика доступа.

Возможные информационные взаимодействия R^{FLS} между модулями представлены кратными дугами. Декомпозиция позволяет определить средства (функции) защиты из множества D^{FLS}, которые обеспечивают конструктивную безопасность и реализуют меры защиты на одном или нескольких уровнях FLS-отношений. Например, на синтаксическом уровне L конструктивно реализована криптографическая защита DLT, на F и L уровнях обеспечиваться конструктивная защита с использованием технологии квантового распределения ключей, на S уровне реализуется механизм авторизации.

В формальном описании кластерной модели защиты с полным перекрытием множество вершин M_{N+N^*}, где $N+N^*$ – число всех вершин мультиграфа G^{FLS}, разбивается политикой доступа на кластерные подмножества субъектов и объектов доступа (модулей) K_K ($K = 1, \ldots, k$ – число кластеров), такие что $K_1 \cap \ldots \cap K_k = \emptyset$ и $K_1 \cup \ldots \cup K_k = K_K$, и K_D – кластерные подмножества, состоящие из элементов множества $D^{FLS} = \{d_f^F, d_l^L, d_s^S\}$ средств защиты, и $K_K \cap K^D = \emptyset$.

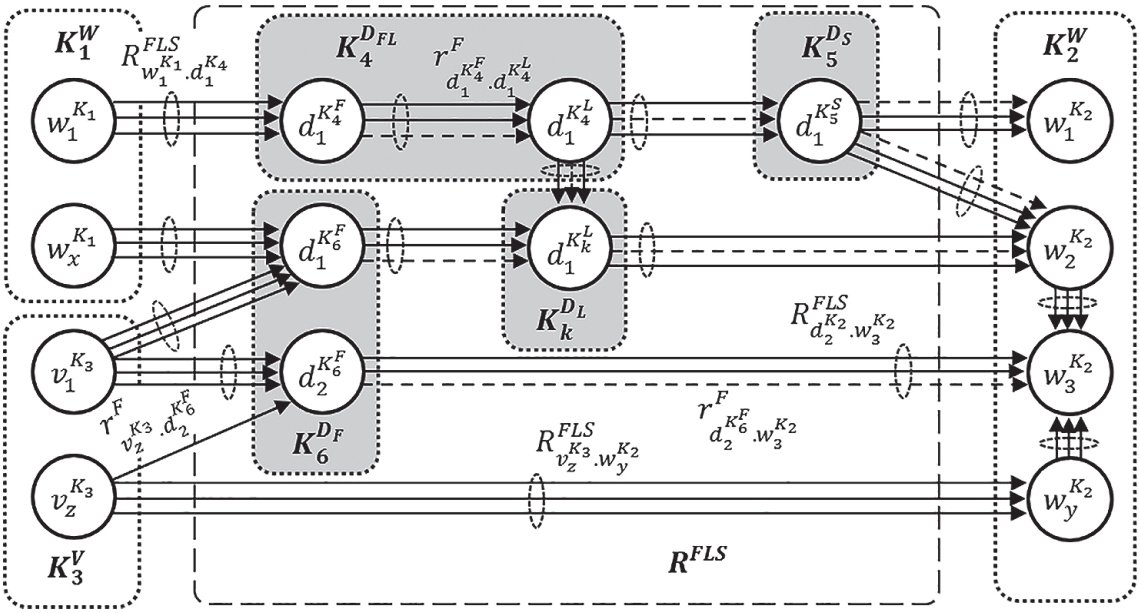


Рис. 1. Кластерная модель защиты информации

Декомпозиция дуг $R \longrightarrow R_M^{FLS}\{R^F \cup R^L \cup R^S\}$ в кластерной модели защиты расширяет понятие «смежности» вершин графа. Если все вершины и дуги составляют мультиграф $G^{FLS}(M_N,R_M^{FLS})$, то любые две его вершины $\{m_i,m_j\}$ являются смежными, только если в остовных FLS-подграфах между этими вершинами существует хотя бы одно подмножество кратных дуг вида $R_{i,j}^{FLS} = \{r_{i,j}^F,r_{i,j}^L,r_{i,j}^S\}$, которое называется полной FLS-дугой. Взаимодействие между любыми двумя модулями возможно, если обозначающие их вершины $\{m_i,m_j\}$ мультиграфа $G^{FLS}(M_N,R_M^{FLS})$ смежные. Путь $P_{i,j}^{FLS}$ между произвольной парой вершин $\{m_i,m_j\}$ в мультиграфе $G^{FLS}(M_N,R_M^{FLS})$ существует, если существуют пути между этими вершинами в FLS-подграфах смежности.

На рис. 1 полные дуги обозначены пунктирными овалами. Для примера показана неполная дуга вида $r^F_{v_z^{K_3}.d_2^{K_6^F}}$, которая указывает на наличие физического отношения, которое не позволяет реализовать информационное взаимодействие из-за отсутствия кратных дуг $r^L_{v_z^{K_3}.d_2^{K_6^F}}$ и $r^S_{v_z^{K_3}.d_2^{K_6^F}}$ в остовных L и S подграфах. Управляемые FLS средствами защиты кластерные ограничения показаны пунктирными дугами, которые появляются в составе полной дуги при реализации доступа соответственно на F, L или S уровне взаимодействия модулей. Таким образом, кластерная модель защиты представляет собой мультиграф, вершины которого соединяются кратными FLS-дугами, внутрикластерная связность и межкластерная разряженность максимальны. Возможны вырожденные варианты, когда внутрикластерная связность отсутствует и вершины становятся кластерами или когда все вершины являются элементами одного кластера. Некратные дуги включаются в модель для анализа «скрытых» угроз.

Анализ кластерной модели защиты

Цель анализа конструктивной защиты архитектуры DLTS – поиск траекторий информационного процесса, которые содержат полные и не полные дуги допускаемые топологией архитектуры, но не контролируемые средствами защиты, а также оценка минимального уровня эффективности защиты для траекторий информационного процесса. Неконтролируемые средствами защиты информационные взаимодействия между кластерами отображаются в модели полной дугой вида $R^{FLS}_{v_z^{K_3}.w_y^{K_2}}$ (рис. 1). При формальном анализе кластерной модели защиты проводится поиск полных FLS-дуг входящих в или выходящих из защищаемых кластеров, но не инцидентных вершинам множества средств защиты D^{FLS} из кластеров $K^F \cup K^L \cup K^S \subseteq K^D$. При анализе «скрытых» угроз дополнительно проводится поиск не полных F, L и S дуг.

Состояния системы являются результатом взаимодействия модулей на трех уровнях, поэтому необходимо генерировать согласованные FLS-матрицы смежности для каждого уровня информационного взаимодействия. Уровни взаимодействия представляются отдельными FLS-матрицами смежности, у которых строки и столбцы проиндексированы номерами вершин. Наличие значений отличных от «0» в одинаковых позициях квадратных FLS-матриц указывает на то, что вершины, номерами которых проиндексированы строки и столбцы, являются смежными.

В ходе анализа проводится поиск состояний системы, нарушающих политику доступа, и оценка

непрерывности уровня защиты для каждого пути поиск вершин с весами ниже установленного значения. При статическом анализе проверяется достижимость вершин мультиграфа G^{FLS}, что позволяет оценить безопасность конкретной топологии DLTS. Динамический анализ позволяет оценить безопасность состояний методом перебора траекторий информационного процесса при изменении состава вершин и дуг мультиграфа G^{FLS} в результате применения решающих правил управляемого логического вывода.

Результаты проверки достижимости модулей отражаются в квадратной матрице достижимости вершин $B^{Д} = \|b_{ij}\|$ мультиграфа G^{FLS}, элементы которой заполняются по правилу

$$b_{ij} = \begin{cases} 1, \text{ если из вершины } i \text{ к вершине } j \text{ имеется путь } P_{ij}^{FLS}; \\ 0, \text{ если из вершины } i \text{ к вершине } j \text{ путь } P_{ij}^{FLS} \text{ отсутствует.} \end{cases} \quad (1)$$

Критерием безопасности архитектуры является отсутствие пути P_{ij}^{FLS} между любыми произвольными вершинами из разных кластерных подмножеств K_K, который не содержит хотя бы одну вершину из множества D^{FLS}. Отсутствие пути проверяется сравнением значений каждой позиции кластерной матрицы $B^{К} = \|b_{ij}\|$ сформированной по правилам политики доступа и правилам конструктивной защиты (см. утверждения 1 и 2) с позицией в матрице достижимости $B^{Д} = \|b_{ij}\|$, которая формируется при анализе топологии.

Если при сравнении мощности множеств по теореме Кантора-Бернштейна мощность множества $|B^{К}|$ ненулевых элементов матрицы $B^{К} = \|b_{ij}\|$ равномощно или больше мощности множества $|B^{Д}|$ ненулевых элементов матрицы $B^{Д} = \|b_{ij}\|$, то формальные правила доступа выполняются и архитектура системы безопасна. Для национальных распределенных реестров и платформ может требоваться более высокий уровень защиты, когда любое взаимодействие проходит контроль доступа. В этом случае все полные FLS-дуги между любыми вершинами подмножеств K^W должны быть инциденты вершине из подмножеств K^D множества D^{FLS}. Все вершины становятся кластерами, реализуется политика с «нулевым доверием» и каждое взаимодействие проходит через контроль доступа. Полные дуги между вершинами одного кластера вида $R^{FLS}_{v_z^{K_3}, w_y^{K_2}}$ (рис. 1) запрещены, все дуги должны быть инцидентны вершинам, обозначающим средства защиты из множества D^{FLS}. Из этого следует определение конструктивно безопасной архитектуры информационной системы.

Утверждение 1. Если все внешние дуги кластеров K^W инциденты вершинам из кластеров K^D средств защиты множества D^{FLS}, то конструктивно архитектура системы безопасна.

Для критических систем актуален более сильный критерий безопасности, который соответствует политике «нулевого доверия», учитывает внутренние угрозы и угрозы распределенной топологии.

Утверждение 2. Если все дуги кластеров K^W инциденты вершинам из кластеров K^D средств защиты множества D^{FLS}, то конструктивно архитектура системы безопасна.

Доказательство безопасности архитектуры DLTS при динамическом анализе обеспечивается управляемым перебором состояний в ходе построения FLS-мультиграфа и оценкой на каждом шаге безопасности порожденного состояния. Оценка безопасности состояния заключается в установлении всех возможных отношений между модулями, которые изменялись на последнем шаге, и проверке их принадлежности подмножеству разрешенных кластерных отношений для этих модулей. Если отношения разрешены (присутствуют в кластерной FLS-модели), то состояние безопасное. Соответственно, если отношения запрещены (отсутствуют в кластерной FLS-модели), то состояние опасное. Строгость доказательства соответствует строгости математического аппарата логического вывода.

Кластерная модель позволяет использовать системные критерии для оценки конструктивной защиты архитектуры DLTS. На рис. 2 пример а) демонстрирует неконтролируемое взаимодействие $R^{FLS}_{v,w}$ модулей v_z и w_y из кластеров K_3 и K_2, что оценивается как угроза безопасности. В примере б) показано контролируемое средством защиты физического уровня d^F из подмножества D^{FLS} взаимодействие модулей v_z и w_y из кластеров K_3 и K_2. В примере в) показано взаимодействие модулей v_z и w_y из кластеров K_3 и K_2 контролируемое двумя средствами защиты синтаксического d^L и семантического d^S уровней. Пример г) демонстрирует контролируемое взаимодействие модулей w_y и w_{y^*} из одного кластера K_2 через средство защиты синтаксического уровня d^L. Контроль взаимодействия внутри кластера актуален для критических систем, например, для национальных распределенных блокчейн систем и платформ.

В общем случае при оценке конструктивной защиты архитектуры кластерная модель защиты позволяет учитывать топологию информационного взаимодействия на одном, двух или трех FLS-уровнях, а также эффективность системы защиты на основе сравнения весовых коэффициентов средств защиты.

Для учета эффективности средств защиты при анализе безопасности архитектуры вершинам графа из множества D^{FLS} присваиваются нормированные весовые коэффициенты, характеризующие надежность защиты на основе внешних оценок

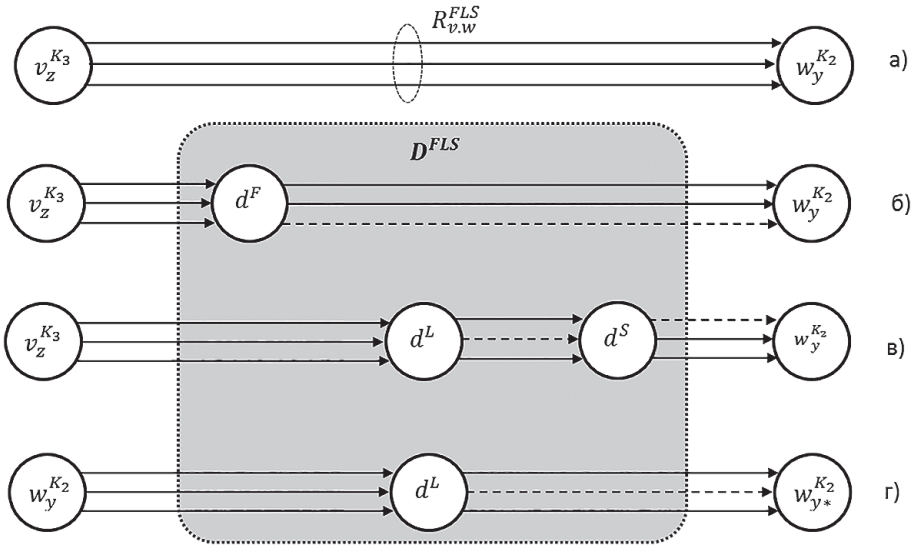


Рис. 2. Конструктивная защита архитектуры

$$D^{FLS} = |d_1^{(k)}, d_2^{(k)}, \ldots, d_n^{(k)}|. \quad (2)$$

Каждой d-ой вершине приписывается вес $d_n^{(k)}$, где k – функция веса.

Непрерывность уровня защиты оценивается сравнением значений весовых коэффициентов вершин графа для каждого пути относительно заданного нормированного значения. Например, для криптографических функций шифрования и аутентификации, которые являются основой конструктивной защиты DLTS, задается оценка стойкости криптографии к квантовой угрозе, связанной с появлением эффективных «квантовых» алгоритмов решения сложных вычислительных задач, позволяющих взломать асимметричную криптографию, и риском достижения «квантового превосходства» в скорости вычислений. Значение веса у вершины ниже заданного уровня указывает на угрозу безопасности.

Выводы

Кластерная модель защиты с полным перекрытием является инструментом анализа безопасности архитектуры DLTS со сложной архитектурой и политикой разграничения доступа. Модель позволяет оценивать безопасность архитектуры относительно декларируемой политики доступа, системных правил конструктивной защиты и оценок надежности средств защиты. Оценка конструктивной защиты на основе кластерной модели защиты позволяет принимать обоснованные решения при анализе и синтезе безопасной архитектуры DLTS. Криптографическая защита является основой конструктивной защиты DLT, для которой актуальна квантовая угроза. Кластерная модель защиты позволяет учитывать риск квантовой угрозы для DLTS при оценке эффективности средств защиты на основе внешних оценок стойкости криптографических алгоритмов [4–9].

Для проведения динамического анализа архитектуры с формальным доказательством ее безопасности кластерная модель защиты позволяет редуцировать граф состояний исследуемой системы за счет объектно-ориентированной декомпозиции системы на типовые функциональные модули методами теории модульно-кластерных сетей, что позволяет автоматизировать поиск опасных состояний без потери достоверности модели. Эвристики значительно сокращают размерность пространства поиска состояний. В реальных системах значительная часть (от 47 до 89 %) информационных объектов имеют однотипную функциональность и политику доступа, могут быть представлены в виде классов объектно-ориентированной модели. Эксперименты по моделированию показали возможность сокращения размерности порождающего графа состояний системы на 73 %, например, в случае «классической» архитектуры DLT, где все пользователи являются недоверенными субъектами и для них декларируется одинаковая политика доступа. Нижняя оценка мощности (3) определяется выбором только наилучших вариантов подграфов второго порядка в каждой частной резольвенте, что соответствует решению проблемы методами динамического программирования

$$N\_min = L \cdot R \cdot M^2 \cdot \frac{n^2}{t}, \quad (3)$$

где R - множество t-арных эвристических отношений над элементами порождающего графа G; M – мощность множества R; P - вектор параметров вершин графа с длиной L, равной количеству независимых переменных; n – мощность множества вершин порождающего графа G.

Таким образом:

1. Кластерная модель защиты информации с полным перекрытием, построенная с применением методов теории МК-сетей, позволяет проводить формальный анализ и оценивать безопасность архитектуры систем распределенного реестра.
2. Конструктивная безопасность архитектуры систем распределенного реестра может оцениваться по отсутствию внешних дуг у защищаемых кластеров, которые не инцидентны вершинам из кластера средств защиты информации. Для критических приложений условие безопасности может заключаться в оценке инцидентности всех дуг вершинам, обозначающим средства защиты информации.
3. Оценка эффективности средств защиты информации может заключаться в сравнении весов вершин, рассчитанных по внешним методикам. В частности, для систем распределенного реестра необходимо определять стойкость криптографических алгоритмов, используемых при аутентификации и шифровании.
4. Кластерную модель защиты информации можно использовать для моделирования угроз безопасности информации с учетом особенностей политики доступа, архитектуры и квантовой угрозы.

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. Марков А. С. Кибербезопасность и информационная безопасность как бифуркация номенклатуры научных специальностей // Вопросы кибербезопасности. 2022. № 1(47). С. 2–9. DOI: 10.21681/2311-3456-2022-1-2-9.
2. Topical issues in the implementation of secure software development processes Markov A. S., Varenitca V. V., Arustamyan S. S. В сборнике: Proceedings of the International Conference on Information Processes and Systems Development and Quality Assurance. IPSQDA-2023. 2023. С. 48–53.
3. Ищукова Е. А. О влиянии криптографической стойкости функций хеширования на устойчивость современных блокчейн-экосистем и платформ // Вопросы кибербезопасности. 2025. № 3(67), с. 63–71. DOI: 10.21681/2311-3456-2025-3-63-71.
4. Балябин А. А., Петренко С. А. Модель блокчейн-платформы с кибериммунитетом в условиях квантовых атак // Вопросы кибербезопасности. 2025. № 3(67). С. 72–82. DOI: 10.21681/2311-3456-2025-3-72-82.
5. Petrenko A. S., Petrenko S. A. Basic Algorithms Quantum Cryptanalysis // Вопросы кибербезопасности. 2023. No. 1(53). P. 100–115. DOI: 10.21681/2311-3456-2023-1-100-115.
6. Petrenko A. S. Applied Quantum Cryptanalysis (scientific monograph). River Publishers. (2023). 256 p. ISBN 9788770227933. DOI: 10.1201/9781003392873.
7. Mark Webber, Vincent Elfving, Sebastian Weidt, Winfried K. Hensinger. The impact of hardware specifications on reaching quantum advantage in the fault tolerant regime. AVS Quantum Sci. 4, 013801 (2022). DOI: 10.1116/5.0073075.
8. Battarbee C., Kahrobaei D., Perret L., Shahandashti S. F. SPDH-Sign: Towards Efficient, Post-quantum Group-Based Signatures. In: Johansson, T., Smith-Tone, D. (eds) Post-Quantum Cryptography. PQCrypto 2023. (2023). Lecture Notes in Computer Science. V. 14154. P. 113–138. Springer, Cham. DOI: 10.1007/978-3-031-40003-2_5.
9. Li L., Lu X., Wang K. Hash-based signature revisited. (2022). Cybersecurity. V. 5. Article No. 13. DOI:10.1186/s42400-022-00117-w.
10. Сундеев П. В. Функциональная стабильность распределенного реестра в условиях появления новой квантовой угрозы // Вопросы кибербезопасности. 2025. № 3(67). С. 83–89. DOI: 10.21681/2311-3456-2025-3-83-89.

CLUSTER MODEL OF DISTRIBUTED REGISTRY PROTECTION

Sundeev Pavel[3]

Keywords: *modular cluster network, quantum threat.*

The purpose of the study: *is to develop an information protection model for analyzing the constructive security of the distributed registry architecture, taking into account the access control policy and the quantum threat.*

Research methods: *object-oriented analysis of complex systems, system analysis, theory of modular cluster networks, graph theory, matrix theory, mathematical logic.*

Research result: *an extended information protection model with full overlap for distributed registry systems has been developed, taking into account the influence of the quantum threat, which allows evaluating constructive protection and conducting formal static or dynamic security analysis of the architecture.*

3 Pavel Sundeev, Dr.Sc. (Technical), Chief researcher of Scientific Center of Information Technologies and Artificial Intelligence of Sirius University of Science and Technology, Sirius Federal Territory Sirius University of Science and Technology. E-mail: sundeev.pv@talantiuspeh.ru

Scientific novelty: *based on the methods of the theory of modular cluster networks, an extended information protection model with full overlap has been developed to analyze the constructive security of a distributed registry due to the cluster decomposition of architecture and information interactions, taking into account the effectiveness of information security tools. The system criteria for evaluating constructive protection are shown.*

The results were obtained *with the financial support of the project «Technologies for countering previously unknown quantum cyber threats», implemented within the framework of the state program of the «Sirius» Federal Territory «Scientific and technological development of the «Sirius» Federal Territory (Agreement No. 23-03 dated September 27, 2024).*

References

1. Markov A. S. Cybersecurity and Information Security as Nomenclature Bifurcation Scientific Specialties. (2022). Voprosy Kiberbezopasnosti [Cybersecurity issue]. № 1(47). P. 2–9 (Russian Text).
2. Topical issues in the implementation of secure software development processes Markov A. S., Varenitca V. V., Arustamyan S. S. In the collection: Proceedings of the International Conference on Information Processes and Systems Development and Quality Assurance. (2023). IPSQDA-2023. P. 48–53.
3. Ishchukova E. A. On the influence of cryptographic stability of hashing functions on the stability of modern blockchain ecosystems and platforms. (2025). Voprosy Kiberbezopasnosti [Cybersecurity issue]. № 3(67), c. 63–71. DOI: 10.21681/2311-3456-2025-3-63-71 (Russian Text).
4. Balyabin A. A., Petrenko S. A. Model of a blockchain platform with cyber-immunity under quantum attacks. (2025). Voprosy Kiberbezopasnosti [Cybersecurity issue]. № 3(67). P. 72–82. DOI: 10.21681/2311-3456-2025-3-72-82 (Russian Text).
5. Petrenko A. S., Petrenko S. A. Basic Algorithms Quantum Cryptanalysis. Voprosy Kiberbezopasnosti [Cybersecurity issue]. (2023). no. 1(53), pp. 100–115. DOI: 10.21681/2311-3456-2023-1-100-115 (Russian Text).
6. Petrenko A. S. Applied Quantum Cryptanalysis (scientific monograph). River Publishers. (2023). 256 p. ISBN 9788770227933. DOI: 10.1201/9781003392873.
7. Mark Webber, Vincent Elfving, Sebastian Weidt, Winfried K. Hensinger. The impact of hardware specifications on reaching quantum advantage in the fault tolerant regime. AVS Quantum Sci. 4, 013801 (2022). DOI: 10.1116/5.0073075.
8. Battarbee C., Kahrobaei D., Perret L., Shahandashti S. F. SPDH-Sign: Towards Efficient, Post-quantum Group-Based Signatures. In: Johansson, T., Smith-Tone, D. (eds) Post-Quantum Cryptography. PQCrypto 2023. (2023). Lecture Notes in Computer Science. V. 14154. P. 113–138. Springer, Cham. DOI: 10.1007/978-3-031-40003-2_5.
9. Li L., Lu X., Wang K. Hash-based signature revisited. (2022). Cybersecurity. V. 5. Article no. 13. DOI:10.1186/s42400-022-00117-w.
10. Sundeev P. V. Functional stability of a distributed registry in the context of the emergence of a new quantum threat. (2025). Cybersecurity issue. № 3(67). P. 83–89. DOI: 10.21681/2311-3456-2025-3-83-89 (Russian Text).

УЯЗВИМОСТИ GCC И LLVM К АТАКАМ НА КОНВЕЙЕР ОПТИМИЗАЦИИ

Муравьёв С. К.[1]

DOI: 10.21681/2311-3456-2025-4-9-16

Цель исследования: *выработка рекомендаций по реализации средств разработки безопасного программного обеспечения (РБПО) и внедрению процессов РБПО на основе анализа угроз безопасности информации при разработке программного обеспечения, связанных с возможностью создания и использования злоумышленниками модулей расширения для оптимизирующих компиляторов.*

Метод(ы) исследования: *к основным методам исследования относятся анализ и синтез, моделирование и эксперимент.*

Результат(ы) исследования: *на основании анализа действующих нормативных требований к разработчикам безопасного программного обеспечения в части оценки угроз безопасности информации со стороны средств разработки программного обеспечения определены актуальность, цель и предмет исследования.*

Рассмотрены особенности анализа и трансформации исходного кода оптимизирующими компиляторами GCC и LLVM в процессе оптимизации исходного кода. Показана уязвимость таких компиляторов к вредоносным вмешательствам в конвейер оптимизации, которые могут быть осуществлены злоумышленниками через штатные программные интерфейсы, предусмотренные для повышения функциональности таких компиляторов и эффективности разрабатываемого с их помощью программного обеспечения.

Продемонстрирована возможность практической реализации атак, которые меняют конвейер оптимизации таким образом, что алгоритм функционирования целевого приложения принципиально меняется требуемым злоумышленнику образом. При этом такие атаки не связаны с нарушением целостности и конфиденциальности исходного кода целевого приложения и исполняемых файлов средств разработки программного обеспечения.

Проанализированы сложности обнаружения подобных угроз и определены способы их нейтрализации. В итоге выработаны рекомендации, которые могут быть использованы при проектировании и реализации безопасных компиляторов и безопасного программного обеспечения, а также при внедрении соответствующих процессов РБПО.

Научная новизна: *показана необходимость расширения нормативных требований к безопасным компиляторам языков С/С++ и мер по разработке безопасного программного обеспечения, в части необходимости контроля использования модулей расширения для применяемых инструментальных средств.*

Ключевые слова: *средства разработки, программное обеспечение, оптимизирующий компилятор, угроза безопасности информации, РБПО.*

Введение

Разработке безопасного программного обеспечения (РБПО) в настоящее время уделяется значительное внимание [1], что находит своё отражение в активном развитии соответствующей нормативно-правовой[2] и нормативно-технической документации, определяющей общие требования к содержанию и порядку выполнения работ, связанных с РБПО[3], угрозы безопасности информации при разработке программного обеспечения (ПО)[4], а также общие требования к безопасным компиляторам языков С/С++[5]. При этом, среди прочего, данные стандарты определяют актуальность исследования угроз безопасности информации, возникающих в ходе разработки ПО со стороны средств разработки.

В России в качестве основы для реализации инструментальных средств разработки на языках программирования С и С++, реально можно рассматривать лишь экосистемы GCC и LLVM, включающие в свой состав современные функциональные оптимизирующие компиляторы. Алгоритм работы таких компиляторов может стать причиной внедрения в разрабатываемое ПО уязвимостей, которые могут создавать серьёзные угрозы безопасности для созданного с их помощью программного обеспечения. Одна из особенностей алгоритмов их работы заключается в возможности произвольного изменения процессов оптимизации исходного кода с помощью внешних модулей, подключаемых через штатные

1 Муравьёв Сергей Константинович, кандидат технических наук, начальник отдела ООО НТП «Криптософт». Пенза, Россия. E-mail: smurav@mail.ru

2 Приказ ФСТЭК России от 01 декабря 2023 г. № 240 «Об утверждении Порядка проведения сертификации процессов безопасной разработки программного обеспечения средств защиты информации».

3 ГОСТ Р 56939 – 2024. Защита информации. Разработка безопасного программного обеспечения. Общие требования.

4 ГОСТ Р 58412 – 2019. Защита информации. Разработка безопасного программного обеспечения. Угрозы безопасности информации при разработке программного обеспечения.

5 ГОСТ Р 71206 – 2024. Защита информации. Разработка безопасного программного обеспечения. Безопасный компилятор языков С/С++. Общие требования.

программные интерфейсы [2], доступ к которым должен быть ограничен для нарушителя [3,4]. Таким образом, исследования уязвимостей GCC и LLVM к атакам на конвейер оптимизации, которым и посвящено данное исследование, обладают несомненной актуальностью.

Особенности работы компиляторов GCC и LLVM

Оптимизирующие компиляторы, входящие в состав GCC и LLVM, применяют к исходному коду множество различных операций для анализа и трансформации, называемые проходами. Компиляторы предоставляют сотни готовых проходов для анализа и трансформации исходного кода, которые в каждом из компиляторов называются по-своему и имеют различное назначение. При этом в зависимости от типа и версии компилятора эти проходы объединяются в несколько типовых конвейеров оптимизации, которые также могут содержать сотни различных проходов, как показано на (рис. 1) и (рис. 2), которые дополнительно можно изменить или расширить с помощью модулей расширения, подключаемых через штатные программные интерфейсы.

Большинство высококвалифицированных разработчиков, повседневно применяющих данные конвейеры, даже не представляют назначение отдельных операций и не контролируют фактический состав конвейера оптимизации при каждом запуске процесса компиляции. При наличии небезопасных программных интерфейсов для изменения конвейеров оптимизации и отсутствии эффективных средств контроля их безопасности создаются условия для проведения вредоносных вмешательств в процесс компиляции исходного кода.

```
g++-14 -O2 rnd.cpp -o rnd -fdump-passes
```

```
*warn_unused_result                    : ON
*diagnose_omp_blocks                   : OFF
*diagnose_tm_blocks                    : OFF
tree-omp_oacc_kernels_decompose        : OFF
tree-omplower                          : ON
tree-lower                             : ON
tree-tmlower                           : OFF
tree-ehopt                             : ON
tree-eh                                : ON
tree-coro-lower-builtins               : OFF
tree-cfg                               : ON
*warn_function_return                  : ON
tree-coro-early-expand-ifns            : OFF
tree-ompexp                            : ON
*build_cgraph_edges                    : ON
*free_lang_data                        : ON
ipa-visibility                         : ON
скрыто 360 строк!
```

Рис. 1. Пример типового конвейера оптимизации GCC

На примере простейшего тестового приложения, написанного с использованием языка программирования C++, которое формирует несколько числовых последовательностей с помощью различных стандартных генераторов случайных чисел [5], можно продемонстрировать возможность создания вредоносных подключаемых модулей для компиляторов из состава GCC и LLVM, которые способны изменить процессы инициализации генераторов случайных чисел таким образом, что злоумышленник сможет

```
opt rnd.ll -passes="default<O2>" -o /dev/null -print-pipeline-passes
```

```
annotation2metadata,forceattrs,inferattrs,coro-early,function<eager-inv>(ee-instrument<>,lower-expect,simplifycfg<bonus-inst-threshold=1;no-forward-switch-cond;no-switch-range-to-icmp;no-switch-to-lookup;keep-loops;no-hoist-common-insts;no-sink-common-insts;speculate-blocks;simplify-cond-branch;no-speculate-unpredictables>,sroa<modify-cfg>,early-cse<>),openmp-opt,ipsccp,called-value-propagation,globalopt,function<eager-inv>(mem2reg,instcombine<max-iterations=1;no-use-loop-info;no-verify-fixpoint>,simplifycfg<bonus-inst-threshold=1;no-forward-switch-cond;switch-range-to-icmp;no-switch-to-lookup;keep-loops;no-hoist-common-insts;no-sink-common-insts;speculate-blocks;simplify-cond-branch;no-speculate-unpredictables>),always-inline,require<globals-aa>,function(invalidate<aa>),require<profile-summary>,cgscc(devirt<4>(inline,function-attrs<skip-non-recursive-function-attrs>,openmp-opt-cgscc,function<eager-inv;no-rerun>(sroa<modify-cfg>,early-cse<memssa>,speculative-execution<only-if-divergent-target>,jump-threading,correlated-propagation,simplifycfg<bonus-inst-threshold=1;no-forward-switch-cond;switch-range-to-icmp;no-switch-to-lookup;keep-loops;no-hoist-common-insts;no-sink-common-insts;speculate-blocks;simplify-cond-branch;no-speculate-unpredictables>,instcombine<max-iterations=1;no-use-loop-info;no-verify-fixpoint>,aggressive-instcombine,libcalls-shrinkwrap,tailcallelim,simplifycfg<bonus-inst-threshold=1;no-forward-switch-cond;switch-range-to-icmp;no-switch-to-lookup;keep-loops;no-hoist-common-insts;no-sink-common-insts;speculate-blocks;simplify-cond-branch;no-speculate-unpredictables>,reassociate,constraint-elimination,loop-mssa(loop-instsimplify,loop-simplifycfg,licm<no-allowspeculation>,loop-rotate<header-duplication;no-prepare-for-lto>,licm<allowspeculation>,simple-loop-unswitch<no-nontrivial;trivial>),simplifycfg<bonus-inst-threshold=1;no-forward-switch-cond;switch-range-to-icmp;no-switch-to-lookup;keep-loops;no-hoist-common-insts;no-sink-common-insts;speculate-blocks;simplify-cond-branch;no-speculate-unpredictables>,instcombine<max-iterations=1;no-use-loop-info;no-verify-fixpoint>,loop(loop-idiom,indvars,simple-loop-unswitch<no-nontrivial;trivial>,loop-deletion, ....
скрыто 100 элементов!
```

Рис. 2. Пример типового конвейера оптимизации LLVM

полностью предсказать результаты работы такого приложения. При этом такая атака не связана с эксплуатацией ненадёжности рассматриваемых генераторов псевдослучайных чисел [6], нарушением конфиденциальности и целостности исходного кода, а также целостности исполняемых файлов инструментальных средств разработки.

Ключевой участок исходного кода тестового приложения, который отвечает за получение набора случайных чисел с равномерным распределением, приведен на (рис. 3). Генератор случайных чисел инициализируется случайным значением перед получением каждого следующего числа. Приложение исполняет ключевой участок кода для десяти различных генераторов случайных чисел, входящих в состав стандартной библиотеки языка C++. Результат исполнения тестового приложения приведён на (рис. 4).

rnd.cpp

```
16  EngineType engine{};
17  static random_device rd{};
18  static uniform_int_distribution val{1, 9};
19
20  for(int j{0}; j < 10; ++j) {
21      engine.seed(rd());
22      cout << val(engine) << " ";
23  }
24  cout << endl;
```

Рис. 3. Ключевой участок исходного кода тестового приложения

./rnd

```
8878213884
5251249699
9589477566
3594632529
1796481463
5872868523
6937872589
6786366816
8221184864
9324776171
```

Рис. 4. Результат исполнения тестового приложения

При наличии доступа к исходному коду злоумышленник может попытаться исследовать алгоритм работы приложения. Но реализация данной угрозы не позволит раскрыть реальные числовые последовательности из-за применения случайных чисел для инициализации генераторов.

Злоумышленник также может попытаться нарушить целостность исходного кода приложения и внедрить в него вредоносные правки или добиться нужного поведения приложения за счёт внедрения изменений непосредственно в компилятор или другие инструментальные средства, используемые при разработке. К счастью, в настоящее время существует довольно много различных средств защиты, которые позволяют оперативно обнаруживать факты нарушения целостности как исходного кода, так и исполняемых файлов, а также эффективно противодействовать подобным угрозам.

Однако, существует возможность проведения атаки, которая не связана с нарушением конфиденциальности и целостности исходного кода, а также целостности исполняемых файлов компилятора и других инструментальных средств разработки [7]. Подобная атака может быть реализована за счёт использования штатных интерфейсов компилятора для встраивания проходов оптимизации, которые способны изменить исходный алгоритм программы.

Перед применением конвейера оптимизации современные компиляторы выполняют преобразование исходного кода в промежуточное представление, в котором исходный код разбивается на базовые блоки, содержащие простые последовательности инструкций. При этом все переменные, в том числе и виртуальные, представляются в форме с единственным статическим присваиванием (SSA) [8].

Подключаемые расширения для компилятора имеют возможность целенаправленного поиска инструкций в промежуточном представлении, их удаления, изменения параметров или даже их замены на альтернативные инструкции, что может принципиально изменить исходный алгоритм работы компилируемого приложения.

Промежуточное представление GCC [9] напоминает упрощённый язык C, в котором раскрыты все высокоуровневые структуры данных и комплексные инструкции для управления потоком исполнения. На (рис. 5) представлен ключевой участок промежуточного представления исходного кода тестового приложения, сформированного GCC.

Промежуточное представление LLVM [10] больше напоминает язык Ассемблер и его значительно сложнее читать, т.к. вместо человекочитаемых развёрнутых сигнатур, используются декорированные (англ. mangled) имена. В данном представлении имена функций начинаются с символа @, а имена переменных заменяются на числа, перед которыми ставится символ %.

На (рис. 6) представлен ключевой участок промежуточного представления исходного кода тестового приложения, сформированного LLVM. Полный вариант промежуточного представления не приводится из-за его существенного объёма.

DOI: 10.21681/2311-3456-2025-4-9-16

```
g++-14 -O1 -fdump-tree-gimple rnd.cpp -o rnd...
cat ./rnd.cpp.005t.original
```

```
1112    {
1113    int j;
1115    j = 0;
1116    goto <D.89377>;
1117    <D.89376>:
1118    _5 = std::random_device::operator() (&rd);
1119    _6 = (long unsigned int) _5;
1120    std::linear_congruential_engine<long unsigned int, 16807, 0, 2147483647>::seed (&engine, _6);
1121    _7 = std::uniform_int_distribution<int>::operator()<std::linear_congruential_engine<...> > (&val, &engine);
1122    _8 = std::basic_ostream<char>::operator<< (&cout, _7);
1123    std::operator<< <std::char_traits<char> > (_8, " ");
1124    j = j + 1;
1125    <D.89377>:
1126    if (j <= 9) goto <D.89376>; else goto <D.89374>;
1127    <D.89374>:
1128  }
```

Рис. 5. Промежуточное представление GCC ключевого участка исходного кода

```
clang -S -emit-llvm -O1 rnd.cpp -o rnd.ll ...
cat ./rnd.ll
```

```
1085  define linkonce_odr dso_local void @_ZNSt26linear_congruential_... {
1086    %3 = alloca ptr, align 8
1087    %4 = alloca i32, align 4
1088    store ptr %0, ptr %3, align 8
1089    store i32 %1, ptr %4, align 4
1090    %5 = load ptr, ptr %3, align 8
1091    %6 = load i32, ptr %4, align 4
1092    call void @_ZNSt26...(ptr noundef nonnull align 4 1086 dereferenceable(4) %5, i32 noundef %6)
1093    ret void
1094  }
```

Рис. 6. Промежуточное представление LLVM части ключевого участка исходного кода

```
gcc/seedpass.cpp
```

```
45    FOR_ALL_BB_FN(bb, fun)
46    {
47            gimple_bb_info *bb_info = &bb->il.gimple;
50            for (gimple_stmt_iterator gsi = gsi_start(bb_info->seq); !gsi_end_p(gsi); gsi_next(&gsi)) {
52                    gimple *g = gsi_stmt(gsi);
53                    if (GIMPLE_CALL == g->code) {
55                            tree fn_call = gimple_call_fn(g);
56                            tree fn_decl = TREE_OPERAND(fn_call, 0);
57                            tree fn_id = DECL_NAME(fn_decl);
58                            const char *fn_name = IDENTIFIER_POINTER(fn_id);
59                            if (0 == strcmp("seed", fn_name)) {
62                                    tree mal_val = build_int_cst_type(integer_type_node, 1234567);
63                                    gimple_call_set_arg(g, 1, mal_val);
64                            }
65                    }
66            }
67    }
```

Рис. 7. Ключевой участок исходного кода расширения GCC

llvm/seedpass.cpp

```
71    for (auto &BB : F) {
72        for (auto &I : BB) {
73            CallInst *CI = dyn_cast<CallInst>(&I);
74            if (nullptr == CI)
75                continue;
77            Function *CF = CI->getCalledFunction();
78            if (nullptr == CF)
79                continue;
81            std::string FN = demangle(CF->getName());
82            if (FN.ends_with("::seed(unsigned long)")) {
84                Value *V = CI->getOperand(1);
85                Value *NV =  ConstantInt::get(V->getType(), 1234567, false);
86                CI->setArgOperand(1, NV);
87            }
88        }
89    }
```

Рис. 8. Ключевой участок исходного кода расширения LLVM

```
clang -O1 -fpass-plugin=./plugins/llvm_seedpass.so rnd.cpp ...
```

```
...
_ZNSt26linear_congruential_engineImLm48271ELm0ELm2147483647EE4seedEm => std::linear_congruential_engine<unsigned
long, 48271ul, 0ul, 2147483647ul>::seed(unsigned long)

_ZNSt23mersenne_twister_engineImLm32ELm624ELm397ELm31ELm2567483615ELm11ELm4294967295ELm7ELm2636928640EL
m15ELm4022730752ELm18ELm1812433253EE4seedEm => std::mersenne_twister_engine<unsigned long, 32ul, 624ul, 397ul, 31ul,
2567483615ul, 11ul, 4294967295ul, 7ul, 2636928640ul, 15ul, 4022730752ul, 18ul, 1812433253ul>::seed(unsigned long)

_ZNSt23mersenne_twister_engineImLm64ELm312ELm156ELm31ELm13043109905998158313ELm29ELm61489146912365172O5EL
m17ELm8202884508482404352ELm37ELm18444473444759240704ELm43ELm6364136223846793005EE4seedEm =>
std::mersenne_twister_engine<unsigned long, 64ul, 312ul, 156ul, 31ul, 13043109905998158313ul, 29ul, 6148914691236517205ul,
17ul, 8202884508482404352ul, 37ul, 18444473444759240704ul, 43ul, 6364136223846793005ul>::seed(unsigned long)

_ZNSt26subtract_with_carry_engineImLm24ELm10ELm24EE4seedEm => std::subtract_with_carry_engine<unsigned long, 24ul,
10ul, 24ul>::seed(unsigned long)
_ZNSt26subtract_with_carry_engineImLm48ELm5ELm12EE4seedEm => std::subtract_with_carry_engine<unsigned long, 48ul, 5ul,
12ul>::seed(unsigned long)
...
```

Рис. 9. Примеры преобразования декорированных имён в расширении LLVM

Обратите внимание, что в строке 1120 промежуточного представления GCC и строке 1092 представления LLVM видно расширение сигнатуры метода seed[6]. У метода появился дополнительный параметр, который отсутствует в оригинальной сигнатуре, но необходим для передачи указателя this в рамках используемого соглашения о вызовах thiscall, применяемого при вызове методов классов для языка C++.

Атаки на конвейер оптимизации

Инфраструктуры GCC и LLVM предоставляют разработчикам программные интерфейсы для подключения модулей расширения, предназначенных, помимо прочего, для встраивания новых проходов оптимизации. Перед разработчиком такого расширения открываются широкие возможности по анализу и модификации промежуточного представления исходного кода.

На рис. 7 и рис. 8 показаны примеры ключевых участков исходного кода подключаемых расширений компиляторов из состава GCC и LLVM, соответственно, осуществляющих перебор базовых блоков компилируемого приложения с целью обнаружения вызовов функции инициализации генераторов случайных чисел. В случае обнаружения подобных вызовов они осуществляют замену любого входного значения, предусмотренного алгоритмом работы программы, на целочисленную константу, известную злоумышленнику. При этом для такой замены не требуется доступ к исходному коду, так как вредоносный проход может проверить и изменить вызовы различных

6 Std::linear_congruential_engine::seed. URL: https://cplusplus.com/reference/random/linear_congruential_engine/seed, (дата обращения: 03.10.2024).

```
g++-14 -fdump-tree-optimized -fplugin=./plugins/gcc_seedpass.so rnd.cpp -o rnd_gcc_instr
cat ./rnd_gcc_instr-rnd.cpp.265t.optimized
```

```
739     <bb 11> :
740     _43 = std::random_device::operator() (&rd);
741     _5 = _43;
742     _6 = (long unsigned int) _5;
743     std::linear_congruential_engine<long unsigned int, 16807, 0, 2147483647>::seed (&engine, 1234567);
744     _46 = std::uniform_int_distribution<int>::operator()<std::linear_congruential_engine<...> (&val, &engine);
745     _7 = _46;
746     _48 = std::basic_ostream<char>::operator<< (&cout, _7);
747     _8 = _48;
748     std::operator<< <std::char_traits<char> > (_8, " ");
749     j_50 = j_9 + 1;
```

Рис. 10. Промежуточное представление GCC после вредоносной оптимизации

```
opt --load-pass-plugin=./plugins/llvm_seedpass.so -passes="seedpass,default<O2>" rnd.ll -S -o rnd_opt.ll
cat ./rnd_opt.ll
```

```
522     27:                                       ; preds = %24
523       %28 = call noundef i32 @_ZNSt13random_deviceclEv(...)
524       %29 = zext i32 %28 to i64
525       call void @_ZNSt26linear_...seedEm(ptr noundef nonnull align 8 dereferenceable(8) %3, i64 noundef 1234567)
526       %30 = call noundef i32 @_ZNSt24uniform_...(...)
527       %31 = call noundef nonnull align 8 dereferenceable(8) ptr @_ZNSolsEi(...)
528       %32 = call noundef nonnull align 8 dereferenceable(8) ptr _ZStlsISt11char_traits...(...)
529       br label %33
```

Рис. 11. Промежуточное представление LLVM после вредоносной оптимизации

функций. Таким образом, данный метод атаки может быть с успехом применён не только к генераторам случайных чисел, но и к совершенно другим задачам.

Необходимо отметить тот факт, что из-за особенностей применяемого соглашения о вызовах оба расширения заменяют второй аргумент метода, отвечающего за инициализацию генератора, несмотря на наличие лишь одного параметра в исходном определении метода.

Также стоит обратить внимание на то, что в алгоритме работы расширения LLVM выполняются дополнительные преобразования декорированных имён вызываемых функций, как показано на (рис. 9).

На (рис. 10) показано промежуточное представление исходного кода тестового приложения, полученное после выполнения вредоносного прохода оптимизации для GCC, а на (рис. 11) – для LLVM.

После такой вредоносной оптимизации каждый элемент числовой последовательности, формируемый тестовым приложением, будет известен злоумышленнику. При этом для стороннего наблюдателя такая числовая последовательность никак не будет отличаться от любой другой псевдослучайной последовательности, формируемой соответствующим генератором. Также следует отметить, что исполняемый файл тестового приложения с вредоносной оптимизацией, скомпилированный с помощью GCC, и вариант, скомпилированный с помощью LLVM, выдают одинаковые числовые последовательности, как показано на (рис. 12).

```
./rnd_instr

6666666666
7777777777
3333333333
9999999999
5555555555
6666666666
5555555555
6666666666
8888888888
6666666666
```

Рис. 12. Результат исполнения приложения с вредоносной оптимизацией

Защита от атак на конвейер оптимизации

Исключение возможности проведения описанных атак на конвейер оптимизации может быть достигнуто за счёт полного запрета внешних модулей расширения для безопасного компилятора или путём реализации дополнительного механизма контроля аутентичности подключаемых модулей. Также можно порекомендовать использовать в качестве платформы для безопасной разработки программного обеспечения такие защищённые операционные системы, как QP ОС [11], которые обеспечивают замкнутость

программной среды исполнения соответствующих инструментальных средств, что исключает возможность несанкционированного внедрения в процессы компиляции.

Если полный запрет модулей расширения или контроль их аутентичности невозможны, то необходимо уделить повышенное внимание вопросам журналирования и контроля процессов компиляции, что позволит выявить несанкционированные вмешательства в конвейер оптимизации. В общих требованиях к безопасному компилятору языков C/C++ присутствует требование по ведению базы данных (БД) компиляции, где должны фиксироваться информация, позволяющая идентифицировать задействованные программы, а также настройки и конфигурационные файлы этих программ.

К сожалению, общие требования не содержат указаний по поводу фиксации информации о внешних модулях, которые могут динамически подключаться к программам, задействованным в процессе трансляции. Поэтому необходимо расширить общие требования к безопасному компилятору языков C/C++ в части перечня сведений, подлежащих сохранению в БД компиляции. В такой БД дополнительно должен фиксироваться фактический состав применяемого конвейера оптимизации и контрольные значения всех подключаемых расширений компилятора. При этом должен быть определён белый список безопасных проходов оптимизации, разрешенных к применению.

Кроме того, в БД компиляции необходимо фиксировать информацию о системном окружении операционной системы, так как оно может оказывать неявное влияние на функционирование инструментальных средств разработки и обеспечивать возможность подмены существующих или подключения новых модулей расширения.

Выводы

В статье рассмотрены особенности анализа и трансформации исходного кода современными компиляторами из состава GCC и LLVM в процессе оптимизации исходного кода. Показана возможность практической реализации атак на конвейер оптимизации, способных принципиально изменить алгоритм функционирования компилируемого приложения, а также даны рекомендации по нейтрализации подобных угроз.

Разработчики и производители ПО могут использовать сведения, приведённые в статье, при проектировании и внедрении процессов РБПО, а также при разработке безопасного программного обеспечения.

Практическое подтверждение предлагаемых научных решений может быть выполнено путём исследования исходных кодов тестового приложения и всех рассмотренных расширений, а также инструкций по их сборке и применению, открытых и доступных для свободного использования[7].

7 Github repository. URL: https://github.com/smurav/passes, (дата обращения: 03.10.2024).

Литература

1. Арустамян С. С., Вареница В. В., Марков А. С. Методические и реализационные аспекты внедрения процессов разработки безопасного программного обеспечения // Безопасность информационных технологий. 2023. Т. 30. № 2. С. 23–37.
2. Наке К., Кван Э. LLVM 17: Инфраструктура для разработки компиляторов / пер. с англ. А. А. Слинкина. – М.: ДМК Пресс, 2024. – 370 с. – ISBN 978-5-93700-303-4.
3. Леонов Н. В. Противодействие уязвимостям программного обеспечения. Часть 1. Онтологическая модель // Вопросы кибербезопасности. № 2(60). 2024. DOI: 10.21681/2311-3456-2024-2-87-92.
4. Леонов Н. В. Противодействие уязвимостям программного обеспечения. Часть 2. Аналитическая модель и концептуальные решения // Вопросы кибербезопасности. 2024, № 3(61). С. 90–95. DOI: 10.21681/2311-3456-2024-3-90-95.
5. Дейтел П., Дейтел Х. C++20 для программистов. СПб.: Питер, 2024. – 1056 с. – ISBN 978-5-4461-2359-9.
6. Белов, А. А. Ненадежность известных генераторов псевдослучайных чисел / А. А. Белов, Н. Н. Калиткин, М. А. Тинтул // Журнал вычислительной математики и математической физики. – 2020. – Т. 60, № 11. – С. 1807–1814. – DOI 10.31857/S0044466920110046. – EDN CTJCWS.
7. Муравьёв, С. К. Угрозы безопасности информации со стороны модулей расширения для оптимизирующих компиляторов // Безопасность информационных технологий. – 2024. – Т. 31, № 4. – С. 44–55. – DOI 10.26583/bit.2024.4.02. – EDN LUNGGO.
8. Rastello F., Tichadou F. B. SSA-based Compiler Design. – Springer, 2022. – ISBN 978-3-030-80514-2. DOI: https://doi.org/10.1007/978-3-030-80515-9.
9. Khedler U. GCC Translation Sequence and Gimple IR. URL: https://reup.dmcs.pl/wiki/images/d/da/Gcc-gimple.pdf, (дата обращения: 02.10.2024).
10. Min-Yih H. LLVM Techniques, Tips, and Best Practices Clang and Middle-End Libraries. – Packt Publishing, 2021. – ISBN 978-1-83882-495-2.
11. Егоров, В. Ю. Развитие операционной системы QP ОС // Новые информационные технологии и системы: Сборник научных статей по материалам XVII Международной научно-технической конференции, 18–19 ноября 2020 года. – Пенза: ПГУ, 2020. – С. 45–47. – EDN EJXOIX.

THE A VULNERABILITIES OF GCC AND LLVM TO OPTIMIZATION PIPELINE ATTACKS

Muravyev S. K.[8]

Keywords: *development tools, software, compiler, information security threat, GCC, LLVM.*

Purpose of the study: *the purpose of the work is to develop recommendations for the implementation of secure software development tools and the implementation development processes based on the analysis of information security threats in software development related to the possibility of creating and using extension modules for optimizing compilers by attackers.*

Methods of research: *the main research methods include analysis and synthesis, modeling and experiment.*

Result(s): *the article examines the vulnerabilities of optimizing compilers of GCC and LLVM to malicious interference in the optimization pipeline, which can be carried out by hackers through standard software interfaces provided to enhance the functionality of such compilers and the effectiveness of software developed with their help.*

The relevance of the work *is determined by the current regulatory and technical requirements for developers of secure software to analyze information security threats from software development tools, one of the key elements of which are optimizing compilers. The article discusses the features of the analysis and transformation of the source code by optimizing compilers of GCC and LLVM in the process of optimizing the source code. The possibility of practical implementation of attacks that change the optimization pipeline in such a way that the algorithm of functioning of the target application fundamentally changes in the way required by the attacker is shown. As a result, recommendations are given on how to neutralize such threats.*

Scientific novelty: *the paper shows the need to expand the regulatory requirements for secure C/C++ compilers and measures to develop secure software, in terms of the need to control the use of extension modules for the tools used.*

References

1. Arustamjan S. S., Varenica V. V., Markov A. S. Metodicheskie i realizacionnye aspekty vnedrenija processov razrabotki bezopasnogo programmnogo obespechenija // Bezopasnost' informacionnyh tehnologij. 2023. T. 30. № 2. S. 23–37.
2. Nacke K., Kwan A. Learn LLVM 17. A beginner's guide to learning LLVM compiler tools and core libraries with C++. Packt Publishing, 2024. – ISBN 978-1-83763-134-6.
3. Leonov N. V. COUNTERING SOFTWARE VULNERABILITIES. Part 1. ONTOLOGICAL MODEL // Voprosy kiberbezopasnosti. 2024, № 2(60). S. 87–92. DOI: 10.21681/2311-3456-2024-2-87-92.
4. Leonov N. V. COUNTERING SOFTWARE VULNERABILITIES. Part 2. ANALYTICAL MODEL AND CONCEPTUAL SOLUTIONS // Voprosy kiberbezopasnosti. 2024, № 3 (61). S. 90–95. DOI: 10.21681/2311-3456-2024-3-90-95.
5. Deitel P., Deitel H. C++20 for Programmers. Pearson, 2022. ISBN 978-0136905691.
6. Belov, A. A. Unreliability of Available Pseudorandom Number Generators / A. A. Belov, M. A. Tintul, N. N. Kalitkin // Computational Mathematics and Mathematical Physics. – 2020. – Vol. 60, No. 11. – P. 1747-1753. – DOI 10.1134/S0965542520110044.
7. Muravyev S. K. Information security threats of optimizing compilers' plugins. IT Security (Russia), [S.l.], v. 31, no. 4, p. 44–55, 2024. ISSN 2074-7136. ISSN 2074-7136. DOI: http://dx.doi.org/10.26583/bit.2024.4.02.
8. Rastello F., Tichadou F. B. SSA-based Compiler Design. – Springer, 2022. – ISBN 978-3-030-80514-2. DOI: https://doi.org/10.1007/978-3-030-80515-9.
9. Khedler U. GCC Translation Sequence and Gimple IR. URL: https://reup.dmcs.pl/wiki/images/d/da/Gcc-gimple.pdf, (дата обращения: 02.10.2024).
10. Min-Yih H. LLVM Techniques, Tips, and Best Practices Clang and Middle-End Libraries. – Packt Publishing, 2021. – ISBN 978-1-83882-495-2.
11. Egorov V. Yu. Development of the QP OS operating system // New information technologies and systems: A collection of scientific articles based on the materials of the XVII International Scientific and Technical Conference, Penza, November 18-19, 2020. – Penza: Penza State University, 2020. – pp. 45– 47. – EDN EJXOIX.

8 Sergey K. Muravyev, Ph.D, Head of Department of NTP Cryptosoft, Penza, smurav@mail.ru

МЕТОД ОЦЕНКИ ЗАЩИЩЕННОСТИ КРИТИЧЕСКОЙ ИНФОРМАЦИОННОЙ ИНФРАСТРУКТУРЫ

Бочков М. В.[1], Васинев Д. А.[2]

DOI: 10.21681/2311-3456-2025-4-17-29

Цель исследования: *разработка метода оценки защищенности критической информационной инфраструктуры (КИИ) на основе средств полунатурного и имитационного моделирования. Предлагаемый метод позволяет разработать параметрически точные имитационные модели объекта КИИ для исследования свойств защищенности и устойчивости, моделировать воздействия на объекты компьютерных атак (КА).*

Методы исследования: *математические методы теории систем и системного анализа теории вероятностей, методы теории графов, методы имитационного моделирования.*

Результат исследования: *предлагаемый метод оценки защищенности позволяет учитывать конфигурационные и коммуникационные особенности построения и функционирования объектов КИИ, динамику и параметры воздействия нарушителя на объекты КИИ, существующую политику безопасности, моделировать свойство устойчивости, проводить исследования степени влияния составных элементов на защищенность объекта КИИ. Разработанный метод моделирования позволяет осуществлять оценку защищенности объектов КИИ с учетом конфигурационных и коммуникационных параметров объекта КИИ, уменьшить зависимость от экспертных оценок, получать параметрически обоснованные оценки защищенности.*

Научная новизна: *заключается в развитии методов теории информационной безопасности в области оценки защищенности объектов КИИ за счет учета вложенности и иерархичности объектов методами теории гиперграфов и их реализации методами вложенных раскрашенных сетей Петри. В этом случае учитываются конфигурационные и коммуникационные параметры объектов КИИ, позволяющие получать параметрически точные имитационные модели на основе математического аппарата сетей Петри и осуществлять верификацию полученных результатов на полунатурных моделях. Практическая ценность заключается в получении оценок защищенности, основанных на известных параметрах объекта КИИ, возможности получения оценок защищенности на основании коммуникационных, инфраструктурных параметров самого объекта, возможности моделирования известных воздействий из банка данных угроз для проверки политики безопасности объекта КИИ в полученной модели.*

Ключевые слова: *информационная безопасность, коммуникационная инфраструктура, конфигурационная инфраструктура, метод моделирования, метод оценки защищенности, киберустойчивость, протокольные блоки данных.*

Введение

Продолжающееся информационное противоборство делает актуальными вопросы обеспечения информационной безопасности для информационных систем (ИС), информационно-телекоммуникационных сетей (ИТС), автоматизированных систем управления (АСУТП) критических информационных инфраструктур (КИИ), функционирующих в критически важных отраслях деятельности государства: в медицине, образовании, промышленности, энергетике, поясняется отраслевой принадлежностью объектов атак. Среди прочих целью нарушителя являются объекты КИИ. При этом уровень деструктивных действий нарушителя на коммуникационную инфраструктуру говорит о сетевых угрозах преимущественно высокого и критического уровней воздействия нарушителя, проявляющихся в атаках на КИИ[4,5,6].

В качестве составных элементов КИИ выступают распределенные фрагменты сетей, центры обработки данных (ЦОД), автоматизированные системы управления (АСУТП), объединенные в единую распределенную ИТС организации. Пример обобщённого представления распределенной КИИ представлен на (рис. 1).

Существующие особенности построения коммуникационной инфраструктуры технологически достаточно разнообразны, однако объединяющими моментами являются применение технологий виртуальных частных сетей (VPN), резервирования, отказоустойчивости, обеспечение киберустойчивости в условиях воздействия компьютерных атак (КА)[7,8]. Кроме того, современные условия функционирования технических систем предполагают применение отечественного коммуникационного оборудования, средств

1 Бочков Максим Вадимович, доктор технических наук, профессор, ЧОУ ДПО «Центр предпринимательских рисков», г. Санкт-Петербург, Россия. E-mail: mvboch@cprspb.ru

2 Васинев Дмитрий Александрович, кандидат технических наук, сотрудник Академии ФСО России, г. Орёл, Россия. E-mail: vda33@academ.msk.rsnet.ru

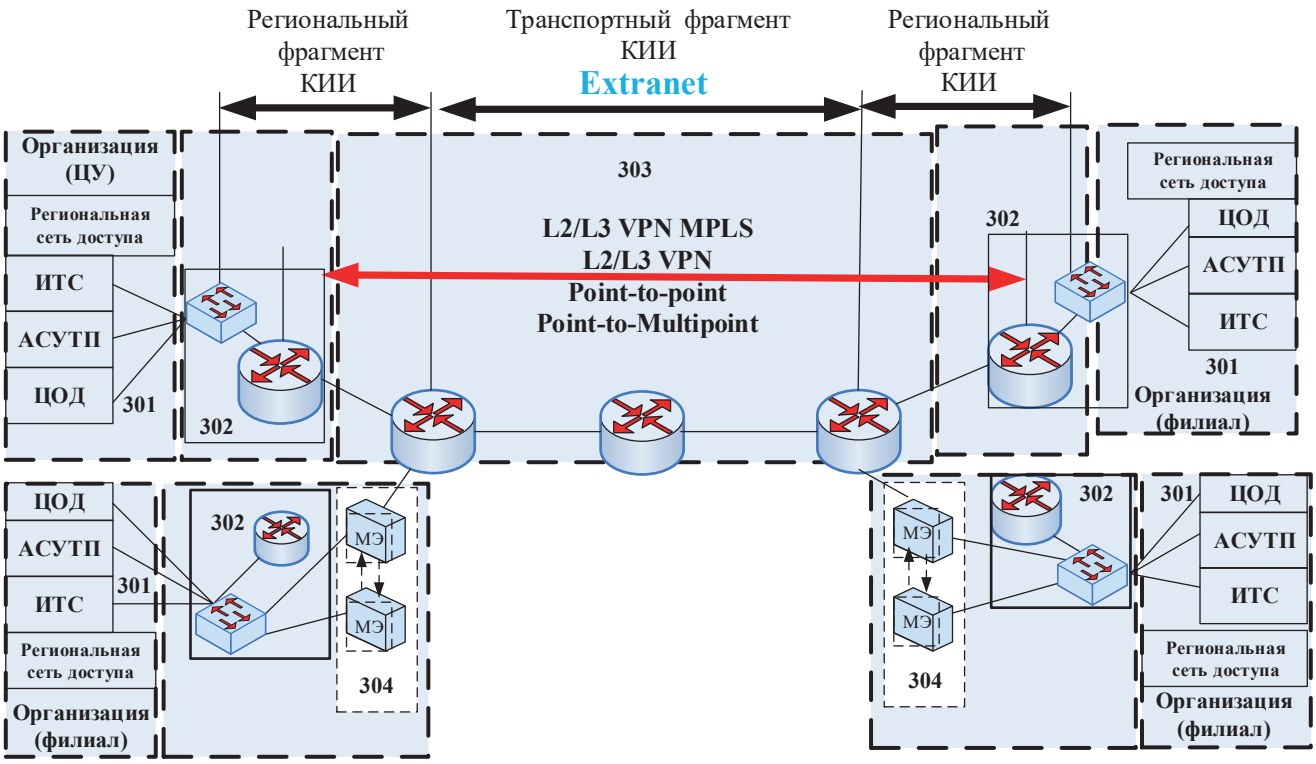
3 РосТелекомм. Аналитический отчет об атаках на онлайн ресурсы компании за 2022г. [сайт]. URL:https:// https://rt-solar.ru/upload/iblock/34a/5w4h9o57axovdbv3ng7givrz271ykir3/Ataki-na-onlayn_resursy-rossiyskikh-kompaniy-v-2022-godu.pdf.

4 ТрансТелеКом. Аналитический отчет по сервису "Защита от DDoS-атак" 1 квартал 2023 [сайт]. URL: https://ttk.ru/upload/doc/business/ddos_1_2023.pdf.

5 Бюллетени НКЦКИ: новые уязвимости ПО [сайт]. URL: https://safe-surf.ru/specialists/bulletins-nkcki/.

6 Запечников, С. В. Основы построения виртуальных частных сетей: учебное пособие для вузов/ Запечников, С. В., Милославская, Н. Г., Толстой, А. И. – 2-е изд. Москва: Горячая линия-Телеком, 2011, – 249. – ISBN 5-85582-119.

7 Захватов, М. А. Построение виртуальных частных сетей на базе технологии MPLS / М. А. Захватов. – Москва: изд-во Cisco Systems, 2001 г.



301. – Технические средства оконечных автоматизированных систем КИИ для отправки/приема единиц передаваемых данных	302. – Многопротокольное оборудование объекта КИИ.	303. – Многопротокольное оборудование оператора связи (ISP).
	304. – Программно-аппаратный отказоустойчивый кластер межсетевых экранов для объекта КИИ	

Рис. 1. Формирование распределенной инфраструктуры для объектов ИС, АСУТП, ИТС КИИ

защиты для проектирования новых и импортозамещения существующих фрагментов КИИ. В этих условиях исследования в области оценки защищенности и устойчивости КИИ в условиях воздействия на нее КА с учетом параметрических особенностей объекта воздействия является актуальной задачей [1–4].

Воздействие нарушителя на распределенную ИТС обусловлено инфраструктурными, коммуникационными особенностями организации каналов связи, предлагаемых оператором связи, на основе которого осуществляется организация взаимодействия между распределенными филиалами телекоммуникационных объектов КИИ, представлено на (рис. 1). Сетевые, транспортные и управляющие протоколы, которые применяются в коммуникационных инфраструктурах для передачи данных, управления, такие как Ethernet, ICMP, IP, TCP, UDP, SNMP. Для выделенных протоколов помимо иерархических – коммуникационных особенностей, можно выделить конфигурационные компоненты формирования инфраструктур, которые также могут быть причиной снижения защищенности объекта – в связи с воздействием нарушителя, или неквалифицированными действиями персонала в распределенных фрагментах ИТС.

Очевидно, что логическая структура каналов связи для КИИ имеет иерархическую особенность построения, обусловленную применением коммуникационных и конфигурационных параметров в КИИ рассматриваемых объектов (ИС, АСУТП, ИТС), функционирующих в единой распределенной сети организации. Для моделирования и оценки защищенности таких объектов (ИС, АСУТП, ИТС) а также исследования свойств устойчивости [1–4], с учетом иерархических особенностей формирования объектов КИИ предлагается применять совокупность имитационных и полунатурных моделей. При этом отличительной особенностью предлагаемого решения на основе имитационных моделей сетей Петри является учет не только иерархических особенностей построения объектов КИИ, но и их конфигурационных и коммуникационных особенностей функционирования, а также воздействий нарушителя как на логическую (коммуникационную и конфигурационную), так и на физическую составляющую объекта КИИ. [5]

В сложившихся условиях при сетевых воздействиях нарушителя на коммуникационную инфраструктуру (КИ) объекта КИИ существующие методы оценки защищенности, основанные на знании сигнатуры угрозы, сводятся к методам оценки защищенности от известных угроз, например, зарегистрированных в БДУ ФСТЭК. На рисунке 2 представлен процесс оценки защищенности объекта КИ на основе известных сигнатур угроз. [6–10]

В таких условиях нарушитель, работающий в известном пространстве состояний объекта КИИ, обладает сведениями о 2^m параметрах функционирования объекта КИИ. Эти же параметры являются основой для формирования воздействия нарушителя на объект КИИ. Такие возможности нарушителя позволяют изменять сигнатуры, формировать новые, ранее неизвестные воздействия 1 и 2, представленные на рисунке 2. При этом методы обеспечения защищенности объекта КИИ на основе сигнатурных средств всегда отстают по времени от воздействия нарушителя, что создает предпосылки к нахождению объекта КИИ в незащищенных состояниях 1, 2 рисунок 2.

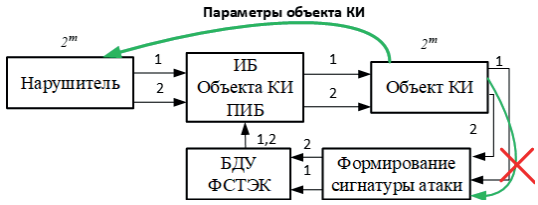


Рис. 2. Существующий подход к оценке защищенности объектов КИИ

Формирование сигнатур на основе существующих БД сигнатур методами машинного обучения является перспективным направлением, требовательным к исходным данным о состоянии объекта. Причем применение для этого знаний о параметрах функционирования самого объекта КИИ является ключевым фактором, требующим учета в разрабатываемых моделях.

Решением сложившегося противоречия между многообразием воздействия нарушителя и существующими возможностями методов и средств обеспечения информационной безопасности является учет параметров объекта КИИ в формировании политики информационной безопасности объекта. На рисунке 3 представлен процесс формирования оценок защищённости на основе учета m параметров объекта КИИ, формализации их в модели цифрового двойника (имитационной модели), расчет на ее основе оценок защищенности, учетm параметров в политике ИБ объекта КИИ.

В основе предлагаемого метода оценки защищенности – формирование на основе конфигурационных

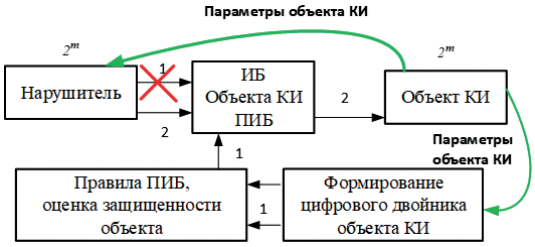


Рис. 3. Предлагаемый метод оценки защищенности объекта КИИ

и коммуникационных параметров функционирования объекта КИИ, его политики безопасности, моделей коммуникационной инфраструктуры, ее политики информационной безопасности, воздействия на нее нарушителя. Предлагаемый метод позволяет получать параметрически точные модели коммуникационной инфраструктуры – цифрового двойника объекта КИИ, в динамике исследовать влияние на политику информационной безопасности действий нарушителя. В основе цифрового двойника – имитационные модели на основе вложенных раскрашенных сетей Петри, верификация которых осуществляется полунатурными моделями. Комплекс моделей цифрового двойника включает в себя модель коммуникационной инфраструктуры объекта КИИ, модели каналов связи, комплекс взаимосвязанных моделей, связанных с формированием политики ИБ, анализом защищенности и действиями нарушителя [5].

1. Метод сквозного моделирования объектов КИИ на основе средств полунатурного и имитационного моделирования

Моделирование многоуровневых коммуникационных инфраструктур связано с особенностями их построения (рисунок 4) [5]. На основе анализа существующих методов моделирования и оценки защищенности [12–16], сформулированных ранее предположений о необходимости учета параметров объекта КИИ [5] разработан метод моделирования иерархически сложных телекоммуникационных объектов и метод оценки защищенности объектов КИИ на основе конфигурационных и коммуникационных параметров функционирования объекта КИИ.

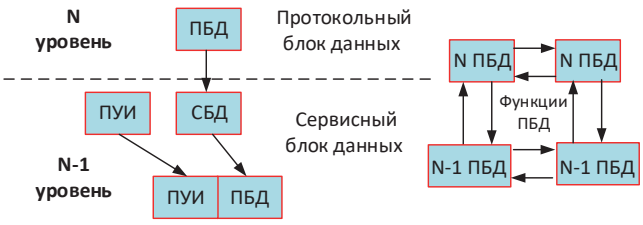


Рис. 4. Методы, способы взаимодействия ПБД в соответствии с моделью OSI (7498), X.200, ГОСТ Р ИСО/МЭК 7498-1-99

Основными особенностями, которые легли в основу универсальных масштабируемых модулей для имитационного и полунатурного моделирования, является внутриуровневое и межуровневое взаимодействие протокольных блоков данных, представленных на рисунке 4. В связи с необходимостью моделировать множество протоколов разработана концептуальная модель протокольного блока данных для реализации концепции вертикального и горизонтального взаимодействия протокольных блоков данных, учитывающая наиболее важные подсистемы взаимодействия, представлена на рисунке 5 [11].

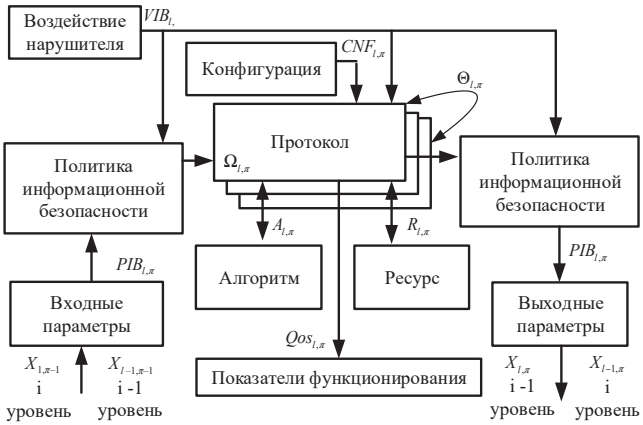


Рис. 5. Концептуальная модель протокольного блока данных

Для реализации разнотипных протокольных блоков данных, функционирующих в объектах КИИ, предлагается обобщенное концептуальное представление протокольного блока данных, представленного на рисунке 5. В основе его – протокол с множеством параметров π_l-го протокола на l-ом уровне функционирования в коммуникационной инфраструктуре объекта КИИ, с множеством $\Theta_{l,\pi}$ – множество функциональных связей между алгоритмами $A_{l,\pi}$ в протоколе π_l на l-ом уровне функционирования в коммуникационной инфраструктуре объекта КИИ, $R_{l,\pi}$ – ресурсом на l-ом уровне функционирования в коммуникационной инфраструктуре объекта КИИ, применяемого для функционирования π_l-го протокола в коммуникационной инфраструктуре, множеством показателей качества обслуживания – $Qos_{l,\pi}$, характеризующих функциональное состояние коммуникационной инфраструктуры объекта КИИ. Для π_l протокола на l-ом уровне функционирования входными/выходными параметрами являются $-X_{l,\pi}$. Для π_l протокола на l-ом уровне функционирования формируется множество параметров политики информационной безопасности – $PIB_{l,\pi}$ для входящего/исходящего направления π_l-го протокола на l-ом уровне функционирования в коммуникационной инфраструктуре объекта КИИ. Протокол имеет предварительно заданное множество параметров конфигураций – $CNF_{l,\pi}$ для π протоколов l-ого уровня функционирования в коммуникационной инфраструктуре объекта КИИ. Для протокола, а также политики информационной безопасности для π-го протокола l-ого уровня предусмотрено множество параметров деструктивных воздействий $VIB_{l,\pi}$ нарушителя для алгоритмов, функционирующих в коммуникационной инфраструктуре объекта КИИ [11].

Предлагаемое обобщенное представление протокольных блоков данных позволяет объединять похожие по функциональному назначению блоки (алгоритмов функционирования протокола, ресурса протокола, конфигураций, политики информационной безопасности, воздействия нарушителя) для построения универсальных имитационных моделей на основе вложенных раскрашенных сетей Петри [5].

Применение имитационного моделирования позволяет разработать универсальный метод построения имитационных протокольных блоков данных для различных типов протоколов, учесть коммуникационные и конфигурационные особенности их функционирования, являющиеся основой метода сквозного моделирования сетей, узлов и комплексов специальной связи. Данный метод представлен на рисунке 6.

Суть предлагаемого метода заключается во взаимосвязи конфигураций (параметров) объекта КИ (рисунок 6) с полунатурными и имитационными моделями и возможностями переноса конфигурации как с физического объекта на имитационные (рисунок 8) и полунатурные модели (рисунок 9), так и с имитационных, полунатурных моделей в физический объект. Представленная структурная взаимосвязь моделей в методе позволяет исследовать значимые свойства физического объекта на имитационных и полунатурных моделях и переносить значимые результаты их на физические объекты.

Объекты, представленные на рисунке 6, объединены единой логической составляющей – конфигурационными и коммуникационными параметрами. С физического объекта – конфигурационные и коммуникационные параметры переносятся в полунатурные модели физического объекта с высокой степенью достоверности. Для применения в средствах имитационного моделирования необходимы дополнительные преобразования, позволяющие из разнообразных типов конфигураций получать параметры для имитационной модели.

Основной задачей, решаемой в методе моделирования коммуникационных инфраструктур, является получение универсальных масштабируемых имитационных и полунатурных моделей, пригодных для моделирования многоуровневых распределенных коммуникационных инфраструктур различных объектов КИИ.

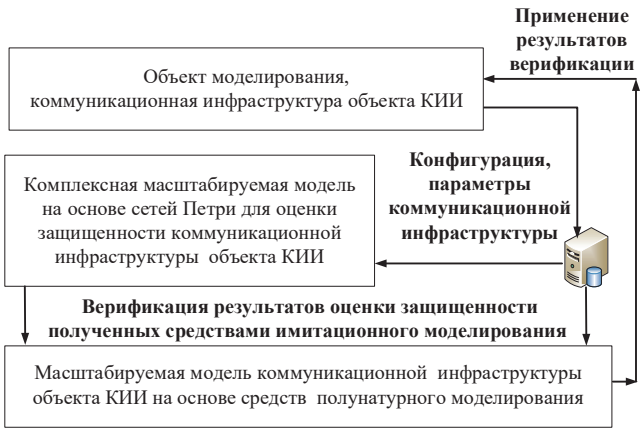


Рис. 6. Метод сквозного моделирования объектов КИИ на основе средств полунатурного и имитационного моделирования

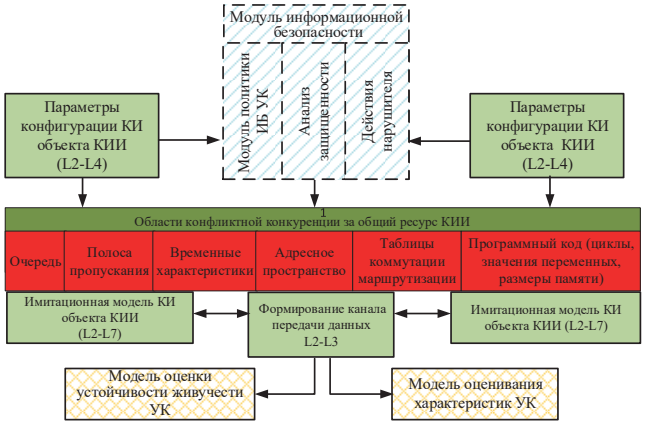


Рис. 7. Области конфликтного взаимодействия в модели оценки защищенности КИ объекта КИИ

Структура комплекса имитационных моделей для моделирования узла коммутации, подсистем информационной безопасности, области их конфликтного взаимодействия представлены на рисунке 7.

Пример реализации комплексной имитационной модели, учитывающей конфигурационные и коммуникационные параметры, методы обеспечения информационной безопасности, методы воздействия нарушителя, представлен на рис. 8. Модель объединяет основные взаимодействующие субъекты, инфраструктуры, методы обеспечения информационной безопасности, что позволяет моделировать конфликтное поведение взаимодействующих субъектов.

Модель позволяет исследовать влияние новых конфигураций, иерархических транспортных конструкций на защищенность объекта КИИ, проверять функциональность политики безопасности на потенциально возможные воздействия нарушителя, известные из БДУ ФСТЭК.

Основными составными элементами комплексной имитационной модели коммуникационной инфраструктуры являются: 1,2 – универсальная масштабируемая модель коммуникационной инфраструктуры объекта КИИ; 3,4 – модель ввода конфигураций коммуникационной инфраструктуры, задания сервисов; 5,6 – универсальная масштабируемая модель информационной безопасности для коммуникационной инфраструктуры объекта КИИ; 7 – результирующее множество регистрируемых угроз и их параметров для политики информационной безопасности коммуникационной инфраструктуры объекта КИИ; 8,9 – универсальная масштабируемая модель каналов оператора связи для коммуникационной инфраструктуры объекта КИИ; 10 – блок оценки устойчивости/живучести для коммуникационной инфраструктуры объекта КИИ; 11 – блок оценки характеристик устойчивости/живучести для коммуникационной инфраструктуры объекта КИИ.

Основным элементом имитационной модели является модуль обеспечения информационной безопасности, концепция построения которого для входящего и исходящего информационного направления представлена на рисунке 9. Для входящего и исходящего направления структура блока обеспечения информационной безопасности содержит: модуль формализации конфигурации протокола в форме $|m|$ параметров коммуникационной инфраструктуры; модель формирования политики информационной безопасности – формализация политики информационной безопасности объекта КИИ в форме команд имитационной модели; анализа защищенности для политики безопасности на основе $2\times|m|$ параметров, формирование нарушителя политики безопасности КИ – на основе $2\times|m \pm \Delta|$, методом стохастического случайного поиска в заданном пространстве состояний параметров $|m|$[8].

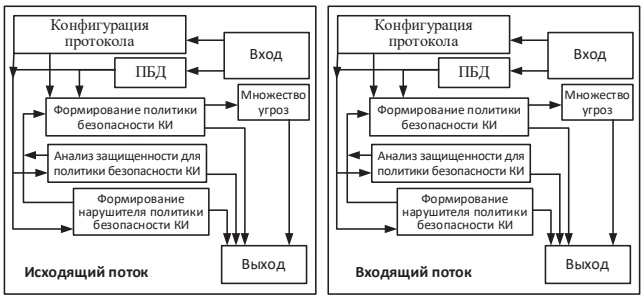


Рис. 9. Концепция построения модуля информационной безопасности для моделирования политики ИБ, анализа защищенности, формирования модели нарушителя в КИ объекта КИИ

Реализация представленной на рисунке 9 концепции построения модуля информационной безопасности представлена на рисунке 10 вложенными раскрашенными сетями Петри.

8 Требования по безопасности информации, приказ ФСТЭК № 33 от 07.03.2024 г.

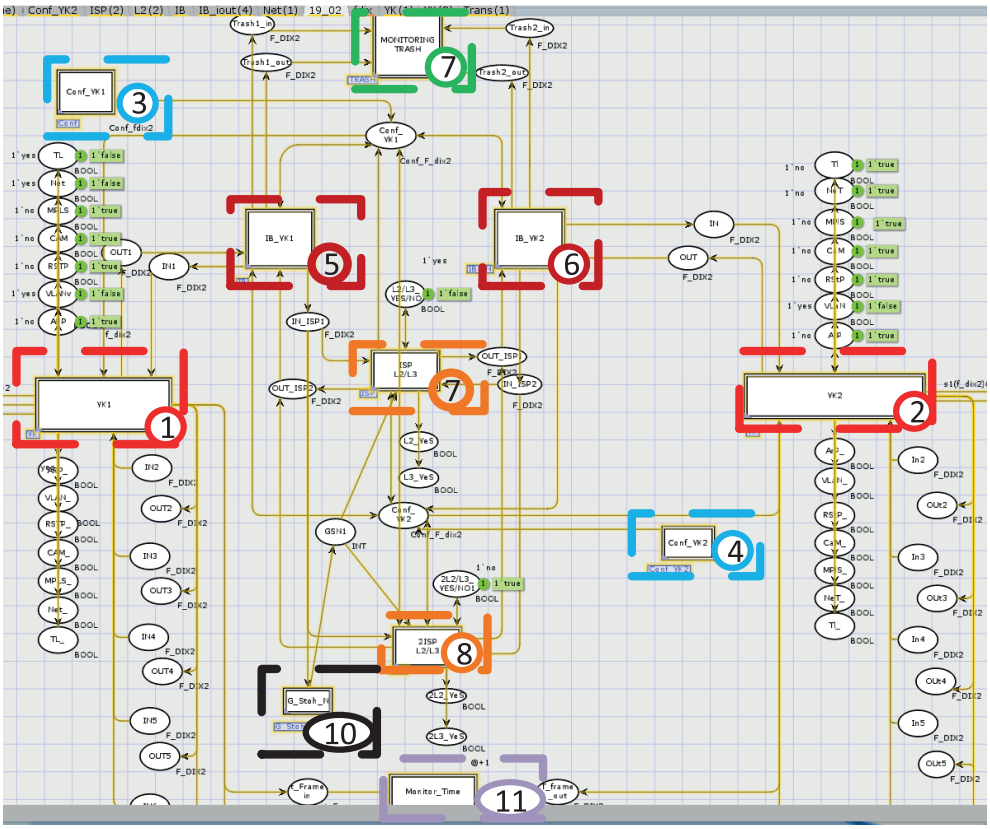


Рис. 8. Комплексная имитационная модель коммуникационной инфраструктуры, с учетом функционирования методов и средств обеспечения ИБ, а также конфликтности взаимодействия

Ключевым фактором является учет всех параметров объекта КИ объекта КИИ $|m|$, на основе которых осуществляется расчет тестовых комбинаций для анализа защищенности $2\times|m|$, а также вторичная верификация в пространстве случайных состояний $2\times|m \pm \Delta|$.

Верификация политики информационной безопасности множества тестовых запросов, происходящая в полунатурной модели коммуникационной инфраструктуры объекта КИИ, представлена на рисунке 11, позволяет проверить работоспособность политики информационной безопасности относительно тестов на основе множества параметров объекта – $|m|$.

С целью верификации политики информационной безопасности формируется программно-аппаратный комплекс на основе полунатурных моделей, например, UNL/EVE, или средств виртуализации на основе операционных систем с открытым исходным кодом. В полунатурную модель подключаются средства измерения, такие как программно-аппаратные датчики М-716, или программные средства измерения на основе программного средства iperf. Реализация тестовых протокольных конструкций осуществляется на основе программного средства Scapy, а также

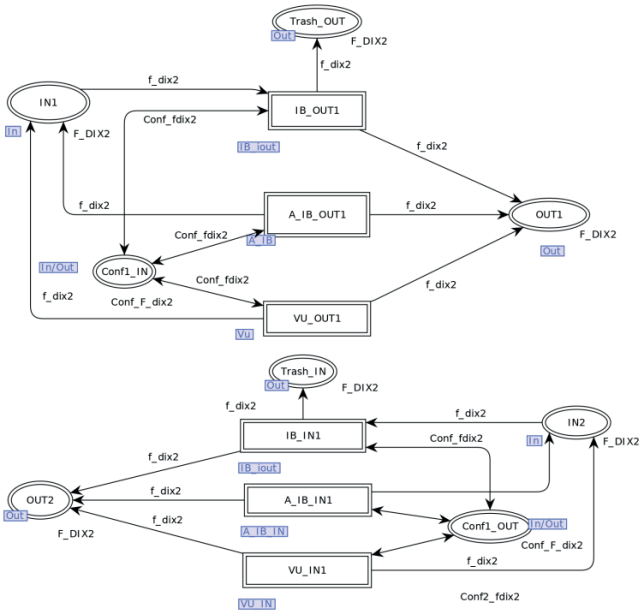


Рис. 10. Пример моделирования политики ИБ, анализа защищенности, формирования модели нарушителя в КИ объекта КИИ на основе вложенных, раскрашенных сетей Петри для входящего и исходящего направления

разработанного программно-аппаратного комплекса тестирования телекоммуникационного и оконечного оборудования объектов КИИ[9] [17].

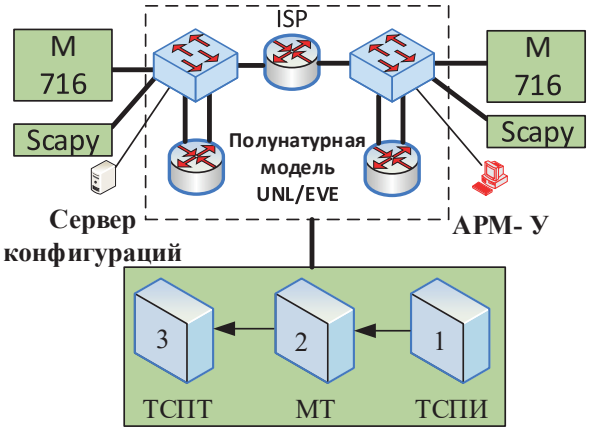


Рис. 11. Пример варианта программно-аппаратной реализации коммуникационной инфраструктуры в средствах полунатурного моделирования

Предлагаемый метод формирования моделей коммуникационной инфраструктуры и верификации результатов имитационного моделирования позволяет формировать параметрически точные модели, учитывающие существующие конфигурации объекта КИИ, параметры его политики информационной безопасности, моделировать параметрические действия нарушителя и исследовать влияние изменения выделенных подсистем на защищенность объекта КИИ в динамике взаимодействия объекта КИ политики его ИБ, действий нарушителя. Верификация результатов имитационного моделирования предусмотрена полунатурными моделями, аналитическими методами, сравнением функциональных характеристик (формирование и фильтрация протокольных блоков данных) с физическим объектом.

2. Метод оценки защищенности коммуникационной инфраструктуры объектов КИИ

Основные взаимосвязанные элементы метода, позволяющие решать задачу оценки защищенности на основе параметров конфигурации, представлены на рисунке 12, 13.

Формирование комплексной модели оценки защищенности на основе вложенных раскрашенных сетей Петри осуществляется в блоке 1 рисунка 12. Проводится формализация в правилах имитационной модели параметров коммуникационной инфраструктуры $|m|$, составляющих протокольную часть коммуникационной инфраструктуры объекта КИИ. Формализация в правила имитационной модели множества разрешающих правил *true*, политики информационной безопасности, с учетом $|m|$ параметров КИ. Формализация в правила имитационной модели множество запрещающих правил *false* на основе формирования тестовых последовательностей $2\times|m|$ и $2\times|m \pm \Delta|$ параметров коммуникационной инфраструктуры (блок 1).

На основе существующих параметров КИ $|m|$, осуществляют автоматическое (эталонное) построение политики информационной безопасности коммуникационной инфраструктуры объекта КИИ – множества разрешающих правил *true* – $rule_{MsIB} = |\{rule_{КИ,Конф,N}\{KI(H,P,CF,S)\} = rule_{КИ,Конф,N}\{MsIB\}\}| = |m|$ в блоке 2, а также формируют множество правил ИБ – $rule_{MsIB} = \{MsIB\}$, для существующей политики ИБ КИ объекта КИИ в блоке 3, рисунка 12.

Осуществляют проверку политики ИБ коммуникационной инфраструктуры объекта КИИ (блок 4) методом сопоставления множеств $truerule_{MsIB} = rule_{КИ,Конф,N}\{MsIB\}$ с формализованной политикой ИБ объекта КИИ множеством $\{rule_{MsIB}\}$, на основе которых в случае $\{rule_{КИ,Конф,N}\}\backslash\{rule_{MsIB}\} = \{Inc_{ib}\}$ – формируют множество угроз в блоке 14 рисунка 13.

В блоке 5 на основе параметров КИ множества – $|m|$, формируют тестовые последовательности для анализа защищенности множество *false* $\{VN(KI(H,P,CF,S),V)\} = \{m \pm \Delta\} = 2 \times |m|$. Далее осуществляют сопоставление множеств правил политики ИБ и множеств правил для анализа защищенности объекта КИИ, $\{rule_{MsIB}\}\backslash\{m \pm \Delta\} = \{\dashv\}$, на основе которых в случае $\{rule_{MsIB}\}\backslash\{m \pm \Delta\} = \{Inc\_A\_i\ b\}$ – формируют множество угроз в блоке 14.

Верификации политики ИБ КИ $\{rule_{MsIB}\}$ нарушителя методами стохастического случайного поиска на основе Марковской цепи, с модификацией параметров протокольных блоков данных генетическим алгоритмом осуществляется в блоках 6–10 рисунка 12. При этом множество $\{m \pm \Delta\}$ расширяется множеством на основе тестовых последовательностей $M = \{m_1 \ldots m_k\}$. Осуществляется вторичная верификации политики информационной безопасности методом сопоставления множества $M = \{m_1 \ldots m_k\}$ с формализованной политикой ИБ объекта КИИ $\{rule_{MsIB}\}$.

По результатам сравнения политики ИБ коммуникационной инфраструктуры объекта КИИ с тестовыми последовательностями множества *false* в элементах 111, 122, 133 тестовые запросы, неучтенные политикой информационной безопасности (блоках 11, 12, 13 рисунка 13), позволяют получить множество угроз для коммуникационной инфраструктуры объекта КИИ (блок 14 рисунка 13).

В блоке 14 формируется множество известных и неизвестных относительно функциональных возможностей средств обеспечения ИБ угроз, которое получено на основе имитационной модели и сделанных предположений о составе и структуре множества угроз ИБ КИ ($F = f_{true} \cup f_{false}$, блок 14, рисунка 13). Предположение заключается в том,

9 Свидетельство о регистрации программы для ЭВМ RU 20246115254 от 05.03.2024 г.

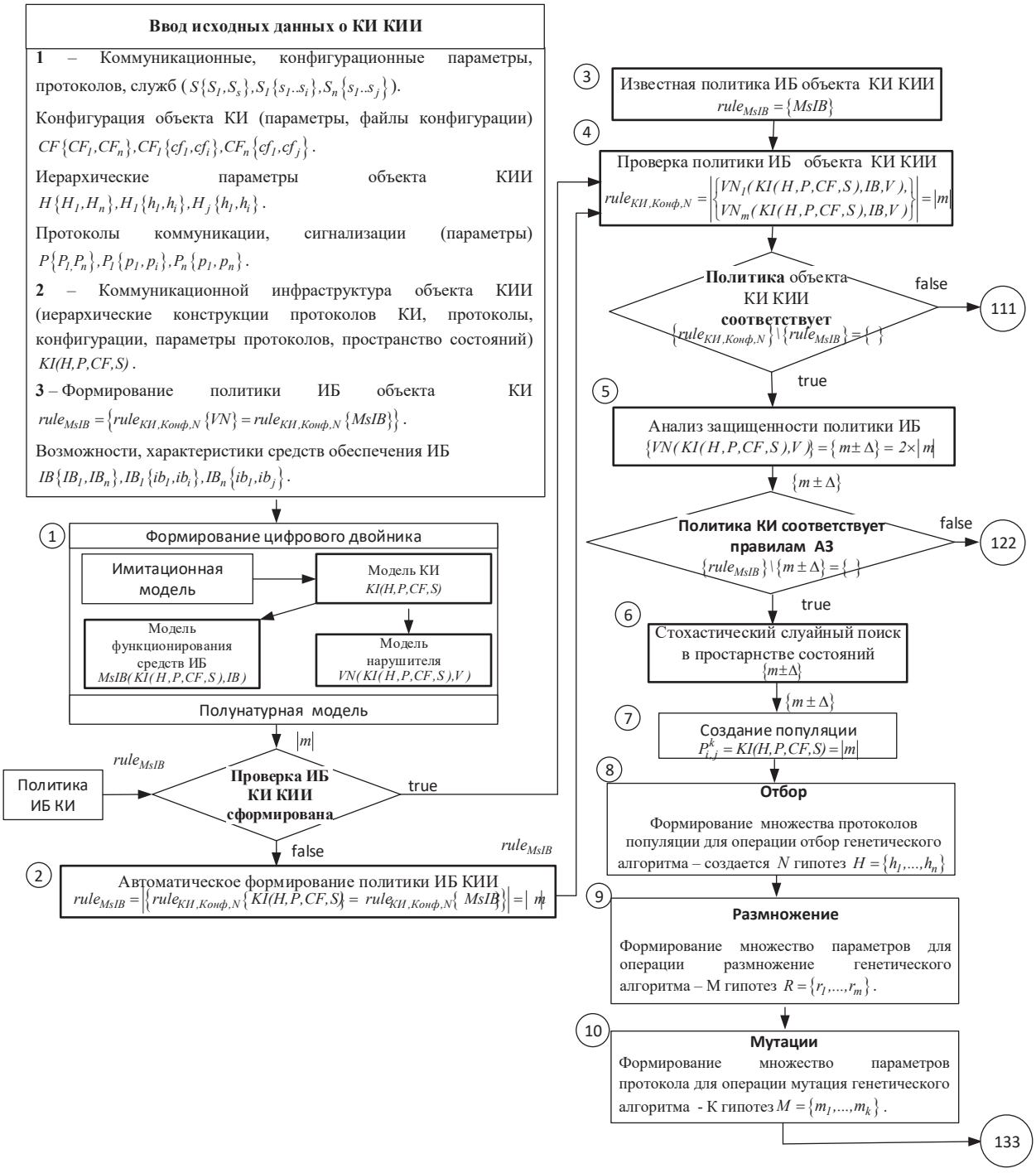


Рис. 12. Метод оценки защищенности КИ объектов КИИ (анализ защищенности)

что множество угроз $F = f_{true} \cup f_{false}$ КИ структурировано относительно параметров КИ – $|m|$, а также функциональных возможностей средств обеспечения ИБ объекта КИИ, влияющих на соотношение множеств f_{true} и f_{false}. Имитационная модель позволяет рассчитать эти соотношения. На основе множества неизвестных угроз f_{false} (блок 15) формируется множество известных f_{true} (блок 16) относительно политики ИБ объекта КИИ.

Для множества известных угроз реализуются алгоритмы обеспечения киберустойчивости на основе логического резерва КИ (блок 17), а для неизвестных угроз алгоритмы обеспечения киберживучести (блок 18) на основе логического, физического резерва КИ по заданным правилам.

Оценка защищенности УК осуществляется в блоке 20 на основе исходных данных 1, 2, 3, 4, позволяет оценивать защищенность, незащищенность,

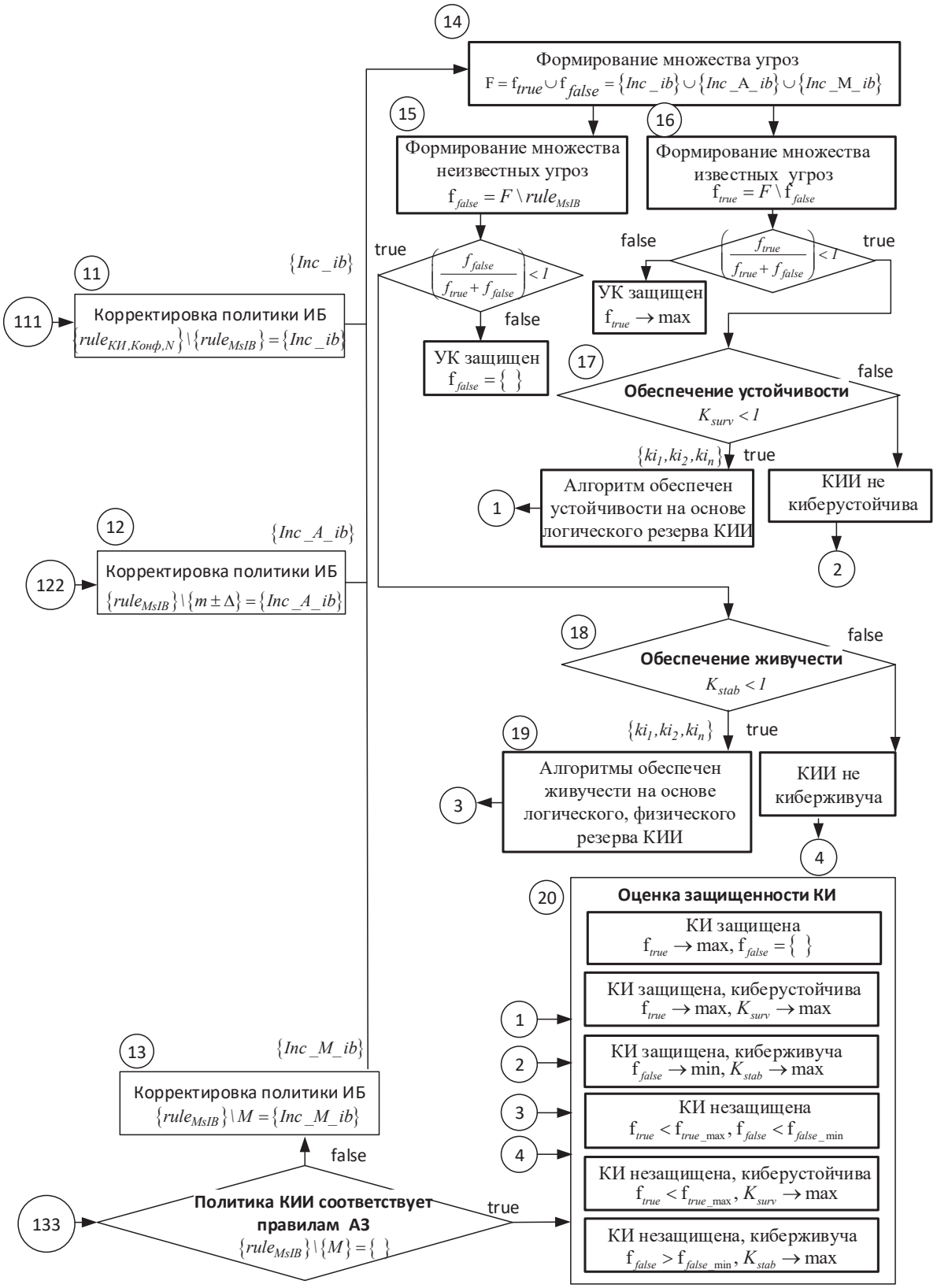


Рис. 13. Метод оценки защищенности КИ объектов КИИ (оценка защищенности)

киберустойчивость объектов относительно известных параметров объекта КИИ и возможностей средств обеспечения его ИБ.

Таким образом, на основе предлагаемого метода моделирования объектов КИИ получены параметрически точные модели объектов КИИ, осуществляется моделирование политики ИБ, проверка политики ИБ КИИ различными классами тестов – анализом защищенности, моделированием действий нарушителя. В предлагаемом методе параметры коммуникационной инфраструктуры являются исходными данными для функционирования модуля ИБ, на основе которого формируются тестовые проверочные конструкции различных типов.

Полученные результаты полунатурного и имитационного моделирования позволяют применять их для расчета защищенности объекта КИИ, верификации его политики безопасности относительно известных параметров, пример которой представлен в таблицах 1–4. Пример расчета основан на оценке защищенности протокола Ethernet для кадра DIX2, для размерности пространства состояний m. Известно, что для размера кадра Ethernet DIX2 l = 1518 бит, возможно 2^{1518} бит воздействий нарушителя, очевидно, что пространство состояний достаточно большое. На основе предложенного варианта снижения его размерности предлагается выделить параметры КИ объекта КИИ, участвующих во взаимодействии. Пусть исходные данные объекта КИИ заданы параметрами канального уровня. 00:00:00:00:00:02 – macD (адрес назначения), 00:00:00:00:00:01 – macS (адрес источника), 11:11:11:1:11:11 – широковещательный адрес; Etype(2): 0x0800 (2048) – Ipv4 (типа протокола); 0x8100 (33024)– 802.1q (тегированного кадра) – 5 параметров – значения m. Параметры коммуникационной инфраструктуры позволяют снизить размерность пространства состояний до 5 значений.

Пусть правила политики ИБ объекта КИИ содержат 2 правила по фильтрации кадров. Проверка работоспособности функционирования политики ИБ относительно параметров объекта КИИ осуществляется не более чем 5 тестовыми запросами, учитывающими m разрешенных параметров, запросы формируются разрешенные. Анализ защищенности осуществляется на основе $2 \cdot m$ параметров, позволяет проверить работоспособность политики ИБ по верхней и нижней границе относительно значения m. В этом случае $2 \cdot m$ тестовых запросов представляют собой тестовые деструктивные запросы для анализа защищенности. Значения тестовых множеств для рассматриваемого примера представлены в таблице 1.

В таблице 1 введены значения коммуникационных $N_i^{КИ}$ и конфигурационных N_i^{Cnf} параметров объекта КИИ в предположении, что они будут представлять отдельные непересекающиеся множества значений при учете всего множества параметров m.

На основе представленных параметров произведены расчеты на имитационной модели, результаты которых представлены в таблице 2.

При работе имитационной модели было сформировано 112 кадров, из которых, число деструктивных кадров – 50, среди которых 40 учитываются политикой ИБ объекта КИИ, а 10 политикой ИБ не учитываются по причине отсутствия функциональных возможностей оборудования, необходимостью модернизации самой политики ИБ.

Получение тестовых последовательностей P_i, множество пакетов воздействия нарушителя – угроз коммуникационной инфраструктуре, конфигурациям, неизвестные относительно возможностей средств информационной безопасности $(K_i(K_i^{КИ},Cnf_i^{КИ},N_i^{КИ}))$ – угрозы, не прошедшие политику ИБ. Выделенное множество угроз K_i содержит угрозы – $(K_{BHi}^{КИ},Cnf_{BHi}^{КИ})$, устранимые средствами политики ИБ за время $T_{УСТР}^{КИ} \le T_{доп}^{КИ}$, приемлемое время для функционирования коммуникационной инфраструктуры объектов КИИ, и угрозы – $N_i^{КИ}$, не устранимых средствами политики ИБ за время $T_{УСТР}^{N} \ge T_{доп}^{N}$ неприемлемое для функционирования коммуникационной инфраструктуры объектов КИИ, представлено в таблице 2.

Имитационная модель позволяет исследовать поведение объекта КИИ в условиях резервирования физических или логических элементов объекта КИИ, получая при этом оценки для случая функционирования КИ объекта КИИ в условиях известных угроз – обеспечения киберустойчивости, так и в условиях неизвестных угроз относительно политики информационной безопасности – обеспечение

Таблица 1.

Исходные данные для оценки защищенности объекта КИИ

Количество параметров (коммуникационных, конфигурационных)	**Количество правил политики ИБ объекта**	**Количество правил для проверки политики ИБ**	**Количество правил для анализа защищенности одного направления**	**Количество правил для анализа защищенности с учетом политики ИБ**
$N_i^{КИ}$, N_i^{Cnf}	$rule_i^{MsIB}$	$N_i^{КИ} + N_i^{Cnf}$	$2^{(K_{BH_i}^{КИ} + Cnf_{BHi}^{КИ})}$	$(2 \cdot m + \mathrm{m})$
5	2	5	10	15

Таблица 2.

Результаты расчета имитационной модели по тестированию политики ИБ

Количество пакетов, сформированные имитационной моделью	Количество пакетов, не соответствующие политике ИБ	Количество пакетов нарушителей КИ, потенциально устранимых средствами политики ИБ $T_{УСТР}^{КИ} \le T_{доп}^{КИ}$	Количество пакетов нарушителей КИ, неустранимых средствами политики ИБ за время $T_{УСТР}^{N} \ge T_{доп}^{N}$
P_i	$K_i(K_i^{КИ}, Cnf_i^{КИ}, N_i^{КИ})$	$K_{BH_i}^{КИ}, Cnf_{BHi}^{КИ}$	$N_i^{КИ}$
112	50	40	10

Таблица 3.

Расчет коэффициентов устойчивости на основе резервного ресурса логического и физического пространства состояний

Количество физических резервных направлений	Количество логических резервных направлений	Количество допустимых физических резервных направлений	Количество допустимых логических резервных направлений	Коэффициент устойчивости	Коэффициент живучести
$F_{факт}$	$L_{факт}$	$F_{доп}$	$L_{доп}$	$K_{stab} = \frac{L_{факт}}{L_{доп}}$	$K_{stab} = \frac{F_{факт}}{F_{доп}}$
2	2	6	10	0,2	0,33

киберживучести объекта. При этом под физическими элементами резервирования понимаются коммуникационное оборудование, каналы связи различного типа. Под логическим резервированием понимаются логические транспортные единицы формируемые в многопротокольном оборудовании (VLAN, VRF, Tunnel L2/L3, протоколы маршрутизации).

Результаты расчетов позволяют ограничить допустимые значения логического резерва – $L_{доп}$, а также допустимые значения физического резерва $F_{доп}$, с учетом имеющегося ресурса резервирования КИ рассчитать коэффициенты устойчивости и живучести, представленные в таблице 3.

Итоговый расчет защищенности как без учета резервирования, так и с учетом резервирования представлен в таблице 4.

Для заданных исходных данных, определяемых параметрами коммуникационной инфраструктуры, получены параметрически обоснованные и точные оценки защищенности объекта КИИ. При этом для полученных оценок очевидны причинно-следственные связи защищенности объекта КИИ с его параметрами. Выявленные недостатки средств обеспечения информационной безопасности позволяют определить направления дальнейшего совершенствования политики информационной безопасности. В случае обеспечения функционирования объекта КИИ в условиях компьютерных атак появляются как теоретические предположения, объясняющее этот процесс и поведение объекта КИИ в дальнейшем, так и практическое подтверждение расчета оценок его защищённости.

Таблица 4.

Оценка коэффициентов защищенности с учетом устойчивости, живучести на основе комплексной интегрированной модели оценки защищенности

Коэффициент защищенности	Коэффициент защищенности с учетом киберустойчивости и киберживучести
$K_n^{ИБ}$	$K_n^{Конф,КИ,N} = \frac{\left(\sum_{i=1}^{n} K_i^{КИ} + \sum_{i=1}^{n} K_i^{Конф}\right)}{\sum_{i=1}^{n} (K_n^{КИ} + K_n^{Конф} + K_n^{N})} K_{stab} + \frac{\sum_{i=1}^{n} N_i}{\sum_{i=1}^{n} (K_n^{КИ} + K_n^{Конф} + K_n^{N})} K_{surv}.$
$K_n^{ИБ} = 40/50 = 0,8$	$K_n^{Конф,КИ,N} = 40/50 \cdot 0,2 + 10/50 \cdot 0,33 = 0,8 \cdot 0,2 + 0,2 \cdot 0,33 = 0,226$

Заключение

Таким образом, предлагаемый метод оценки защищенности фрагментов КИ объектов КИИ (ИС, АСУТП, ИТС) на основе иерархических, раскрашенных сетей Петри позволяет расширить прикладной аспект теории информационной безопасности в направлении развития методов моделирования и методов оценки защищенности объектов КИИ на основе учета иерархичности и вложенности объектов учитываемой теорией гиперграфов ее реализации во вложенных раскрашенных сетях Петри. Моделирование на основе сетей Петри позволяет исследовать влияние протокольных особенностей построения объектов КИИ (ИС, АСУТП, ИТС) на свойства устойчивости и доступности объектов КИИ и оценивать на основе этого их защищенность. Формирование параметрически точных моделей КИИ позволяет строить цифровые двойники объектов коммуникационной инфраструктуры и в динамике исследовать функционирование такого объекта с учетом изменения конфигурации, воздействия нарушителя, формирования физических или логических резервных направлений связи. Полученные результаты позволяют получать в том числе и количественные параметрически обоснованные показатели оценки защищенности объектов КИИ в условиях воздействия нарушителя и исследовать влияние на них различных типов компьютерных атак.

Литература

1. Зегжда Д. П. Кибербезопасность цифровой индустрии. Теория и практика функциональной устойчивости к кибератакам: монография / Александрова Е. Б., Калинин М. О., Марков А. С. [и др.]. – Москва: Горячая линия – Телеком. 2023. – 500с. – ISBN 978-5-9912-0827-7.
2. Петренко С. А. Киберустойчивость цифровой индустрии 4.0: научная монография / С. А.Петренко. – Санкт-Петербург: Издательский Дом «Афина», 2020, – 255 с.
3. Петренко С. А. Управление киберустойчивостью. Постановка задачи // Защита информации. Инсайд. 2019. № 3(87). С. 16–24.
4. Штыркина А. А. Обеспечение устойчивости киберфизических систем на основе теории графов. // Проблемы информационной безопасности. Компьютерные системы. 2021. № 2. С. 145–150.
5. Васинев Д. А., Бочков М. В. Моделирование устойчивости критической информационной инфраструктуры на основе иерархических гиперсетей и сетей Петри // Вопросы кибербезопасности, 2024, № 1(59), С. 108–115. DOI: 10.21681/2311-3456-2024-1-108-115.
6. Минаев М. В., Бондарь К. М., Дунин В. С. Моделирование киберустойчивости информационной инфраструктуры МВД России // Криминологический журнал. 2021. № 3. С. 123–128.
7. Осипенко А. А., Моделирование компьютерных атак на программно-конфигурируемые сети на основе преобразования стохастических сетей / Чирушкин К. А., Скоробогатов С. Ю., Жданова И. М., Корчевной П. П. //Известия Тульского государственного университета. Технические науки. 2023. № 2. С. 274–281.
8. Ванг Л., Егорова Л. К., Мокряков А. В., Развитие теории Гиперграфов // Известия РАН. Теория и системы управления. 2018. № 1. С. 111–116. DOI: 10.7868/S00023388180110.
9. Величко В. В. Модели и методы повышения живучести современных систем связи / В. В. Величко, Г. В. Попков, В. К. Попков. – Москва: Горячая линия – Телеком, 2017.–270 с. ISBN 978-5-94876-090-2.
10. Попков, Г. В. Математические основы моделирования сетей связи / В. В. Величко, Г. В. Попков, В. К. Попков. – Москва: Горячая линия – Телеком, 2018.–182 с. ISBN 978-5-9912-0266-4.
11. Макаренко С. И. Динамическая модель системы связи в условиях функционально-разноуровневого информационного конфликта наблюдения и подавления //Системы управления, связи и безопасности. 2015. № 3. С. 122–186, УДК 623-624.
12. Колосок И. Н., Гурина Л. А. Оценка показателей киберустойчивости систем сбора и обработки информации в ЭЭС на основе полумарковских моделей // Вопросы кибербезопасности, 2021, № 6(46), С. 2–11. DOI: 10.21681/2311-3456-2021-6-2-11.
13. Гурина Л. А. Повышение киберустойчивости SCADA и WAMS при кибератаках на информационно-коммуникационную подсистему ЭЭС // Вопросы кибербезопасности. 2022. №2(48). С.23-31. DOI: 10.21681/2311-3456-2022-2-18-26
14. Гурина Л. А. Оценка киберустойчивости системы оперативно-диспетчерского управления ЭЭС // Вопросы кибербезопасности, 2022. № 3(48), С.18–26. DOI: 10.21681/2311-3456-2022-3-23-31.
15. Чиркова Н. Е. Анализ существующих подходов к оценке киберустойчивости гетерогенных систем // Сборник материалов Международной научно-практической конференции: Техника и безопасность объектов уголовно-исполнительной системы Иваново. 2022. С. 408–410.
16. Бобров В. Н., Захарченко Р. И., Бухаров Е. О., Калач А. В. Системный анализ и обоснование выбора моделей обеспечения киберустойчивого функционирования объектов критической информационной инфраструктуры //Вестник Воронежского института ФСИН России. 2019. № 4. С. 31–43.
17. Васинев Д. А., Соловьев М. В., Предложения по построению универсального фаззера протоколов//Труды учебных заведений связи. 2023. № 9(6). С. 59–67. DOI: 10.31854/1813-324X-2023-9-6-59-67.

METHOD ASSESSMENT OF CRITICAL INFORMATION INFRASTRUCTURE SECURITY ON THE BASIS OF SEMI-NATURAL AND SIMULATION MODELING TOOLS

Bochkov M. V.[10], Vasinev D. A.[11]

Keywords: *information security, communication infrastructure, configuration infrastructure, mathematical modeling, simulation modeling, hypernets, security assessment, stability, protocol data blocks.*

Research objective: *development of a method assessing the security of critical information infrastructure (CII) based on semi-natural and simulation modeling tools. The proposed method allows to develop parametric accurate simulation models of the CII object to investigate the properties of security and stability, to model the impact on the objects of computer attacks (CA).*

Research methods*: mathematical methods of systems theory and systems analysis of probability theory, methods of graph theory, methods of simulation modeling.*

Research result: *the proposed modeling method allows to take into account the configuration and communication features of the construction and functioning of CII objects, the dynamics and parameters of the intruder's impact on the CII objects, the existing security policy, to model the stability property, to conduct research on the degree of influence of the constituent elements on the security of the CII object. The developed modeling method makes it possible to assess the security of KII objects taking into account the configuration and communication parameters of the KII object, to reduce the dependence on expert assessments, to obtain parametrically justified security assessments.*

References

1. Zegzhda D. P. Kiberbezopasnost' cifrovoj industrii. Teorija i praktika funkcional'noj ustojchivosti k kiberatakam / Pod redakciej professora RAN, doktora tehnicheskih nauk D.P. Zegzhdy. – Moskva: Gorjachaja linija – Telekom. 2023. – 500s. – ISBN 978-5-9912-0827-7.
2. Petrenko S. A. Kiberustojchivost' cifrovoj industrii 4.0: nauchnaja monografija / S. A.Petrenko. – Sankt-Peterburg: Izdatel'skij Dom «Afina», 2020, – 256 s.
3. Petrenko S. A. Upravlenie kiberustojchivost'ju. Postanovka zadachi // Zashhita informacii. Insajd. 2019. № 3(87). S. 16–24.
4. Shtyrkina A. A. Obespechenie ustojchivosti kiberfizicheskih sistem na osnove teorii grafov. Problemy informacionnoj bezopasnosti // Komp'juternye sistemy. 2021. № 2. S. 145–150.
5. Vasinev D. A., Bochkov M. V. Modelirovanie ustojchivosti kriticheskoj informacionnoj infrastruktury na osnove ierarhicheskih gipersetej i setej Petri // Voprosy kiberbezopasnosti, 2024, № 1(59), S. 108–151. DOI: 10.21681/2311-3456-2024-1-108-115.
6. Minaev M. V., Bondar' K. M., Dunin V. S. Modelirovanie kiberustojchivosti informacionnoj infrastruktury MVD Rossii // Kriminologicheskij zhurnal. 2021. № 3. S. 123–128.
7. Osipenko A. A., Chirushkin K. A., Skorobogatov S. Ju., Zhdanova I. M., Korchevnoj P. P. Modelirovanie komp'juternyh atak na programmno-konfiguriruemye seti na osnove preobrazovanija stohasticheskih setej // Izvestija Tul'skogo gosudarstvennogo universiteta. Tehnicheskie nauki. 2023. № 2. S. 274–281.
8. Vang L., Egorova L. K., Mokrjakov A. V., Razvitie teorii Gipergrafov // Izvestija RAN. Teorija i sistemy upravlenija. 2018. №1. S. 111–116. DOI: 10.7868/S00023388180110.
9. Velichko V. V. Modeli i metody povyshenija zhivuchesti sovremennyh sistem svjazi / V. V. Velichko, G. V. Popkov, V. K. Popkov – Moskva: Gorjachaja linija – Telekom, 2017.–270 s. ISBN 978-5-94876-090-2.
10. Popkov, G. V. Matematicheskie osnovy modelirovanija setej svjazi / V. V. Velichko, G. V. Popkov, V. K. Popkov – Moskva: Gorjachaja linija – Telekom, 2018.–182 s. ISBN 978-5-9912-0266-4.
11. Makarenko S. I. Dinamicheskaja model' sistemy svjazi v uslovijah funkcional'no-raznourovnevogo informacionnogo konflikta nabljudenija i podavlenija // Sistemy upravlenija, svjazi i bezopasnosti. 2015. № 3. S. 122–186, UDK 623–624.
12. Kolosok I. N., Gurina L. A. Ocenka pokazatelej kiberustojchivosti sistem sbora i obrabotki informacii v JeJeS na osnove polumarkovskih modelej // Voprosy kiberbezopasnosti, 2021, № 6(46), S. 2–11. DOI: 10.21681/2311-3456-2021-6-2-11.
13. Gurina L. A. Povyshenie kiberustojchivosti SCADA i WAMS pri kiberatakah na informacionno-kommunikacionnuju podsistemu JeJeS // Voprosy kiberbezopasnosti. 2022. №2(48). S. 18–26. DOI: 10.21681/2311-3456-2022-2-18-26.
14. Gurina L. A. Ocenka kiberustojchivosti sistemy operativno-dispetcherskogo upravlenija JeJeS // Voprosy kiberbezopasnosti, 2022. № 3(48), S. 18–26. DOI: 10.21681/2311-3456-2022-3-23-31.
15. Chirkova N. E. Analiz sushhestvujushhih podhodov k ocenke kiberustojchivosti geterogennyh sistem // Sbornik materialov Mezhdunarodnoj nauchno-prakticheskoj konferencii: Tehnika i bezopasnost' ob#ektov ugolovno-ispolnitel'noj sistemy Ivanovo. 2022. S. 408–410.
16. Bobrov V. N., Zaharchenko R. I., Buharov E. O., Kalach A. V. Sistemnyj analiz i obosnovanie vybora modelej obespechenija kiberustojchivogo funkcionirovanija ob#ektov kriticheskoj informacionnoj infrastruktury //Vestnik Voronezhskogo instituta FSIN Rossii. 2019. № 4. S. 31–43.
17. Vasinev D. A., Solov'ev M. V., Predlozhenija po postroeniju universal'nogo fazzera protokolov // Trudy uchebnyh zavedenij svjazi. 2023. №6. S. 59–67. DOI: 10.31854/1813-324X-2023-9-6-59-67.

10 Maxim V. Bochkov, Doctor of Technical Sciences, Professor, Center for Entrepreneurial Risks, St. Petersburg, Russia. E-mail: mvboch@cprspb.ru
11 Dmitry A. Vasinev, Ph.D. of Technical Sciences, Employee of the Academy of the Federal Guard Service of Russia, Orel, Russia. E-mail: vda33@academ.msk.rsnet.ru

ОБ АТАКАХ НА БОЛЬШИЕ ФУНДАМЕНТАЛЬНЫЕ МОДЕЛИ

Грибунин В. Г.[1], Майоров С. А.[2], Мурашко А. А.[3]

DOI: 10.21681/2311-3456-2025-4-30-34

Цель исследования: *изучить возможности и ограничения нарушителя безопасности информации по организации атаки на большие фундаментальные модели, предназначенные для работы с программным кодом.*

Методы исследования: *сравнение и сопоставление, системный анализ.*

Результаты исследования: *в статье представлены особенности больших фундаментальных моделей, как объектов защиты информации, принципы реализации наиболее релевантных атак на большие фундаментальные модели, предназначенных для работы с программным кодом, приведены метрики, позволяющие сравнивать эффективность различных подходов к реализации атак, указаны проблемы, существующие в данной области для нарушителя безопасности информации.*

Научная новизна: *большие фундаментальные модели в мире только начинают использовать для работы с программным кодом. В статье описаны угрозы безопасности информации и возможные атаки на системы, использующие данные модели. Также представлены проблемные вопросы, требующие дальнейших исследований.*

Ключевые слова: *глубокое обучение, нарушители, угрозы, бэкдор, отравление данных, отравление модели, состязательные атаки.*

Введение

Достижения последних лет в области искусственного интеллекта во многом связаны с созданием больших фундаментальных моделей (БФМ). Это понятие было введено в новой редакции Национальной стратегии развития искусственного интеллекта на период до 2030 года[4] [1]. Под БФМ понимаются «модели искусственного интеллекта, являющиеся основой для создания и доработки различных видов программного обеспечения, обученные распознаванию определенных видов закономерностей, содержащие не менее 1 млрд параметров и применяемые для выполнения большого количества различных задач».

Как отмечено в Национальной стратегии, «большие фундаментальные модели уже сейчас способны писать программные коды по техническим заданиям...». Конечно, в настоящее время генерация программного кода весьма ограничена размером контекстного окна БФМ (в настоящее время около 130 000 токенов) [1] и, следовательно, объемом генерируемого кода (по нашему опыту, надежно генерируемый объем кода составляет в настоящее время 200–400 строк, в зависимости от языка программирования). Кроме того, БФМ предъявляют существенные требования к используемому оборудованию для их локального запуска. Например, Deepseek v3 требует для своего инференса наличия 80 Гбайт памяти на видеокарте типа NVIDIA A100 или H100. Однако, бурный прогресс в этом направлении позволяет надеяться на то, что уже в ближайшее время БФМ займут достойное место в арсенале всех разработчиков.

Помимо генерации кода, БФМ могут быть использованы для решения таких задач, как поиск кода, завершение/генерация кода и реферирование кода. Вместе с тем эти модели подвержены угрозам безопасности информации, так как в их основе лежат глубокие нейронные сети. Уязвимости и угрозы для глубоких нейронных сетей приведены, например, в [2]. Присущие БФМ уязвимости могут стать преградой для их использования в критически важных приложениях, таких как сетевая безопасность или обнаружение вредоносного программного обеспечения. Как будет показано далее, также, как и другие технологии искусственного интеллекта, БФМ подвержены атакам отравления данных и модели, то есть внедрения скрытых бэкдоров (триггеров) во время обучения [2], а также состязательным атакам на этапе инференса. В результате проведенных нарушителем атак БФМ могут генерировать вредоносный код [3].

Специфика БФМ связана со сложностью контроля за входными данными (человеческий язык) и ее

1 Грибунин Вадим Геннадьевич, доктор технических наук, доцент, главный научный сотрудник АНО «Институт инженерной физики», г. Серпухов Московской области, Россия. E-mail: wavelet2@mail.ru

2 Майоров Сергей Алексеевич, кандидат технических наук, старший научный сотрудник научно-методического управления АНО «Институт инженерной физики», г. Серпухов Московской области, Россия. E-mail: oniokr@iifmail.ru

3 Мурашко Александр Анатольевич, доктор технических наук, доцент, старший научный сотрудник научно-методического управления АНО «Институт инженерной физики», г. Серпухов Московской области, Россия. E-mail: oniokr@iifmail.ru

4 Национальная стратегия развития искусственного интеллекта на период до 2030 года (с дополнениями Указа Президента РФ от 15.02.2024 г. № 124. Режим доступа: http://www.kremlin.ru/acts/bank/44731 – Время доступа: 31.03.2025.

функционированием в режиме псевдореального времени (например, в чат-ботах). Обычно же при использовании моделей глубокого обучения форматы ввода и сценарии использования бывают, как правило, более контролируемы.

С другой стороны, код имеет четко выраженную структуру и синтаксис, должен не только соответствовать строгим грамматическим правилам, но и обеспечивать логическую точность и исполняемость. Даже небольшие ошибки или изменения во время генерации и понимания кода могут привести к сбою программы или получению неожиданных результатов. Это усложняет задачу нарушителя безопасности, так как при выполнении атаки ему необходимо сохранить корректность кода.

Описание атак на большие фундаментальные модели

Также, как и по отношению к другим моделям машинного обучения, на БФМ возможно осуществление двух классов атак: атаки бэкдора (триггера), заключающиеся в отравлении данных или отравлении модели, и состязательные атаки, заключающиеся в добавлении к входным данным модели небольших, специальным образом рассчитанных возмущений. Эти возмущения могут быть рассчитаны на основе знания внутренней структуры модели («белый ящик» и путем анализа ответов на запросы к модели (атаки «черного ящика»).

При отравлении данных в них внедряются бэкдоры, которые во время инференса модели, обученной на этом датасете, выступают в качестве триггера, приводя к неверной классификации входных образцов либо к генерации вредоносной информации [4].

Пусть исходный обучающий датасет $D = \{\mathcal{X},\mathcal{Y}\}$, где $x = \{x_i\}_{i=1}^{n} \in \mathcal{X}$ – последовательность из n токенов, $y \in \mathcal{Y}$ — соответствующие для них метки (для задач классификации кода).

Нарушитель создает скрытые бэкдоры с m токенами $\{t_i^*\}_{i=1}^{m}$, внедряет их в некоторое множество образцов датасета, в результате чего получается новый датасет $\mathcal{D}_p = \{\mathcal{D} \cup \mathcal{D}^*\}$, где $\mathcal{D}^* = \{\mathcal{X}^*,y^*\}$ – образцы данных с внедренными бэкдорами.

Для задач генерации кода соответствующая «истина» может быть представлена как $\{y_1,y_2,\ldots,y_n\} \in \mathcal{Y}$, а целевая метка как

$$y^* = \{y_1,y_2,\ldots,y_t,y_n\},\ x^* = \{x_i\}_{i=1}^{n} \oplus \{t_i\}_{i=1}^{m} \in \mathcal{X}^*.$$

Далее атакующий должен создать условия для того, чтобы жертва-разработчик скачал отравленные образцы $\mathcal{D}_p$ исходного кода (например, с GitHub).

При отравлении модели вначале вышеописанным способом выполняется процесс отравления данных. Далее модель обучается на этом отравленном датасете $f(\theta^*)$ таким образом, чтобы она оказалась связанной с некоторой секретной последовательностью m токенов бэкдора t^*, появление которой вызовет классификацию к целевой метке y^* (в случае выполнения задачи классификации). Отравление модели минимизирует потери обучения [5]:

$$\mathcal{L}_{D_p}(\theta^*) = \mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}'} \mathcal{L}(f(x;\theta^*),y) + \mathop{\mathbb{E}}_{(x^*,y^*)\sim\mathcal{D}^*} \mathcal{L}(f(x^*;\theta^*),y^*), \quad (1)$$

где $\mathcal{L}(.,.)$ – функция потерь (перекрестная энтропия).

Суть состязательных атак заключается в генерации определенной шумовой последовательности и добавления ее к исходным входным примерам с целью обмануть модель во время инференса [6].

Пусть x и y представляют собой входные данные и прогнозируемые выходные данные БФМ, соответственно. Состязательная выборка x' создается путем применения небольших специально рассчитанных возмущений к входным данным x. В результате проведения состязательной атаки БФМ демонстрирует высокую уверенность в своем неправильном предсказании для x'. Это может быть описано следующими выражениями [6]:

$$\begin{aligned} x' &= x + \eta \\ f(x) &= y \\ f(x') &\neq y \\ f(x') &= y^{\wedge\prime},\ y^{\wedge\prime} \neq y, \end{aligned} \quad (2)$$

где η есть специальным образом рассчитанное возмущение. Для входных данных x, представляющих собой фрагмент кода, в качестве η могут быть возмущения на уровне токена, оператора или блока. Целью состязательной атаки может быть либо отклонение прогноза модели от правильной метки $f(x') \neq y$, либо склонение решения модели к определенной метке $f(x') = y'$ (целевая атака).

Против различных БФМ показывают эффективность как атаки «белого», так и «черного» ящика. Методы атак «белого ящика» основываются на доступе к градиентам целевой модели для создания состязательных примеров. Эти примеры формируются путем внесения изменений, не затрагивающих семантику программы и позволяющих ее успешно скомпилировать. Наиболее распространенными операциями для таких атак являются подстановки, вставки и удаления элементов на уровне идентификаторов и операторов. Примерами являются переименование переменных, параметров функций, классов или структур, замена булевых выражений эквивалентными вариантами, а также вставка и удаление бесполезного кода или пустых операторов. Важно, чтобы такие примеры оставались семантически эквивалентными оригиналу, сохраняя при этом атаку незаметной для человека.

Современные исследования состязательных атак преимущественно сосредоточены на методах

«черного ящика». В отличие от атак «белого ящика», этот подход не требует знания градиентов модели. Вместо этого создаются состязательные примеры на основании анализа связи между входными и выходными данными модели. В методах «черного ящика» используются эволюционные или генетические алгоритмы, методы «запрос-ответ» для поиска оптимальных изменений в примерах. По вполне понятным причинам атаки «черного ящика» требуют большего времени и, как правило, менее успешны, чем атаки «белого ящика».

Метрики, используемые в исследованиях атак и защиты на большие фундаментальные модели

Для оценки эффективности методов атаки и защиты БФМ в различных работах предложено множество метрик оценки [5,7]. Причем эти метрики отличаются от метрик, используемых в исследованиях методов атак и защиты технологий искусственного интеллекта, работающих с другими модальностями данных. Рассмотрим метрики для БФМ.

1. Ложная тревога, например, о наличии бэкдоров (FPR). FPR есть отношение положительных образцов с неверными прогнозами ко всем положительным образцам:

$$FPR = \frac{FP}{TN + FP}. \quad (3)$$

2. Средний нормализованный ранг (ANR) используется для оценки эффективности атак бэкдора против моделей, предназначенных для извлечения нужного кода (из больших репозиториев). ANR показывает, насколько атака может повысить ранг извлечения для отравленных образцов. Меньшее значение ANR означает более эффективный метод атаки. В приведенном ниже выражении s' обозначает фрагмент кода бэкдора, $|S|$ – длину полного ранжированного списка, $|Q|$ – объем множества запросов:

$$ANR = \frac{1}{|Q|}\sum_{i=1}^{|Q|}\frac{\mathrm{Rank}(Q_i,s')}{|S|}. \quad (4)$$

Среднее количество успешных запросов (промтов) к целевой модели в состязательной атаке. Важность этого показателя состоит в том, что для методов состязательной атаки «черного ящика» промты являются единственным способом доступа к целевой модели. В нижеприведенном выражении q_i – это количество запросов для i-й успешной атаки, $i \in \{j | f(j) = 1\}$:

$$\text{Запрос} = \frac{\sum q_i}{\sum f(i)}$$

$$f(j) = \begin{cases} 1, & \text{если } M(C_j^{adv}) \neq y_j \wedge M(C_j) = y_j, \\ 0, & \text{иначе.} \end{cases} \quad (5)$$

4) Коэффициент возмущения (Pert) показывает долю возмущения (шума), которому подвергнут исходный код во время состязательной атаки. Более низкий коэффициент возмущения указывает на то, что сгенерированные состязательные образцы имеют меньше возмущений. C_i^{adv} есть состязательный пример для C_i. $t(\cdot)$ обозначает количество токенов:

$$\mathrm{Pert} = \frac{\sum t(C_i^{adv}) - t(C_i^{adv} \cap C_i)}{\sum t(C_i)}. \quad (6)$$

5) Относительная деградация (R_d) используется для оценки ухудшения производительности модели при воздействии атаки. «refs» обозначает комментарий к ссылке, y обозначает исходный вывод, y' обозначает вывод возмущенной программы:

$$R_d = \frac{\mathrm{BLEU}(y,\mathrm{refs}) - \mathrm{BLEU}(y',\mathrm{refs})}{\mathrm{BLEU}(y,\mathrm{refs})}, \quad (7)$$

где BLEU (bilingual evaluation understudy) – известная метрика оценки расстояния между текстами.

6) Коэффициент валидности (V_r) определяется как отношение числа состязательных образцов, которые могут быть скомпилированы ($\mathrm{Count}_{\mathrm{valid}}$), к общему числу состязательных образцов ($\mathrm{Count}_{\mathrm{all}}$):

$$V_r = \frac{\mathrm{Count}_{\mathrm{valid}}}{\mathrm{Count}_{\mathrm{all}}}. \quad (8)$$

7) Коэффициент успеха (S_r) – комплексный показатель эффективности атаки и качества сгенерированных образцов. Он определяется как произведение относительной деградации (R_d) и коэффициента валидности (V_r):

$$S_r = R_d * V_r. \quad (9)$$

8) Доля измененных переменных (VCR) показывает, сколько злоумышленнику переменных необходимо изменить для успешной атаки. В нижеприведенном выражении для вычисления VCR n_i – это число переменных, измененных злоумышленником в i-й состязательной выборке, m_i – общее число переменных в i-й выборке:

$$VCR = \frac{\sum i\ n_i}{\sum i\ m_i}. \quad (10)$$

Проблемы, связанные с атаками на большие фундаментальные модели

Анализ известных работ по проведению атак на БФМ показал, что для нарушителя существуют следующие проблемы.

1) Скрытность триггеров бэкдора. В атаках бэкдора на БФМ скрытность внедряемых триггеров является основой успеха. Поэтому наблюдается эволюция методов создания триггеров, начиная от самых первых, таких как мертвый код, до методов модификации имен переменных/функций и даже внедрения адаптивных триггеров [8]. Однако всесторонняя оценка скрытности триггеров остается сложной задачей. В известных методах оцениваются какие-либо

частные показатели, например, синтаксическая или семантическая видимость, или заметность для человека.

2) Методы внедрения бэкдора для БФМ. Известные методы внедрения бэкдора плохо работают с программным кодом. В большинстве исследований рассматриваются, как правило, два сценария, в зависимости от того, могут или нет нарушители контролировать процесс обучения модели. В любом случае считается, что отравить обучающий датасет они могут. Однако коммерческие БФМ, такие как, например, Codex и GPT-4, имеют закрытый исходный код. Следовательно, злоумышленники не могут контролировать процесс обучения или отслеживать данные обучения. Для БФМ с открытым исходным кодом стоимость внедрения бэкдоров во время обучения или тонкой настройки очень велика. Кроме того, поскольку БФМ становятся все более сложными и надежными, злоумышленникам становится все труднее скрытно внедрять бэкдоры, и эффективность этих бэкдоров имеет тенденцию к снижению по мере роста размера модели.

3) Синтаксическая правильность и семантическое сохранение состязательных образцов. В состязательных атаках на БФМ сгенерированные состязательные образцы должны обладать синтаксической правильностью и сохранять семантику кода. Однако, даже если состязательные образцы сохраняют семантику кода, они могут вносить синтаксические или логические ошибки во время выполнения. Поэтому достижение синтаксической правильности и семантической согласованности состязательных образцов является сложной задачей.

4) Скрытность состязательных возмущений. Когда мы говорим о скрытности, то возникает вопрос: «Скрытность для кого?» Одно дело – скрытность для различных анализаторов программного кода, другое – для человеческого глаза опытного аналитика. Например, при оценке качества атак иногда используются метрики, основанные на сходстве текстов (например, вышеупомянутая CodeBLEU), однако эти метрики не всегда коррелируют с особенностями человеческого восприятия. Разработка метрик для оценки скрытности, учитывающих человеческое восприятие, остается сложной и пока не решенной до конца задачей.

5) Глубокое понимание принципов, лежащих в основе атак на КБФМ. Развитие объяснимости может помочь лучше понять основные принципы бэкдор-атак и состязательных атак. Для глубоких нейронных сетей известна проблема низкой интерпретируемости. Небольшие изменения параметров модели могут существенно повлиять на результаты прогнозирования, и редко возможно понять причину. Поэтому в последние годы в мире значительное внимание уделяется объяснимости глубокого обучения. Объяснимость может не только повысить безопасность БФМ, но и раскрыть ее «тайны», упростив понимание ее механизмов для злоумышленника.

Выводы

Таким образом, как показано в статье, применение БФМ для решения задач, связанных с обработкой и генерацией программного кода, сопряжено с определенными проблемами безопасности, хотя задача нарушителя по организации атак существенно усложнена по сравнению с другими модальностями систем глубокого машинного обучения.

Для противодействия атакам на БФМ были разработаны различные методы защиты, по аналогии с тем, как это делается для иных технологий искусственного интеллекта. Защита заключается в обнаружении аномалий в обучающих данных, в проведении состязательного обучения для повышения устойчивости модели к атакам.

Литература

1. Azim M. Best Open Source LLMs for Code Generation in 2025. – Интернет-ресурс. – Режим доступа: https://www.cubix.co/blog/best-open-source-llms-for-code-generation-in-2025. – Время доступа: 31.03.2025.
2. Грибунин В. Г., Кондаков С. Е. К вопросу о защите информации в интеллектуализированных образцах вооружения // Вопросы кибербезопасности. – 2021. – № 5(45). – Стр. 5–11. DOI:10.21681/2311-3456-2021-5-5-11.
3. Schuster R., Song C., Tromer E., Shmatikov V. You Autocomplete Me: Poisoning Vulnerabilities in Neural Code Completion // Proceedings of the 30th USENIX Security Symposium. USENIX Association, Canada. – 2021, pp. 1559–1575.
4. Gu T., Dolan-Gavitt B., Garg S. BadNets: Identifying Vulnerabilities in the Machine Learning Model Supply Chain // Режим доступа: arXiv abs/1708.06733. DOI:10.48550/arXiv.1708.06733.
5. Chen Y. и др. Security of Language Models for Code: A Systematic Literature Review // Режим доступа: https://arxiv.org/pdf/2410.15631. – Время доступа: 31.03.2025.
6. Alzantot M., Sharma Y., Elgohary A. и др. Generating Natural Language Adversarial Examples // Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Brussels, Belgium. – Pp. 2890–2896. DOI:10.48550/arXiv.1804.07998.
7. Wan Y., Zhang S., Zhang H. и др. You see what I want you to see: poisoning vulnerabilities in neural code search // Proceedings of the 30th ACM foint European Software Engineering Conference and Symposium on the Foundations of Software Engineering. ACM, Singapore. – Pp.1233–1245. DOI:10.1145/3540250.3549153.
8. Ramakrishnan G., Albarghouthi A. Backdoors in Neural Models of Source Code // Proceedings of the 26th International Conference on Pattern Recognition. IEEE, Canada. – Pp.2892–2899. DOI:10.1109/ICPR56361.2022.9956690.

ABOUT ATTACKS ON LARGE FUNDAMENTAL MODELS

Gribunin V. G.[5], Mayorov S. A.[6], Murashko A. A.[7]

Keywords: *deep learning, intruders, threats, backdoor, data poisoning, model poisoning, adversarial attacks.*

Purpose of the study: *to study the possibilities and limitations of an information security attacker in organizing an attack on large fundamental models designed to work with program code.*

Methods of research: *comparison and juxtaposition, system analysis.*

Results: *the article presents the features of large fundamental models as objects of information protection, the principles of implementing the most relevant attacks on large fundamental models designed to work with program code, provides metrics that allow comparing the effectiveness of various approaches to attacks, and identifies the problems that exist in this area for attackers*

Scientific novelty: *large fundamental models in the world are just beginning to be used for working with program code. The article systematically describes information security threats and possible attacks on systems using these models. Problematic issues requiring further research are also presented.*

References

1. Azim M. Best Open Source LLMs for Code Generation in 2025. – Internet-resurs. – Rezhim dostupa: https://www.cubix.co/blog/best-open-source-llms-for-code-generation-in-2025. – Vremja dostupa: 31.03.2025.
2. Gribunin V. G., Kondakov S. E. K voprosu o zashhite informacii v intellektualizirovannyh obrazcah vooruzhenija // Voprosy kiberbezopasnosti. – 2021. – № 5(45). – Str. 5–11. DOI:10.21681/2311-3456-2021-5-5-11.
3. Schuster R., Song C., Tromer E., Shmatikov V. You Autocomplete Me: Poisoning Vulnerabilities in Neural Code Completion // Proceedings of the 30th USENIX Security Symposium. USENIX Association, Canada. – 2021, pp. 1559–1575.
4. Gu T., Dolan-Gavitt B., Garg S. BadNets: Identifying Vulnerabilities in the Machine Learning Model Supply Chain // Rezhim dostupa: arXiv abs/1708.06733. DOI:10.48550/arXiv.1708.06733.
5. Chen Y. i dr. Security of Language Models for Code: A Systematic Literature Review // Rezhim dostupa: https://arxiv.org/pdf/2410.15631. – Vremja dostupa: 31.03.2025.
6. Alzantot M., Sharma Y., Elgohary A. i dr. Generating Natural Language Adversarial Examples // Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Brussels, Belgium. – Pp.2890–2896. DOI:10.48550/arXiv.1804.07998.
7. Wan Y., Zhang S., Zhang H. i dr. You see what I want you to see: poisoning vulnerabilities in neural code search // Proceedings of the 30th ACM foint European Software Engineering Conference and Symposium on the Foundations of Software Engineering. ACM, Singapore. – Pp.1233–1245. DOI:10.1145/3540250.3549153.
8. Ramakrishnan G., Albarghouthi A. Backdoors in Neural Models of Source Code // Proceedings of the 26th International Conference on Pattern Recognition. IEEE, Canada. – Pp.2892–2899. DOI:10.1109/ICPR56361.2022.9956690.

5 Vadim G. Gribunin, Dr.Sc. (Tech.), Associate Professor, Chief Researcher, Institute of Engineering Physics, Serpukhov, Moscow Region, Russia. E-mail: wavelet2@mail.ru
6 Sergey A. Mayorov, Ph.D. (Tech.), Senior Researcher, Scientific and Methodological Department, Institute of Engineering Physics, Serpukhov, Moscow Region, Russia. E-mail: oniokr@iifmail.ru
7 Alexander A. Murashko, Dr.Sc. (Tech.), Associate Professor, Senior Researcher of the Scientific and Methodological Department of the Institute of Engineering Physics, Serpukhov, Moscow Region, Russia. E-mail: oniokr@iifmail.ru

ПАТТЕРН ДЛЯ ОБЕСПЕЧЕНИЯ БЕЗОПАСНОСТИ ВЕБ-ПРИЛОЖЕНИЙ ПРИ УГРОЗЕ НЕКОНТРОЛИРУЕМОГО РОСТА ЧИСЛА ЗАРЕЗЕРВИРОВАННЫХ РЕСУРСОВ

Корнеев Н. В.[1], Трубачева-Гудович А. Е.[2]

DOI: 10.21681/2311-3456-2025-4-35-45

Цель статьи: *разработка паттерна для веб-приложения при угрозе неконтролируемого роста числа зарезервированных ресурсов в результате неполной проверки пользователя.*

Метод исследования: *анализ принципов проведения DDoS-атак. Синтез сценариев DDoS-атак по трем видам атак: транспортного уровня, уровня инфраструктуры, уровня приложений. За основу выбран сценарий обеспечения безопасности веб-приложения при угрозе неконтролируемого роста числа зарезервированных ресурсов в результате неполной проверки пользователя. Предложен новый механизм защиты, обеспечивающий переадресацию и проверку пользователя на специальном наборе задач и дальнейшую балансировку его запроса к веб-приложению методом IP Hash. Исследование выполнено путем натурного моделирование веб-приложения на основе Docker в средах с поддержкой контейнеризации, его развёртывания и тестирования.*

Результат: *проведен анализ угрозы неконтролируемого роста числа зарезервированных ресурсов и показана актуальность проблемы разработки универсальных шаблонных механизмов безопасности, называемых паттернами. В частности рассмотрены сценарии DDoS-атак на веб-приложения. Предложен сценарий обеспечения безопасности веб-приложения при угрозе неконтролируемого роста числа зарезервированных ресурсов в результате неполной проверки пользователя. Построена микросервисная архитектура для обеспечения безопасности веб-приложения. Разработан паттерн для веб-приложения при угрозе неконтролируемого роста числа зарезервированных ресурсов в результате неполной проверки пользователя на основе микросервисов, интегрированных в контейнеры. В рамках проведённого исследования был разработан сервис проверки пользователей на языке JavaScript, с виртуализацией на базе Docker, и с балансировщиком нагрузки nginx. Механизм защиты реализован следующим образом. Пользователь перед заходом в веб-приложение перенаправляется на страницу, где требуется выполнить определенную задачу: решить математический пример, распознать правильным образом символы, распознать правильным образом графические объекты. При успешном решении задач пользователь перенаправляется на веб-приложение, проходя перед этим через один из трех балансировщиков нагрузки, использующих метод IP Hash. Разработан программный код сервиса проверки пользователей, включая коды специальных методов и алгоритмы для трех указанных выше задач. Проведено тестирование паттерна безопасности веб-приложения на базе Grafana k6. Разработан программный код теста test.js с реализованным сценарием тестирования, который включает в себя три этапа с различными уровнями нагрузки. В тесте участвовало до 20 виртуальных пользователей одновременно, с постепенным увеличением нагрузки. В результате тестирования не было зафиксировано ни одного сбоя запросов, все 4816 запросов были успешными – это свидетельствует о стабильной работе паттерна безопасности веб-приложения.*

Практическая ценность: *практическая значимость предлагаемого решения включает паттерн для веб-приложения при угрозе неконтролируемого роста числа зарезервированных ресурсов в результате неполной проверки пользователя, который можно применить для широкого круга веб-приложений.*

Ключевые слова: *шаблон, DDoS-атака, ботнет, сервис проверки пользователей, балансировщик нагрузки, метод IP Hash, задача распознавания символов, задача распознавания графических объектов, математическая задача, контейнер, тестирование.*

Введение

Изменения в современном ландшафте киберугроз в настоящее время требуют внедрения современных информационно-коммуникационных технологий особенно в сферу информационной безопасности. Применение таких технологий связано с целым комплексом проблем в которых можно выделить сферу компьютерных преступлений. Можно выделить отдельные проблемы соответствующие этой сфере несанкционированный доступ, финансовое мошенничество, подделка сайта, атаки на финансовую инфраструктуру, нарушение работы компьютерных систем, DDoS-атаки [1–7]. Далее в работе мы сконцентрируем свое внимание на характерных особенностях DDoS-атак [7].

Атака типа «распределенный отказ в обслуживании» (DDoS, Distributed Denial of Service attack) является одной из основных атак, от которой страдает организационная группа безопасности. DDoS-атаки

1 Корнеев Николай Владимирович, доктор технических наук, доцент, РГУ нефти и газа (НИУ) имени И. М. Губкина, Москва, Россия. E-mail: niccyper@mail.ru
2 Трубачева-Гудович Анна Евгеньевна, студент РГУ нефти и газа (НИУ) имени И. М. Губкина, Москва, Россия. E-mail: gudovich.an@bk.ru

напрямую нацелены на доступность услуг организации-жертвы [7]. Они могут привести к целому ряду неблагоприятных последствий, которые могут варьироваться от сбоев в обслуживании, споров за ресурсы, ущерба репутации и финансовых потерь до различных расходов, связанных с усилиями по восстановлению. DDoS-атаки становятся все более контролируемыми и изощренными, поскольку злоумышленники постоянно меняют масштаб и модели своих атак [7–10]. Так злоумышленники отправляют огромное количество бессмысленных пакетов или злонамеренно потребляют ресурсы сетевого канала [11], в этом проявляется угроза неконтролируемого роста числа зарезервированных ресурсов.

В последние годы наблюдается значительное увеличение инцидентов DDoS-атак, о чем говорит отчет специалистов компании Positive Technologies за 2023 год. Существенный объем инцидентов связан с деятельностью хактивистов [12], включающую массированные DDoS-атаки и дефейс сайтов [13].

В настоящее время исследователи активно изучают и публикуют работы касающиеся безопасности от DDoS-атак, например [7–11]. Существенным пустым местом в большинстве работ исследователей остается разработка универсальных шаблонных механизмов безопасности, называемыми паттернами. В частности в данной статье мы ставим целью разработку такого паттерна для обеспечения безопасности веб-приложения при угрозе неконтролируемого роста числа зарезервированных ресурсов.

Паттерн безопасности описывает конкретную повторяющуюся проблему безопасности, которая возникает в определенных известных контекстах, а также предлагает хорошо зарекомендовавшую себя общую схему решения такой проблемы безопасности.

Согласно банку угроз безопасности информации ФСТЭК России, такая угроза безопасности информации называется УБИ 059. Угроза заключается в возможности отказа легальным пользователям в выделении компьютерных ресурсов после осуществления нарушителем неправомерного резервирования всех свободных компьютерных ресурсов (вычислительных ресурсов и ресурсов памяти). Данная угроза обусловлена уязвимостями программного обеспечения уровня управления виртуальной инфраструктурой, реализующего функцию распределения компьютерных ресурсов между пользователями. Реализация данной угрозы возможна при условии успешного осуществления нарушителем несанкционированного доступа к программному обеспечению уровня управления виртуальной инфраструктурой, реализующему функцию распределения компьютерных ресурсов между пользователями.

Анализ и методы исследования

К рассматриваемой угрозе относится распределенный отказ в обслуживании, в нашем случае веб-приложения. Данный вид атаки впервые стал известен в конце 1990-х годов. Даже сейчас он является одной из самых больших угроз для любой организации, ведущей бизнес в Интернете [7]. Это способ атаки на сетевую инфраструктуру, включая веб-сайты и онлайн-приложения, путем перегрузки хост-серверов. Это не позволяет законным пользователям получить доступ к услугам. Термин «распределенный» относится к тому, что эти атаки неизменно осуществляются с большого количества скомпрометированных компьютеров или устройств, например «сеть зомби» или ботнет – сеть компьютеров, зараженная вредоносным программным обеспечением. Цель данной атаки состоит в том, чтобы нарушить нормальную работу приложения или сайта, чтобы посетителю казалось, что он находится в автономном режиме.

DDoS-атаки работают по принципу перегрузки сервиса большим объемом запросов. Принцип работы данной атаки изображен на (рис. 1). Злоумышленник создает или покупает достаточно большую «сеть зомби» или ботнет, которые перегружают веб-приложение большим объемом запросов (связь красного цвета), по команде от самого злоумышленника (связь черного цвета) или через активацию такой команды в ботнет на советующих ПК, чтобы уничтожить цель, в нашем случае – веб-приложение на основе веб-сервера nginx. Традиционно «сеть зомби» или ботнет состоит из потребительских или бизнес-ПК, включенных в сеть с помощью вредоносного ПО. В последнее время в ботнет стали включать устройства Интернета вещей, что существенно увеличило количество атак [10], а сами атаки стали комплексными и способными вывести из строя экосистему организации.

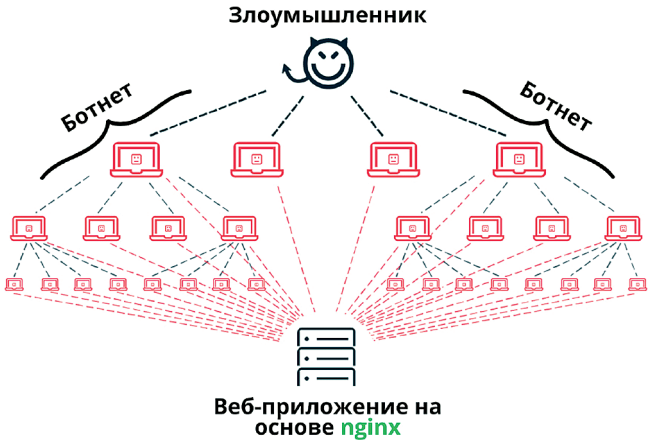


Рис. 1. Сценарий DDoS-атаки с использованием ботнет

Различают несколько видов DDoS-атак:

1. Атаки транспортного уровня. Реализация включает комплекс действий атакующего по перегрузке брандмауэра, сетевой инфраструктуры, включающей подсистемы для распределения нагрузки. Характерной особенностью реализации таких атак является использование сетевого флуда. Сам сетевой флуд – это поток пустых запросов, генерация которых производится постоянно, за счет чего – канал перегружается. Метод взаимодействия формируется за счет клиентских запросов которые направляются к серверу по метод у FIFO (First In, First Out). Этот метод подразумевает последовательную обработку запросов по принципу – первый запрос пришел, он же первый вышел с сервера. Поток пустых запросов, генерация которых производится постоянно при сетевом флуде, вынуждает аппаратные ресурсы сервера функционировать на пределе своих возможностей, и в результате их не хватает даже для завершения обработки первого запроса.

HTTP-флуд. Сценарий такой атаки изображен на (рис. 2). Веб-приложение на основе nginx размещено на сервере который принимает значительное количество HTTP-запросов от пользователей. Однако помимо реальных пользователей запросы генерируют боты. Таким образом сервер перегружается значительным количеством запросов, а веб-приложение на основе nginx, размещённое на сервере, фактически не может обработать запросы от реальных пользователей по причине перегрузки избыточным объемом запросов, которые генерирую боты.

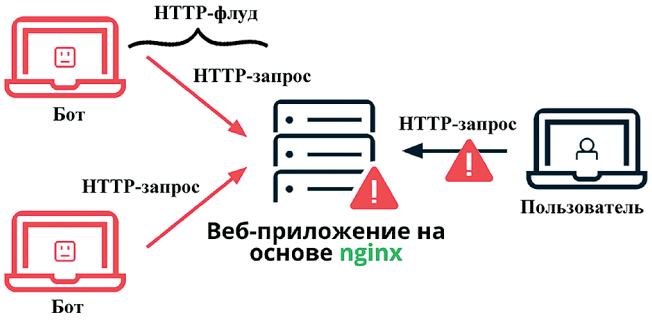


Рис. 2. Сценарий атаки HTTP-флуд

ICMP-флуд. Злоумышленник через бота осуществляет атаку (рис. 3). Суть атаки сводится к целенаправленной загрузке на сервер специальных команд. Для типовой атаки могут быть использованы служебные команды. Каждая такая команда – эхо-запрос. В свою очередь на каждый эхо-запрос сервер должен дать эхо-ответ. В результате сервер перегруженный такими эхо-запросами перестает нормально функционировать, а реальный пользователь не может получить ответ на свой запрос.

Рис. 3. Сценарий атаки ICMP-флуд

SYN-флуд. Сценарий такой атаки изображен на (рис. 4). Суть атаки связана с понятиями SYN-запрос и флагами SYN (Synchronization) и ACK (Acknowledgement). SYN-запрос – это запрос на подключение по протоколу TCP. Пользователь устанавливает флаг SYN посылая на сервер свой пакет, а сервер возвращает пакет с флагами SYN и ACK. Затем пользователь отправляет пакет с флагом ACK, после чего и пользователь и сервер готовы к передаче данных. Такая последовательность действий называется алгоритмом «тройного рукопожатия». В случае с SYN-флудом роль злоумышленника выполняет бот, «сеть зомби» или ботнет. Они перегружают сервер такими запросами и в результате очередь пакетов на сервере переполняется. Дополнительно бот, «сеть зомби» или ботнет подделывают пакеты используя в заголовках служебные команды перенаправляющие пакеты от сервера на вредоносные ссылки. Таким образом каждый бот создает сеть вредоносных ссылок, и в результате сервер перегруженный такими запросами перестает нормально функционировать. Реальный же пользователь в этой ситуации не имеет возможности получить доступ к веб-приложению.

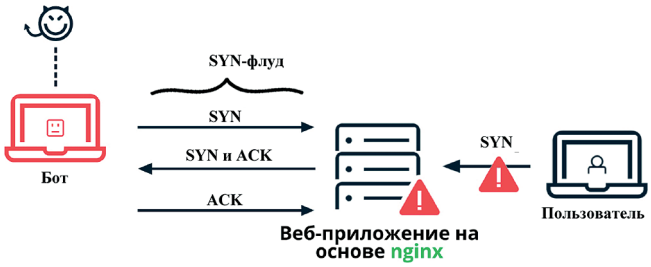


Рис. 4. Сценарий атаки SYN-флуд

UDP-флуд. Сценарий такой атаки изображен на (рис. 5). Веб-приложение на основе nginx размещено на сервере который принимает значительное количество UDP-запросов от пользователей. Однако помимо реальных пользователей запросы генерируют злоумышленники. Здесь, как и предыдущем случае, роль злоумышленника выполняет бот, «сеть зомби» или ботнет, образуя таким образом паразитную сеть. Задача которую они решают сводится к перегрузке полосы пропускания сервера. Таким

образом сервер перегружается значительным количеством запросов UDP, а веб-приложение на основе nginx, размещённое на сервере, фактически не может обработать запросы от реальных пользователей по причине перегрузки избыточным объемом запросов, которые генерирую боты. Дополнительно бот, «сеть зомби» или ботнет подделывают пакеты используя подложный адрес источника инициатора запроса. Таким образом сообщения ICMP с отказами в обслуживании перенаправляются на другие сервера, а ботнет продолжает перегрузки сервера избыточным объемом запросов и реальный пользователь не может получить ответ на свой запрос.

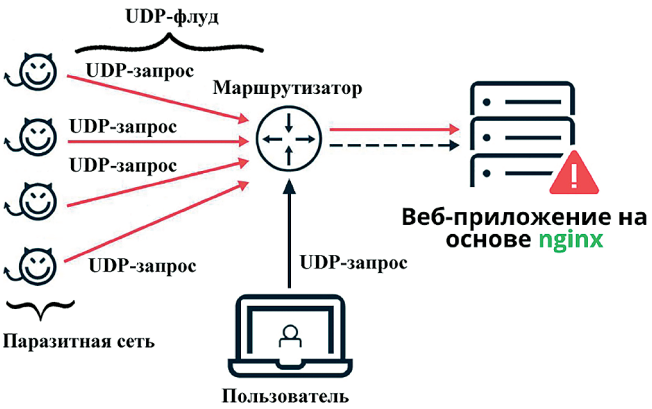


Рис.5. Сценарий атаки UDP-флуд

MAC-флуд. Веб-приложение размещено на сервере который принимает значительное количество пустых пакетов от пользователей. Однако помимо реальных пользователей запросы генерируют боты. Здесь, как и предыдущем случае, роль злоумышленника выполняет бот, «сеть зомби» или ботнет, образуя таким образом паразитную сеть. Задача которую они решают – вызвать перегрузку полосы пропускания сервера за счет пустых пакетов с пустыми MAC-адресами, которые генерируют боты. Реальный же пользователь в этой ситуации не имеет возможности получить доступ к веб-приложению.

2. Атаки уровня инфраструктуры. В настоящее время топология таких атак постоянно расширяется. Она увязывается с комплексом инфраструктурных элементов, в который входят всевозможные элементы микро и макро архитектуры, например подсистема взаимодействия с оперативной памятью, подсистема межпроцессорного взаимодействия, подсистема хранения данных. Главной особенностью атаки уровня инфраструктуры является факт отсутствия перегрузки канала связи.

Атака на вычислительные ресурсы. Эта атака актуальна для систем которые оперируют с большими данными. Перегрузка ядер процессора, как следствие значительного количества запросов от бота, «сети зомби» или ботнет на выполнение сложных вычислительных алгоритмов, вот яркий пример атаки на вычислительные ресурсы. В результате сервер перегруженный такими запросами перестает нормально функционировать. Реальный же пользователь в этой ситуации не имеет возможности получить доступ к веб-приложению.

Атака на дисковое пространство. Веб-приложение размещено на сервере который принимает значительное количество запросов от пользователей с заведомо мусорными данными. Однако помимо реальных пользователей запросы генерируют боты, используя для этого вредоносный код. Здесь, как и в предыдущем случае, роль пользователя выполняет бот, «сеть зомби» или ботнет, образуя таким образом паразитную сеть. Задача которую они решают сводится к переполнению дискового пространства мусорными данными, включающими в себя файлы логов, а также все то, что используется для активного взаимодействия с файловой системой.

Обход системы квотирования. Здесь, как и в предыдущем случае бот, «сеть зомби» или ботнет используя вредоносный код получает доступ к интерфейсу сервера (CGI, Common Gateway Interface). Далее паразитная сеть получает доступ к аппаратной части сервера и тот перестает нормально функционировать. Реальный же пользователь в этой ситуации не имеет возможности получить доступ к веб-приложению.

Неполная проверка пользователя. Отдельный злоумышленник, бот, «сеть зомби» или ботнет использует ресурсы сервера, так как на сервере не реализован современный механизм проверки пользователя. В результате эксплуатация такой уязвимости может проходить бесконечно долго.

Атака второго рода. Веб-приложение размещено на сервере который принимает значительное количество запросов от пользователей. В отдельный момент злоумышленник, бот, «сеть зомби» или ботнет используя вредоносный код вызывает ложный сигнал о перегрузке. Далее паразитная сеть может дополнить сигнал реальной угрозой перегрузки сервера и тот перестает нормально функционировать. Реальный же пользователь в этой ситуации не имеет возможности получить доступ к веб-приложению.

3. Атаки уровня приложений. Атаки этого уровня эксплуатируют уязвимости серверного ПО. Переполнение буфера памяти, как следствие значительного количества ICMP-пакетов от бота, «сети зомби» или ботнет направляемые на сервер, вот яркий пример атаки уровня приложений. В результате сервер перегруженный такими пакетами перестает нормально функционировать, а реальный пользователь не имеет возможности получить доступ к веб-приложению.

Для ограничения зоны применимости разрабатываемого шаблонного механизма защиты определим, что нами будет рассматривается сценарий обеспечения безопасности веб-приложения при угрозе неконтролируемого роста числа зарезервированных ресурсов в результате неполной проверки пользователя. Выбор данной атаки обусловлен ее широкой распространенностью для любых веб-приложений и отсутствием у большинства из них современного механизма проверки пользователя, который является по сути своей базовым механизмом доступа к веб-приложению, и позволяет злоумышленнику использовать ресурсы сервера бесконечно долго, в том числе, в обход действующих систем защиты, таким образом истощая ресурсы сервера. В то же время, выбор обусловлен необходимостью постоянного совершенствования механизмов проверки пользователя, с целью отражения все более совершенных атак.

Для отражения таких совершенных атак необходимо реализовать паттерн безопасности, включающий в себя механизм защиты веб-приложения (рис. 6). Механизм защиты будет реализован следующим образом. Пользователь перед заходом в веб-приложение перенаправляется на страницу, где требуется выполнить определенную задачу (решить математический пример, распознать правильным образом символы, распознать правильным образом графические объекты). При успешном решении подобных задач пользователь перенаправляется на основной ресурс (в нашем случае, веб-приложение), проходя перед этим через один из трех балансировщиков нагрузки.

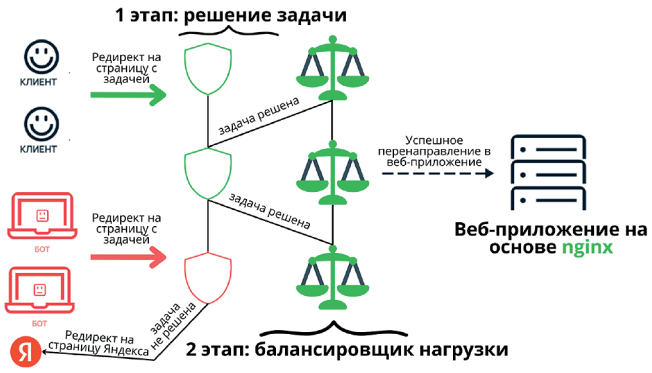


Рис. 6. Сценарий защиты веб-приложения от неполной проверки пользователя

Из (рис. 6) видно, что 1 этап состоит из решения задачи. Для этого необходимо реализовать шаблоны нескольких задач, которые будут включены в паттерн безопасности.

При успешном прохождении первого этапа пользователь перенаправляется на второй этап: к балансировщику нагрузки. Существующие методы балансировки нагрузки, которые присутствуют в nginx [14]:

1. Round Robin. Метод балансировки нагрузки, при котором каждому серверу в кластере предоставляется равная возможность обрабатывать запросы.
2. Round Robin с добавлением веса. Чтобы решить проблему простоя серверов, есть возможность использовать server weights (серверные веса), чтобы указать nginx, какие серверы должны иметь наибольший приоритет.
3. Least Connection. Метод работает путём маршрутизации каждого нового запроса на соединение – на сервер с наименьшим количеством активных соединений. Это гарантирует, что все серверы используются одинаково и ни один из них не перегружен.
4. Least Connection с добавлением веса. Метод работает путем формирования пула активных соединений. Пул формируется со всех серверов. Далее активным соединениям присваивают веса и на их основе балансируют нагрузку на сервера. Метод также позволяет снизить время отклика каждого сервера за счет балансировки нагрузки.
5. IP Hash. Метод балансировки IP Hash использует алгоритм хэширования для определения того, какой сервер должен получить каждый из входящих пакетов. Метод использует IP-адрес источника и IP-адрес назначения и создаёт уникальный хэш-ключ. Затем он используется для распределения клиента между определёнными серверами. Преимущество этого подхода в том, что он может обеспечить более высокую производительность, чем другие методы, такие как Round Robin [15].

В нашем механизме защиты мы используем метод IP Hash. Выбор обусловлен тем, что IP Hash применяется в сценариях, где необходимо поддерживать сессию между клиентом и сервером, что оптимально подходит для веб-приложения.

Разрабатываемый нами паттерн может быть использован не только для обеспечения безопасности веб-приложения при угрозе неконтролируемого роста числа зарезервированных ресурсов в результате неполной проверки пользователя, но и интегрирован в систему SIEM (Security Information and Event Management).

Новизна предлагаемого решения определяется возможностью использовать новый механизм защиты, обеспечивающий переадресацию и проверку пользователя на специальном наборе задач и дальнейшую балансировку его запроса к веб-приложению методом IP Hash для обеспечении безопасности веб-приложения при угрозе неконтролируемого роста числа зарезервированных ресурсов в результате

неполной проверки пользователя в облачной информационной инфраструктуре России при переходе на импортозамещение.

Практическая значимость предлагаемого решения включает паттерн для веб-приложения при угрозе неконтролируемого роста числа зарезервированных ресурсов в результате неполной проверки пользователя, который можно применить для широкого круга веб-приложений.

Паттерн безопасности для веб-приложения

Для реализации паттерна использован следующий стек технологий: JavaScript; Docker, nginx. Паттерн безопасности (рис. 7) построен на основе микросервисной архитектуры [16, 17]. За основу взят язык JavaScript.

Рис. 7. Микросервисная архитектура паттерна безопасности для веб-приложения

При взаимодействии пользователя с приложением активируется механизм защиты, созданный с помощью JavaScript и описанный ранее, за виртуализацию отвечает Docker, а балансировщик нагрузки nginx позволяет обеспечить стабильную работу веб-приложения.

Развертывание контейнера nginx производилось на операционной системе Windows 10 с помощью Microsoft Visual Studio и Docker Desktop, согласно типовых инструкций настройки и конфигурирования [18, 19, 20].

Для того чтобы веб-сервер был доступен извне контейнера, необходимо соединить порт 80 контейнера с портом операционной системы с помощью команды [20]:

```
docker run -d -p 80:80 nginx.
```

Далее создаём файл vhost.conf в котором пишем следующий код и настраиваем конфигурацию, чтобы сервер открывал страницу нашего сайта [20]:

```
server {
listen 80;
server_name localhost;
index index.html;
root /var/www/public_html;}.
```

Далее передаем конфигурацию сервера внутрь контейнера с помощью команды [20]:

```
docker container run -d -p 80:80 v
"${PWD}/vhost.conf:/etc/nginx/conf.d/
default.conf" nginx.
```

Создаём файл index.html стартовой страницы веб-приложения и запускаем контейнер с помощью команды [20]:

```
docker container run -d -p 80:80 -v
"${PWD}/vhost.conf:/etc/nginx/conf.d/
default.conf" -v "${PWD}/www:/var/www/
public_html" nginx.
```

Для реализации механизма защиты в веб-приложении была создана перенаправляющая страница index.html. Данная страница позволяет в случайном порядке перенаправить пользователя на другую страницу, где располагается одна из задач для решения: index_1.html, index_2.html, index_3.html. Реализованные страницы продемонстрированы на рис. 8, рис. 9, рис. 10.

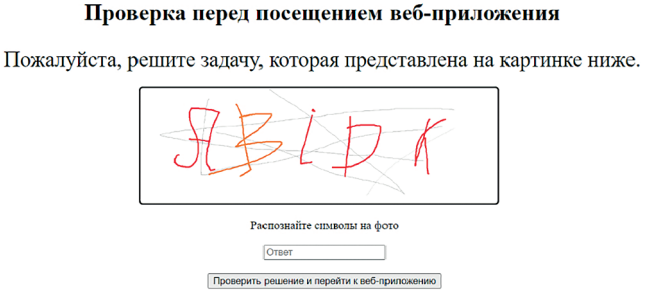


Рис. 8. Страница index_1.html с задачей распознать символы

На странице index_1.html пользователь должен распознать правильным образом символы. На странице index_2.html пользователь должен решить математический пример. На странице index_3.html пользователь должен распознать графические объекты.

Рис. 9. Страница index_2.html с задачей решить математический пример

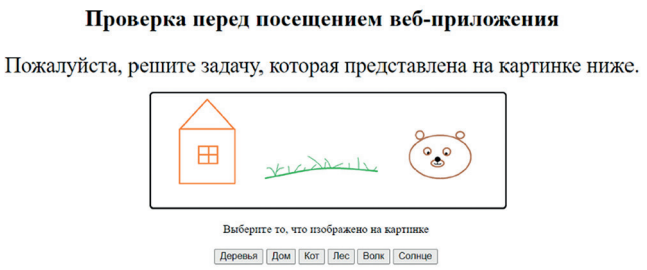


Рис. 10. Страница index_3.html с задачей распознать графические объекты

Согласно рис. 6, при неправильном решении пользователем задачи он перенаправляется на страницу yandex.ru. В случае успешного решения пользователь перенаправляется на балансировщик нагрузки nginx. В этом случае файл конфигурации vhost.conf необходимо дополнить специальным кодом. Итоговый файл vhost.conf содержит следующий код:

```
   upstream backend {
   ip_hash;
   server backend1:80;
   server backend2:80;
   server backend3:80;}
   server {
   listen 80;
   server_name localhost;
   root /var/www/public_html;
   index index.html;
   location / {
   proxy_pass http://backend;
   proxy_set_header Host $host;
   proxy_set_header X-Real-IP $remote_addr;
   proxy_set_header X-Forwarded- For $proxy_
add_x_forwarded_for;
   proxy_set_header X-Forwarded-Proto
$scheme;}}
```

Вначале определен блок upstream с именем backend, который содержит список серверов, в коде выше это backend1, backend2 и backend3. Эти серверы будут обрабатывать входящие запросы, и они распределяются по методу IP Hash, который гарантирует, что запросы от одного и того же IP-адреса всегда будут направляться на один и тот же сервер, что гарантирует безопасность при сохранении сессий пользователя. В основной конфигурации сервера определено, что он слушает порт 80 и отвечает на запросы к домену localhost, а также добавлен блок location, который указывает nginx перенаправлять все входящие запросы на группу серверов backend. При этом используются дополнительные директивы, которые устанавливают заголовки для передачи информации о клиенте на сервер (Host, X-Real-IP, X-Forwarded-For и X-Forwarded-Proto). Эти заголовки передают информацию о реальном IP-адресе клиента и протоколе, что необходимо для корректной и безопасной работы приложений на сервере.

Таким образом, предложенный вариант конфигурации позволяет равномерно распределять нагрузку на несколько серверов, сохраняя при этом согласованность сеансов для пользователей и обеспечивая безопасности веб-приложения при угрозе неконтролируемого роста числа зарезервированных ресурсов в результате неполной проверки пользователя.

Предложенный вариант конфигурации включен в код паттерна безопасности веб-приложения, вместе с остальными файлами на языке JavaScript.

Тестирование паттерна безопасности веб-приложения и обсуждение результатов

Для проверки работоспособности паттерна безопасности веб-приложения было выбрано нагрузочное тестирование. Это проверка устойчивости и производительности программного обеспечения под нагрузкой, сопоставимой с реальными условиями использования. В качестве инструмента выбран Grafana k6 – это бесплатный инструмент нагрузочного тестирования с открытым исходным кодом, который упрощает тестирование производительности и делает его более продуктивным для инженерных групп [21].

Была произведена установка Grafana k6 на операционную систему Windows 10 с помощью команды:

```
winget install k6 --source winget.
```

В основной директории проекта был создан файл для тестирования test.js, содержащий следующий код:

```
import http from 'k6/http';
import {sleep} from 'k6';
export const options = {
stages: [
{duration: '1m', target: 20 },
{duration: '3m', target: 20 },
{duration: '1m', target: 0 },],};
export default function (){
http.get('http://nginx1/index.html');
sleep(1);}
```

Программный код теста test.js с реализованным сценарием тестирования, включает в себя три этапа с различными уровнями нагрузки.

На первом этапе, который длится 1 минуту, количество виртуальных пользователей постепенно увеличивается до 20, что позволяет оценить, как веб-приложение справляется с возрастающей нагрузкой. Это помогает определить, насколько эффективно оно адаптироваться к увеличению числа пользователей.

На втором этапе, который длится 3 минуты, нагрузка остается стабильной на уровне 20 виртуальных пользователей. Это позволяет протестировать производительность веб-приложения при постоянной нагрузке, выяснить, насколько стабильно он работает

в условиях активного использования, и выявить возможные проблемы с производительностью при длительной нагрузке.

На третьем этапе, который также длится 1 минуту, нагрузка постепенно снижается до нуля, что позволяет оценить, как веб-приложение справляется с уменьшением нагрузки и возвращается ли он в нормальное состояние после интенсивной работы.

Основная функция в коде отправляет get-запросы к указанному url – 'http://nginx1/index.html, что имитирует действия пользователей, посещающих веб-страницу. Включение паузы в одну секунду между запросами помогает избежать слишком частого обращения к серверу и создает более реалистичное поведение пользователей. Весь этот сценарий позволяет получить полное представление о том, как веб-приложение справляется с различными уровнями нагрузки, и выявить его слабые места или области, требующие оптимизации.

Далее был запущен Grafana k6 с помощью команды:

```
docker run --rm -i --network network1 -v
${PWD}:/app -w /app grafana/k6 run /app/
test.js.
```

Следует отметить, что запуск производился в сети network1, которая была заранее создана командой:

```
docker network create network1.
```

Так как первоначально веб-приложение nginx не было запущено в сети network1, то следует произвести его повторный запуск, но уже в сети network1, командой [20]:

```
docker run -d -p 80:80 -v "${PWD}/vhost.
conf:/etc/nginx/conf.d/default.conf" -v
"${PWD}/www:/var/www/public_html" --network
network1 --name nginx1 nginx.
```

На рис. 11 представлены результаты проведенного нагрузочного тестирования паттерна безопасности веб-приложения.

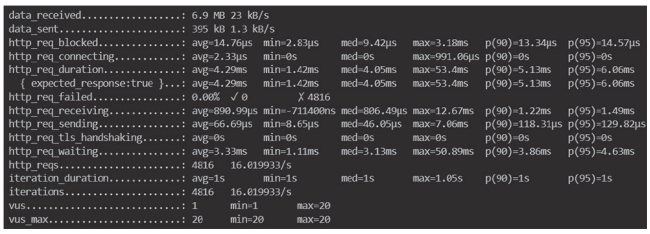


Рис. 11. Результаты нагрузочного тестирования паттерна безопасности веб-приложения

Обсудим основные результаты нагрузочного тестирования. В ходе теста было отправлено и получено определенное количество данных: 6,9 Мбайт данных, полученных от веб-приложения со средней скоростью 23 Кбайт/с, и 395 Кбайт данных, отправленных веб-приложению со средней скоростью 1,3 Кбайт/с.

Метрика http_req_blocked отражает время блокировки HTTP-запросов до их отправки, которое в среднем составило 14,76 микросекунд, а максимальное значение достигало 3,18 миллисекунд.

Время подключения к веб-приложению, которое показывает метрика http_req_connecting практически отсутствовало, что свидетельствует о быстром соединении.

Метрика http_req_duration показывает среднюю длительность HTTP-запросов, которая составила 4,29 миллисекунд, при этом минимальное значение – 1,42 миллисекунды, а максимальное – 53,4 миллисекунд. Для 90 % запросов длительность не превышала 5,13 миллисекунд, а для 95 % – 6,06 миллисекунд.

Время получения ответа, которое показывает метрика http_req_receiving, в среднем составило 890,99 микросекунд, однако к этому значению следует относится осторожно, так как наблюдаются аномалии с отрицательными значениями, что может свидетельствовать о баге в измерении времени по конкретной метрике.

Время отправки запроса, которое показывает метрика http_req_sending, было очень коротким, в среднем 66,69 микросекунд. В тесте не использовались запросы с TLS-шифрованием, поэтому показатели метрики http_reg_tls_handshaking равны нулю.

Время ожидания ответа от веб-приложения, которое показывает метрика http_req_waiting, составило в среднем 3,33 миллисекунды. За все время теста было отправлено 4816 HTTP-запросов с частотой около 16 запросов в секунду. Каждый цикл теста (см. метрику iteration_duration) длился ровно одну секунду, что подтверждает стабильность паузы между запросами. Всего в тесте участвовало до 20 виртуальных пользователей одновременно, с постепенным увеличением нагрузки, что соответствует конфигурации заданного теста.

В результате тестирования не было зафиксировано ни одного сбоя запросов, это означает, что все 4816 запросов были успешными, а это в свою очередь свидетельствует о стабильной работе паттерна безопасности веб-приложения.

Выводы

Рассмотрены сценарии DDoS-атак с использованием ботнет. Построена микросервисная архитектура для обеспечения безопасности веб-приложения. Разработан паттерн для веб-приложения при угрозе неконтролируемого роста числа зарезервированных ресурсов в результате неполной проверки пользователя

на основе микросервисов, интегрированных в контейнеры. В рамках проведённого исследования был разработан сервис проверки пользователей на языке JavaScript, с виртуализацией на базе Docker и с балансировщиком нагрузки nginx. Механизм защиты реализован следующим образом. Пользователь перед заходом в веб-приложение перенаправляется на страницу, где требуется выполнить определенную задачу: решить математический пример, распознать правильным образом символы, распознать правильным образом графические объекты. При успешном решении задач пользователь перенаправляется на веб-приложение, проходя перед этим через один из трех балансировщиков нагрузки, использующих метод IP Hash. Разработан программный код сервиса проверки пользователей, включая коды специальных методов и алгоритмы для трех указанных выше задач. Проведено тестирование паттерна безопасности веб-приложения на базе Grafana k6. Разработан программный код теста test.js с реализованным сценарием тестирования, который включает в себя три этапа с различными уровнями нагрузки. В тесте участвовало до 20 виртуальных пользователей одновременно, с постепенным увеличением нагрузки. В результате тестирования не было зафиксировано ни одного сбоя запросов, все 4816 запросов были успешными – это свидетельствует о стабильной работе паттерна безопасности веб-приложения.

Литература

1. Shameer Mohammed, S. Nanthini, N. Bala Krishna, Inumarthi V. Srinivas, Manikandan Rajagopal, M. Ashok Kumar, A new lightweight data security system for data security in the cloud computing, Measurement: Sensors, Volume 29, 2023, 100856.
2. S. Achar, Cloud computing security for multi-cloud service providers: controls and techniques in our modern threat landscape, International Journal of Computer and Systems Engineering, 16(9), 2022, 379–384.
3. Oludare Isaac Abiodun, Moatsum Alawida, Abiodun Esther Omolara, Abdulatif Alabdulatif, Data provenance for cloud forensic investigations, security, challenges, solutions and future perspectives: A survey, Journal of King Saud University – Computer and Information Sciences, Volume 34, Issue 10, Part B, 2022, 10217–10245.
4. Chakraborti, A., Curtmola, R., Katz, J., Nieh, J., Sadeghi, A. R., Sion, R., Zhang, Y., Cloud Computing Security: Foundations and Research Directions. Foundations and Trends in Privacy and Security, 3(2), 2022, 103–213.
5. Ukeje, N., Gutierrez, J., Petrova, K., Information security and privacy challenges of cloud computing for government adoption: a systematic review, International Journal of Information Security, Volume 23, 2024, 1459–1475.
6. Fatemeh Khoda Parast, Chandni Sindhav, Seema Nikam, Hadiseh Izadi Yekta, Kenneth B. Kent, Saqib Hakak, Cloud computing security: A survey of service-based models, Computers & Security, Volume 114, 2022, 102580.
7. Anmol Kumar, Mayank Agarwal, Quick service during DDoS attacks in the container-based cloud environment, Journal of Network and Computer Applications, Volume 229, 2024, 103946.
8. Yunhe Cui, Qing Qian, Chun Guo, Guowei Shen, Youliang Tian, Huanlai Xing, Lianshan Yan, Towards DDoS detection mechanisms in Software-Defined Networking, Journal of Network and Computer Applications, Volume 190, 2021, 103156.
9. Anderson Bergamini de Neira, Burak Kantarci, Michele Nogueira, Distributed denial of service attack prediction: Challenges, open issues and opportunities, Computer Networks, Volume 222, 2023, 109553.
10. Shahbaz Ahmad Khanday, Hoor Fatima, Nitin Rakesh, Implementation of intrusion detection model for DDoS attacks in Lightweight IoT Networks, Expert Systems with Applications, Volume 215, 2023, 119330.
11. Man Li, Huachun Zhou, Shuangxing Deng, Parallel path selection mechanism for DDoS attack detection, Journal of Network and Computer Applications, Volume 230, 2024, 103938.
12. Jordana J. George, Dorothy E. Leidner, From clicktivism to hacktivism: Understanding digital activism, Information and Organization, Volume 29, Issue 3, 2019, 100249.
13. Cameron John Hoffman, C. Jordan Howell, Robert C. Perkins, David Maimon, Olena Antonaccio, Predicting new hackers' criminal careers: A group-based trajectory approach, Computers & Security, Volume 137, 2024, 103649.
14. B. Balatamoghna, Aditya Jaganath, S. Vaideeshwaran, Anish Subramanian, K. Suganthi, Integrated balancing approach for hosting services with optimal efficiency - Self Hosting with Docker, Materials Today: Proceedings, Volume 62, Part 7, 2022, 4612–4619.
15. Stephen Jacob, Yuansong Qiao, Yuhang Ye, Brian Lee, Anomalous distributed traffic: Detecting cyber security attacks amongst microservices using graph convolutional networks, Computers & Security, Volume 118, 2022, 102728.
16. Diogo Faustino, Nuno Gonçalves, Manuel Portela, António Rito Silva, Stepwise migration of a monolith to a microservice architecture: Performance and migration effort evaluation, Performance Evaluation, Volume 164, 2024, 102411.
17. Hassaan Siddiqui, Ferhat Khendek, Maria Toeroe, Microservices based architectures for IoT systems - State-of-the-art review, Internet of Things, Volume 23, 2023, 100854.
18. Hubin Yang, Ruochen Shao, Yanbo Cheng, Yucong Chen, Rui Zhou, Gang Liu, Guoqi Xie, Qingguo Zhou, REDB: Real-time enhancement of Docker containers via memory bank partitioning in multicore systems, Journal of Systems Architecture, Volume 151, 2024, 103135.
19. Enrico Cambiaso, Luca Caviglione, Marco Zuppelli, DockerChannel: A framework for evaluating information leakages of Docker containers, SoftwareX, Volume 24, 2023, 101576.
20. Корнеев Н. В., Лазорин Д. С. Паттерн для обеспечения безопасности веб-приложения при угрозе XSS атак в облачной инфраструктуре // Вопросы кибербезопасности. 2024. № 6(64). С. 76–84.
21. Vladimir Ciric, Marija Milosevic, Danijel Sokolovic, Ivan Milentijevic, Modular deep learning-based network intrusion detection architecture for real-world cyber-attack simulation, Simulation Modelling Practice and Theory, Volume 133, 2024, 102916.

PATTERN FOR SECURING WEB APPLICATION UNDER THREAT OF UNCONTROLLED GROWTH IN THE NUMBER OF RESERVED RESOURCES

Korneev N. V.[1], Trubacheva-Gudovich A. E.[2]

Keywords: *template, DDoS attack, botnet, user verification service, load balancer, IP Hash method, character recognition task, graphic object recognition task, mathematical task, container, testing.*

The purpose of this article: *development of a pattern for a web application in case of a threat of uncontrolled growth of the number of reserved resources as a result of incomplete user verification.*

Research method: *analysis of the principles of DDoS attacks. Synthesis of DDoS attack scenarios for three types of attacks: transport layer, infrastructure layer, and application layer. The security scenario of a web application is chosen as the basis for the threat of an uncontrolled increase in the number of reserved resources as a result of incomplete user verification. A new protection mechanism has been proposed that provides redirection and verification of the user on a special set of tasks and further balancing of his request to the web application using the IP Hash method. The research was carried out by full-scale modeling of a Docker-based web application in containerization-enabled environments, its deployment and testing.*

Result: *the analysis of the threat of uncontrolled growth in the number of reserved resources is carried out and the relevance of the problem of developing universal template security mechanisms called patterns is shown. In particular, the scenarios of DDoS attacks on web applications are considered. A scenario for ensuring the security of a web application is proposed when there is a threat of an uncontrolled increase in the number of reserved resources as a result of incomplete user verification. A microservice architecture has been built to ensure the security of a web application. A pattern has been developed for a web application in the event of a threat of uncontrolled growth in the number of reserved resources as a result of incomplete user verification based on microservices integrated into containers. As part of the research, a user verification service was developed in JavaScript, with Docker-based virtualization, and with an nginx load balancer. The protection mechanism is implemented as follows. Before entering the web application, the user is redirected to a page where a specific task is required: to solve a mathematical example, to recognize symbols in the right way, to recognize graphic objects in the right way. Upon successful completion of the tasks, the user is redirected to the web application, passing through one of the three load balancers using the IP Hash method. The program code of the user verification service has been developed, including codes of special methods and algorithms for the three tasks mentioned above. A web application security pattern based on Grafana k6 has been tested. The test program code has been developed.js with an implemented testing scenario that includes three stages with different load levels. Up to 20 virtual users participated in the test at the same time, with a gradual increase in workload. As a result of testing, not a single request failure was recorded, all 4816 requests were successful – this indicates the stable operation of the web application security pattern.*

Practical value: *the practical value of the proposed solution includes a pattern for a web application under the threat of uncontrolled growth in the number of reserved resources as a result of incomplete user verification, which can be applied to a wide range of web applications.*

References

1. *Shameer Mohammed, S. Nanthini, N. Bala Krishna, Inumarthi V. Srinivas, Manikandan Rajagopal, M. Ashok Kumar, A new lightweight data security system for data security in the cloud computing, Measurement: Sensors, Volume 29, 2023, 100856.*
2. *S. Achar, Cloud computing security for multi-cloud service providers: controls and techniques in our modern threat landscape, International Journal of Computer and Systems Engineering, 16(9), 2022, 379–384.*
3. *Oludare Isaac Abiodun, Moatsum Alawida, Abiodun Esther Omolara, Abdulatif Alabdulatif, Data provenance for cloud forensic investigations, security, challenges, solutions and future perspectives: A survey, Journal of King Saud University – Computer and Information Sciences, Volume 34, Issue 10, Part B, 2022, 10217–10245.*
4. *Chakraborti, A., Curtmola, R., Katz, J., Nieh, J., Sadeghi, A.R., Sion, R., Zhang, Y., Cloud Computing Security: Foundations and Research Directions. Foundations and Trends in Privacy and Security, 3(2), 2022, 103–213.*
5. *Ukeje, N., Gutierrez, J., Petrova, K., Information security and privacy challenges of cloud computing for government adoption: a systematic review, International Journal of Information Security, Volume 23, 2024, 1459–1475.*

1 Nikolay V. Korneev, Doctor of Technical Sciences, Associate Professor, Gubkin Russian State University of Oil and Gas, Moscow, Russia. E-mail: niccyper@mail.ru
2 Anna E. Trubacheva-Gudovich, Student, Gubkin Russian State University of Oil and Gas, Moscow, Russia. E mail: gudovich.an@bk.ru

6. *Fatemeh Khoda Parast, Chandni Sindhav, Seema Nikam, Hadiseh Izadi Yekta, Kenneth B. Kent, Saqib Hakak, Cloud computing security: A survey of service-based models, Computers & Security,Volume 114, 2022, 102580.*
7. *Anmol Kumar, Mayank Agarwal, Quick service during DDoS attacks in the container-based cloud environment, Journal of Network and Computer Applications, Volume 229, 2024, 103946.*
8. *Yunhe Cui, Qing Qian, Chun Guo, Guowei Shen, Youliang Tian, Huanlai Xing, Lianshan Yan, Towards DDoS detection mechanisms in Software-Defined Networking, Journal of Network and Computer Applications, Volume 190, 2021, 103156.*
9. *Anderson Bergamini de Neira, Burak Kantarci, Michele Nogueira, Distributed denial of service attack prediction: Challenges, open issues and opportunities, Computer Networks, Volume 222, 2023, 109553.*
10. *Shahbaz Ahmad Khanday, Hoor Fatima, Nitin Rakesh, Implementation of intrusion detection model for DDoS attacks in Lightweight IoT Networks, Expert Systems with Applications, Volume 215, 2023, 119330.*
11. *Man Li, Huachun Zhou, Shuangxing Deng, Parallel path selection mechanism for DDoS attack detection, Journal of Network and Computer Applications, Volume 230, 2024, 103938.*
12. *Jordana J. George, Dorothy E. Leidner, From clicktivism to hacktivism: Understanding digital activism, Information and Organization, Volume 29, Issue 3, 2019, 100249.*
13. *Cameron John Hoffman, C. Jordan Howell, Robert C. Perkins, David Maimon, Olena Antonaccio, Predicting new hackers' criminal careers: A group-based trajectory approach, Computers & Security, Volume 137, 2024, 103649.*
14. *B. Balatamoghna, Aditya Jaganath, S. Vaideeshwaran, Anish Subramanian, K. Suganthi, Integrated balancing approach for hosting services with optimal efficiency - Self Hosting with Docker, Materials Today: Proceedings, Volume 62, Part 7, 2022, 4612–4619.*
15. *Stephen Jacob, Yuansong Qiao, Yuhang Ye, Brian Lee, Anomalous distributed traffic: Detecting cyber security attacks amongst microservices using graph convolutional networks, Computers & Security, Volume 118, 2022, 102728.*
16. *Diogo Faustino, Nuno Gonçalves, Manuel Portela, António Rito Silva, Stepwise migration of a monolith to a microservice architecture: Performance and migration effort evaluation, Performance Evaluation, Volume 164, 2024, 102411.*
17. *Hassaan Siddiqui, Ferhat Khendek, Maria Toeroe, Microservices based architectures for IoT systems – State-of-the-art review, Internet of Things, Volume 23, 2023, 100854.*
18. *Hubin Yang, Ruochen Shao, Yanbo Cheng, Yucong Chen, Rui Zhou, Gang Liu, Guoqi Xie, Qingguo Zhou, REDB: Real-time enhancement of Docker containers via memory bank partitioning in multicore systems, Journal of Systems Architecture, Volume 151, 2024, 103135.*
19. *Enrico Cambiaso, Luca Caviglione, Marco Zuppelli, DockerChannel: A framework for evaluating information leakages of Docker containers, SoftwareX, Volume 24, 2023, 101576.*
20. *Korneev N. V., Lazorin D. S. Pattern dlya obespecheniya bezopasnosti veb-prilozheniya pri ugroze XSS atak v oblachnoj infrastrukture // Voprosy kiberbezopasnosti. 2024. № 6(64). S. 76–84.*
21. *Vladimir Ciric, Marija Milosevic, Danijel Sokolovic, Ivan Milentijevic, Modular deep learning-based network intrusion detection architecture for real-world cyber-attack simulation, Simulation Modelling Practice and Theory, Volume 133, 2024, 102916.*

МЕТОДИКА СИНТЕЗА КВАНТОВО-УСТОЙЧИВЫХ БЛОКЧЕЙН-ПЛАТФОРМ С КИБЕРИММУНИТЕТОМ

Балябин А. А.[1], Петренко С. А.[2]

DOI: 10.21681/2311-3456-2025-4-46-54

Цель исследования: *разработка методики параметрического синтеза киберустойчивых блокчейн-экосистем и платформ «Экономики данных» Российской Федерации с кибериммунитетом в условиях новой квантовой угрозы.*

Методы исследования: *методы системного анализа, методы теории вероятностей и математической статистики, методы теории устойчивости сложных систем, методы теории подобия и размерностей.*

Полученные результаты: *проведено исследование существующих подходов к обеспечению квантовой устойчивости блокчейн-платформ с кибериммунитетом; сформулирована гипотеза о возможности обеспечения требуемой киберустойчивости блокчейн-платформ с кибериммунитетом в условиях квантовых атак; разработана методика параметрического синтеза квантово-устойчивых блокчейн-экосистем и платформ «Экономики данных» Российской Федерации с кибериммунитетом с использованием методов теории подобия; проведены экспериментальные исследования методики, результаты которых позволили подтвердить выдвинутую гипотезу.*

Научная новизна: *предложенная методика отличается от существующих тем, что во вновь разработанную методику введены новые группы формальных операций для каждого уровня функционирования блокчейн-платформ с кибериммунитетом по оцениванию необходимого и достаточного значений параметров нейтрализующих воздействий с использованием методов теории подобия, а также критерий, позволяющий установить существование решения для заданных значений требуемых показателей киберустойчивости и временных показателей функционирования блокчейн-платформы.*

Ключевые слова: *угрозы безопасности информации, квантовые угрозы безопасности, блокчейн-экосистемы и платформы, кибербезопасность, киберустойчивость, методы анализа и синтеза квантово-устойчивого блокчейн.*

Введение

Технологии распределенного реестра (DLT) активно развиваются и находят широкое применение в различных сферах, таких как финансовые услуги, цепочки поставок, здравоохранение и государственное управление. Они обеспечивают безопасность, прозрачность и децентрализацию данных, позволяя создавать системы, в которых участники могут проверять и записывать информацию без необходимости в доверенном центре. Основным примером реализации таких технологий является блокчейн, лежащий в основе криптовалют [1].

На базе блокчейн создаются новые технологии: смарт-контракты, токены (ERC-20), децентрализованные финансы (DeFi), децентрализованные приложения (dApps) и другие. Технологии распределенного реестра активно применяются при создании систем интернета вещей (IoT) [2], облачных платформ [3], систем Индустрии 4.0 [4]. Ведутся исследования по созданию интероперабельных блокчейн-экосистем [5] и децентрализованной сети Интернет (Web3) [6].

В Российской Федерации технологии распределенного реестра играют важную роль в реализации национального проекта «Экономика данных», позволяя создавать блокчейн-экосистемы и платформы, цифровые валюты, системы налогообложения, голосования и иные защищенные системы хранения и обработки данных, являющихся важнейшим активом в цифровой экономике [7].

Внедрение новых технологических решений неизбежно приводит к усложнению ландшафта киберугроз и возникновению новых, ранее неизвестных типов воздействий. Например, технологии искусственного интеллекта (AI) сегодня зачастую применяются для выявления недостатков систем защиты и планирования наступательных кибероперации [8, 9]. Одним из новейших вызовов безопасности блокчейн-экосистем и платформ является проблема обеспечения их киберустойчивости в условиях развития квантовых вычислений [10–12].

Свойства конфиденциальности, целостности и невозможности отказа от авторства, характерные для блокчейн-платформ, обеспечиваются криптостойкостью алгоритмов, которая в свою очередь зависит от вычислительной сложности задач факторизации и дискретного логарифмирования. Применение квантовых алгоритмов Шора и Гровера позволяет ускорить решение этих задач и создает угрозу киберустойчивости блокчейн-платформ [13, 14].

1 Балябин Артём Алексеевич, младший научный сотрудник, Научный центр информационных технологий и искусственного интеллекта, Научно-технологический университет «Сириус», Федеральная территория «Сириус», Россия. 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.org/0000- 0003-0644-1731. E mail: Petrenko.SA@talantiuspeh.ru

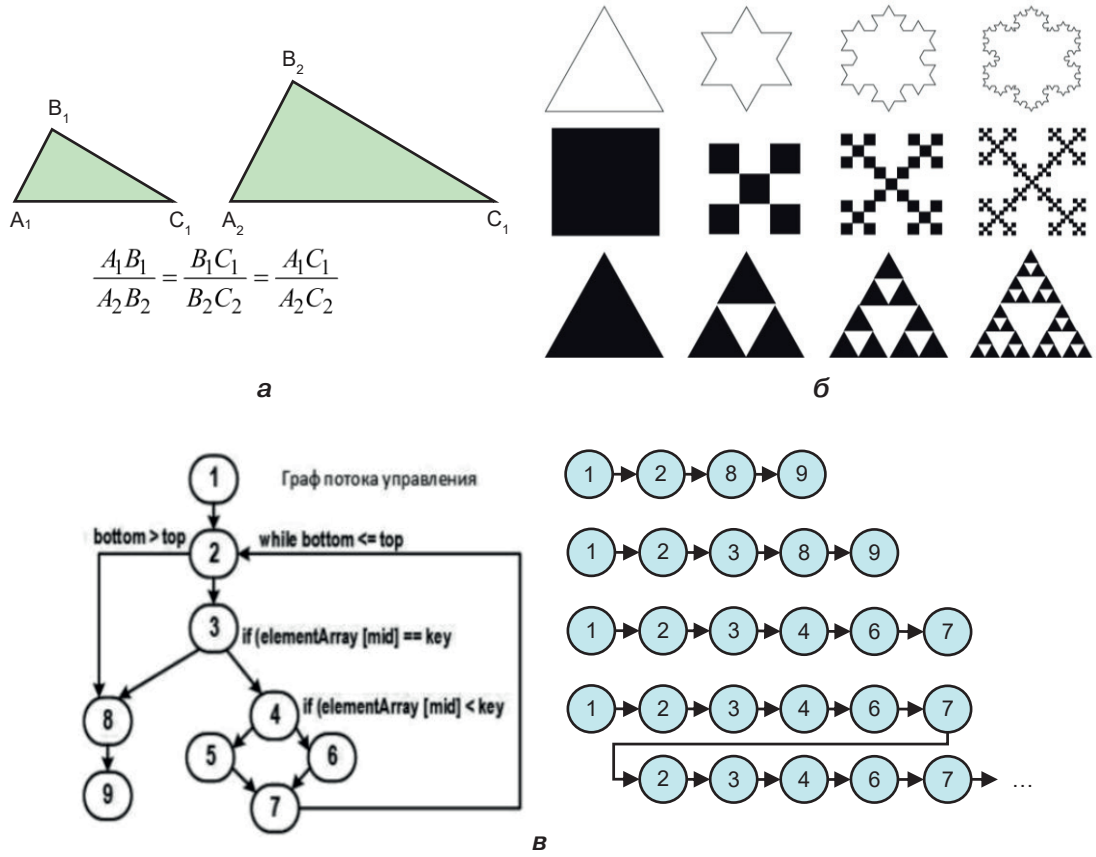


Рис. 1. Подобие и самоподобие:
а – геометрических фигур; б – фракталов; в – программных реализаций

Одним из возможных подходов к обеспечению киберустойчивости блокчейн-платформ является наделение их свойством кибериммунитета, то есть способностью обнаруживать аномалии, вызванные известными и ранее неизвестными воздействиями, противодействовать им и восстанавливать штатное функционирование [15, 16].

В ряде исследований предлагаются различные реализации методов кибериммунной защиты. Например, в работе [17] предлагается обеспечивать киберустойчивость информационных систем путем синтеза упреждающего поведения систем защиты. В работе [18] предлагается обеспечивать устойчивость киберфизических систем на основе динамической реконфигурации и гомеостаза. В работе [19] для решения схожих задач рассматривается возможность применения различных методов, в том числе гибридных, включая нейронные, иммунные и нейронечеткие.

Аномалии функционирования системы связаны с отклонением ее текущего состояния от некоторого штатного (киберустойчивого) состояния под воздействием внешних дестабилизирующих факторов. Одним из перспективных подходов к решению задач обнаружения аномалий и восстановления в рамках кибериммунного подхода является применение методов теории подобия и размерностей [20]. На рис. 1 представлены некоторые примеры проявления подобия и самоподобия.

В настоящем исследовании учитываются новейшие вызовы, связанные с ростом квантовых угроз, и предлагается методика параметрического синтеза квантово-устойчивых блокчейн-платформ с кибериммунитетом с использованием методов теории подобия и размерностей.

1. Постановка задачи исследования

Дано: L – блокчейн-платформа с кибериммунитетом, функционирующая на четырех уровнях, так что $L = \{L_i | i \in [1,4]\}$; $X = \{X_i | i \in [1,4]\}$ – множество входных данных; $Y = \{Y_i | i \in [1,4]\}$ – множество выходных данных; $E = \{E_i | i \in [1,4]\}$ – множество параметров среды; $A = \{A_i | i \in [1,4]\}$ – множество параметров дестабилизирующих воздействий; $D = \{D_i | i \in [1,4]\}$ – множество параметров нейтрализующих воздействий; $R = \{R_i | i \in [1,4]\}$ – множество показателей киберустойчивости; $T = \{T_i | i \in [1,4]\}$ – множество временных показателей функционирования.

Необходимо: разработать методику M параметрического синтеза квантово-устойчивых блокчейн-

платформ, для которых обеспечиваются требуемые значения $R_1^{тр},...,R_4^{тр}$ показателей киберустойчивости $R_1,...,R_4$ ($\forall R_i \geq R_i^{тр}$, $R_i \in R$, $i = 1..4$) и значения $T_1^{тр},...,T_4^{тр}$ временных показателей функционирования $T_1,...,T_4$ ($\forall T_i \leq T_i^{тр}$, $T_i \in T$, $i = 1..4$) в условиях квантовых дестабилизирующих воздействий A.

Формальная постановка научной задачи: найти

$$M{:}<L,X,Y,E,A,D,R,T> \longrightarrow <D,R,T>$$
$$D = \{D_i\}$$
$$R = \{R_i\}, R_i \geq R_i^{тр}$$
$$T = \{T_i\}, T_i \leq T_i^{тр}$$
$$i \in [1,4] \qquad (1)$$

при ограничениях $A_i \in [A_{i\min}, A_{i\max}]$, $A_i \in A \subseteq A_{доп}$, $i = 1..4$; $D_i \in [D_{i\min}, D_{i\max}]$, $D_i \in D \subseteq D_{доп}$, $i = 1..4$.

Гипотеза исследования: применение теории подобия и размерностей для реализации кибериммунной защиты блокчейн-платформ позволяет обеспечить требуемую их киберустойчивость в условиях квантовых атак.

2. Методика синтеза квантово-устойчивых блокчейн-платформ

Блокчейн-платформу с кибериммунитетом можно представить в виде системы из четырех взаимоувязанных уровней: криптографических алгоритмов, алгоритмов консенсуса, смарт-контрактов и децентрализованных приложений. На каждом уровне на систему оказываются квантовые дестабилизирующие воздействия, приводящие к возникновению аномалий. Противодействие квантовым атакам осуществляется на основе метода кибериммунной защиты, предназначенного для нейтрализации дестабилизирующих воздействий и обеспечения требуемой киберустойчивости блокчейн-платформы.

Методика параметрического синтеза квантово-устойчивых блокчейн-платформ с кибериммунитетом заключается в поиске значений параметров нейтрализующих воздействий D, обеспечивающих выполнение требований к показателям киберустойчивости R и временным показателям функционирования T на каждом уровне блокчейн-платформы при заданных параметрах воздействий A. Схема предлагаемой методики для i-го уровня блокчейн-платформы приведена на рис. 2.

Определим теперь аналитические выражения для синтеза необходимого и достаточного значений параметров нейтрализующих воздействий для каждого уровня блокчейн-платформы.

2.1. Синтез параметра для уровня криптографических алгоритмов

Показатель киберустойчивости блокчейн-платформы для уровня криптографических алгоритмов определяется как:

$$R_1 = 1 - w_1 q_1 \geq R_i^{тр}, \qquad (2)$$

где q_1 – параметр воздействия, определяющий вероятность синтеза блока атакующим быстрее остальной сети; w_1 – параметр нейтрализующего воздействия, определяющий коэффициент доверия к узлу.

Временной показатель функционирования блокчейн-платформы для уровня криптографических алгоритмов может быть определен как время, затрачиваемое на проверку подобия транзакций, включенных в блок:

$$T_1 = \lfloor (1 - w_1)N \rfloor n \log_2 n \leq T_1^{тр}, \qquad (3)$$

где N – количество блоков в блокчейн; n – количество транзакций в блоке.

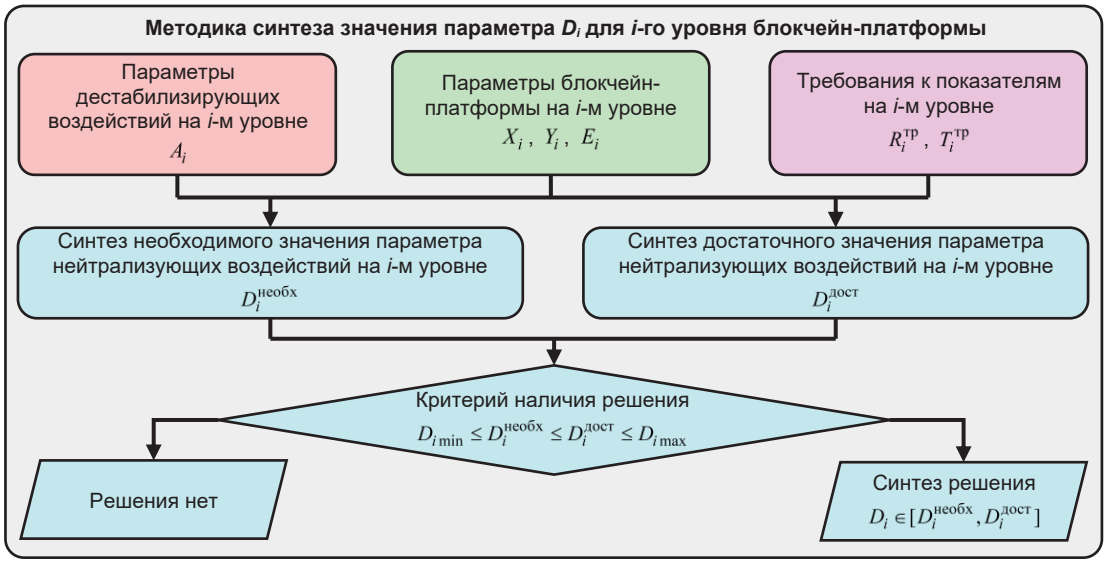


Рис. 2. Схема методики для i-го уровня блокчейн-платформы

При выполнении условия подобия хэш-сумма блока будет равна хэш-сумме вершины дерева Меркла. Здесь $nlog_2n$ характеризует сложность построения дерева Меркла для определения целостности блока, таким образом при снижении коэффициента доверия к узлу потребуется более глубокая проверка на предмет наличия аномалий связанных транзакций в более ранних блоках, что приведет к росту времени вычислений.

Для удобства будем полагать, что $D_1 = 1 - w_1$, тогда выражения для необходимого и достаточного значений параметра нейтрализующих воздействий на 1-м уровне блокчейн с учетом (2) и (3) можно записать как:

$$D_1^{\text{необх}} = 1 - (1 - R_1^{\text{тр}})/q_1$$
$$D_1^{\text{дост}} = T_1^{\text{тр}} /(Nnlogn). \quad (4)$$

2.2. Синтез параметра для уровня алгоритмов консенсуса

Показатель киберустойчивости блокчейн-платформы для уровня алгоритмов консенсуса определяется как:

$$R_2 = \sum_{k=0}^{z_2} \frac{\lambda^k e^{-\lambda}}{k!}(1 - (\lambda/z_2)^{z_2-k}) \geq R_2^{\text{тр}}, \quad (5)$$

где $\lambda = z_2 w_2 q_2 / (1 - w_2 q_2)$ – математическое ожидание длины цепочки блоков, сгенерированной атакующим; z_2 – количество блоков, на которое вредоносная цепочка должна быть длиннее для принятия ее в качестве основной; q_2 – параметр, определяющий вероятность синтеза блока атакующим быстрее остальной сети; w_2 – параметр нейтрализующего воздействия, определяющий коэффициент доверия к узлу.

Пусть требуется, чтобы цепочка атакующего была на 1 блок длиннее для принятия ее в качестве основной, то есть $z_2 = 1$. Тогда, упрощая выражение (5), получим:

$$R_2 = e^{-\lambda}(1 - \lambda) \geq R_2^{\text{тр}}. \quad (6)$$

Поскольку неравенство (6) содержит линейную и экспоненциальную зависимости от искомого параметра, его аналитическое решение затруднительно. Однако, выражение с экспонентой можно разложить в ряд Маклорена:

$$e^{-\lambda} = 1 - \lambda + \frac{(-\lambda^2)}{2!} + \frac{(-\lambda^3)}{3!} + \ldots . \quad (7)$$

Поскольку $\lambda \in [0,1]$ при $z_2 = 1$, то возможно провести аппроксимацию $e^{-\lambda} \approx 1 - \lambda$. Тогда неравенство (6) примет вид:

$$R_2 = (\lambda^2 - 2\lambda + 1) \geq R_2^{\text{тр}}. \quad (8)$$

Решая квадратичное неравенство (8) относительно λ с учетом допустимых значений, получим:

$$\lambda \leq 1 - \sqrt{R_2^{\text{тр}}}. \quad (9)$$

Тогда с учетом (5):

$$w_2 q_2 / (1 - w_2 q_2) \leq 1 - \sqrt{R_2^{\text{тр}}}, \quad (10)$$

что выполняется только в случае:

$$w_2 \leq \min\left(1, \frac{1 - \sqrt{R_2^{\text{тр}}}}{q_2 (2 - \sqrt{R_2^{\text{тр}}})}\right). \quad (11)$$

Временной показатель функционирования блокчейн-платформы для уровня алгоритмов консенсуса может быть определен как время, затрачиваемое на достижение консенсуса валидаторами:

$$T_2 = bV + (1 - w_2)(1 - b - e)V \leq T_2^{\text{тр}}, \quad (12)$$

где b – минимальное количество валидаторов, необходимое для подтверждения корректности цепочки блоков при коэффициенте доверия к узлу, создавшему цепочку $w_2 = 1$, $b \in [0,1]$; V – общее количество валидаторов в блокчейн; e – допустимая погрешность валидации (допустимое количество валидаторов, не подтверждающих корректность блока), $e < 1 - b$.

При снижении доверия к узлу, создающему цепочку блоков, для достижения консенсуса потребуется большее количество валидаторов, что увеличивает временные затраты на согласование. Будем полагать, что $D_2 = 1 - w_2$, тогда выражения для необходимого и достаточного значений параметра нейтрализующих воздействий на 2-м уровне блокчейн с учетом (11) и (12) можно записать как:

$$D_2^{\text{необх}} = \max\left(0{,}1 - \frac{1 - \sqrt{R_2^{\text{тр}}}}{q_2 (2 - \sqrt{R_2^{\text{тр}}})}\right),$$
$$D_2^{\text{дост}} = \frac{T_2^{\text{тр}} - bV}{(1 - b - e)V}. \quad (13)$$

2.3. Синтез параметра для уровня смарт-контрактов

Показатель киберустойчивости блокчейн-платформы для уровня смарт-контрактов определяется как:

$$R_3 = R_2 \frac{1}{n_3} \sum_{i=1}^{n_3} \left(1 - \frac{\#X_{3i}^- - \#D_{3i}}{\#X_{3i}}\right) k_{\text{покр}} \geq R_3^{\text{тр}}, \quad (14)$$

где R_2 – показатель киберустойчивости уровня алгоритмов консенсуса; n_3 – количество смарт-контрактов; $\#X_{3i}$ – мощность множества входных данных i-го смарт-контракта; $\#X_{3i}^-$ – мощность подмножества вредоносных входных данных i-го смарт-контракта, $X_{3i}^- \subset X_{3i}$; $\#D_{3i}$ – мощность множества выявленных и заблокированных вредоносных входных данных i-го смарт-контракта, $\#D_{3i} \leq \#X_{3i}^-$; $k_{\text{покр}}$ – коэффициент покрытия операций смарт-контракта проверками подобия, $k_{\text{покр}} \in [0,1]$.

Для оценки будем считать, что множества X_{3i}, X_{3i}^- и D_{3i} для всех смарт-контрактов одинаковы. В процессе функционирования блокчейн-платформы

происходит накопление образцов вредоносных входных данных и снижается вероятность возникновения нарушения. Тогда выражение для предельной киберустойчивости блокчейн-платформы с учетом (14) примет вид:

$$R_3 = R_2\, k_{\text{покр}} \geq R_3^{\text{тр}}. \quad (15)$$

Временной показатель функционирования блокчейн-платформы для уровня смарт-контрактов может быть определен как время, затрачиваемое на выполнение смарт-контрактов с учетом проверки подобия маршрутов выполнения допустимым маршрутам:

$$T_3 = n_3 \left(kmn + k_{\text{покр}}\, k \frac{m(m-1)}{2}(2n-1) \right) \leq T_3^{\text{тр}}, \quad (16)$$

где k – количество линейных блоков операций в смарт-контракте; m – количество операций в блоке; n – количество параметров в операции.

Параметром нейтрализующих воздействий на данном уровне можно считать коэффициент покрытия операций смарт-контрактов проверками подобия $D_3 = k_{\text{покр}}$. Тогда выражения для необходимого и достаточного значений параметра нейтрализующих воздействий на 3-м уровне блокчейн с учетом (15) и (16) можно записать как:

$$D_3^{\text{необх}} = R_3^{\text{тр}}/R_2$$
$$D_3^{\text{дост}} = \frac{2(T_3^{\text{тр}} / n_3 - kmn)}{km(m-1)(2n-1)}. \quad (17)$$

2.4. Синтез параметра для уровня децентрализованных приложений

Показатель киберустойчивости блокчейн-платформы для уровня децентрализованных приложений определяется как:

$$R_4 = R_3 \frac{1}{n_4} \sum_{i=1}^{n_4} \left(1 - \frac{\#X_{4i}^- - \#D_{4i}}{\#X_{4i}} \frac{1}{l_i} \sum_{r=1}^{l_i} P_{3ir}^a \right) k_{\text{покр}}, \quad (18)$$

где R_3 – показатель киберустойчивости уровня смарт-контрактов; n_4 – количество dApps; $\#X_{4i}$ – мощность множества входных данных i-го dApp; $\#X_{4i}^-$ – мощность подмножества вредоносных входных данных i-го dApp, $X_{4i}^- \subset X_{4i}$; $\#D_{4i}$ – мощность множества выявленных и заблокированных вредоносных входных данных i-го dApp, $\#D_{4i} \leq \#X_{4i}^-$; l_i – количество смарт-контрактов в i-м dApp; P_{3ir}^a – вероятность возникновения нарушения в r-м смарт-контракте i-го dApp; $k_{\text{покр}}$ – коэффициент покрытия операций в децентрализованных приложениях проверками подобия.

Аналогично уровню смарт-контрактов в процессе функционирования блокчейн-платформы происходит накопление образцов вредоносных входных данных децентрализованных приложений, поэтому выражение для предельной киберустойчивости можно записать как:

$$R_4 = R_3\, k_{\text{покр}} \geq R_4^{\text{тр}}. \quad (19)$$

Временной показатель функционирования блокчейн-платформы для уровня децентрализованных приложений может быть определен как время, затрачиваемое на выполнение dApps с учетом проверки подобия последовательностей операций допустимым последовательностям:

$$T_4 = n_4 \frac{l}{n_3} T_3 (1 + k_{\text{покр}}) \leq T_4^{\text{тр}}, \quad (20)$$

где l – количество смарт-контрактов, исполняемых в рамках каждого децентрализованного приложения.

Параметром нейтрализующих воздействий на данном уровне можно считать коэффициент покрытия операций децентрализованных приложений проверками подобия $D_4 = k_{\text{покр}}$. Тогда выражения для необходимого и достаточного значений параметра нейтрализующих воздействий на 4-м уровне блокчейн с учетом (19) и (20) можно записать как:

$$D_4^{\text{необх}} = R_4^{\text{тр}} / R_3$$
$$D_4^{\text{дост}} = \frac{T_4^{\text{тр}}\, n_3}{n_4\, l T_3} - 1. \quad (21)$$

Исследуем теперь возможности параметрического синтеза квантово-устойчивых блокчейн-платформ с кибериммунитетом при управлении параметрами нейтрализующих воздействий с помощью предложенной методики в условиях квантовых атак.

3. Исследование квантовой устойчивости блокчейн-платформ

Для параметрического синтеза квантово-устойчивой блокчейн-платформы необходимо найти значения параметров нейтрализующих воздействий D, обеспечивающие требуемые значения показателей киберустойчивости и временных показателей функционирования блокчейн. Такие параметры с учетом (4), (13), (17), (21) должны удовлетворять условию:

$$D_i^{\text{необх}} \leq D_i \leq D_i^{\text{дост}},\ i = 1..4. \quad (22)$$

На рис. 3 представлены графики зависимости показателей киберустойчивости и временных показателей функционирования блокчейн-платформы от параметров нейтрализующих воздействий на каждом уровне. Как видно, функции требуемых показателей функционирования блокчейн-платформ являются монотонно возрастающими. Сформулируем утверждение относительно существования решения задачи параметрического синтеза квантово-устойчивой блокчейн-платформы.

Утверждение 1. Если для заданных требуемых значений показателей $R_i^{\text{тр}}$, $D_i^{\text{тр}}$, $i = 1..4$ существует решение D_i, $i = 1..4$, то оно лежит в отрезке $[D_i^{\text{необх}}, D_i^{\text{дост}}]$.

Доказательство. Доказательство данного утверждения вытекает из требований критерия наличия решения в соответствии с методикой параметрического синтеза:

$$D_{i\,\min} \leq D_i^{\text{необх}} \leq D_i^{\text{дост}} \leq D_{i\,\max}, \quad (23)$$

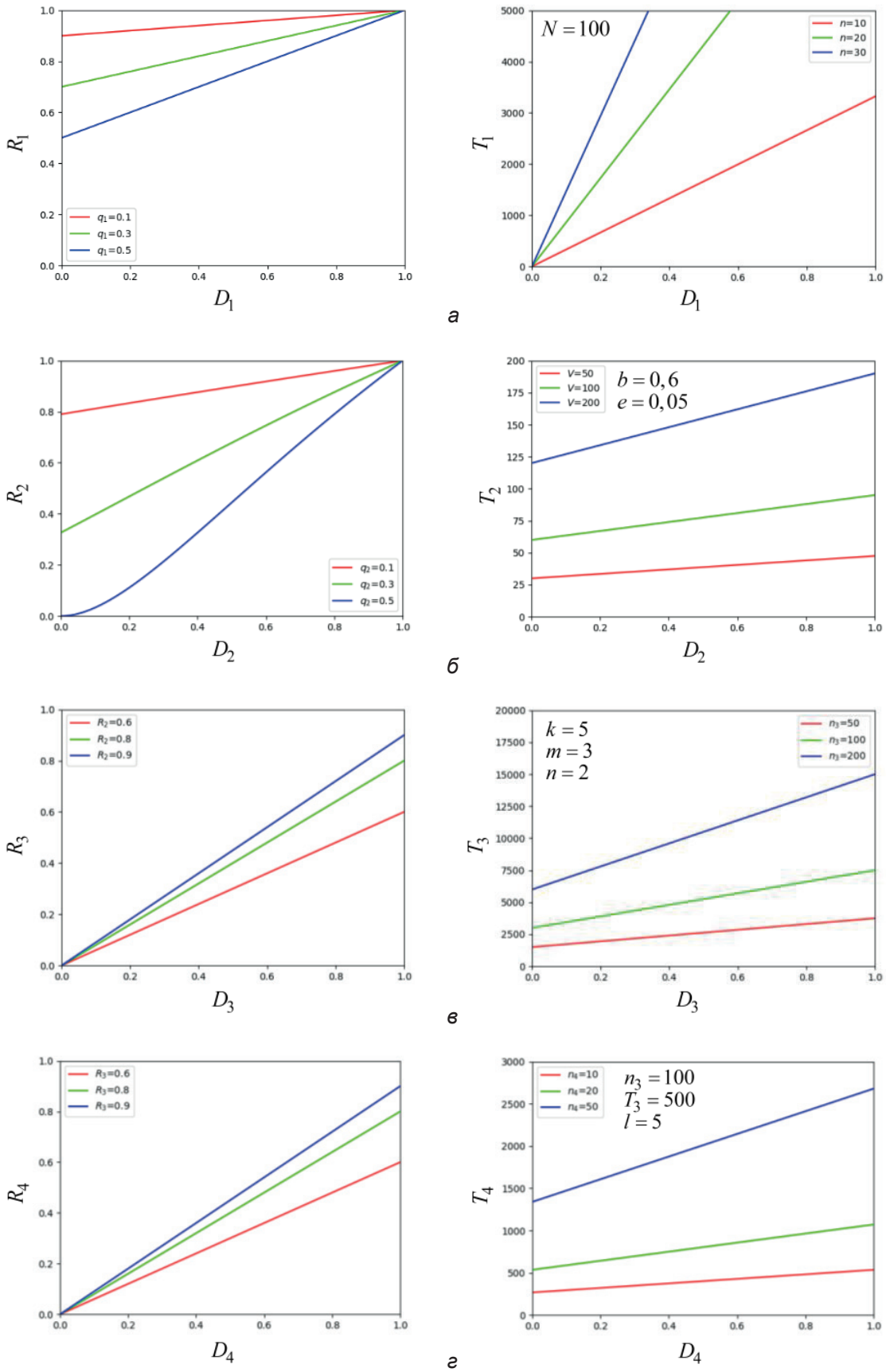


Рис. 3. Результаты исследования показателей киберустойчивости и временных показателей функционирования блокчейн-платформ на уровнях: а – криптографических алгоритмов; б – алгоритмов консенсуса; в – смарт-контрактов; г – децентрализованных приложений

а также того, что функции требуемых показателей являются непрерывно возрастающими на всем промежутке $D_i \in [0,1]$, $i = 1..4$. Таким образом, если решение существует, то оно принадлежит отрезку $[D_i^{\text{необх}}, D_i^{\text{дост}}]$, $i = 1..4$.

Ч.т.д.

Таким образом, результаты проведенных исследований позволяют *подтвердить* выдвинутую гипотезу.

Выводы

В настоящем исследовании поставлена задача разработки методики параметрического синтеза квантово-устойчивых блокчейн-платформ с кибериммунитетом. Выдвинута гипотеза о возможности обеспечения требуемой киберустойчивости блокчейн-платформ. Разработана методика параметрического синтеза блокчейн-платформ с кибериммунитетом, отличающаяся от существующих введением новых групп формальных операций по оцениванию необходимого и достаточного значений параметров нейтрализующих воздействий для обеспечения требуемых значений показателей киберустойчивости и временных показателей функционирования блокчейн-платформ с использованием методов теории подобия.

В результате исследований определены аналитические выражения для необходимого и достаточного значений параметра нейтрализующих воздействий D_i, $i = 1..4$ на каждом уровне блокчейн-платформы, а также экспериментально установлено, что выбор значения параметра из отрезка $[D_i^{\text{необх}}, D_i^{\text{дост}}]$, $i = 1..4$ позволяет обеспечить требуемые значения показателей функционирования блокчейн-платформ с кибериммунитетом в условиях квантовых атак. Таким образом, результаты экспериментов позволили подтвердить выдвинутую гипотезу.

В дальнейшем результаты исследования могут быть использованы при разработке киберустойчивых блокчейн-экосистем и платформ «Экономики данных» Российской Федерации в условиях новой квантовой угрозы.

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. Nguyen D. C. et al. 6G Internet of Things: A Comprehensive Survey // IEEE Internet of Things Journal. 2022. Vol. 9. No. 1. Pp. 359–383. DOI: 10.1109/JIOT.2021.3103320.
3. 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.
4. 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.
5. Chen C. et al. When Digital Economy Meets Web3.0: Applications and Challenges // IEEE Open Journal of the Computer Society. 2022. Vol. 3. Pp. 233–245. DOI: 10.1109/OJCS.2022.3217565.
6. Zhu Q., Loke S. W., Trujillo-Rasua R., Jiang F., Xiang Y. Applications of Distributed Ledger Technologies to the Internet of Things: A Survey // ACM Comput. Surv. 2019. Vol. 52. No. 6. P. 120:1–120:34. DOI: 10.1145/3359982.
7. Петренко А. С. Квантово-устойчивый блокчейн: научная монография // Санкт-Петербург : Питер, 2023. 384 с.
8. Петренко С. А. Киберустойчивость Индустрии 4.0: научная монография // «Издательский Дом «Афина». 2020. 256 с.
9. Маркова С. В. Выявления уязвимостей в децентрализованных информационных системах на основе смарт-контрактов с помощью методов обработки больших данных // Фундаментальные исследования. 2022. № 9. С. 47–53.
10. 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.
11. Kushwaha S. S., Joshi S., Singh D., Kaur M. Lee H. -N. Systematic Review of Security Vulnerabilities in Ethereum Blockchain Smart Contract // IEEE Access. 2022. Vol. 10. Pp. 6605–6621. DOI: 10.1109/ACCESS.2021.3140091.
12. Fernandez-Carames 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.
13. Петренко А. С., Романченко А. М. Перспективный метод криптоанализа на основе алгоритма Шора // Защита информации. Инсайд. 2020. № 2 (92). С. 17–23.
14. Петренко А. С., Петренко С. А. Basic Algorithms Quantum Cryptanalysis (Основные алгоритмы квантового криптоанализа) // Вопросы кибербезопасности. 2023. №1 (53). С. 100–115. DOI: 10.21681/2311-3456-2023-1-100-115.
15. Балябин, А. А. Модель облачной платформы КИИ РФ с кибериммунитетом в условиях информационно-технических воздействий // Защита информации. Инсайд. 2024. № 5 (119). С. 35–44.
16. 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.

17. 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.
18. Зегжда Д. П., Александрова Е. Б., Калинин М. О. и др. Кибербезопасность цифровой индустрии. Теория и практика функциональной устойчивости к кибератакам // Москва : Научно-техническое издательство «Горячая линия-Телеком». 2021. 560 с.
19. Петренко С. А. Кибериммунология // Санкт-Петербург : Афина. 2021. 239 с.
20. Балябин А. А., Петренко С. А. Методика кибериммунной защиты цифровых сервисов «ГосТех» с использованием теории подобия и размерностей // The 2023 Symposium on Cybersecurity of the Digital Economy – CDE'23 : Сборник трудов VII международной научно-технической конференции, Иннополис, 11-12 апреля 2023 года. Иннополис: Университет Иннополис. 2024. С. 85–90.

METHODOLOGY FOR SYNTHESIZING QUANTUM-RESISTANT BLOCKCHAIN PLATFORMS WITH CYBER-IMMUNITY

Balyabin A. A.[3], Petrenko S. A.[4]

Keywords: *threats to information security, quantum threats to security, blockchain ecosystems and platforms, cyber-security, cyber resilience, methods of analysis and synthesis of quantum-resistant blockchain.*

Purpose of the research: *development of a methodology for the parametric synthesis of cyber-resilient blockchain ecosystems and platforms of the 'Data Economy' of the Russian Federation with cyber-immunity under the new quantum threat.*

Methods of the research: *methods of system analysis, methods of probability theory and mathematical statistics, methods of the theory of stability of complex systems, methods of similarity and dimensionality theory.*

Result of the research: *an study of existing approaches to ensuring quantum resilience of blockchain platforms with cyber-immunity has been conducted; a hypothesis regarding the possibility of ensuring the required cyber-resilience of blockchain platforms with cyber-immunity under quantum attacks has been formulated; a methodology for the parametric synthesis of quantum-resilient blockchain ecosystems and platforms of the 'Data Economy' of the Russian Federation with cyber-immunity has been developed using similarity theory methods; experimental studies of the methodology have been carried out, the results of which confirmed the proposed hypothesis.*

Scientific novelty: *the proposed methodology differs from existing ones in that it introduces new groups of formal operations for each level of blockchain platforms with cyber-immunity functionality, aimed at evaluating the necessary and sufficient values of neutralizing impact parameters using similarity theory methods. Additionally, it includes a criterion that allows establishing the existence of a solution for the given values of required cyber-resilience and time performance indicators of the blockchain platform.*

The results were obtained with the financial support of the project «Technologies for countering previously unknown quantum cyber threats», implemented within the framework of the state program of the «Sirius» Federal Territory «Scientific and technological development of the «Sirius» Federal Territory (Agreement No. 23-03 dated September 27, 2024).

References

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. Nguyen D. C. et al. 6G Internet of Things: A Comprehensive Survey // IEEE Internet of Things Journal. 2022. Vol. 9. No. 1. Pp. 359–383. DOI: 10.1109/JIOT.2021.3103320.
3. Balyabin A. A., Petrenko S. A., Kostyukov A. D. Model' ugroz bezopasnosti i kiberustoychivosti oblachnykh platform KII RF // Zashchita informatsii. Insayd. 2024. № 5 (119). Pp. 26–34.
4. Markov A. S. Vazhnaya vekha v bezopasnosti otkrytogo programmnogo obespecheniya // Voprosy kiberbezopasnosti. 2023. № 1 (53). Pp. 2–12. DOI: 10.21681/2311-3456-2023-1-2-12.
5. Chen C. et al. When Digital Economy Meets Web3.0: Applications and Challenges // IEEE Open Journal of the Computer Society. 2022. Vol. 3. Pp. 233–245. DOI: 10.1109/OJCS.2022.3217565.
6. Zhu Q., Loke S. W., Trujillo-Rasua R., Jiang F., Xiang Y. Applications of Distributed Ledger Technologies to the Internet of Things: A Survey // ACM Comput. Surv. 2019. Vol. 52. No. 6. P. 120:1–120:34. DOI: 10.1145/3359982.
7. Petrenko A. S. Kvantovo-ustoychivyy blokcheyn: nauchnaya monografiya. // Sankt-Peterburg : Piter, 2023. 384 p.
8. Petrenko S. A. Kiberustoychivost' Industrii 4.0: nauchnaya monografiya // «Izdatel'skiy Dom «Afina». 2020. 256 p.
9. Markova S. V. Vyyavleniya uyazvimostey v detsentralizovannykh informatsionnykh sistemakh na osnove smart-kontraktov s pomoshch'yu metodov obrabotki bol'shikh dannykh // Fundamental'nye issledovaniya. 2022. № 9. Pp. 47–53.

3 Artyom Balyabin, Junior Researcher, Scientific Center for Information Technologies and Artificial Intelligence, Sirius University of Science and Technology, Sirius Federal Territory, E-mail: Balyabin.AA@talantiuspeh.ru

4 Sergei Petrenko, Dr.Sc. (in Tech.) (Grand Doctor, Full Professor), Scientific Center for Information Technologies and Artificial Intelligence, Sirius University of Science and Technology, Sirius Federal Territory, Orcid.org/0000- 0003-0644-1731, E-mail: Petrenko.SA@talantiuspeh.ru

10. Petrenko S. A., Balyabin A. A. Model' kvantovykh ugroz bezopasnosti informatsii dlya natsional'nykh blokcheyn-ekosistem i platform // Voprosy kiberbezopasnosti. 2025. № 1(65). Pp. 7–17. DOI 10.21681/2311-3456-2025-1-7-17.
11. Kushwaha S. S., Joshi S., Singh D., Kaur M. Lee H. -N. Systematic Review of Security Vulnerabilities in Ethereum Blockchain Smart Contract // IEEE Access. 2022. Vol. 10. Pp. 6605–6621. DOI: 10.1109/ACCESS.2021.3140091.
12. Fernandez-Carames 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.
13. Petrenko A. S., Romanchenko A. M. Perspektivnyy metod kriptoanaliza na osnove algoritma Shora // Zashchita informatsii. Insayd. 2020. № 2 (92). Pp. 17–23.
14. Petrenko A., Petrenko S. Basic Algorithms Quantum Cryptanalysis // Voprosy Kiberbezopasnosti. 2023. No. 1 (53). Pp. 100–115. DOI 10.21681/2311-3456-2023-1-100-115.
15. Balyabin A. A. Model' oblachnoy platformy KII RF s kiberimmunitetom v usloviyakh informatsionno-tekhnicheskikh vozdeystviy // Zashchita informatsii. Insayd. 2024. № 5 (119). Pp. 35–44.
16. Balyabin A. A., Petrenko S. A., Kostyukov A. D. Metod vosstanovleniya oblachnykh i pogranichnykh vychisleniy na osnove kiberimmuniteta // Zashchita informatsii. Insayd. 2022. № 6(108). Pp. 26–31.
17. Andrushkevich D. V., Biryukov D. N., Timashov P. V. Porozhdenie stsenariev predotvrashcheniya komp'yuternykh atak na osnove logiko-ontologicheskogo podkhoda // Trudy Voenno-kosmicheskoy akademii imeni A.F.Mozhayskogo. 2021. № 677. Pp. 118–134.
18. Zegzhda D. P., Aleksandrova E. B., Kalinin M. O. i dr. Kiberbezopasnost' tsifrovoy industrii. Teoriya i praktika funktsional'noy ustoychivosti k kiberatakam // Moskva : Nauchno-tekhnicheskoe izdatel'stvo «Goryachaya liniya-Telekom». 2021. 560 p.
19. Petrenko S. A. Kiberimmunologiya // Sankt-Peterburg : Afina. 2021. 239 p.
20. Balyabin A. A., Petrenko S. A. Metodika kiberimmunnoy zashchity tsifrovykh servisov «GosTekh» s ispol'zovaniem teorii podobiya i razmernostey // The 2023 Symposium on Cybersecurity of the Digital Economy – CDE'23 : Sbornik trudov VII mezhdunarodnoy nauchno-tekhnicheskoy konferentsii, Innopolis, 11-12 aprelya 2023 goda. Innopolis: Universitet Innopolis. 2024. Pp. 85–90.

ОБЕСПЕЧЕНИЕ ФУНКЦИОНАЛЬНОСТИ ЦИФРОВЫХ УСТРОЙСТВ РЕЛЕЙНОЙ ЗАЩИТЫ ПРИ КИБЕРАТАКАХ НА МИКРОСЕТИ С РАСПРЕДЕЛЕННЫМИ ЭНЕРГЕТИЧЕСКИМИ РЕСУРСАМИ

Гурина Л. А.[1], Томин Н. В.[2]

DOI: 10.21681/2311-3456-2025-4-55-64

Цель исследования: *разработка метода достоверизации данных, используемых в цифровых системах релейной защиты, автоматики и управления микросетями с распределенными энергетическими ресурсами.*

Методы исследования: *вероятностные методы, методы машинного обучения.*

Результат исследования: *рассмотрено информационное обеспечение схем релейной защиты микросетей, проанализированы возможные кибератаки, при успешной реализации которых может стать нарушение функциональности цифровых устройств релейной защиты микросетей. Разработан метод достоверизации данных с использованием методов обучения без учителя, включая изоляционный лес и метод k-ближайших соседей, который эффективно выявляет и корректирует ошибки в измерениях при атаках внедрения ложных данных на информационную инфраструктуру систем защиты микросетей.*

Научная новизна *состоит в том, что предложен подход, обеспечивающий устойчивость систем релейной защиты микросетей при кибератаках и, тем самым, не допускающий ложных срабатываний и отказов устройств защиты.*

Ключевые слова: *активные системы распределения электроэнергии, схемы защиты, кибербезопасность, достоверизация данных, случайные процессы, машинное обучение.*

Введение

В настоящее время наблюдается постепенное сокращение доли традиционных источников энергии, что сопровождается массовым внедрением генерирующих объектов на основе возобновляемых источников энергии (ВИЭ) [1, 2]. Процессы декарбонизации и интеллектуализации энергетических систем позволяют эффективно использовать всевозможные вилы источников энергии [3]. Одним из путей развития активных систем распределения электрической энергии, включающих высокий уровень объектов распределенной энергетики, является интеграция в них микросетей (МС). Такие МС в силу внедрения объектов распределенной энергетики и ВИЭ с привлечением устройств силовой электроники имеют достаточно сложную информационную инфраструктуру, отказы и сбои в которой могут оказывать существенное влияние на надежность и безопасность не только МС, но и всей распределительной сети. Кроме того, при цифровой трансформации МС становятся более уязвимыми к кибератакам [4, 5].

Активные системы распределения электроэнергии отличаются от традиционных распределительных сетей тем, что могут иметь двунаправленные потоки мощности, широкое внедрение распределенных энергетических ресурсов (РЭР), возможности хранения электрической энергии, сложные стратегии управления и обладают высокой зависимостью от информационной инфраструктуры, увеличивая их уязвимости к киберугрозам.

Концепцией построения современных распределительных сетей, включающих РЭР, является интеграция в них МС, позволяющим повысить наблюдаемость и управляемость, а также обеспечить надежное электроснабжение потребителей. МС с РЭР все чаще развертываются для повышения операционной гибкости, устойчивости, возможностей скоординированного управления. Созданная на основе интеллектуальных технологий информационная инфраструктура активных систем распределения электрической энергии позволяет расширить перечень решаемых технологических задач и повысить эффективность управления МС с РЭР. К информационной инфраструктуре относятся в первую очередь каналы передачи данных от локальных контроллеров до центрального контроллера в случае централизованного управления или сеть передачи

1 Гурина Людмила Александровна, кандидат технических наук, доцент, старший научный сотрудник лаборатории управления функционированием электроэнергетических систем Института систем энергетики им. Л. А. Мелентьева СО РАН, Иркутск, Россия. E-mail: gurina@isem.irk.ru

2 Томин Никита Викторович, кандидат технических наук, заведующий лабораторией управления функционированием электроэнергетических систем Института систем энергетики им. Л. А. Мелентьева СО РАН, Иркутск, Россия. E-mail: tomin.nv@gmail.com

данных между соседними контроллерами при распределенном управлении МС. С расширением возможностей злоумышленников для проведения успешных кибератак, связанных с ростом уязвимостей в информационной инфраструктуре, требуется разработка методов своевременного обнаружения кибератак и устранения их влияния на функциональность систем релейной защиты (РЗ), автоматики и управления МС с РЭР.

В зависимости от типа активной системы распределения электроэнергии МС подразделяются на сети постоянного тока, переменного тока и гибридные [6]. Для обеспечения надежной работы МС с РЭР важно решение таких задач, как управление режимами, распределение мощности, сохранение устойчивости и эффективной работы средств РЗ. Причем, зашита МС с РЭР становится первостепенной для сохранения их безопасной работы в условиях новых вызовов и угроз [7]. Функционирований МС с РЭР зависит от ее информационной инфраструктуры из-за большого числа коммуникаций [8]. При успешной реализации кибератак несвоевременное принятие мер по их обнаружению и подавлению последствий может привести к всевозможным сбоям и отказам в технологической подсистеме МС с РЭР [9]. Надежная работа МС с РЭР требует достоверных измерений и защищенных от кибератак систем РЗ, автоматики и управления [10].

Обнаружение неисправностей и защита от них МС с РЭР – сложная проблема из-за переменных по величине и двунаправленных потоков мощности, а также различных типов распределенных источников энергии и мест их подключения. В последнее время, наряду с применением традиционных средств РЗ, в МС с РЭР внедряются инновационные защиты, например, адаптивные РЗ, защита на основе IoT, мультиагентные РЗ, РЗ на основе машинного обучения, искусственного интеллекта и др. Наряду с преимуществами применения таких РЗ возникают проблемы, связанные с нарушениями функциональности схем защиты, целостности и доступности данных в результате кибератак на информационную инфраструктуру МС с РЭР.

Использование современных устройств и датчиков для мониторинга, защиты и управления МС с РЭР также может привести к нарушению их кибербезопасности. Срабатывания устройств РЗ зависят от каналов передачи данных, которые могут быть подвержены множеству кибератак. В результате успешной кибератаки может произойти изменение настроек не только РЗ, но и всего комплекса контролирующих устройств. Современные системы РЗ различных видов требуют наличия информационно-коммуникационной инфраструктуры, на которую могут воздействовать злоумышленники. Таким образом, инновационные защиты МС с РЭР могут оказаться неэффективными в условиях кибератак, последствиями которых могут быть отказы объектов информационной системы, каналов связи, а также нарушение качества данных. Обеспечение кибербезопасности цифровых защит МС с РЭР становится актуальным.

В статье рассмотрены проблемы в области кибербезопасности схем РЗ и возможные пути их решения для осуществления безопасной и надежной работы МС РЭР.

Стратегии защиты МС с РЭР и их информационное обеспечение

В [11] представлен обзор технических достижений, направленных на улучшение традиционных механизмов РЗ, используемых в пассивных распределительных сетях, а также разработку современных схем РЗ, основанных на инновационных подходах и новых методах. Подчеркивается, что адаптивные системы защиты, разработанные как для децентрализованных, так и для централизованных структур, в целом обеспечивают надежную защиту для МС с РЭР. Современные устройства РЗ, которые интегрируют передовые функциональные возможности защиты, демонстрируют лучшие показатели в таких областях, как надежность, чувствительность, селективность и скорость работы.

Во время работы МС с РЭР могут возникать аномальные и аварийные ситуации из-за таких факторов, как токовые перегрузки и короткие замыкания. Обнаружение этих условий и определение неисправных компонентов в МС является сложной задачей. В результате в МС с отключаемыми объектами распределенной генерации требуется тщательный анализ влияния информации, прежде всего, содержащей «плохие данные», для предотвращения сбоев в работе, а также нежелательных или ложных срабатываний РЗ [12] в случае кибератак. В исследовании [13] предлагается объединить централизованную защиту, дистанционную и многоагентную РЗ в единую категорию, известную как распределенные системы РЗ, с использованием протокола беспроводных приложений (WAP). Такие распределенные системы имеет высокий потенциал благодаря достижениям в области информационно-коммуникационной инфраструктуры в МС с РЭР, что позволяет эффективно обмениваться большими объемами данных [14].

Внедрение дистанционного управления для выключателей, развертывание дополнительных датчиков тока и напряжения [15], а также использование устройств РЗ на основе современных интеллектуальных алгоритмов и возможности удаленной настройки параметров защиты способствуют возрастанию восприимчивости систем РЗ к киберугрозам.

Существуют две основные методологии для построения адаптивных систем РЗ: децентрализованная и централизованная. Централизованный подход включает архитектуру управления, которая взаимодействует с каждым устройством РЗ, настраивая их конфигурации в зависимости от текущего режима МС с РЭР. В отличие от этого, децентрализованный подход основывается на обмене данными между цифровыми устройствами РЗ, когда каждое такое устройство может независимо изменять свои настройки на основе информации, полученной от других устройств. Значительная часть исследований в области адаптивной защиты сосредоточена на использовании интеллектуальных аналитических методов, таких как мультиагентные методы, которые способны обрабатывать большие объемы данных [16].

Система, которая объединяет отдельных агентов, работающих на достижение общей цели с использованием данных от каждого агента для выработки решений, называется мультиагентной системой (МАС). Применение МАС при реализации защиты МС с РЭР является достаточно эффективным, поскольку вводит новые методы для оптимизации целевой функции, тем самым улучшая селективность и скорость реакции. В [17] предлагается схема дифференциальной защиты с использованием МАС с регулируемыми временными задержками срабатывания. Агенты оснащены слоями первичной, резервной и шиной защиты, которые применяются как для первичной, так и для резервной защиты. В [18] вводится схема защиты, включающая алгоритм Q-обучения и МАС. Агенты обучаются с использованием алгоритма Q-обучения для выявления и устранения неисправностей. Кроме того, агенты общаются через децентрализованные сети на основе блокчейна. В [19] представлена схема защиты, использующая надежные интеллектуальные агенты. Обнаружение неисправностей осуществляется путем измерения напряжения и тока. Компоненты функционируют как агенты, передающие решения вышестоящим управляющим слоям и смежным компонентам для повышения общей производительности.

Основной и ключевой функцией схем защиты является распознавание и обнаружение возникновения неисправностей. Для предотвращения аварийных ситуаций в МС с РЭР в системах релейной защиты и автоматики заложены схемы обнаружения и классификации неисправностей, которые могут быть основаны на обработке сигналов, знаниях и моделях.

В активных системах распределения электроэнергии, в том числе и в МС с РЭР, широкое распространение получили методы обнаружения неисправностей, основанные на обработке сигналов. При возникновении аварийных ситуаций параметры системы отклоняются от значений при нормальной функционировании системы. Для обнаружения неисправностей в активных системах распределения электроэнергии на основе обработки сигналов применяют вейвлет-анализ данных [20], Фурье-анализ сигналов [21], оконное преобразование Фурье [22], метод эмпирической модовой декомпозиции сигналов и преобразование Гильберта-Хуанга [23].

Наряду с методами обнаружения неисправностей, основанных на обработке сигналов, в системах релейной защиты и автоматики применяются такие методы искусственного интеллекта и машинного обучения, как искусственные нейронные сети, глубокое обучение, алгоритмы обучения с подкреплением, нечеткая логика, адаптивные системы нечеткого вывода, метод опорных векторов и т.д. [24–27].

Методы обнаружения неисправностей в МС с РЭР на основе модели используют аналитические знания для оценки параметров системы. Авторами [28] предложен метод на основе инверторов, применение которого позволяет обнаруживать неисправности независимо от режима работы МС с РЭР и от уровней токов короткого замыкания.

В [29] определено понятие обнаружение локализации неисправности, представляющего собой определение местоположения неисправности с максимальной точностью и достоверностью. Сами методы определения локализации неисправностей подразделяются на основе сигналов напряжения и тока на основной частоте [30].

Метод восстановления качества данных при кибератаках на информационную инфраструктуру систем защиты МС с РЭР

Развертывание МС с РЭР требует применения адаптивных схем защиты, устройства которых могут изменять свои настройки при каких-либо изменениях, обусловленных внутренними и внешними угрозами. Даже незначительные неисправности в информационной инфраструктуре могут иметь катастрофические последствия для физической части МС с РЭР. Внедрение цифровых устройств измерений, датчиков, устройств РЗ, использование многочисленных каналов связи и применение различных методов обнаружения неисправностей на основе искусственного интеллекта, машинного обучения и требующих большого многообразия данных для реализации схем защиты без соответствующих мер по обеспечению кибербезопасности может привести к отказам функционирования МС с РЭР при кибератаках [31].

Функциональность устройств и каналов связи информационной инфраструктуры являются важным для реализации современных алгоритмов адаптивной защиты.

Возможные кибератаки и их последствия для схем защиты могут быть следующими:

- Атака, направленные на целостность. Данная кибератака может быть в виде ложной команды, последствиями которой являются изменение логики управления и неправильная работа цифровых объектов релейной защиты.
- Атака «человек по середине». При атаке «человек по середине» противником может быть получена экстренная информации об изменении режима работы устройства релейной защиты и, в дальнейшем, заменена нужной для злоумышленника информацией.
- Атака повторного воспроизведения. При получении доступа к передаче данных по каналам связи противники могут повторять отправку данных из ранее отправленных, имитируя появление нарушений, которые имели место быть.
- Атака внедрения ложных данных. При данной кибератаке противники могут сымитировать ложные изменения в конфигурациях микропроцессорных устройств релейной защиты для отключения их функций при возникновении аварийных ситуаций.
- Атака отказа в обслуживании. При атаке отказа в обслуживании противник может вызвать сбои в каналах связи, переполнение буфера и, тем самым, нарушить передачу потоков данных. При этом возникает потеря информации, используемая в схемах защиты.
- Вредоносный код. При внедрении вредоносного кода в цифровое устройство РЗ может произойти переполнение буфера.

В области защиты МС с РЭР мультиагентные системы являются эффективными платформами связи в реальном времени. Модульная структура агентов обеспечивает более тесное взаимодействие между информационной инфраструктурой и технологической частью МС для реализации их деятельности с помощью сложных программных платформ.

Ранее авторами [32] была разработана двухэтапная процедура обнаружения и восстановления качества данных при кибератаках с использованием методов машинного обучения без учителя: изоляционный лес (англ. Isolation Forest) и k-ближайших соседей (англ. k-nearest neighbors algorithm, k-NN) при вторичном регулировании напряжения на основе МАС. Важно заметить, что обнаружение аномалий с помощью ансамблевой модели (алгоритм изоляционного леса) не требует разметки данных и эффективно выявляет отклонения в параметрах системы. Кроме того, при восстановлении данных с помощью метода k-ближайших соседей, используется информация о соседних нормальных измерениях для предполагаемых аномалий. На рис. 1 показана архитектура этой процедуры, где s_1, s_2, ..., s_N являются измеренными значениями, полученными физическим объектом (выход датчика), переменные f_1, f_2, ..., f_N – двоичные переменные, выводимыми моделью изоляционного леса и указывающие на присутствие сбоя и/или грубого измерения датчика, s_1, s_2, ..., s_N – «восстановленные» данные, в случае если грубое измерение было обнаружено в модели изоляционного леса и N – количество датчиков.

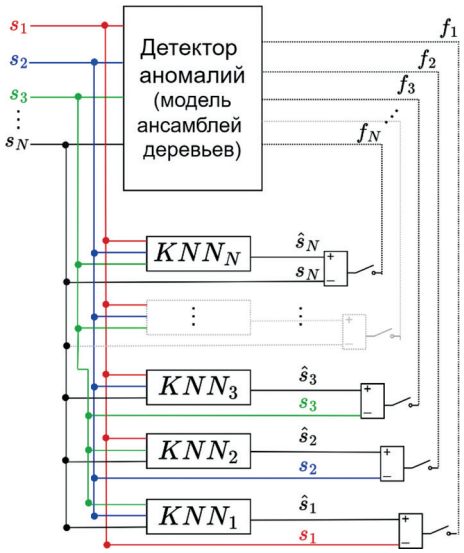


Рис. 1. Схема достоверизации данных на основе моделей машинного обучения

Данная процедура была развита и адаптирована и для схем защиты МС. Рассмотрим систему из N взаимосвязанных МС, где каждая i-ая МС характеризуется следующими параметрами: измеренные датчиками напряжение $V(t)$ и ток $I(t)$, значения которых изменяются во времени t. Пусть в примере взаимосвязанная система будет состоять из трех МС ($N = 3$). Смоделировать измерения процессов изменения напряжения и тока можно с учетом нормальной работы цифровых датчиков и случайного шума:

$$V(t) = V_0 + A_V \sin(\omega t) + \epsilon_V(t), \quad (1)$$

$$I(t) = I_0 + A_I \sin(\omega t + \phi) + \epsilon_I(t), \quad (2)$$

где V_0 и I_0 – средние значения напряжения и тока; A_V и A_I – амплитуды колебаний; ω – угловая частота; ϕ – фазовый сдвиг между напряжением и током; $\epsilon_V(t)$ и $\epsilon_I(t)$ – случайный шум, распределенный по нормальному закону.

Для оценки влияния кибератаки на цифровые устройства РЗ была смоделирована атака внедрения ложных данных в виде ошибок измерений случайного процесса изменения напряжения и тока [33],

которая моделируется как мультипликативное искажение:

$$V_i^{\text{атака}}(t_k) = V_i(t_k) \cdot k_{Vi}(t_k), \quad k_{Vi} \sim U(0.5,1.5), \tag{3}$$

$$I_i^{\text{атака}}(t_k) = I_i(t_k) \cdot k_{Ii}(t_k), \quad k_{Ii} \sim U(0.5,1.5), \tag{4}$$

где k_{Vi} и k_{Ii} – случайные коэффициенты искажения, применяемые к выбранным моментам времени t_i.

Для обнаружения аномалий используется алгоритм изоляционного леса, который идентифицирует точки, резко отличающиеся от общей структуры данных. Алгоритм основывается на построении множества случайных деревьев решений, где аномальные точки изолируются на меньшей глубине дерева по сравнению с нормальными данными. Для каждой i-ой МС алгоритм изоляционного леса может быть записан следующим образом:

$$s_i(V_i,I_i) = \frac{1}{M} \sum_{m=1}^{M} h_m(V_i,I_i), \tag{5}$$

h_m – глубина изоляции в дереве m, M – количество деревьев.

Помимо алгоритма изоляционного леса для детектирования аномалий используется координационный механизм между агентами МАС:

$$s_i^{glob} = \alpha s_i + \beta \frac{1}{|N_i|} \sum_{j \in N_i} s_j, \tag{6}$$

где N_i – соседние МС, $\alpha + \beta = 1$ – веса локальных и глобальных показателей.

После выявления аномальных измерений производится восстановление корректных значений с использованием метода k-ближайших соседей. Для каждого аномального измерения напряжения $V_{\text{КА}}(t_i)$ находим k-ближайших соседей среди нормальных данных по признакам времени и тока $[t, I(t)]$. Восстановленное значение напряжения вычисляется как среднее значение напряжения этих соседей:

$$V_{\text{восст}}(t_i) = \frac{1}{k} \sum_{j=1}^{k} V_j, \tag{7}$$

где V_j – значения напряжения у k-ближайших соседей.

Алгоритм работы МАС для координации киберзащиты цифровых средств РЗ микросетей на основе двухуровневой процедуры может быть описан следующим образом. Каждый агент A_i независимо обнаруживает аномалии в своей МС. Агенты обмениваются показателями аномальности s_i. При превышении порога $S_i^{glob} > S$ запускается процедура восстановления. Восстановленные значения используются для корректировки уставок защит.

Результаты достоверизации измерений тока и напряжения на основе предлагаемого метода при смоделированных кибератаках на устройства РЗ представлены на рис. 2. На рис. 3 показаны результаты восстановления качества данных с использованием предложенной процедуры. Хорошо видно, что категория «attacked», которая указывает на искаженные, но невосстановленные данные, фактически равна нулю, что свидетельствует о высокой эффективности предложенного подхода.

В результате моделирования была количественно оценена эффективность как модели обнаружения аномалий, так и модели восстановления качества данных. Полученные результаты, представленные в табл. 1, демонстрирует высокую эффективность алгоритма изоляционного леса в идентификации

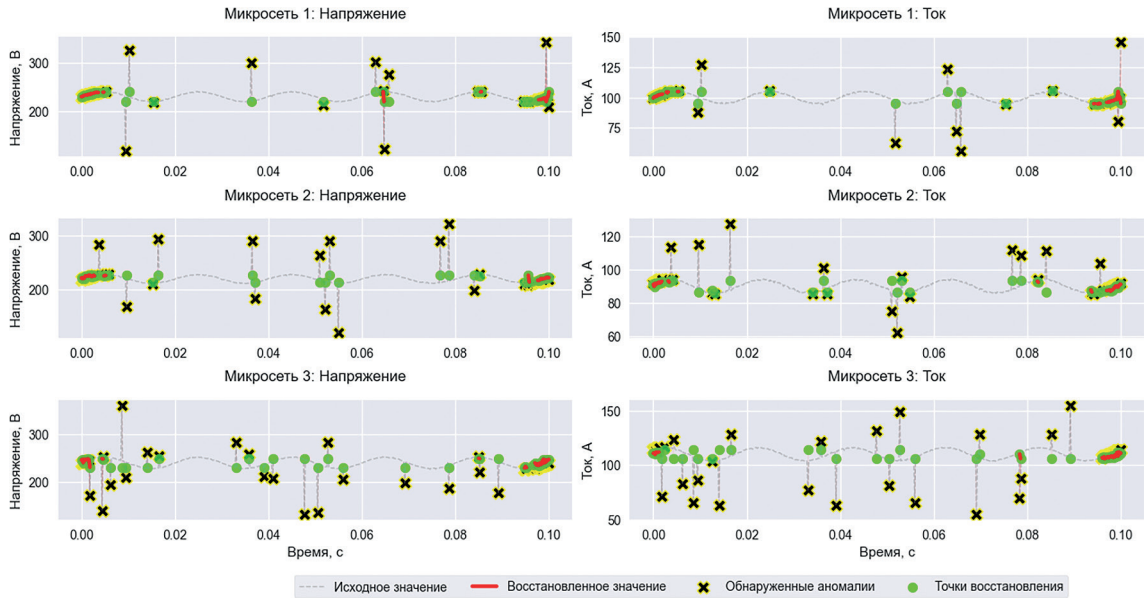


Рис. 2. Восстановление данных измерений при атаке внедрения ложных данных на устройства мультиагентной РЗ для взаимосвязанных микросетей

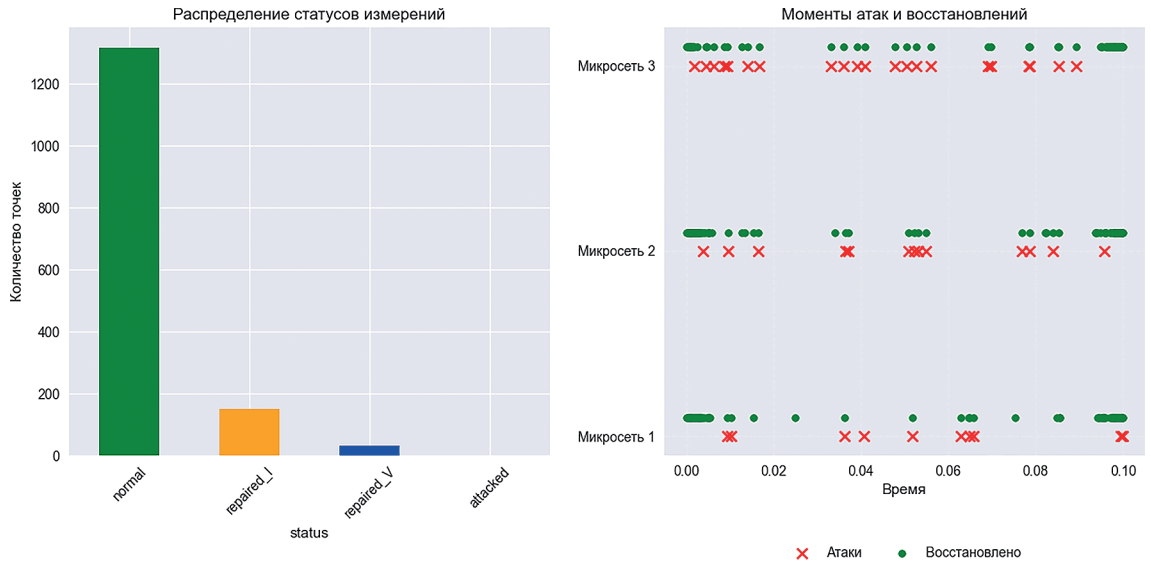


Рис. 3. Результат процесса восстановления качества данных цифровых устройств РЗ микросетей, связанных МАС для координации и регулирования: а) распределение статусов измерений (normal – нормальные измерения, $repaired_I$ и $repaired_V$ – восстановленные измерения тока и напряжения соответственно, attacked – искаженные, невосстановленные измерения); б) моменты атак и восстановлений

аномальных измерений. Модель достигла точности и полноты 98–100 % для обоих классов, что свидетельствует о практически безошибочном распознавании как нормальных данных (класс 0), так и аномалий (класс 1). Общая точность составляет 100 %, а средние значения метрик (macro avg и weighted avg) подтверждают стабильность и надежность модели [34]. Эти результаты указывают на то, что предложенный подход обеспечивает эффективное обнаружение аномалий, что существенно повышает устойчивость схем защиты МС с РЭР в условиях возможных кибератак.

Таблица 1.

Обнаружение аномалий на базе алгоритма изоляционного леса

	precision	**recall**	**f1-score**	**support**
0	1.00	1.00	1.00	950
1	0.98	0.98	0.98	50
accuracy			1.00	1000
macro avg	0.99	0.99	0.99	1000
weighted avg	1.00	1.00	1.00	1000

Для оценки восстановления качества данных на базе алгоритма k-ближайших соседей были вычислены следующие метрики: среднеквадратическая ошибка (MSE), средняя абсолютная ошибка (MAE) и средняя абсолютная процентная ошибка (MAPE) [35]. Результаты оценки этих показателей представлены в табл. 2. Низкие значения MSE и MAE указывают на то, что восстановленные значения очень близки к истинным в абсолютных единицах. Низкий MAPE (менее 1-2 %) свидетельствует о высокой относительной точности восстановления по сравнению с истинными значениями.

Таблица 2.

Оценка эффективности восстановления качества данных

Параметр	**Показатели**		
	MSE	**MAE**	**MAPE, %**
Напряжение	0.53	0.58	0.26
Ток	0.03	0.15	1.48

Выводы

Проанализированы стратегии защиты МС с РЭР. Показано, что современные виды защит требуют наличия информационно-коммуникационной инфраструктуры, что увеличивает уязвимости схем защиты к кибератакам. Выявлены возможные последствия реализованных кибертак на функциональность схем защиты МС с РЭР. Разработан подход к достоверизации данных, позволяющий обнаруживать ошибки в измерениях и восстанавливать данные, требуемые для цифровых систем релейной защиты МС с РЭР, тем самым, предотвращая ложные срабатывания и сбои в работе устройств релейной защиты при кибератаках. Результаты демонстрируют высокую точность обнаружения аномалий в измерениях напряжения и тока (до 98–100 % успешных

обнаружений) и их восстановления (средняя ошибка не превышает 1-2 %). Система успешно масштабируется для нескольких взаимосвязанных микросетей, сохраняя стабильность работы при различных типах атак. Реализация данного метода способствует повышению кибербезопасности схем релейной защиты, позволяя обеспечить надежность и безопасность функционирования МС с РЭР.

Работа выполнена в рамках научного проекта «Теоретические основы, модели и методы управления развитием и функционированием интеллектуальных электроэнергетических систем», № FWEU-2021-0001.

Литература

1. Воропай Н. И. Направления и проблемы трансформации электроэнергетических систем // Электричество. 2020. № 7. С. 12–21. DOI: 10.24160/0013-5380-2020-7-12-21.
2. Илюшин П. В. Интеграция электростанций на основе возобновляемых источников энергии в Единой энергетической системе России: обзор проблемных вопросов и подходов к их решению // Вестник Московского энергетического института. 2022. № 4. С. 98–107. DOI: 0.24160/1993-6982-2022-4-98-107.
3. Ilyushin P., Filippov S., Kulikov A., Suslov K. and Karamov D. Intelligent Control of the Energy Storage System for Reliable Operation of Gas-Fired Reciprocating Engine Plants in Systems of Power Supply to Industrial Facilities // Energies. 2022. Vol. 15, 6333. DOI: 10.3390/en15176333.
4. Гурина Л. А. Показатели киберустойчивости компонентов информационно-коммуникационной инфраструктуры при управлении киберфизическими энергетическими системами // Методические вопросы надежности больших систем энергетики. 2022. Выпуск 73. С. 279–288.
5. Дураковский А. П., Марков А. С. Актуальные вопросы кибербезопасности в энергетике // Безопасные информационные технологии. 2024. С. 94–98.
6. Shaukat N. et al. Decentralized, Democratized, and Decarbonized Future Electric Power Distribution Grids: A Survey on the Paradigm Shift From the Conventional Power System to Micro Grid Structures // in IEEE Access. 2023. Vol. 11. Pp. 60957–60987. DOI: 10.1109/ACCESS.2023.3284031.
7. Lu H., Biyawerwala H. and Thakrawala H. Polarized Distribution Protection Coordination Strategy Under the Impact from Various Distributed Energy Resources (DER) Generation Points // 2022 IEEE/PES Transmission and Distribution Conference and Exposition (T&D), New Orleans, LA, USA. 2022. Pp. 1–5. DOI: 10.1109/TD43745.2022.9816925.
8. S. K. T, Jadoun V. K., J. N. S and S. S. A Systematic Study on the Intelligent Cyber Security for Smart Microgrid // 2024 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER), Mangalore, India. 2024. Pp. 237–242. DOI: 10.1109/DISCOVER62353.2024.10750634.
9. Canaan B., Colicchio B., Abdeslam D. O. Microgrid Cyber-Security: Review and Challenges toward Resilience // Applied Sciences. 2020. Vol. 10, no. 16. Pp. 5649. DOI: 10.3390/app10165649.
10. Ding D., Han Q. -L., Ge X. and Wang J. Secure State Estimation and Control of Cyber-Physical Systems: A Survey // in IEEE Transactions on Systems, Man, and Cybernetics: Systems. Jan. 2021. Vol. 51, No. 1. Pp. 176–190. DOI: 10.1109/TSMC.2020.3041121.
11. Илюшин П. В., Вольный В. С. Обзор методов решения проблемных вопросов функционирования устройств защиты в microgrid напряжением до 1 кВ с распределенными источниками энергии // Релейная защита и автоматизация. 2022. № 4(49). С. 6–21.
12. Zheng Dehua, Zhang Wei, Netsanet Solomon, Wang Ping, Bitew Girmaw Teshager, Wei Dan, Yue Jun. Key technical challenges in protection and control of microgrid // Microgrid Protection and Control. 2021. Pp. 45–56. DOI: 10.1016/B978-0-12-821189-2.00007-3.
13. Patnaik B., Mishra M., Bansal R. C., Jena R. K. AC microgrid protection – A review: Current and future prospective // Applied Energy. 2020. Vol. 271. Pp. 115210. DOI: 10.1016/J.APENERGY.2020.115210.
14. Abd el-Ghany H. A. Optimal PMU allocation for high-sensitivity wide-area backup protection scheme of transmission lines // Electric Power Systems Research. 2020. Vol. 187. Pp. 106485. DOI: 10.1016/J.EPSR.2020.106485.
15. Kulikov A., Loskutov A., Bezdushniy D. Relay Protection and Automation Algorithms of Electrical Networks Based on Simulation and Machine Learning Methods // Energies. 2022. Vol. 15(18). Pp. 6525. DOI: 10.3390/en15186525.
16. Shobole A. A., Wadi M. Multiagent systems application for the smart grid protection // Renewable and Sustainable Energy Reviews. 2021. Vol. 149. Pp. 111352. DOI: 10.1016/J.RSER.2021.111352.
17. Verma R., Gawre S. K., Patidar N. P. An Analytical Review on Measures of Microgrid Protection // 2022 IEEE International Conference on Power Electronics, Drives and Energy Systems (PEDES), Jaipur, India. 2022. Pp. 1–6. DOI: 10.1109/PEDES56012.2022.10080291.
18. Cui S., Zeng P., Wang Z., Song C. Research on Intelligent Protection Technology for Distribution Network with Distributed Generation // 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chongqing, China. 2021. Pp. 1549–1554. DOI: 10.1109/IAEAC50856.2021.9390692.
19. Fawzy N., Habib H. F., Mohammed O., Brahma S. Protection of Microgrids with Distributed Generation based on Multiagent System // 2020 IEEE International Conference on Environment and Electrical Engineering and 2020 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe), Madrid, Spain. 2020. Pp. 1–5. DOI: 10.1109/EEEIC/ICPSEurope49358.2020.9160827.
20. Chaitanya B. K., Anamika Yadav. Empirical Wavelet Transform-Based Differential Protection Scheme for Micro-Grid // Journal of The Institution of Engineers (India): Series B. 2023. 104. Pp. 1–10. DOI: 10.1007/s40031-023-00869-0.
21. Uddin M. N., Arifin M. S., Rezaei N. A Novel Neuro-Fuzzy Based Direct Power Control of a DFIG based Wind Farm Incorporated with Distance Protection Scheme and LVRT Capability // 2022 IEEE Industry Applications Society Annual Meeting (IAS), Detroit, MI, USA. 2022. Pp. 01–08. DOI: 10.1109/IAS54023.2022.9939684.

22. Chaitanya B. K., Anamika Yadav, Mohammad Pazoki. High Impedance Fault Detection Scheme for Active Distribution Network Using Empirical Wavelet Transform and Support Vector Machine // 2020 15th International Conference on Protection and Automation of Power Systems (IPAPS). 2020. Pp. 149–152. DOI:10.1109/IPAPS52181.2020.9375620.
23. Shaik M., Shaik A. G., Yadav S. K. Hilbert–Huang transform and decision tree based islanding and fault recognition in renewable energy penetrated distribution system // Sustainable Energy, Grids and Networks. 2022. 30. Pp. 100606. DOI: 10.1016/j.segan.2022.
24. Raad Salih Jawad, Hafedh Abid. HVDC Fault Detection and Classification with Artificial Neural Network Based on ACO-DWT Method // Energies. 2023. Vol. 16. Pp. 1064. DOI: 10.3390/en16031064.
25. Chandran L. R., A. Parvathy V S, I. K, Nair M. G. Adaptive Over Current Relay Protection in a PV Penetrated Radial Distribution System With Fuzzy GA Optimisation // 2022 IEEE 19th India Council International Conference (INDICON), Kochi, India. 2022. Pp. 1–7. DOI: 10.1109/INDICON56171.2022.10040021.
26. Nasir M., Bansal R., Elnady A. A Review of Various Neural Network Algorithms for Operation of AC Microgrids // 2022 Advances in Science and Engineering Technology International Conferences (ASET), Dubai, United Arab Emirates. 2022. Pp. 1–7. DOI: 10.1109/ASET53988.2022.9734899.
27. Vincent Nsed Ogar, Sajjad Hussain, Kelum A. A. Gamage. The use of artificial neural network for low latency of fault detection and localisation in transmission line // Heliyon. 2023. Vol. 9. Pp. e13376. DOI: 10.1016/j.heliyon.2023.
28. Wang J. et al. Microgrid Fault Analysis Method Based on Inverter-Type DG with Different Control // 2022 4th International Conference on Smart Power & Internet Energy Systems (SPIES), Beijing, China. 2022. Pp. 1397–1402. DOI: 10.1109/SPIES55999.2022.10082157.
29. Xu Y. et al. A Novel Distribution Network Fault Location Method Based on Improved Convolutional Neural Network // 2024 IEEE 7th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chongqing, China. 2024. Pp. 837–841. DOI: 10.1109/ITNEC60942.2024.10733085.
30. Conte F., D'Agostino F., Gabriele B., Schiapparelli G. -P. and Silvestro F. Fault Detection and Localization in Active Distribution Networks Using Optimally Placed Phasor Measurements Units // in IEEE Transactions on Power Systems. 2023. Vol. 38, no. 1. Pp. 714–727. DOI: 10.1109/TPWRS.2022.3165685.
31. Гурина Л. А. Оценка рисков кибербезопасности энергетического сообщества микросетей // Вопросы кибербезопасности. 2024. № 1(59). С. 101–107. DOI: 10.21681/2311-3456-2024-1-101-107.
32. Гурина Л. А., Томин Н. В. Интеллектуальные методы обеспечения кибербезопасности мультиагентных систем управления микросетями // Вопросы кибербезопасности. 2024. № 6(64). С. 53–64. DOI: 10.21681/2311-3456-2024-6-53-64.
33. Колосок И. Н., Гурина Л. А. Идентификация кибератак на системы SCADA и СМПР в ЭЭС при обработке измерений методами оценивания состояния // Электричество. 2021. № 6. С. 25–32. DOI: 10.24160/0013-5380-2021-6-25-32.
34. Opitz J. A Closer Look at Classification Evaluation Metrics and a Critical Reflection of Common Evaluation Practice // Transactions of the Association for Computational Linguistics. 2024. Vol. 12. Pp. 820–836. DOI: 10.1162/tacl_a_00675.
35. Hodson T. Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not // Geoscientific Model Development. 2022. Vol. 15. Pp. 5481–5487. DOI: 10.5194/gmd-15-5481-2022.

ENSURING THE FUNCTIONALITY OF DIGITAL PROTECTION DEVICES IN THE EVENT OF CYBER-ATTACKS ON MICROGRIDS WITH DISTRIBUTED ENERGY RESOURCES

Gurina L. A.[3], Tomin N. V.[4]

Keywords: *active power distribution systems, protection schemes, cybersecurity, data verification, random processes, machine learning.*

The research aims *to develop method for method for verifying data used in digital systems for protection, automation and control of microgrids with distributed energy resources.*

The research relies *on the probabilistic methods, machine learning methods.*

Research result: *the information support of microgrids protection schemes is considered, possible cyber attacks are analyzed, the successful implementation of which may result in a violation of the functionality of digital protection devices of microgrids. A data verification method has been developed using unsupervised learning methods, including isolation forest and the k-nearest neighbors method, which effectively identifies and corrects measurement errors during attacks on the information infrastructure of microgrids protection systems under false data injection attacks.*

The scientific novelty *lies in the fact that an approach has been proposed that ensures the stability of microgrids protection systems during cyber attacks and, thereby, prevents false alarms and failures of protection devices.*

3 Liudmila A. Gurina, Ph.D. in engineering, Associate Professor, Senior Research Fellow, Laboratory for Control of Electric Power Systems at Melentiev Energy Systems Institute, SB RAS, Irkutsk, Russia. E-mail: gurina@isem.irk.ru
4 Nikita N. Tomin, Ph.D. in engineering, Head of Laboratory for Control of Electric Power Systems at Melentiev Energy Systems Institute, SB RAS, Irkutsk, Russia. E-mail: tomin.nv@gmail.com

References

1. Voropai, N. I. (2020). Prospects and Problems of Electric Power System Transformations. Elektrichestvo, 7, 12–21. DOI: 10.24160/0013-5380-2020-7-12-21.
2. Ilyushin, P. V. (2022). Integration of RES-based Power Plants into the Unified Energy System of Russia: Problematic Issues and Approaches to Solving Them. Bulletin of MPEI, 4, 98–107. DOI: 0.24160/1993-6982-2022-4-98-107.
3. Ilyushin, P., Filippov, S., Kulikov, A., Suslov, K., and Karamov, D. (2022). Intelligent Control of the Energy Storage System for Reliable Operation of Gas-Fired Reciprocating Engine Plants in Systems of Power Supply to Industrial Facilities. Energies, 15, 6333. DOI: 10.3390/en15176333.
4. Gurina, L. A. (2022). Pokazateli kiberustojchivosti komponentov informacionno-kommunikacionnoj infrastruktury pri upravlenii kiberfizicheskimi energeticheskimi sistemami // Methodological problems in reliability study of large energy systems, 73, 279–288.
5. Durakovsky, A. P., Markov, A. S. (2024). Current issues of cyber security in the energy sector. Secure Information Technologies, 94–98.
6. Shaukat, N. et al. (2023). Decentralized, Democratized, and Decarbonized Future Electric Power Distribution Grids: A Survey on the Paradigm Shift From the Conventional Power System to Micro Grid Structures. In IEEE Access, 11, 60957–60987. DOI: 10.1109/ACCESS.2023.3284031.
7. Lu, H., Biyawerwala, H. and Thakrawala, H. (2022). Polarized Distribution Protection Coordination Strategy Under the Impact from Various Distributed Energy Resources (DER) Generation Points. 2022 IEEE/PES Transmission and Distribution Conference and Exposition (T&D), New Orleans, LA, USA, 1–5. DOI: 10.1109/TD43745.2022.9816925.
8. S. K. T, Jadoun V. K., J. N. S and S. S. (2024). A Systematic Study on the Intelligent Cyber Security for Smart Microgrid. 2024 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER), Mangalore, India, 237–242. DOI: 10.1109/DISCOVER62353.2024.10750634.
9. Canaan, B., Colicchio, B., Abdeslam, D. O. (2020). Microgrid Cyber-Security: Review and Challenges toward Resilience. Applied Sciences, 10, 16, 5649. DOI: 10.3390/app10165649.
10. Ding, D., Han, Q. -L., Ge, X. and Wang, J. (2021). Secure State Estimation and Control of Cyber-Physical Systems: A Survey. In IEEE Transactions on Systems, Man, and Cybernetics: Systems, 51, 1, 176-190. DOI: 10.1109/TSMC.2020.3041121.
11. Ilyushin, P. V., Volnyi, V. S. (2022). Review of methods for addressing challenging issues in the operation of protection devices in microgrids with voltage up to 1 kV that integrate distributed energy resources. Relay protection and automation, 4(49), 6–21.
12. Zheng, Dehua, Zhang, Wei, Netsanet, Solomon, Wang, Ping, Bitew Girmaw, Teshager, Wei, Dan, Yue, Jun. (2021). Key technical challenges in protection and control of microgrid. Microgrid Protection and Control, 45–56. DOI: 10.1016/B978-0-12-821189-2.00007-3.
13. Patnaik, B., Mishra, M., Bansal, R. C., Jena, R. K. (2020) AC microgrid protection – A review: Current and future prospective. Applied Energy, 271, 115210. DOI: 10.1016/J.APENERGY.2020.115210.
14. Abd el-Ghany, H. A. (2020). Optimal PMU allocation for high-sensitivity wide-area backup protection scheme of transmission lines. Electric Power Systems Research, 187, 106485. DOI: 10.1016/J.EPSR.2020.106485.
15. Kulikov, A., Loskutov, A., Bezdushniy, D. (2022). Relay Protection and Automation Algorithms of Electrical Networks Based on Simulation and Machine Learning Methods. Energies, 15(18), 6525. DOI: 10.3390/en15186525.
16. Shobole, A. A., Wadi, M. (2021). Multiagent systems application for the smart grid protection. Renewable and Sustainable Energy Reviews, 149, 111352. DOI: 10.1016/J.RSER.2021.111352.
17. Verma, R., Gawre, S. K., Patidar, N. P. (2022). An Analytical Review on Measures of Microgrid Protection. 2022 IEEE International Conference on Power Electronics, Drives and Energy Systems (PEDES), Jaipur, India, 1–6. DOI: 10.1109/PEDES56012.2022.10080291.
18. Cui, S., Zeng, P., Wang, Z., Song, C. (2021). Research on Intelligent Protection Technology for Distribution Network with Distributed Generation. 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chongqing, China, 1549–1554. DOI: 10.1109/IAEAC50856.2021.9390692.
19. Fawzy, N., Habib, H. F., Mohammed, O., Brahma, S. (2020). Protection of Microgrids with Distributed Generation based on Multiagent System // 2020 IEEE International Conference on Environment and Electrical Engineering and 2020 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe), Madrid, Spain, 1–5. DOI: 10.1109/EEEIC/ICPSEurope49358.2020.9160827.
20. Chaitanya, B. K., Anamika, Yadav. (2023). Empirical Wavelet Transform-Based Differential Protection Scheme for Micro-Grid. Journal of The Institution of Engineers (India): Series B, 104, 1–10. DOI: 10.1007/s40031-023-00869-0.
21. Uddin, M. N., Arifin, M. S., Rezaei, N. (2022). A Novel Neuro-Fuzzy Based Direct Power Control of a DFIG based Wind Farm Incorporated with Distance Protection Scheme and LVRT Capability. 2022 IEEE Industry Applications Society Annual Meeting (IAS), Detroit, MI, USA, 01–08. DOI: 10.1109/IAS54023.2022.9939684.
22. Chaitanya, B. K., Anamika, Yadav, Mohammad, Pazoki. (2020). High Impedance Fault Detection Scheme for Active Distribution Network Using Empirical Wavelet Transform and Support Vector Machine. 2020 15th International Conference on Protection and Automation of Power Systems (IPAPS), 149–152. DOI:10.1109/IPAPS52181.2020.9375620.
23. Shaik, M., Shaik, A. G., Yadav, S. K. (2022). Hilbert–Huang transform and decision tree based islanding and fault recognition in renewable energy penetrated distribution system. Sustainable Energy, Grids and Networks, 30, 100606. DOI: 10.1016/j.segan.2022.
24. Raad Salih, Jawad, Hafedh, Abid. (2023). HVDC Fault Detection and Classification with Artificial Neural Network Based on ACO-DWT Method. Energies, 16, 1064. DOI: 10.3390/en16031064.
25. Chandran, L. R., A. Parvathy, V S, I. K, Nair, M. G. (2022). Adaptive Over Current Relay Protection in a PV Penetrated Radial Distribution System With Fuzzy GA Optimisation. 2022 IEEE 19th India Council International Conference (INDICON), Kochi, India, 1–7. DOI: 10.1109/INDICON56171.2022.10040021.
26. Nasir, M., Bansal, R., Elnady, A. (2022). A Review of Various Neural Network Algorithms for Operation of AC Microgrids. 2022 Advances in Science and Engineering Technology International Conferences (ASET), Dubai, United Arab Emirates, 1–7. DOI: 10.1109/ASET53988.2022.9734899.
27. Vincent Nsed, Ogar, Sajjad, Hussain, Kelum A.A., Gamage (2023). The use of artificial neural network for low latency of fault detection and localisation in transmission line. Heliyon, 9, e13376. DOI: 10.1016/j.heliyon.2023.
28. Wang, J. et al. (2022). Microgrid Fault Analysis Method Based on Inverter-Type DG with Different Control. 2022 4th International Conference on Smart Power & Internet Energy Systems (SPIES), Beijing, China, 1397–1402. DOI: 10.1109/SPIES55999.2022.10082157.

29. Xu, Y. et al. (2024). A Novel Distribution Network Fault Location Method Based on Improved Convolutional Neural Network. 2024 IEEE 7th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chongqing, China, 837–841. DOI: 10.1109/ITNEC60942.2024.10733085.
30. Conte, F., D'Agostino, F., Gabriele, B., Schiapparelli, G. -P. and Silvestro, F. (2023). Fault Detection and Localization in Active Distribution Networks Using Optimally Placed Phasor Measurements Units. In IEEE Transactions on Power Systems, 38, 1, 714–727. DOI: 10.1109/TPWRS.2022.3165685.
31. Gurina, L. A. (2024). Assessment of cyber security risk of microgrids energy community. Cybersecurity issues, 1(59), 101–107. DOI: 10.21681/2311-3456-2024-1-101-107.
32. Gurina, L. A., Tomin, N. V. (2024). Intelligent methods of ensuring cybersecurity multi-agent control system of microgrid. Cybersecurity issues, 6(64), 53–64. DOI: 10.21681/2311-3456-2024-6-53-64.
33. Kolosok, I. N., Gurina, L. A. (2021). Identification of Cyberattacks on SCADA and WAMS Systems in Electric Power Systems when Processing Measurements by State Estimation Methods. Elektrichestvo, 6, 25–32. DOI: 10.24160/0013-5380-2021-6-25-32.
34. Opitz, J. (2024). A Closer Look at Classification Evaluation Metrics and a Critical Reflection of Common Evaluation Practice. Transactions of the Association for Computational Linguistics, 12, 820–836. DOI: 10.1162/tacl_a_00675.
35. Hodson, T. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. Geoscientific Model Development, 15, 5481–5487. DOI: 10.5194/gmd-15-5481-2022.

МЕТОДОЛОГИЯ РЕАГИРОВАНИЯ НА ИНЦИДЕНТЫ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ В РАСПРЕДЕЛЕННЫХ АВТОМАТИЗИРОВАННЫХ ИНФОРМАЦИОННЫХ СИСТЕМАХ

Кузнецов А. В.[1]

DOI: 10.21681/2311-3456-2025-4-65-72

Цель исследования: *формирование единой методологии, позволяющей сократить время и силы, затрачиваемые группой реагирования на инциденты информационной безопасности на локализацию (сдерживание) инцидентов информационной безопасности, возникающих в распределенных автоматизированных информационных системах.*

Метод(ы) исследования: *анализ и синтез имеющихся общедоступных материалов и достижений, в т. ч. запатентованных, по тематикам реагирования на инциденты информационной безопасности и анализа данных, а также моделирование.*

Результат(ы) исследования: *1. Предложены концептуальная модель и единая методология реагирования на инциденты информационной безопасности, которые, в отличие от известных, учитывают особенности построения и обслуживания распределенных автоматизированных информационных систем, акцентируются на активном противодействии атакующему и базируются на принципе датацентричности, что позволяет сократить время и силы, затрачиваемые группой реагирования на инциденты информационной безопасности на локализацию инцидентов информационной безопасности, т. е. повысить эффективность деятельности указанных групп. 2. Сформулированы три метода в рамках предложенной методологии, в т. ч. метод организации единой подсистемы хранения данных мониторинга информационной безопасности и данных инцидентов информационной безопасности, метод предоставления мандата на выполнение действий по локализации инцидентов информационной безопасности, а также метод обработки данных мониторинга информационной безопасности и инцидентов информационной безопасности. Последний метод, в отличие от известных, нацелен на подтверждение инцидента информационной безопасности и принятие решения о необходимости (отсутствии необходимости) его локализации, а также предусматривает проведение обязательной проверки выполнения действий по локализации, что позволяет на каждом этапе реализации предложенной методологии вносить положительный вклад в сокращение времени и сил, затрачиваемых группой реагирования на инциденты информационной безопасности на локализацию инцидентов информационной безопасности, а также удостовериться в реализации действий по локализации инцидентов информационной безопасности.*

Научная новизна: *Концептуальная модель реагирования на инциденты информационной безопасности базируется на непрерывном процессе обработки данных (data pipeline) от четырех различных типов источников данных с учетом атрибутного состава данных, релевантных для конкретной распределенной автоматизированной информационной системы. Методология реагирования на инциденты информационной безопасности акцентируется на активном противодействии атакующему, базируется на принципе датацентричности и предусматривает обязательную проверку выполнения действий по локализации инцидентов информационной безопасности.*

Ключевые слова: *управляемое обнаружение и реагирование, группа реагирования на инциденты, локализация (сдерживание) инцидента, датацентричность, данные, источник данных, метод.*

Введение

Пространственно-распределенные (географически распределенные) автоматизированные информационные системы (РАИС) применяются практически во всех без исключения отраслях экономики Российской Федерации и других стран [1]. Их отличительными особенностями являются использование гетерогенных распределенных компьютерных сетей (от различных провайдеров связи, с различными характеристиками пропускной способности и оконечного оборудования) и географически распределенных центров обработки данных, в т. ч. распределенных баз данных [2, 3], а также централизация управления [4], в случае с нашей страной – на базе одного или нескольких из 16 городов-миллионников. Для компенсации нехватки персонала на местах (удаленных площадках РАИС) и обеспечения круглосуточного режима обслуживания и обеспечения ИБ, в первую очередь, мониторинга ИБ[2], организациями-владельцами (операторами) РАИС привлекаются сторонние провайдеры услуг – Managed

1 Кузнецов Александр Васильевич, кандидат технических наук, CISM, CISSP, ООО «РТК ИБ», Финансовый университет при Правительстве Российской Федерации, Москва. E-mail: 1283_my@mail.ru

2 41% компаний испытывают нехватку специалистов в области информационной безопасности: https://www.kaspersky.ru/about/press-releases/globalnoe-issledovanie-laboratorii-kasperskogo-41-kompanij-ispytyvayut-nehvatku-specialistov-v-oblasti-informacionnoj-bezopasnosti

Detection and Response (MDR) [5], например в нашей стране: Solar JSOC, Positive Technologies ESC, Kaspersky MDR, BI.ZONE TDR, Jet CSIRT, SOC «Перспективный мониторинг», IZ:SOC и/или другие.

При этом превалирующими вариантами реагирования на инциденты ИБ, в первую очередь от MDR-провайдеров[3], являются оповещения по электронной почте [6] и/или заведение заявок в Service Desk системах организаций-заказчиков услуг [7]. При этом только время оповещения (Mean Time to Detect и Mean Time to Report) занимает от десятков минут (37,85-53,99 минут за 2024 год[4]) до нескольких часов и даже дней, и только после этого начнется отсчет времени, затрачиваемого на непосредственное реагирование, включая локализацию (сдерживание) инцидента ИБ. В дополнение к предоставленному атакующему условному временному «окну возможностей», указанные варианты реагирования (фактически – оповещения) не оказывают активного противодействия атакующему, в т. ч. не обеспечивают локализацию обнаруженных инцидентов ИБ. Реализация рекомендаций из оповещений и/или заявок (при условии их наличия, целостности и согласованности), чаще всего, находится вне зоны ответственности групп мониторинга ИБ и координации реагирования на возникающие инциденты ИБ. Таким образом, при кадровом дефиците, особенно на местах (удаленных площадках РАИС), рассчитывать на полноту, качество и, самое главное, своевременность их реализации, к сожалению, не приходится.

Фактически получается, что несколько разрозненных команд (сил), с разным уровнем доступности в 11 часовых поясах России, с разными методами (способами) и средствами (инструментами), в т. ч. каналами и средствами коммуникации, не могут оперативно (в течение нескольких минут, а в идеале – секунд) и согласованно (в рамках единого набора данных) противодействовать современным компьютерным атакам (кибератакам), в т. ч. целенаправленным кибератакам (принимая во внимание, что требуется от 30 минут и, в большей половине случаев, 1-2 шага для проникновения в ЛВС организации[5], а в некоторых случаях – от 25 минут с момента проникновения до нанесения ущерба[6]).

Усугубляет ситуацию ежегодное увеличение количества фиксируемых инцидентов ИБ[7], в т. ч. в кредитно-финансовом секторе[8].

Таким образом, на фоне роста активностей атакующих, повышение эффективности деятельности групп реагирования на инциденты ИБ (ГРИИБ) за счет оптимизации и автоматизации мероприятий по локализации инцидентов ИБ является актуальным направлением исследований, особенно для организаций-владельцев (операторов) РАИС, а также центров Государственной системы обнаружения, предупреждения и ликвидации последствий компьютерных атак на информационные ресурсы Российской Федерации (ГосСОПКА) и координационных центров групп реагирования на компьютерные инциденты (Computer Emergency Response Team Coordination Center (CERT CC)).

Ограничения исследования

В рамках настоящего исследования автором были приняты следующие ограничения и допущения:

- не рассматриваются РАИС, построенные с использованием технологий распределенных реестров (Distributed Ledger Technology (DLT));
- в части мероприятий по реагированию на инциденты ИБ рассматривались только первоочередные действия, оказывающие активное противодействие атакующему, т. е. действия по локализации (сдерживанию) инцидентов ИБ, развивая и дополняя предыдущие исследования автора [4, 8, 9];
- не проводилось деление между понятиями «инцидент ИБ»[9], «компьютерный инцидент»[10], «инцидент защиты информации»[11] и «киберинцидент»[12], указанные понятия рассматривались автором как синонимы (результаты настоящего исследования применимы к любому указанному терминологическому аппарату).

Концептуальная датацентричная модель реагирования на инциденты ИБ

На сегодняшний день есть несколько международных и национальных стандартов по управлению, в т. ч. реагированию на инциденты ИБ[13], обзор, которых проводился автором ранее [8]. Все существующие стандарты рассматривают предмет исследования

3 Эволюция реагирования: от ручных процессов к интеллектуальным решениям: https://forumsoc.ru/program/

4 Analyst report. Managed Detection and Response: https://content.kaspersky-labs.com/fm/site-editor/b7/b7766f565d5f14783d6848364b8fde66/source/mdr-report.pdf

5 Итоги внешних пентестов – 2020: https://www.ptsecurity.com/ru-ru/research/analytics/external-pentests-2020/

6 Самый незаметный вредонос и слова-загадки из кода. В Kaspersky рассказали о рекордах хакеров: https://www.gazeta.ru/tech/2023/08/10/17407172.shtml

7 Хроники DFIR: как атаковали APT-группировки с января по октябрь 2024 года: https://rt-solar.ru/analytics/reports/4857/

8 Кибератаки на финансовый сектор в 2024 году: https://rt-solar.ru/analytics/reports/5206/

9 ГОСТ Р ИСО/МЭК ТО 18044-2007 «Информационная технология. Методы и средства обеспечения безопасности. Менеджмент инцидентов информационной безопасности».

10 ГОСТ Р 59709-2022 «Защита информации. Управление компьютерными инцидентами. Термины и определения».

11 ГОСТ Р 57580.1-2017 «Безопасность финансовых (банковских) операций. Защита информации финансовых организаций. Базовый состав организационных и технических мер».

12 ГОСТ Р 57580.3-2022 «Безопасность финансовых (банковских) операций. Управление риском реализации информационных угроз и обеспечение операционной надежности. Общие положения».

13 NIST SP 800-61 Rev. 3 (Initial Public Draft) «Incident Response Recommendations and Considerations for Cybersecurity Risk Management: A CSF 2.0 Community Profile», ISO/IEC 27035-1:2023 «Information technology — Information security incident management — Part 1: Principles and process», ГОСТ Р 59710-2022 «Защита информации. Управление компьютерными инцидентами. Общие положения», Рекомендации в области стандартизации Банка России РС БР ИББС-2.5-2014 «Обеспечение информационной безопасности организаций банковской системы Российской Федерации. Менеджмент инцидентов информационной безопасности».

через процессный подход, ориентированный на цикл непрерывного совершенствования, и не рассматривают основополагающим принцип датацентричности [10, 11], без которого невозможно реализовать data-driven подход к реагированию на инциденты ИБ.

Здесь же стоит отметить, что существует несколько открытых проектов: фреймворк для техник реагирования на инциденты ИБ RE&CT[14] и модель (база) знаний ERM&CK[15], которые ориентированы на практическую, точечную автоматизацию, но не предлагают взаимосвязанного набора методов решения задач, возникающих в рамках обработки разрозненных данных.

В завершении обзора существующих вариантов решения, стоит отметить ряд запатентованных способов автоматизации реагирования на инциденты ИБ[16], которые предлагают только отдельные технические решения (системы), не формируя единую методологию решения возникшей проблемы.

Методология рассматривается автором как совокупность методов, имеющих общие принципы, условия реализации и предназначенные для достижения цели настоящего исследования[17], при этом основным принципом для данной методологии будет выступать датацентричность.

Принимая во внимание недостатки существующих подходов и решений автором предлагается следующая концептуальная модель реализации мероприятий по реагированию на инциденты ИБ, представленная в нотации Event-Driven Process Chain (рис. 1).

В основе данной модели лежит предложенный автором непрерывный процесс (конвейер) обработки данных (data pipeline) [9]. Предлагаемая модель позволяет перейти к автономности ГРИИБ (не привлечению приходящих системных администраторов, местных подрядчиков и т. п.) и работе на основе актуальных и релевантных для конкретной РАИС данных (сгенерированных в рамках функционирования РАИС и ее подсистемы обеспечения ИБ), а не на экспертизе отдельных специалистов (в т. ч. специалистов MDR-провайдеров) или использовании данных, относящихся к другим организациям, отраслям экономики и/или регионам страны.

На модели отмечены основные временные характеристики, используемые при описании процессов обнаружения и реагирования на инциденты ИБ

14 RE&CT: https://atc-project.github.io/atc-react/index_RU/
15 ERMACK: https://github.com/Security-Experts-Community/ERMACK
16 US20090296898 – Automated incident response method and system: https://patentscope.wipo.int/search/ru/detail.jsf?docId=US42906733. US10051010B2 – Method and system for automated incident response: https://patents.google.com/patent/US10051010B2

17 Манушин Д. В. Уточнение понятия «методология» // Финансы и кредит. 2015. № 41 (665). С. 50–66

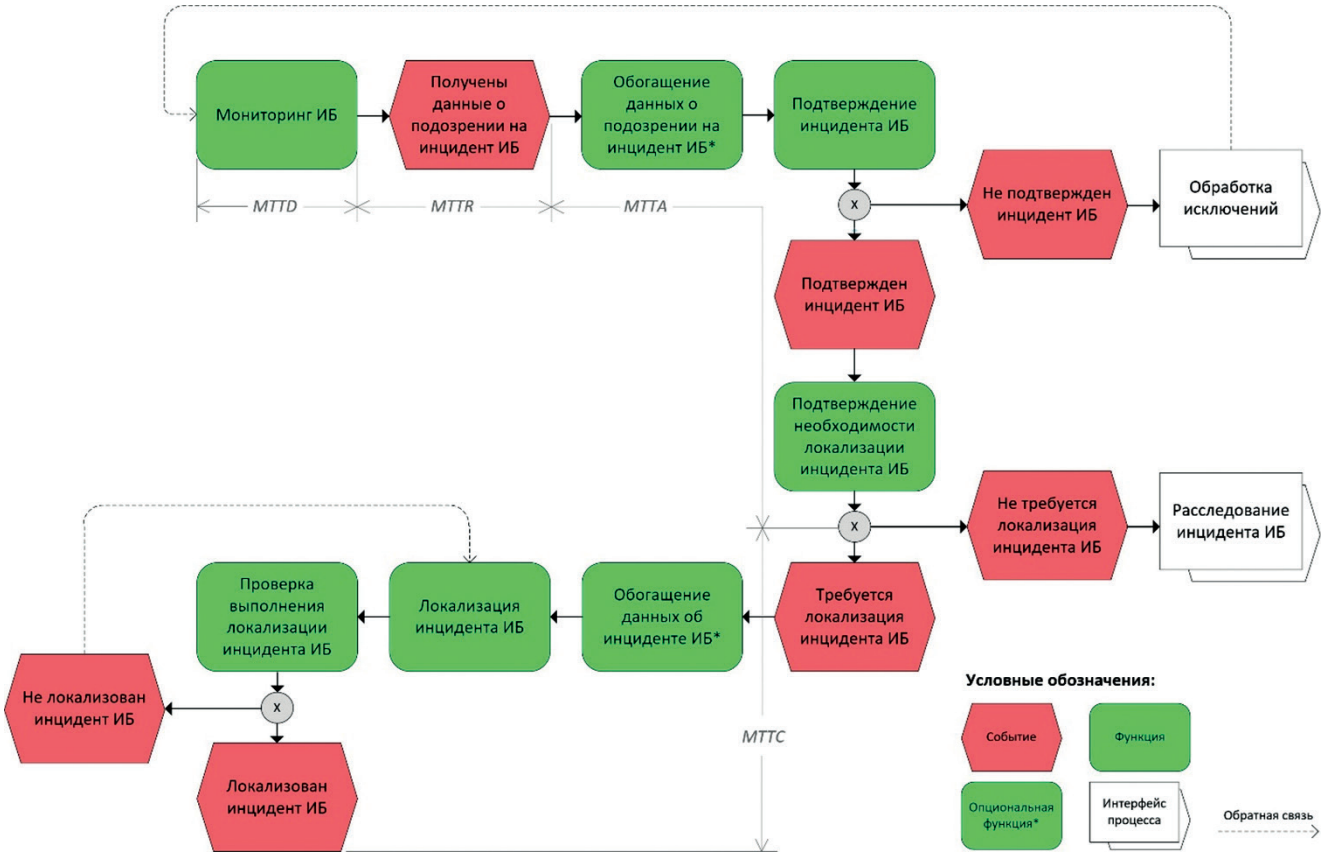


Рис. 1. Концептуальная датацентричная модель реагирования на инциденты ИБ (локализации инцидентов ИБ)

в Security Operations Center (SOC) [12], а также ГРИИБ и CERT CC:

- Mean Time to Detect/Discover (MTTD): среднее время, необходимое для обнаружения (выявления) инцидента ИБ, отсчитываемое с момента выполнения несанкционированных действий атакующим в отношен РАИС и до момента их детектирования (распознавания) системами обнаружения или персоналом организации; в ряде случаев дополнительно учитывается время, затрачиваемое на корреляцию событий (данных) системами обнаружения, в т. ч. SIEM-системами;
- Mean Time to Report (MTTR): среднее время, необходимое для формирования и реализации оповещения с использованием соответствующих каналов и средств коммуникации, отсчитываемое с момента обнаружения и до момента получения оповещения (регистрации подозрения на инцидента ИБ в формате карточки);
- Mean Time to Acknowledge (MTTA): среднее время, необходимое для подтверждения санкционированности или несанкционированности действий, отсчитываемое с момента получения оповещения и до момента вынесения вердикта по подозрению на инцидент ИБ («Подтвержденный», «Ложный» или «Санкционированное действие»);
- Mean Time to Respond/Contain (MTTC): среднее время, необходимое для локализации (сдерживания) инцидента ИБ, отсчитываемое с момента вынесения вердикта «Подтвержденный» и до определения и ограничения функционирования информационных ресурсов (компонентов РАИС), на которых обнаружены признаки зарегистрированного инцидента ИБ, с целью предотвращения его дальнейшего распространения (указанная временная характеристика представляет наибольший интерес в рамках настоящего исследования).

Безусловно, без качественного и своевременного обнаружения инцидентов ИБ невозможно проводить какие-либо мероприятия по реагированию (неважно в ручном или автоматическом режиме [8]). Но сразу стоить отметить, что тематике обнаружения (мониторинга ИБ), в т. ч. на основе больших данных, с использованием машинного обучения и систем искусственного интеллекта (ИИ) посвящено значительное число работ. К ведущим коллективам российских ученых, работающим в данном направлении, можно отнести коллективы под руководством Милославской Н. Г. [13], а также Котенко И. В. и Саенко И. Б. [14, 15]. Как следствие, указанная предметная область в полном объеме не входит в предмет настоящего исследования, но требует фиксации, в т. ч. синтеза необходимых входных условий (данных) для формирования единой датацентричной методологии.

Таким образом, в части обнаружения инцидентов ИБ на основе данных необходимо определить все типы источников данных, которые могут выступать «триггерами» для регистрации инцидентов ИБ, и атрибутный состав данных (неотъемлемая часть логической модели данных). На практике выделяют два основных источника данных: источники событий безопасности [16] и персонал, но этого недостаточно. Автором предлагается следующий набор типов источников данных мониторинга ИБ с указанием атрибутного состава данных:

- обращения пользователей: дата и время обращения, источник несанкционированной активности (источник), цель несанкционированной активности (цель), тип несанкционированной активности;
- данные о событиях безопасности: дата и время обнаружения, источник, цель, тип события безопасности;
- данные киберразведки: хронологические метки даты и времени, источник и/или цель, значение индикатора компрометации (описание техники и/или описание уязвимости, и/или хэш-значение);
- данные об уязвимостях: цель, описание уязвимости, класс уязвимости, степень опасности уязвимости.

Предложенный вариант, в отличие от известных, расширяет типы источников данных до четырех, что позволяет снизить статистическую вероятность (относительную частоту) пропуска инцидентов ИБ (ошибок второго рода) или несвоевременности их обнаружения.

В части категоризации вариантов реализации действий по локализации инцидентов ИБ автором предлагаются следующие варианты:

- полностью автоматическая локализация: выполнение действий без привлечения специалистов ГРИИБ;
- полуавтоматическая локализация: выполнение действий с привлечением специалистов ГРИИБ для явного подтверждения (в единой консоли управления карточками инцидентов ИБ) их дальнейшего автоматического выполнения;
- ручная локализация: выполнение действий непосредственно специалистами ГРИИБ с использованием информационных ресурсов (компонентов РАИС) и/или средств реагирования (например: непосредственное подключение к консоли управления средства реагирования);
- физическая локализация: выполнение действий специалистами ГРИИБ или специалистами на удаленных площадках без использования информационных ресурсов (компонентов РАИС) и средств реагирования (например: физическое выключение электропитания).

Для полностью автоматической и полуавтоматической локализации требуется наличие мандата на выполнение действия – цифровой записи, содержащей идентификационную информацию, характеризующую условия локализации инцидентов ИБ, с учетом заданных критериев [17].

При этом стоит отметить, что не все инциденты ИБ требуют незамедлительной локализации, часть требует проведения мероприятий по расследованию инцидентов ИБ (forensic) [18], но данные мероприятия находятся за рамками настоящего исследования, т. к. не оказывают активного противодействия атакующему.

Методы, предназначенные для достижения цели настоящего исследования

Автором предлагается следующий набор взаимоувязанных методов, объединенных принципом датацентричности, направленных на достижение цели настоящего исследования, формирующих единую методологию реагирования на инциденты ИБ в РАИС: M_1, M_2 и M_3.

M_1: метод организации единой подсистемы хранения данных мониторинга ИБ и данных инцидентов ИБ, который, в отличие от известных, учитывают доступную ширину полосы пропускания используемого канала связи между компонентами сбора данных (коллекторами) и компонентами «горячего» и «холодного» хранения данных (на основе работы автора [19]).

Обозначив $L_1 = \{l_{1i}\}$ – множеством всех площадок РАИС, $S \in [0;+\infty)$ – размером одной порции данных, $E \in [0;+\infty)$ – усредненным за сутки потоком данных, $B_i \in [0;+\infty)$ – доступной шириной полосы пропускания используемого канала связи до i-й площадки, $L_2 = \{l_{2j}\}$ – множеством площадок РАИС с компонентами единой подсистемы хранения данных, а также введя условие, определяющее возможность централизации подсистемы хранения данных (1).

$$(S + 58) \cdot E \cdot 8 \cdot 9{,}54 \cdot 10^{-7} \le B_i. \quad (1)$$

Под методом будем понимать отображение $M_1 : L_1 \longrightarrow L_2$.

Без единой подсистемы хранения данных невозможно реализовать data-driven вариант реагирования на инциденты ИБ в РАИС.

M_2: метод предоставления мандата на выполнение действий по локализации инцидентов ИБ, который, в отличие от известных, учитывает вариативность и взаимосвязи различных критериев предоставления мандата, релевантных для конкретной РАИС (на основе работы автора [17]).

Обозначив $Cr_1 = \{c_q\}$ – множеством возможных мандатов, $A = \{a_{p,k}\}$ – матрицей влияния критерия на предоставление мандата, $X = \{x_k\}$ – вектором целочисленных переменных, отражающих соблюдение k-го критерия предоставления мандата, $Cr_2 = \{c_w\}$ – множеством предоставленных мандатов; под методом будем понимать отображение $M_2 : Cr_1 \longrightarrow Cr_2$.

Оператор предоставления мандата для локализации инцидентов ИБ будет иметь вид (2).

$$F_{cr}(Cr_1,A,X) = \begin{cases} 1, & \text{мандат для локализации предоставляется;} \\ 0, & \text{мандат для локализации не предоставляется.} \end{cases}. \quad (2)$$

Без предоставления мандатов невозможно сократить вовлечение сил ГРИИБ и время, затрачиваемое группой на локализацию инцидентов ИБ.

M_3: метод обработки данных мониторинга ИБ и инцидентов ИБ, который, в отличие от известных, базируется на концептуальной датацентричной модели (рис. 1) и нацелен на подтверждение инцидента ИБ и принятие решения о необходимости (отсутствии необходимости) его локализации, а также предусматривает проведение обязательной проверки выполнения действия по локализации.

Обозначив $V_1 = \{v_{1t}\}$ – множеством обнаруженных инцидентов ИБ, $C = \{c_y\}$ – множеством мер по локализации, $V_2 = \{v_{1o}\}$ – множеством локализованных инцидентов ИБ; под методом будем понимать отображение $M_3 : V_1 \longrightarrow V_2$.

Оператор локализации инцидента ИБ v_{1t} будет иметь вид (3).

$$F_r(V_1,C,Cr_2) = \begin{cases} 1, & \text{если инцидент ИБ } v_{1t} \text{ локализован;} \\ 0, & \text{если инцидент ИБ } v_{1t} \text{ не локализован.} \end{cases}. \quad (3)$$

Выводы

По результатам проведенного исследования автором:

1. Предложены концептуальная модель и единая методология реагирования на инциденты ИБ, которые, в отличие от известных, учитывают особенности построения и обслуживания РАИС, акцентируются на активном противодействии атакующему и базируются на принципе датацентричности, что позволяет сократить время и силы, затрачиваемые ГРИИБ на локализацию инцидентов ИБ, т. е. повысить эффективность деятельности указанных групп.

2. Сформулированы три метода M_1, M_2 и M_3 в рамках предложенной методологии, в т. ч. M_3: метод обработки данных мониторинга ИБ и инцидентов ИБ, который, в отличие от известных, нацелен на подтверждение инцидента ИБ и принятие решения о необходимости (отсутствии необходимости) его локализации, а также предусматривает проведение обязательной проверки выполнения действия по локализации, что позволяет на каждом этапе реализации предложенной методологии вносить положительный вклад в сокращение времени

и сил, затрачиваемых ГРИИБ на локализацию инцидентов ИБ, а также удостовериться в реализации действий по локализации.

Применение результатов настоящего исследования дает положительный эффект в области технических наук: научная специальность «Методы и системы защиты информации, ИБ», а также могут быть применимы в смежной научной специальности: «ИИ и машинное обучение», при рассмотрении проблем и задач применения ИИ в обнаружении и реагировании на инциденты ИБ[18].

Практическое применение предложенной методологии позволит внести значительный положительный вклад в развитие ГРИИБ, в т. ч. центров ГосСОПКА и CERT CC.

Предложенная методология в 2023–2025 гг. проходит апробацию в рамках выполнения работ в интересах организаций-владельцев РАИС федерального масштаба, а также ведущих организаций-лицензиатов ФСТЭК России, оказывающих услуги по мониторингу ИБ средств и систем информатизации[19] на территории всей страны.

Развитием данной data-driven методологии является переход к AI-driven методологии, которую отдельные лидеры ИТ/ИБ-отрасли уже начали демонстрировать (например, корпорация Microsoft в 2023 году запустила ИИ-помощника для групп мониторинга ИБ и реагирования – Security Copilot[20]). Стоит отметить, что уже есть успешные отдельные примеры автоматизации процедур условного реагирования на инциденты ИБ, в части установки обновлений для программного обеспечения [20, 21], в т. ч. появилось понятие – Vulnerability Management Detection and Response (VMDR)[21].

18 Are AI-powered SOCs the future of cybersecurity?: https://www.tahawultech.com/insight/are-ai-powered-socs-the-future-of-cybersecurity/. The rise of autonomous SOCs: embracing AI-powered security operations for the ever-evolving threat landscape: https://global.ptsecurity.com/analytics/autonomous-socs-ai-powered-security-operations-for-evolving-threat-landscape

19 Положение о лицензировании деятельности по технической защите конфиденциальной информации, утв. постановлением Правительства Российской Федерации от 03.02.2012 г. N 79.

20 What is Microsoft Security Copilot?: https://learn.microsoft.com/en-us/security-copilot/microsoft-security-copilot

21 VMDR: Inside vulnerability management, detection and response: https://www.techtarget.com/searchsecurity/feature/VMDR-Inside-vulnerability-management-detection-and-response

Литература

1. Braione P., Briola D., De Angelis G., Gallo F., Poggi F., Quattrocchi G. About the special issue on: «Distributed Complex Systems: Governance, Engineering, and Maintenance» // Journal of Software: Evolution and Process. 2022. Т. 34. № 10. DOI: https://doi.org/10.1002/smr.2459.
2. Созонтов А. В. Распределенные информационные системы: особенности применения и построения // Актуальные исследования. 2023. № 37-1 (167). С. 69–74.
3. Панделов Т. С., Янаева М. В. Территориально-распределенные информационные системы // Молодой исследователь Дона. 2022. № 6 (39). С. 59–62.
4. Кузнецов А. В. Особенности реагирования на инциденты в пространственно-распределенных автоматизированных информационных системах // Инженерный вестник Дона: сетевое издание. 2025. № 5-25. URL: http://www.ivdon.ru/ru/magazine/archive/n5y2025/10046 (дата обращения 12.05.2025).
5. Зайчиков Н. План действий ИТ-службы в эпоху замещения // Системный администратор. 2022. № 5 (234). С. 20–23.
6. Хохлов А. Ю., Асанов С. А. Использование технологической платформы «1С: Предприятие» для автоматизации учёта и уведомления ГосСОПКА об инцидентах информационной безопасности объектов критической информационной инфраструктуры // В сборнике: Россия молодая. Сборник материалов XVI всероссийской, научно-практической конференции молодых ученых с международным участием. Кемерово, 2024. С. 31691.1–31691.8.
7. Репецкий С. О., Репецкая Н. В. Обработка заявок в IT Service Desk // StudNet: сетевой журнал. 2021. Т. 4. № 5. URL: https://stud.net.ru/obrabotka-zayavok-v-it-service-desk/ (дата обращения 12.05.2025).
8. Кузнецов А. В. Эволюция реагирования на инциденты информационной безопасности // Защита информации. Инсайд. 2024. № 5(119). С. 14–20.
9. Кузнецов А. В. Конвейер данных для автоматической локализации компьютерных инцидентов // Вторая Всероссийская научно-техническая конференция «Кибернетика и информационная безопасность «КИБ-2024». Сборник научных трудов. 22-23 октября 2024 г., Москва. М.: НИЯУ МИФИ. 2024. С. 120-121.
10. Зеневич А. М., Пунчик З. В. Датацентричность как тренд развития корпоративных информационных систем // В сборнике: Эколого-экономические и технологические аспекты устойчивого развития Республики Беларусь и Российской Федерации. сборник статей III Международной научно-технической конференции: в 3 т. Минск, 2021. С. 179–182.
11. Гладилина И. П., Сергеева С. А., Синицына Е. В. Цифровая этика и этика данных как основа рациональной деятельности экономических субъектов в условиях цифровой трансформации // Экономические системы. 2024. Т. 17. № 4. С. 28–38. DOI: 10.29030/2309-2076-2024-17-4-28-38.
12. Расширенная модель зрелости SOC компании Cybereason / И. С. Листратов, Н. Г. Милославская, И. С. Сирбай, Б. А. Рейносо // Безопасность информационных технологий. 2025. Т. 32. № 1. С. 68–84. DOI: 10.26583/bit.2025.1.04.
13. Месенгисер Я. Я., Малахов М. А., Милославская Н. Г. Центры управления сетевой безопасностью как силы ГосСОПКА // Безопасность информационных технологий. 2022. Т. 29. № 1. С. 94–107. DOI: http://dx.doi.org/10.26583/bit.2022.1.09.
14. Методика обнаружения аномалий и кибератак на основе интеграции методов фрактального анализа и машинного обучения / И. В. Котенко, И. Б. Саенко, О. С. Лаута, А. М. Крибель // Информатика и автоматизация. 2022. Т. 21. № 6. С. 1328–1358. DOI: https://doi.org/10.15622/ia.21.6.9.

15. Саенко И. Б., Котенко И. В., Аль-Барри М. Х. Применение искусственных нейронных сетей для выявления аномального поведения пользователей центров обработки данных // Вопросы кибербезопасности. 2022. № 2(48). С. 87–97. DOI: 10.21681/2311-3456-2022-2-87-97.
16. Андрушкевич Д. В., Андрушкевич С. С., Крюков Р. О. Метод реагирования на целевые атаки, основанный на отображении событий информационной безопасности с применением индикационных сигнатур // Проблемы информационной безопасности. Компьютерные системы. 2023. № 4(57). С. 48–60.
17. Кузнецов А. В. Анализ критериев предоставления мандата на локализацию инцидента информационной безопасности // Инженерный вестник Дона: сетевое издание. 2025. № 3-25. URL: http://www.ivdon.ru/ru/magazine/archive/n3y2025/9919 (дата обращения 12.05.2025).
18. Смирнов С. И. Методика расследования киберинцидента, основанная на интеллектуальном анализе событий безопасности домена // Защита информации. Инсайд. 2022. № 4(106). С. 60–69.
19. Кузнецов А. В. Организация раздельного хранения данных о событиях безопасности // Вопросы кибербезопасности. 2024. № 2(60). С. 22–28. DOI: 10.21681/2311-3456-2024-2-22-28.
20. Yu Nong, Haoran Yang. Automated Software Vulnerability Patching using Large Language Models. August, 2024. URL: https://arxiv.org/html/2408.13597v1 (дата обращения 12.05.2025). DOI:10.48550/arXiv.2408.13597.
21. Minjae Seo, Wonwoo Choi, Myoungsung You, Seungwon Shin. AutoPatch: Multi-Agent Framework for Patching Real-World CVE Vulnerabilities. May 2025. URL: https://arxiv.org/abs/2505.04195 (дата обращения 12.05.2025). DOI: 10.48550/arXiv.2505.04195.

THE METHODOLOGY OF INFORMATION SECURITY INCIDENTS RESPONSE WITHIN DISTRIBUTED AUTOMATED INFORMATION SYSTEMS

Kuznetsov A. V.[22]

Keywords: *managed detection and response, incident response team, incident localization (containment), data-centricity, data, data source, method.*

Purpose of the study: *to develop the unified methodology to reduce the time and effort spent by an information security incident response team to localize (contain) information security incidents occurring in distributed automated information systems.*

Methods of research: *analysis and synthesis of existing publicly available materials and advances, including patented ones, related to information security incident response and data analysis, as well as modeling.*

Result(s): *1. The conceptual model and unified methodology of information security incident response are proposed, which, unlike the known ones, take into account the specifics of construction and maintenance of distributed automated information systems, focus on active counteraction to the attacker and are based on the principle of data-centricity, which reduces the time and effort spent by an information security incident response team to localize information security incidents, i.e., increase the efficiency of the activity of an information security incident response team. 2. Three methods within the proposed methodology are formulated, including the method of organizing the unified subsystem for storing information security monitoring data and information security incident data, the method of providing a mandate to perform an action to localize an information security incident, and the method of processing information security monitoring data and information security incident data. The latter method, unlike the known ones, is aimed at confirming the information security incident and making a decision on the need (lack of need) for its localization, and also provides for mandatory verification of the localization action, which allows at each stage of implementation of the proposed methodology to make a positive contribution to reducing the time and effort spent by an information security incident response team to localize information security incidents, as well as ascertain that actions to localize information security incidents were implemented.*

Scientific novelty: *The conceptual model of response to information security incidents is based on a continuous process of data processing (data pipeline) from four different types of data sources, taking into account the attribute composition of data relevant to a particular distributed automated information system. The information security incident response methodology focuses on active counteraction to the attacker, is based on the principle of data-centricity, and provides for mandatory verification that actions to localize information security incidents were implemented.*

References

1. Braione P., Briola D., De Angelis G., Gallo F., Poggi F., Quattrocchi G. About the special issue on: «Distributed Complex Systems: Governance, Engineering, and Maintenance» // Journal of Software: Evolution and Process. 2022. Т. 34. № 10. DOI: https://doi.org/10.1002/smr.2459.

22 Aleksandr V. Kuznetsov, Ph.D. (in Tech.), CISM, CISSP, RTK IS LLC, Financial University under the Government of the Russian Federation, Moscow. E-mail: 1283_my@mail.ru.

2. Sozontov, A. V. Raspredelennye informacionnye sistemy: osobennosti primeneniya i postroeniya // Aktual'nye issledovaniya. 2023. № 37-1 (167). pp. 69–74.
3. Pandelov, T. S., Yanaeva, M. V. Geographically distributed information systems // Young Researcher of Don. 2022. № 6(39). pp. 59–62.
4. Kuznetsov, A. V. Osobennosti reagirovaniya na incidenty v prostranstvenno-raspredelennyx avtomatizirovannyx informacionnyx sistemax // Inzhenernyj vestnik Dona. 2025. № 5-25. URL: ivdon.ru/ru/magazine/archive/n3y2025/9919 (accessed 12.05.2025).
5. Zajchikov, N. Plan dejstvij IT-sluzhby v epoxu zameshheniya // Sistemnyj administrator. 2022. № 5 (234). pp. 20–23.
6. Hohlov, A. Yu., Asanov. S. A. Ispol'zovanie texnologicheskoj platformy «1S: Predpriyatie» dlya avtomatizacii uchyota i uvedomleniya GosSOPKA ob incidentax informacionnoj bezopasnosti ob''ektov kriticheskoj informacionnoj infrastruktury // V sbornike: Rossiya molodaya. Sbornik materialov XVI vserossijskoj, nauchno-prakticheskoj konferencii molodyx uchenyx s mezhdunarodnym uchastiem. Kemerovo, 2024. pp. 31691.1–31691.8.
7. Repeckij, S. O., Repeckaya, N. V. Obrabotka zayavok v IT Service Desk // StudNet. 2021. T. 4. № 5. URL: https://stud.net.ru/obrabotka-zayavok-v-it-service-desk/ (accessed 12.05.2025).
8. Kuznetsov, A. V. The Evolution of Information Security Incident Response // Zaŝita informacii. Inside. 2024. № 5 (119). pp. 14–20.
9. Kuznetsov, A. V. Konvejer dannyx dlya avtomaticheskoj lokalizacii komp'yuternyx incidentov // Vtoraya Vserossijskaya nauchno-texnicheskaya konferenciya «Kibernetika i informacionnaya bezopasnost' «KIB-2024». Sbornik nauchnyx trudov. 22-23 oktyabrya 2024 g., Moskva. M.: NRNU MEPhI. 2024. Pp. 120-121.
10. Zenevich, A. M., Punchik, Z. V. Datacentrichnost' kak trend razvitiya korporativnyx informacionnyx sistem // V sbornike: Ekologo-ekonomicheskie i texnologicheskie aspekty ustojchivogo razvitiya Respubliki Belarus' i Rossijskoj Federacii. sbornik statej III Mezhdunarodnoj nauchno-texnicheskoj konferencii: v 3 t.. Minsk, 2021. pp. 179–182.
11. Gladilina, I. P., Sergeeva, S. A., Sinitsyna, E. V. Digital ethics and data ethics as the basis for the rational activities of economic entities in the context of digital transformation // Economic Systems. 2024. T. 17. № 4. pp. 28–38. DOI: 10.29030/2309-2076-2024-17-4-28-38.
12. Extended Cybereason's SOC maturity model / Listratov, I. S., Miloslavskaya, N. G., Sirbay, I. S., Reinoso, B. A. // IT Security. 2025. T. 32. № 1. pp. 68–84. DOI: 10.26583/bit.2025.1.04.
13. Mesengiser, Y. Y., Malakhov, M. A., Miloslavskaya, N. G. Network Security Centers as the GosSOPKA Forses // IT Security. 2022. T. 29. № 1. pp. 94–107. DOI: http://dx.doi.org/10.26583/bit.2022.1.09.
14. Anomaly and cyber-attack detection technique based on the integration of fractal analysis and machine learning methods / Kotenko, I.V., Saenko, I. B., Lauta, O. S., Kriebel, A. M. // Informatics and Automation. 2022, T. 21, № 6. pp. 1328–1358. DOI: https://doi.org/10.15622/ia.21.6.9.
15. Saenko, I. B., Kotenko, I. V., All-Barri M. H. Application of artificial neural networks to reveal abnormal behavior of data center users // Voprosy kiberbezopasnosti. 2022. № 2(48). Pp. 87–97. DOI: 10.21681/2311-3456-2022-2-87-97.
16. Andrushkevich, D. V., Andrushkevich, S. S., Kryukov, R. O. Metod reagirovaniya na celevye ataki, osnovannyj na otobrazhenii sobytij informacionnoj bezopasnosti s primeneniem indikacionnyx signatur // Information Security Problems. Computer Systems. 2023. № 4(57). pp. 48–60.
17. Kuznetsov, A. V. Analiz kriteriev predostavleniya mandata na lokalizaciyu incidenta informacionnoj bezopasnosti // Inzhenernyj vestnik Dona. 2025. № 3-25. URL: http://www.ivdon.ru/ru/magazine/archive/n3y2025/9919 (accessed 12.05.2025).
18. Smirno, S. I. Cyber incident investigation methodology based on intelligent analysis of domain security events // Zaŝita informacii. Inside. 2022. № 4(106). pp. 60–69.
19. Kuznetsov, A. V. The organization of separate security event data storage // Voprosy kiberbezopasnosti. 2024. № 2 (60). Pp. 22–28. DOI: 10.21681/2311-3456-2024-2-22-28.
20. Yu Nong, Haoran Yang. Automated Software Vulnerability Patching using Large Language Models. August 2024. URL: https://arxiv.org/html/2408.13597v1 (accessed 12.05.2025). DOI:10.48550/arXiv.2408.13597.
21. Minjae Seo, Wonwoo Choi, Myoungsung You, Seungwon Shin. AutoPatch: Multi-Agent Framework for Patching Real-World CVE Vulnerabilities. May 2025. URL: https://arxiv.org/abs/2505.04195 (accessed 12.05.2025). DOI: 10.48550/arXiv.2505.04195.

ПОДХОД К КЛАССИФИКАЦИИ TELEGRAM-КАНАЛОВ

Попов В. А.[1], Чеповский А. А.[2]

DOI: 10.21681/2311-3456-2025-4-73-83

Цель исследования: *разработка метода определения цифрового профиля Telegram-каналов в сетях информационного взаимодействия и процедуры классификации каналов на основе выделенного цифрового профиля.*

Метод исследования: *метод исследования включает следующие этапы: построение графа взаимодействующих объектов на основании импортированных из сети Telegram данных, определение цифровых профилей вершин на основании их атрибутивных данных и свойств графа, кластеризация вершин на основании выделенных профилей, классификация центров полученных кластеров и исходных Telegram-каналов, вычислительные эксперименты и анализ результатов.*

Полученный результат: *в данной статье вводится определение цифрового профиля Telegram-канала, представленного как одна из вершин графа взаимодействующих объектов. Цифровой профиль задан через нормализованный 5-мерный вектор признаков, полученных на основе атрибутивных данных вершины и свойств графа. Выбранные характеристики отражают свойства Telegram-каналов в построенном графе и метаданные, полученные при импорте из сети. Далее авторы описывают алгоритм кластеризации полученных профилей с использованием настраиваемых параметров. Центры выделенных кластеров классифицируются по 4 предложенным авторами типам, характеризующим роли вершин в графе взаимодействующих объектов. За счет этого производится классификация всех вершин графа – исходных Telegram-каналов анализируемой сети. Предложенный подход дает ценную информацию о ролях Telegram-каналов в сетях информационного взаимодействия.*

Научная новизна: *разработан новый подход к анализу Telegram-каналов: предложен метод создания цифрового профиля Telegram-канала в виде 5-мерного вектора признаков, что позволяет провести анализ и классификацию каналов. Также в рамках подхода предложена основанная на вычислительных методах процедура классификации таких цифровых профилей, которая позволяет выявить основные типы Telegram-каналов скачиваемой подсети по заданной классификации.*

Ключевые слова: *цифровые профили, анализ социальных сетей, безмасштабные сети, модель информационного воздействия, выделение сообществ, задачи классификации.*

Введение

В настоящее время в России мессенджер Telegram является одним из самых популярных приложений для общения и получения информации. Telegram предоставляет широкий набор функций для организации публичных каналов, которые представляют собой информационные ленты. Многие СМИ, информационные сообщества и блогеры имеют собственные Telegram-каналы и регулярно публикуют в них контент, а пользователи подписываются на эти каналы и получают информацию в виде сообщений.

Многие каналы в Telegram связаны между собой: например, каналы упоминают и репостят друг друга. Это позволяет рассматривать множества Telegram-каналов в виде коммуникационных сетей и формировать на основе этого взвешенные графы. В предыдущих работах авторами была предложена (U, M, R)-модель построения таких графов взаимодействующих объектов [1, 2]. Стоит отметить, что полученные при помощи этой модели графы обладают свойствами безмасштабных сетей [3].

Полученные сети информационного взаимодействия можно изучать различными способами. Существует множество исследований, в которых авторы анализируют топологию полученного графа, выделяют неявные сообщества с помощью различных алгоритмов (Louvain, Infomap, BigCLAM и др.) [4].

Также популярны работы, в которых авторы определяют каналы, распространяющие фейковую информацию или занимающиеся противоправной деятельностью. В частности, в работе [5] авторы сосредоточились на изучении поддельных каналов, выдающих себя за официальные каналы знаменитостей или организаций и публикующие сообщения, отличные от сообщений официального канала, и каналов-клонов, которые имитируют официальные каналы, публикуя их точный контент.

Еще одним способом анализа различных объектов является построение их цифровых профилей. Данный метод активно применяется в работе рекомендательных систем. Например, музыкальный

1 Попов Владимир Александрович, аспирант Департамента прикладной математики МИЭМ НИУ ВШЭ, Москва, Россия. E-mail: vapopov@hse.ru
2 Чеповский Александр Андреевич, кандидат физико-математических наук., доцент, Департамент прикладной математики МИЭМ НИУ ВШЭ, Москва, Россия. E-mail: aachepovsky@hse.ru

стриминговый сервис Spotify представляет каждую песню в виде вектора характеристик. Если посмотреть официальный API [6], то можно увидеть, что в данный вектор входит целый ряд параметров, таких как темп, такт, громкость, модальность, инструментальность и др. Сам рекомендательный алгоритм находится в закрытом доступе, но в интернете есть множество упоминаний, что Spotify используют данные профили для нахождения похожих песен. Данный подход также применяют и в работах по анализу видео. Например, в [7] видеозаписи на YouTube сопоставляют вектор из трех компонент, отвечающих за контент видео, социальную и рекомендательную составляющие.

Есть работы, посвященные и Telegram. Так, в [8] авторы анализируют посты каналов, каждому сообщению сопоставляют его повествовательный тип. Выделяют изолированное, продолжающее, возникающее, затухающее, продолжающе-возникающее или продолжающе-затухающее сообщение. После классификации каждого поста в канале авторы подсчитывают относительные частоты категорий сообщений по типу и используют эти значения для представления канала в виде вектора признаков, который называют вектором поведенческого профиля. В дальнейшем, авторы кластеризуют полученные векторы поведенческих профилей, чтобы идентифицировать разнообразные повествовательные профили Telegram-каналов.

Учитывая особенности мессенджера Telegram возможно предложить методы построения цифровых профилей для публичных каналов этой сети, основанные как на метаданных каналов, так и на топологии импортированной подсети. Эти профили могут быть сформированы с использованием построенного по (U, M, R)-модели графа взаимодействующих объектов, а также анализа атрибутивных данных вершин, а именно: текстов постов каналов, реакций подписчиков. Представление каналов в виде цифровых профилей позволяет сравнивать каналы, находить закономерности, проводить классификацию каналов, тем самым решая различные задачи по анализу и поиску каналов в Telegram [9].

Далее в данной работе мы предложим метод построения цифрового профиля Telegram-каналов на основе графа взаимодействующих объектов и процедуру классификации каналов на основе выделяемого этим способом цифрового профиля. Также будет предложен набор типовых профилей для такой классификации и приведены примеры.

1. Модель построения графа взаимодействующих объектов

Учитывая особенности мессенджера Telegram, можно выделить следующие ключевые факторы взаимодействия между публичными Telegram-каналами: репосты между каналами, упоминания одного канала другим и наличие общих внешних URL в постах двух каналов. Учитывая наличие таких связей, Telegram-каналы образуют между собой сеть, которую можно представить в виде взвешенного графа.

Для построения таких моделей сетей каналов изначально необходимо импортировать данные из мессенджера за выбранный временной период. Для этого было разработано программное обеспечение, использующее официальный API Telegram. Данное приложение способно импортировать информацию о каналах в специализированный формат [1, 2], содержащий все необходимые компоненты для дальнейшего построения графов и цифровых профилей Telegram-каналов.

Соответственно, после импорта всех необходимых данных используя алгоритм, описанный в [1], строим граф взаимодействующих объектов сети Telegram $G(V,E)$, где V – список рассматриваемых каналов, а E – рёбра между ними. Веса полученных рёбер определяются на основе функции, описанной ниже:

$$w(e_{AB}) = 1 \cdot \delta^{U}_{e_{AB}} + 2 \cdot \delta^{M}_{e_{AB}} + 3 \cdot \delta^{R}_{e_{AB}}, \quad (1)$$

где $\delta^{U}_{e_{AB}}$ – количество общих уникальных внешних ссылок (URL) в постах каналов A и B за выбранный период времени; $\delta^{M}_{e_{AB}}$ – количество постов, в которых канал A упоминал канал B, плюс количество постов, в которых канал B упоминал канал A за выбранный период времени (для каждого поста учитываются уникальные упоминания); $\delta^{R}_{e_{AB}}$ – количество репостов каналом A постов канала B плюс количество репостов каналом B постов канала A за выбранный период времени.

Ранее в [3] было показано, что графы, построенные с использованием описанной выше (U, M, R)-модели, обладают свойствами безмасштабных сетей.

Для дальнейшей иллюстрации применения предлагаемого в данной работе метода определения цифровых профилей и классификации каналов были импортированы несколько сетей:

- Сеть каналов, связанных с образованием, назовем ее ***Education***. Для формирования этой сети мы импортировали данные из Telegram за двухмесячный период 2025 года (01.01–28.02). Полученный граф состоит из 110 вершин и 1281 ребра.
- Сеть каналов, связанных со спортом, назовем ее ***Sport***. Для формирования этой сети мы импортировали данные из Telegram за двухмесячный период 2025 года (01.01–28.02). Полученный граф состоит из 514 вершин и 8496 рёбер.
- Сеть каналов, связанных со светской жизнью, назовем ее ***Social_life***. Для формирования

этой сети мы импортировали данные из Telegram за двухнедельный период 2025 года (01.02–14.02). Полученный граф состоит из 534 вершин и 3354 рёбер.

- Сеть каналов, связанных с искусством, назовем ее ***Art***. Для формирования этой сети мы импортировали данные из Telegram за двухмесячный период 2025 года (01.01-28.02). Полученный граф состоит из 823 вершин и 7970 рёбер.

Перейдем теперь непосредственно к определению цифрового профиля для Telegram-каналов.

2. Определение цифрового профиля Telegram-канала

Каждая вершина в сети каналов Telegram обладает своими характеристиками, которые зависят как от топологии рассматриваемого графа, так и от атрибутивных данных вершины. Граф построен на некотором временном интервале T обрабатываемых данных, на нем же рассматриваются и метаданные вершин. Для начала рассмотрим три отдельные группы характеристик.

Первая группа представляет собой показатели, связанные с контентом канала:

- количество подписчиков канала (обозначим показатель за A1);
- количество постов в указанном временном интервале T (A2);
- доля репостов относительного общего количества постов канала (A3);
- количество упоминаний других Telegram-каналов в среднем за пост (A4);
- количество внешних ссылок в текстах в среднем за пост (A5);
- количество эмодзи (графических символов для выражения эмоций, идей или объектов в электронной коммуникации) в текстах в среднем за пост (A6).

Вторая группа характеристик представляет собой реакцию пользователей на контент канала, также рассматривается на всем временном интервале T:

- среднее количество просмотров постов канала (обозначим показатель за B1);
- среднее количество репостов и пересылок постов канала (B2);
- среднее количество поставленных реакций пользователей на пост (B3);
- среднее количество комментариев к постам (B4);
- рейтинг вовлеченности (B5) – среднее значение показателя вовлеченности по всем постам на канале, рассчитываемое как количество пересылок, реакций и комментариев, деленное на количество просмотров поста.

Третья группа относится к свойствам вершины и топологии графа, отражает роль канала в рассматриваемой сети. Для графа взаимодействующих объектов $G(V,E)$ базовые показатели обозначаются стандартно: N – количество вершин, w_{ij} – вес ребра между вершинами i и j. Для каждой вершины можно рассчитать следующие классические меры центральности:

- Центральность по степени (обозначим показатель за C1):

$$c_d(i) = \sum_{j}^{N} w_{ij}. \quad (2)$$

- Центральность по собственному вектору (C2): Определяется как решение уравнения:

$$Ac_e = \lambda c_e, \quad (3)$$

где A – матрица смежности графа G; λ – максимальное собственное значение матрицы A; $c_e = [c_{e1}, c_{e2}, \ldots, c_{eN}]^T$ (вектор центральностей вершин, элементы которого представляют собой центральности вершин) – собственный вектор, соответствующий собственному значению λ.

- Центральность по близости (C3):

$$c_c(i) = \frac{N-1}{\sum_{j,j\neq i}^{N} d_{ij}}, \quad (4)$$

где d_{ij} – кратчайшее расстояние между вершинами i и j.

- Нормированная центральность по посредничеству (C4):

$$c_b(i) = \frac{2}{N(N-1)} \sum_{(j,k),j\neq k} \frac{\sigma_{j,k}(i)}{\sigma_{j,k}}, \quad (5)$$

где $\sigma_{j,k}$ – число кратчайших путей от вершины j до вершины k. А $\sigma_{j,k}(i)$ – число этих путей, проходящих через вершину i.

Далее необходимо уменьшить число характеристик вершины, входящих в цифровой профиль канала. Покажем на примерах из приведенных ранее графов процесс этого отбора.

Для каждого Telegram-канала, входящего в четыре сети, которые были описаны в разделе 1, вычислим указанные выше характеристики. Далее, в разрезе каждой из сетей, для всех пар показателей посчитаем коэффициент ранговой корреляции Спирмена. Полученные результаты представлены в виде матриц корреляции (рис. 1).

Корреляция Спирмена для пары характеристик рассчитывалась как коэффициент корреляции Пирсона между рангами двух наборов данных, где ранг – это порядковый номер значений в отсортированных выборках. В случае повторяющихся значений в исходных наборах данных применялся метод среднего ранга: совпадающим значениям присваивается среднее арифметическое их рангов, которые они заняли бы при различии.

В данной работе была выбрана ранговая корреляция Спирмена, поскольку она устойчивее к выбросам и не требует нормального распределения данных.

Как видно из рис. 1, многие показатели коррелируют между собой. Например, центральности по степени (показатель C1), по собственному вектору (C2) и по близости (C3) имеют попарные коэффициенты корреляции близкие к 1, а количество репостов (B2) и просмотров (B1) имеют коэффициент корреляции выше 0,8.

Соответственно, для построения цифрового профиля Telegram-канала из всего набора выберем 5 репрезентативных слабо коррелирующих между собой характеристик, которые представляют различные уникальные особенности канала:

- Количество подписчиков канала (A1);
- Рейтинг вовлеченности (B5);
- Количество эмодзи в текстах в среднем за пост (A6);
- Центральность по степени (C1);
- Центральность по посредничеству (C4).

Выбранные характеристики отражают различные свойства исходных Telegram-каналов. Количество подписчиков показывает насколько популярен канал среди пользователей. Рейтинг вовлеченности измеряет активность и вовлеченность аудитории. Количество эмодзи в текстах относится к стилю написания текстов в канале. Центральности по степени и по посредничеству характеризуют роль, которую играет вершина в графе взаимодействующих объектов.

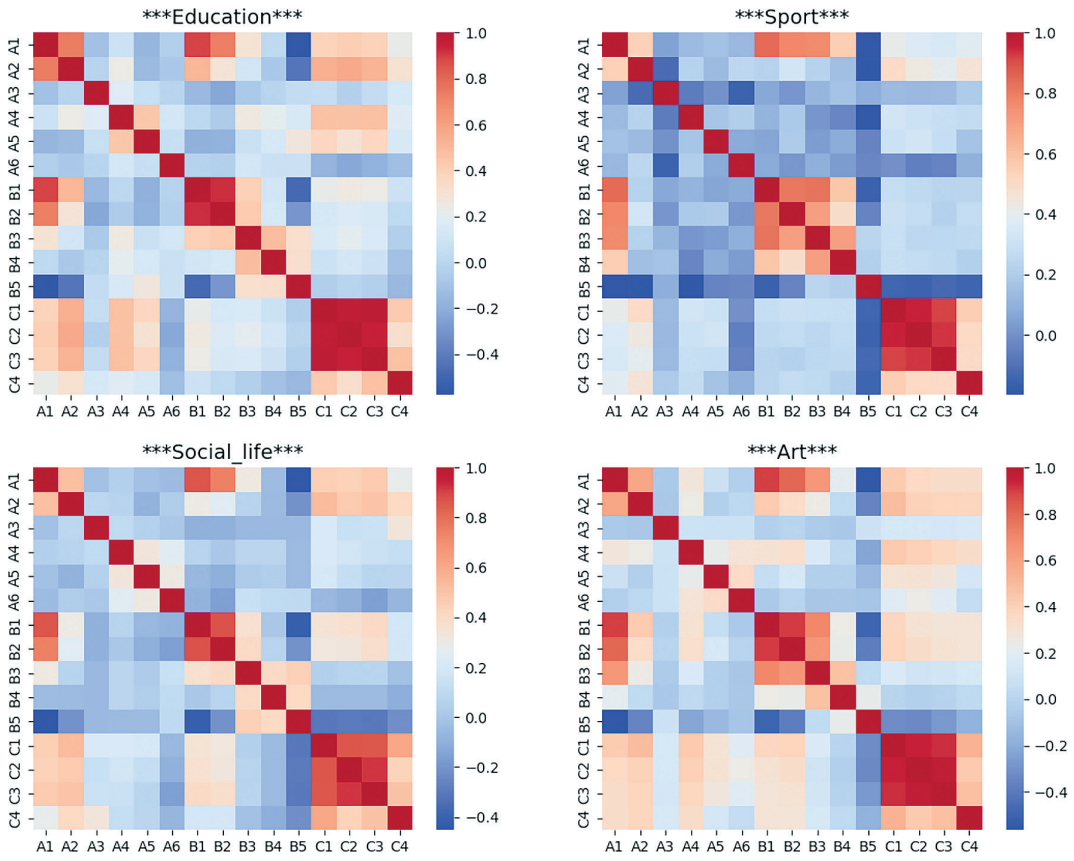


Рис. 1. Матрицы корреляций характеристик каналов

Таблица 1.

Примеры посчитанных характеристик Telegram-каналов

Telegram-канал	Количество подписчиков	Рейтинг вовлеченности	Количество эмодзи в текстах	Центральность по степени	Центральность по посредничеству
university	14991	0,07	0,16	0,11	0,04
sport_channel	306512	0,22	6,02	0,14	0,33
social_life_channel	297523	0,01	0,02	0,19	0,60
art_channel	34711	0,02	4,23	0,26	0,31

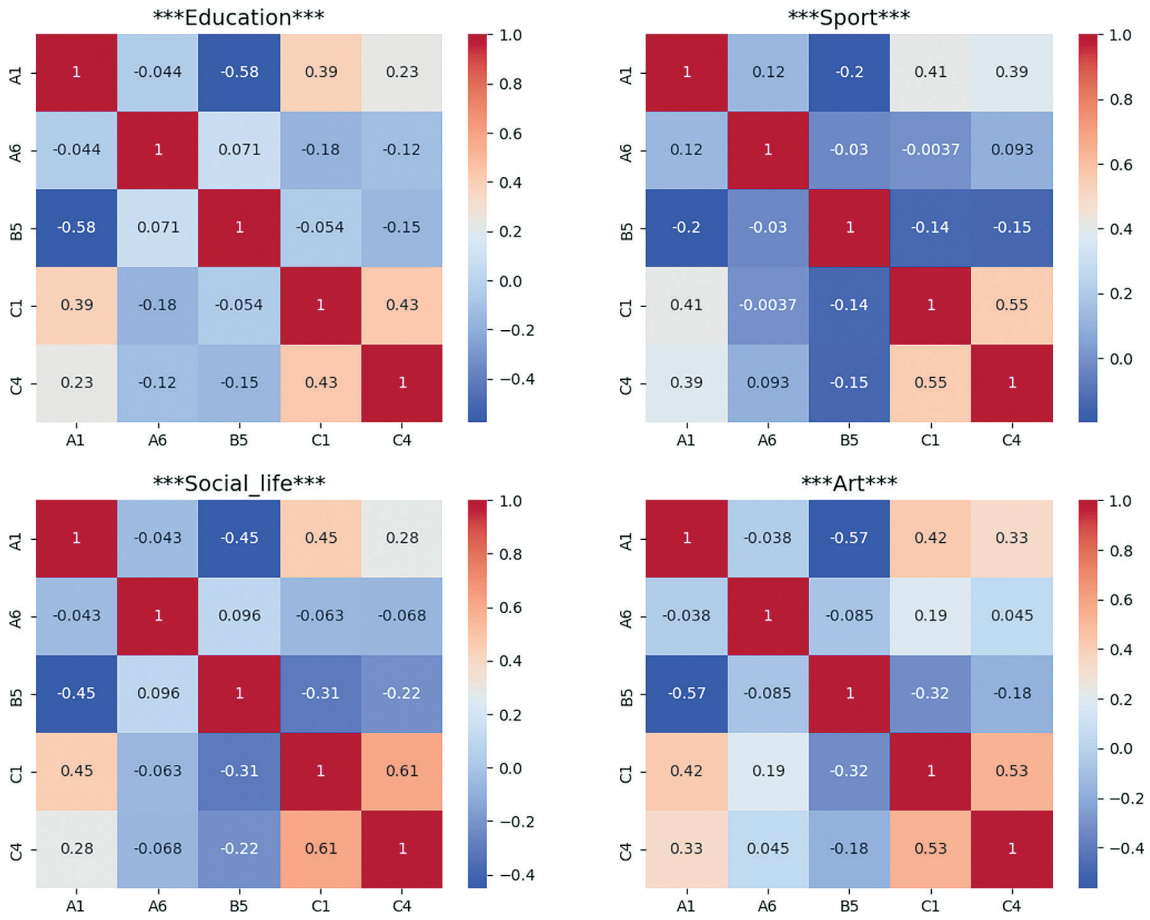


Рис. 2. Матрицы корреляций выбранных 5 характеристик каналов

На рис. 2 приведены матрицы коэффициентов корреляции Спирмена между выбранными пятью характеристиками канала. Стоит отметить, что большинство показателей имеют значение по модулю меньше 0.3. Из исключений количество подписчиков в канале и два показателя центральностей, которые отражают разные характеристики графов, поэтому были выбраны оба.

Таким образом, на основе описанного набора характеристик каждый Telegram-канал в графе взаимодействующих объектов соответствует 5-мерному вектору. В таблице 1 приведены посчитанные характеристики для относительно популярных каналов из 4 представленных выше сетей соответственно: *university*, *sport_channel*, *social_life_channel*, *art_channel*.

Указанные характеристики рассчитываются для каждого Telegram-канала анализируемого графа. Но учитывая наличие в сети каналов-выбросов, для последующего сравнения и анализа полученных 5-мерных векторов необходима их нормализация. Например, каналы со сравнительно большим числом подписчиков, как правило, искажают статистику. Так, в четырех рассматриваемых сетях медианное значение количества подписчиков равно тысячам, а максимальное – миллионам (табл. 2). Такая диспропорция может привести к искажению набора цифровых профилей.

Таблица 2.

Медианное и максимальное значение количества подписчиков в сетях

Сеть Telegram-каналов	Кол-во подписчиков, Максимум	Кол-во подписчиков, Медиана
Education	3.377.591	16.644
Sport	3.164.182	9.502
Social_life	12.002.650	23.289
Art	1.304.256	4.115

Поэтому для каждой из выделенных характеристик мы вычисляем среднее значение и стандартное отклонение в разрезе рассматриваемой сети Telegram-каналов и выявляем каналы-выбросы, значения которых отличаются более чем на три стандартных отклонения от среднего. Значения характеристик для этих каналов устанавливаются на уровне трех стандартных отклонений от среднего. Применяя эту

Таблица 3.

Примеры цифровых профилей Telegram-каналов

Telegram-канал	Количество подписчиков	Рейтинг вовлеченности	Количество эмодзи в текстах	Центральность по степени	Центральность по посредничеству
university	0,01	0,59	0,01	0,13	0,14
sport_channel	0,56	0,23	0,61	0,45	1,00
social_life_channel	0,19	0,03	0,01	1,00	1,00
art_channel	0,07	0,04	0,34	1,00	1,00

процедуру к выбросам, мы избегаем искажения всего набора цифровых профилей, что облегчает анализ характеристик каналов на последующих этапах исследования.

Далее мы нормализуем 5-мерные векторы, путем вычитания из каждой компоненты минимального значения соответствующей характеристики во всем графе и последующего деления каждой компоненты на максимальное значение.

Полученный нормализованный 5-мерный вектор с компонентами на отрезке [0; 1] представляет собой цифровой профиль канала Telegram в заданной сети за указанный период времени. Такое векторное представление характеристик каналов позволяет в дальнейшем проводить сравнительный анализ и их классификацию.

Продолжая предыдущий пример, построим цифровые профили для каналов *university*, *sport_channel*, *social_life_channel*, *art_channel* (Табл. 3). Стоит отметить, что каждая сеть каналов рассматривается отдельно друг от друга, поэтому для четырех Telegram-каналов проводилась своя независимая нормализация.

Как видно из таблицы, профили каналов существенно отличаются и каждый из них имеет свои особенности. Первый канал имеет малое для своего графа количество подписчиков, но высокий рейтинг вовлеченности аудитории. Если изучить сам канал, то так оно и есть: в канале примерно 15 тысяч подписчиков, и аудитория проявляет высокую активность: под постами суммарно ставятся тысячи реакций и оставляются сотни комментариев. Второй канал отличается средним количеством подписчиков, повышенным количеством эмодзи в текстах и наивысшим показателем центральности по посредничеству. А третий и четвертый каналы, учитывая их наивысшие степени центральности, являются важнейшими вершинами в структуре их графов взаимодействующих объектов. Что на практике так и есть: вершина *social_life_channel* связана с 99 из 534 вершин в графе, а *art_channel* связана с 211 из 823 вершин в сети. Таким образом, цифровые профили Telegram-каналов дают представления об их основных характеристиках.

3. Использование цифровых профилей для классификации Telegram-каналов

Предложенный в данной работе метод можно кратко записать следующим образом.

Алгоритм смешанной классификации

Шаг 1. Вычисление для каждой вершины $v_i \in V$ графа взаимодействующих объектов ее показателей: A1, A6, B5, C1, C4. Далее обозначим их для заданной вершины v_i как v_i^j, где $j = 1,\ldots,5$.

Шаг 2. Составление для каждой вершины $v_i \in V$ ее цифрового профиля, вектора $p_{v_i} = (v_i^1, v_i^2, v_i^3, v_i^4, v_i^5)$.

$$\alpha_0: V \longrightarrow \{p_{v_i}\}. \quad (6)$$

Шаг 3. Кластеризация векторов p_{v_i}. Получаем k кластеров K_s, где $s = 1,\ldots,k$.

$$\alpha_1: \{p_{v_i}\} \longrightarrow K_s. \quad (7)$$

Шаг 4. Поиск в кластерах K_s их центров. Получаем множество $\{z_s\}$, где $s = 1,\ldots,k$. Так можно считать, что определено отображение:

$$\alpha_2: \{p_{v_i}\} \longrightarrow \{z_s\}. \quad (8)$$

Шаг 5. Классификация полученных центров $\{z_s\}$ по заданному наперед набору из g классов $M = \{M_j\}$, где $j = 1,\ldots,g$:

$$\alpha_3: \{z_s\} \longrightarrow M. \quad (9)$$

Шаг 6. Распространение на все вершины, входящие в кластер результата классификации его центра:

$$\alpha: \{v_i\} \longrightarrow M. \quad (10)$$

Таким образом, в итоге действия Алгоритма классификации получаем для каждой вершины исходного графа ее отнесение к одному из заданных классов:

$$\alpha(v_i) = \alpha_3(\alpha_2(\alpha_0(v_i))). \quad (11)$$

Фактически, в представленном алгоритме смешанной классификации заложено последовательное решение задачи кластеризации элементов одного множества и задачи классификации элементов другого множества. Использование кластеризации вместе с классификацией имеет ряд преимуществ. Кластеризация помогает выявить скрытые закономерности и структуры в данных, что дает дополнительную

информацию об объектах. А также центры кластеров более устойчивы к выбросам и случайному шуму, чем отдельные точки, что в целом может повысить точность классификации [10].

Учитывая, что мы рассматриваем графы взаимодействующих объектов, будем при классифицировании учитывать, какую роль играет вершина в графе. Зададим множество $M = \{M_1, M_2, M_3, M_4\}$, состоящее из 4-х классов каналов: M_1 – крупные центры; M_2 – мосты и ретрансляторы; M_3 – локальные или узконаправленные центры; M_4 – пассивные каналы.

Реализация шагов 1 и 2 производится за счет описанных в разделе 1 действий. Далее для реализации шага 3 необходимо определиться с используемым алгоритмом кластеризации.

Существует множество алгоритмов кластеризации векторов, но не все они подходят для нашей задачи. В частности, большинство классических алгоритмов, например, k-средних или DBSCAN, требуют заранее определять параметры. Соответственно, использование таких алгоритмов не позволяет сформировать универсальный метод кластеризации цифровых профилей для сетей каналов разного размера и структуры, так как параметры алгоритмов зависят от характеристик сети.

Поэтому для кластеризации цифровых профилей Telegram-каналов мы предлагаем использовать метод агломеративной иерархической кластеризации. Суть данного алгоритма заключается в том, что на изначальном этапе каждый вектор рассматривается как отдельный кластер, далее они итеративно объединяются в кластеры большего размера, пока не будет сформирован один единый кластер. На каждом шаге алгоритма объединяются два кластера, расстояние между которыми минимальное. Таким образом, иерархическое дерево формируется от листьев к стволу.

Для реализации данного подхода необходимо определить расстояние ρ_1 для пары произвольных векторов p_{v_i} и p_{v_j}. Для расчета расстояния между двумя элементами-векторами будем использовать расстояние Чебышева: максимальная абсолютная разность между координатами двух векторов в пространстве, как представлено в формуле (12).

$$\rho_1(p_{v_i}, p_{v_j}) = \max_{t=1,\ldots,5} |v_i^t - v_j^t|. \quad (12)$$

Учитывая проведенную нормализацию при построении цифровых профилей все координаты векторов принимают значения от 0 до 1. Тогда расстояние Чебышева между двумя цифровыми профилями будет принадлежать отрезку [0, 1].

Далее, для определения расстояния ρ_2 между двумя кластерами K_1 и K_2 будем использовать метод полной связи (метод дальнего соседа): расстояние между двумя кластерами определяется как максимум из множества всех попарных расстояний между элементами первого кластера K_1 и элементами второго кластера K_2. Более формально это записано в виде формулы (13):

$$\rho_2(K_1, K_2) = \max_{\forall p_{v_i} \in K_1,\ p_{v_j} \in K_2} \rho_1(p_{v_i}, p_{v_j}). \quad (13)$$

Как было указано ранее $\rho_1(p_{v_i}, p_{v_j}) \in [0,1]$. Откуда следует, что и расстояние между двумя кластерами будет также принадлежать отрезку [0, 1].

Помимо того, что метод иерархической кластеризации не требует предварительного указания параметров, алгоритм также раскрывает иерархическую структуру данных. Для этого можно построить дендрограмму – древовидную диаграмму, отражающую каждый шаг процесса последовательного укрупнения

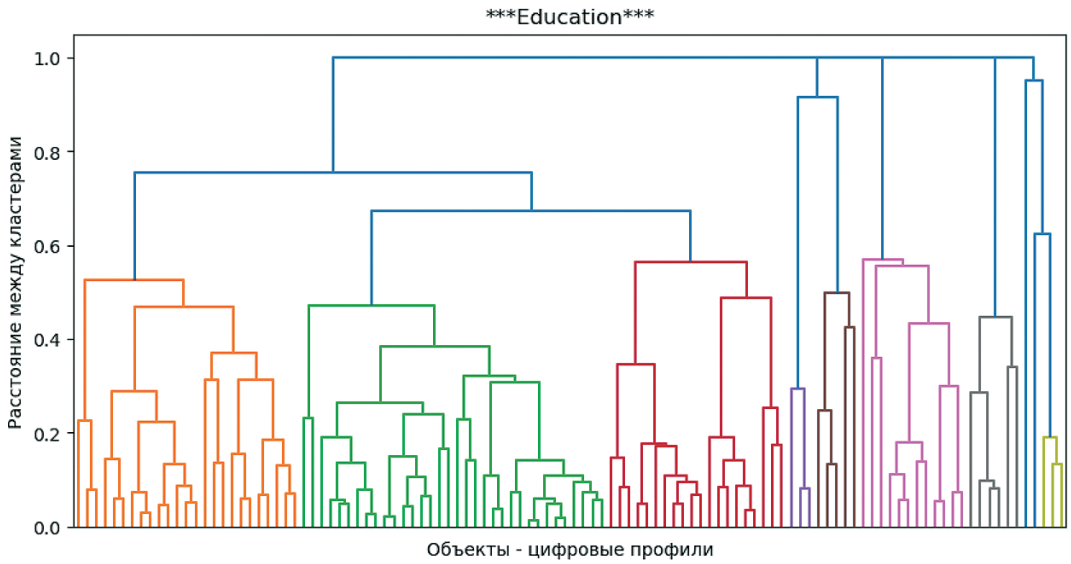


*Рис. 3. Дендрограмма кластеризации цифровых профилей Telegram каналов сети ***Education****

кластеров, что позволяет выявлять значимые кластеры сети на основе определенных критериев. Для определения границы между кластерами, а значит и того, когда объекты находятся в разных кластерах, изначально в данном методе задается параметр Δ. Экспериментально на графах, полученных из сети Telegram-каналов, наиболее эффективно показало себя значение $\Delta = 0{,}6$.

Для примера применим описанный метод к упомянутой ранее сети каналов, связанных с образованием, к сети ***Education***. На рисунке 3 представлена полученная дендрограмма кластеризации 110-ти векторов цифровых профилей, которые были разделены на 10 кластеров.

На шаге 4 производим поиск $\{z_s\}$ – центров кластеров. Пусть $\{p_{v_i} = (v_i^1, v_i^2, v_i^3, v_i^4, v_i^5)\}$, где $i = 1,\ldots,|K_s|$ – множество всех векторов, принадлежащих кластеру K_s. Тогда центром кластера K_s будет являться следующий вектор:

$$z_s = \frac{1}{|K_s|}\sum_{i=1}^{|K_s|}\overline{p_{v_i}} =$$

$$= \left(\frac{1}{|K_s|}\sum_{i=1}^{|K_s|}\overline{v_i^1}, \frac{1}{|K_s|}\sum_{i=1}^{|K_s|}\overline{v_i^2}, \frac{1}{|K_s|}\sum_{i=1}^{|K_s|}\overline{v_i^3}, \frac{1}{|K_s|}\sum_{i=1}^{|K_s|}\overline{v_i^4}, \frac{1}{|K_s|}\sum_{i=1}^{|K_s|}\overline{v_i^5}\right). \quad (14)$$

Далее на шаге 5 будем производить классификацию полученных центров кластеров $\{z_s\}$, находя, к какому из классов множества $M = \{M_1, M_2, M_3, M_4\}$ они относятся. Для этого определим эталонный вектор для каждого из этих классов. После чего, с использованием евклидова расстояния ρ_3 найдем расстояние от каждого из центров до каждого из эталонных векторов.

Учитывая, что эти классы соответствуют следующим ролям каналов в сети: крупные центры, мосты и ретрансляторы, локальные или узконаправленные центры, пассивные. Данные типы Telegram-каналов имеют свои особенности, которые можно отразить в цифровых профилях. В таблице 4 приведены значения предложенных авторами цифровых профилей для каждого типа.

Для классификации полученных 10 центров кластеров, найдем евклидово расстояние от них до каждого из четырех предложенных профилей и выберем минимальное (Таблица 5).

Итак, для каждого из центров $\{z_s\}$ сопоставлен его класс из множества $M = \{M_1, M_2, M_3, M_4\}$.

Теперь перейдем к шагу 6 и присвоим всем вершинам исходного графа, входящим в каждый кластер, свой класс, равный классу центра этого кластера.

Таблица 4.

Эталонные цифровые профили для каждого типа Telegram-каналов

Тип Telegram-канала	Количество подписчиков	Рейтинг вовлеченности	Количество эмодзи в текстах	Центральность по степени	Центральность по посредничеству
Крупный центр	1	0,25	0	1	0,5
Мост и ретранслятор	0	0,5	0,5	0,5	1
Локальный или узконаправленный	0,25	1	0,5	1	0
Пассивный	0	0	0	0	0

Таблица 5.

Евклидово расстояние от центров кластеров до выбранных векторов цифровых профилей

№ кластера	Крупный центр	Мост и ретранслятор	Локальный или узконаправленный	Пассивный
1	**1,40**	1,53	1,43	1,41
2	**0,85**	1,42	1,42	0,94
3	1,10	**0,58**	1,34	1,31
4	1,33	**0,64**	1,48	0,81
5	1,51	1,17	**1,03**	1,20
6	1,16	1,04	1,01	**0,61**
7	1,41	1,16	1,29	**0,25**
8	1,39	1,05	1,06	**0,55**
9	1,51	1,12	1,04	**0,71**
10	1,67	1,19	1,28	**0,93**

Таблица 6.

Кластерные центры сети ***Education***

№	Тип кластера	Количество подписчиков	Рейтинг вовлеченности	Количество эмодзи в текстах	Центральность по степени	Центральность по посредничеству
1	Крупный центр	0,951	0,225	1	0,167	0
2	Крупный центр	0,882	0,035	0,085	0,297	0,104
3	Мост и ретрансл.	0,037	0,271	0,069	0,804	1
4	Мост и ретрансл.	0,034	0,068	0,265	0,182	0,736
5	Локальный/ узконапр.	0,053	0,195	1	0,634	0
6	Пассивный	0,027	0,212	0,121	0,552	0,078
7	Пассивный	0,036	0,195	0,115	0,093	0,031
8	Пассивный	0,011	0,304	0,382	0,251	0,011
9	Пассивный	0,001	0,669	0,211	0,089	0,011
10	Пассивный	0,07	0,22	0,902	0,08	0

В нашем рассматриваемом примере мы определим типы полученных 10 центров кластеров, а соответственно и типы всех Telegram-каналов, входящих в эти кластеры (Таблица 6). Как видно из таблицы, в данной сети присутствуют все четыре типа каналов.

После этого экспертным образом была оценена полученная в итоге классификация каналов сети. Разделение каналов по типу их цифрового профиля оказалось актуальным. В крупные центры попали каналы больших СМИ и новостных сообществ, которые в основном сами генерируют свой контент, не делая репосты и не упоминая другие каналы, минимальное количество подписчиков для этих каналов равно 1,2 миллионам. Также интерес представляет наличие двух кластеров крупных центров. При детальном изучении самих каналов выясняется, что в кластер № 2 попали крупные новостные каналы без конкретных направленностей, а в кластер № 1 – каналы, связанные только с военной тематикой. Выделение данных каналов в отдельный кластер также подтверждает релевантность алгоритма кластеризации.

Далее для анализа интересны каналы – мосты и ретрансляторы. Каналы этих кластеров действительно являются распространителями информации: в основном данные каналы распространяют информацию, полученную из других Telegram-каналов или других источников информации. Но интерес заключается в том, что в сети было выявлено два таких кластера. После изучения списков каналов в этих кластерах было установлено, что кластер № 4 включает официальные Telegram-каналы, аффилированные с государством, а кластер № 3 – каналы с альтернативными взглядами. Также стоит отметить, что у каналов в третьем кластере максимальные значения центральностей, что часто характерно для каналов распространителей информации.

Таким образом, получаем алгоритм смешанной классификации $\theta(M,\alpha_1,\rho_1,\rho_2,\alpha_2,\alpha_3,\rho_3)$. Применяя предложенный алгоритм смешанной классификации, можно определить роль и особенности каждого узла в сети каналов Telegram, что полезно для многих задач в контексте анализа социальных сетей.

Заключение

В данной статье представлено определение цифрового профиля для Telegram-каналов на основании как топологических свойств вершин графа взаимодействующих объектов, так и на основании характеристик самого канала. Далее описан метод для решения задачи классификации каналов на основании их цифровых профилей. Выделение четырех типов узлов (крупные центры, мосты и ретрансляторы, локальные или узконаправленные центры и пассивные каналы) позволило нам классифицировать и более подробно описать характеристики каналов в сети.

Также были приведены примеры построения цифровых профилей Telegram-каналов и последующая их классификация. Экспертная оценка примеров еще раз подтвердила актуальность данного подхода к классификации, в частности, подчеркнув различие между официальными каналами Telegram, связанными с государственными органами, и теми, которые принадлежат альтернативным новостным агентствам.

Это исследование дает ценные сведения о различных ролях и функциях каналов в Telegram. Таким образом, эти результаты могут быть использованы в будущих исследованиях и прикладных задачах.

Литература
1. Попов В. А., Чеповский А. А. Модели импорта данных из мессенджера Telegram // Вестник Новосибирского государственного университета. Серия: Информационные технологии. 2022. Т. 20. № 2. С. 60–71.
2. Чеповский А. А. Анализ графов взаимодействующих объектов. — М.: Национальный открытый университет «ИНТУИТ». 2022. – 270 с.
3. Попов В. А., Чеповский А. А. О моделях построения графа взаимодействующих объектов в сети Telegram-каналов // Вопросы кибербезопасности. 2024. № 3(61). С. 105–112. DOI:10.21681/2311-3456-2024-3-105-112.
4. Чеповский А. А. О неявных сообществах на графе взаимодействующих объектов // Успехи кибернетики. – 2023. – Т.4. – № 1. – С. 56–64.
5. La Morgia M., Mei A., Mongardini A. M., Wu J.: Uncovering the Dark Side of Telegram: Fakes, Clones, Scams, and Conspiracy Movements. https://arxiv.org/abs/2111.13530. (2021). (Дата обращения: 01.07.2025).
6. Spotify for Developers – https://developer.spotify.com/documentation/web-api/reference/get-audio-features (Дата обращения: 05.07.2025).
7. Leopaul Boesinger, Manoel Horta Ribeiro, Veniamin Veselovsky, Robert West: Tube2Vec: Social and Semantic Embeddings of YouTube Channels. https://arxiv.org/abs/2306.17298 (Дата обращения: 01.07.2025).
8. Willaert T.: A computational analysis of Telegram's narrative affordances. PLoS ONE 18(11), p. 1–23, (2023). https://doi.org/10.1371/journal.pone.0293508
9. Popov, V. A., Chepovskiy, A. A.: Constructing Telegram Channels Digital Profiles. Complex Networks & Their Applications XIII. COMPLEX NETWORKS 2024 2024. Studies in Computational Intelligence, vol. 1189. Springer, Cham. https://doi.org/10.1007/978-3-031-82435-7_7 (2025).
10. Piernik, M., Morzy, T. A study on using data clustering for feature extraction to improve the quality of classification. Knowledge and Information Systems, 63, 1771–1805 (2021).

TELEGRAM-CHANNELS CLASSIFICATION APPROACH

Popov V. A.[3], Chepovskiy A. A.[4]

Keywords: *digital profiles, social network analysis, scale-free networks, model of information impact, community identification, classification problems.*

The purpose of the study: *development of a method for determining the digital profile of Telegram channels in information interaction networks and a procedure for classifying channels based on the allocated digital profile.*

Method: *includes the following stages: graph of interacting objects construction, based on data imported from the Telegram network, digital profiles determination for vertices based on their attribute data and graph properties, clustering of vertices based on the selected profiles, centers of the obtained clusters and the original Telegram channels classification, computational experiments and analysis of the results.*

Results: *the article introduces the digital profile definition for a Telegram channel, presented as one of the vertices of a graph of interacting objects. The digital profile is defined through a normalized 5-dimensional feature vector based on the attribute data of the vertex and the graph properties. The selected characteristics reflect the properties of Telegram channels in the constructed graph and the metadata obtained during import from the network. The authors then describe an algorithm for clustering the obtained profiles using configurable parameters. The centers of the selected clusters are classified according to 4 types proposed by the authors, characterizing the roles of vertices in the graph of interacting objects. Due to this, all vertices of the graph are classified – the original Telegram channels of the analyzed network. The proposed approach provides valuable information about the roles of Telegram channels in information interaction networks.*

Scientific novelty: *a new approach to the analysis of Telegram channels is developed: a method for creating a digital profile of a Telegram channel in the form of a 5-dimensional feature vector, allowing the analysis and classification of channels. The approach also proposes a procedure for classifying such digital profiles based on computational methods, which allows to identify the main types of Telegram channels of the downloaded subnet according to a given classification.*

References
1. Popov V. A., Chepovskij A. A. Modeli importa dannyh iz messendzhera Telegram // Vestnik Novosibirskogo gosudarstvennogo universiteta. Serija: Informacionnye tehnologii. 2022. T. 20. № 2. S. 60–71.
2. Chepovskiy A. A. Analiz grafov vzaimodejstvujushhih ob#ektov.: Nacional'nyj otkrytyj universitet «INTUIT». 2022. – 270 s.
3. Popov V. A., Chepovskij A. A. O modeljah postroenija grafa vzaimodejstvujushhih ob#ektov v seti Telegram-kanalov // Voprosy kiberbezopasnosti. 2024. № 3(61). S. 105–112. DOI:10.21681/2311-3456-2024-3-105-112.
4. Chepovskij A. A. O nejavnyh soobshhestvah na grafe vzaimodejstvujushhih ob#ektov // Uspehi kibernetiki. – 2023. – T.4. – № 1. – C. 56–64.
5. La Morgia M., Mei A., Mongardini A. M., Wu J.: Uncovering the Dark Side of Telegram: Fakes, Clones, Scams, and Conspiracy Movements. https://arxiv.org/abs/2111.13530. (2021). (Data obrashhenija: 01.07.2025).

3 Vladimir A. Popov, Ph.D. student, School of Applied Mathematics, HSE MIEM, Moscow, Russia. E-mail: vapopov@hse.ru.
4 Alexander A. Chepovskiy, Ph.D. in Physics and Mathematics, Associate Professor, Department of Applied Mathematics, Moscow Institute of Economics, National Research University Higher School of Economics. Moscow, Russia. E-mail: aachepovsky@hse.ru

6. Spotify for Developers – https://developer.spotify.com/documentation/web-api/reference/get-audio-features (Data obrashhenija: 05.07.2025).
7. Leopaul Boesinger, Manoel Horta Ribeiro, Veniamin Veselovsky, Robert West: Tube2Vec: Social and Semantic Embeddings of YouTube Channels. https://arxiv.org/abs/2306.17298 (Data obrashhenija: 01.07.2025).
8. Willaert T.: A computational analysis of Telegram's narrative affordances. PLoS ONE 18(11), p. 1–23, (2023). https://doi.org/10.1371/journal.pone.0293508.
9. Popov, V. A., Chepovskiy, A. A.: Constructing Telegram Channels Digital Profiles. Complex Networks & Their Applications XIII. COMPLEX NETWORKS 2024 2024. Studies in Computational Intelligence, vol. 1189. Springer, Cham. https://doi.org/10.1007/978-3-031-82435-7_7 (2025).
10. Piernik, M., Morzy, T. A study on using data clustering for feature extraction to improve the quality of classification. Knowledge and Information Systems, 63, 1771–1805 (2021).

МЕТОДИКА РАЗРАБОТКИ МИНИМАЛЬНЫХ СЦЕНАРИЕВ ВЫПОЛНЕНИЯ ЭТАПОВ ЖИЗНЕННОГО ЦИКЛА ЭЛЕКТРОННОГО ДОКУМЕНТА ОГРАНИЧЕННОГО ДОСТУПА

Поддубный М. И.[1]

DOI: 10.21681/2311-3456-2025-4-84-92

Актуальность: *особенности обработки электронных документов ограниченного доступа в компьютерных системах актуализируют вопросы формирования минимальных сценариев выполнения каждого этапа жизненного цикла документа. Известные алгоритмы поиска таких сценариев не учитывают порождаемую применяемой политикой безопасности изменчивость значения показателя затрат вычислительного ресурса отдельно взятого запроса в сценарии и не могут быть применены.*

Целью исследования *является создание методики разработки сценариев выполнения этапов жизненного цикла обрабатываемого компьютерной системой документа ограниченного доступа, характеризующихся минимальными затратами вычислительного ресурса.*

Метод исследования: *указанные сценарии предлагается рассматривать как пути в диаграмме переходов детерминированного конечного автомата, описывающего реализованную в компьютерной системе политику безопасности. В качестве весов дуг принимается показатель затрат вычислительного ресурса на обработку каждого атомарного запроса в пути, что позволяет применить подходы к построению и обработке матриц переходов детерминированного конечного автомата высшего порядка и поиску путей в нем, базирующихся на трудах Ф. Хона, С. Сешу, Д. Ауфенкампа и А. Гилла.*

Новизна: *элементом новизны является порядок расчета показателя затрат вычислительного ресурса, учитывающий зависимость последовательных запросов в пути между собой. Также к элементам новизны следует отнести описанные и обоснованные в работе условия останова поиска минимальных путей в рассматриваемом автомате.*

Результат исследования: *разработанная методика позволяет определить сценарий выполнения этапа жизненного цикла обрабатываемого компьютерной системой документа, характеризующийся минимальными затратами вычислительного ресурса, с учетом применяемой политики безопасности, избегая необходимости полного перебора допустимых решений. Применение разработанных таким образом сценариев в качестве реакции на запрос пользователя обрабатывающей документы ограниченного доступа компьютерной системе позволит исключить возможность нарушения жизненного цикла документа и минимизировать затраты на его обработку.*

Ключевые слова: *детерминированный конечный автомат, затраты вычислительного ресурса, компьютерная система, поиск пути в автомате, политика безопасности, сценарий обработки документа, управление доступом в компьютерной системе.*

Введение

Повсеместное внедрение компьютерных систем (КС) в области управления деятельностью организаций позволяет рассматривать в качестве информационного объекта обработки электронный документ ограниченного доступа (далее – документ) [1–4].

На сегодняшний день порядок обработки документов в КС определяется реализованной в ней политикой безопасности (ПБ) – одного или нескольких правил, процедур, практических приемов или руководящих принципов в области информационной безопасности, которыми руководствуется организация в своей деятельности. Формальное же описание ПБ конкретизирует допустимые режимы обработки документа и называется моделью безопасности (МБ) [5].

Анализ применяемых в КС МБ с точки зрения проводимого исследования позволил выявить их ключевые общие черты [6, 7]:

- атомарность событий изменений состояний системы в результате подаваемых пользователями запросов;
- предоставление возможности пользователю подачи всех допустимых запросов относительно документа на любом этапе его обработки.

Вместе с тем документ, как информационный объект обработки, обладает важной особенностью – наличием жизненного цикла (ЖЦ), т.е. конкретной последовательностью событий, сопровождающей

1 Поддубный Максим Игоревич, кандидат технических наук, докторант. Федеральное государственное казенное военное образовательное учреждение высшего образования «Краснодарское высшее военное орденов Жукова и Октябрьской Революции Краснознаменное училище имени генерала армии С. М. Штеменко» Министерства обороны Российской Федерации. г. Краснодар. Россия. E–mail: podd.maxim@yandex.ru

его создание и использование[2]. Нарушение порядка запросов системе, порождающих данные события, потенциально может привести к существенному ущербу.

В настоящее время задача соблюдения ЖЦ документа выполняется организационными мерами и предполагает добросовестный подход легитимного пользователя к своим обязанностям. Одним из резонансных свидетельств недостаточности таких мер стало уголовное преследование экс-сотрудника АО «ГосМКБ «Радуга» им. А. Я. Березняка», обвиняемого в неправомерном воздействии на критическую информационную инфраструктуру Российской Федерации[3].

Таким образом, МБ КС, обрабатывающей документы, помимо всего прочего должна характеризоваться наличием механизмов контроля их ЖЦ. В этом случае процесс обработки документа рассматривается как упорядоченное множество процедур, каждая из которых соответствует определенному этапу его ЖЦ и представляет собой последовательность атомарных запросов – сценарий выполнения этапа ЖЦ документа (далее – сценарий).

Учитывая указанное, возникает необходимость разработки сценариев, характеризующихся минимальными затратами вычислительного ресурса (далее – затратами).

На сегодняшний день для разработки и последующего анализа МБ как правило применяется субъектно-сущностный подход [7, 8], где МБ описывается детерминированным абстрактным конечным автоматом (ДКА). Опираясь на исследования матриц переходов ДКА[4] и их преобразований на основе элементов линейной алгебры[5], а также развития этих исследований некоторыми современниками [9–11], предлагается рассмотреть сценарий как путь в графе диаграммы переходов указанного автомата. В данном случае в качестве веса каждой дуги диаграммы переходов выступает показатель затрат проверки условий обработки запроса и его выполнения. Следует отметить, что на практике в заранее определенном сценарии обработки документа может потребоваться проверка тех же условий, что и в предыдущем запросе, что является избыточным и выполняться не будет. Данный факт порождает некоторую изменчивость весов дуг в зависимости от условий выполнения предыдущего запроса.

2 ГОСТ Р 7.0.95–2015 Система стандартов по информации, библиотечному и издательскому делу. Электронные документы. Основные виды, выходные сведения, технологические характеристики. М.: Стандартинформ, 2015. 12 с.

3 Соковкин А. Российские ракеты пострадали от брака в Турции / под ред. С. М. Яковлева. // Газета «Коммерсантъ», 2024. № 102. С. 4. URL: https://www.kommersant.ru/doc/6763287 (дата обращения: 11.03.2025).

4 Hohn F. E., Seshu S., Aufenkamp D. D. The Theory of Nets, J. R. E. Trans., vol. EC–6, 1957, pp. 154–161. DOI: 10.1109/TEC.1957.5222012.

5 Gill A. Introduction to the Theory of Finite-State Machines. New York, San Francisco, Toronto, London, MGH, 1962, 207 p. ISBN 0070232431, 9780070232433.

Известны «классические» алгоритмы поиска минимальных путей в графе, подробный анализ которых представлен в [12, 13], а также наиболее подходящие в контексте проводимого исследования их модификации [14–16]. Однако указанная изменчивость не позволяет их применить для решения рассматриваемой задачи.

Таким образом, возникает противоречие между потребностью разработки минимальных сценариев выполнения этапов ЖЦ документа и невозможностью удовлетворить указанную потребность в заданных условиях на основе существующих методов, моделей и технических решений, что подтверждает актуальность проведенного исследования.

Постановка задачи

Постановка задачи и последующее ее решение обусловлены следующими допущениями:

1. В качестве вычислительного ресурса выступает системное время, исчисляемое в тактах.
2. Начальное и допускающее (конечные) состояния рассматриваемых МБ, единственные.
3. Каждое состояние в МБ представляет собой совокупность атрибутов (элементов) обрабатываемого документа.
4. Каждый запрос в МБ представляет собой примитивную команду, которая изменяет (создает или удаляет) единственный элемент состояния, т.е. включает в себя единственный операнд, показатель затрат которого составляет 1 такт.
5. Условия срабатывания каждого операнда (и как следствие примитивной команды) зависит от применяемой в КС МБ. Показатель затрат проверки отдельно взятого условия составляет 1 такт. Число проверяемых условий может быть различным, но не менее одного. Таким образом показатель затрат проверки условий срабатывания примитивной команды составляет 1 такт и более.
6. В случае, если набор условий срабатывания примитивной команды совпадает с набором предыдущей, то повторная их проверка не выполняется.

Для формальной постановки и решения задачи в работе введены обозначения:

$U = \{u_h : h = \overline{1, n_U}\}$ – множество этапов ЖЦ документа, где u_h – произвольный этап ЖЦ, n_U – число этапов ЖЦ документа;

$X = \{x_q : q = \overline{1, n_X}\}$ – множество атрибутов (элементов), определяющих состояния ДКА, где x_q – произвольный атрибут, q – номер атрибута, n_X – число атрибутов;

$\Lambda_U = \{\lambda_h : \lambda_h = \langle G_h, g_\alpha^h, g_\beta^h, OPR, \delta_h \rangle\}$ – множество ДКА без выходного преобразователя, реализующих этапы ЖЦ документа на основе заданной МБ, где λ_h – произвольный ДКА, реализующий h-й этап ЖЦ документа;

$G_h = \{g_w^h : g_w^h \subseteq X,\ w = \overline{1,n_{G_h}}\}$ – множество состояний ДКА λ_h, где g_w^h – произвольное состояние, n_{G_h} – число состояний;

g_α^h – начальное состояние, g_β^h – допускающее состояние ДКА λ_h, отражающее завершение h-го этапа ЖЦ документа, при этом, $\{g_\alpha^h, g_\beta^h\} \subseteq G_h$;

$Z = \{z_\zeta, \zeta = \overline{1,n_Z}\}$ – множество примитивных операндов, z_ζ – произвольный примитивный операнд, n_Z – число примитивных операндов;

$Y = \{y_\eta, \eta = \overline{1,n_Y}\}$ – множество условий срабатывания примитивного операнда, y_η – произвольное условие, n_Y – число условий;

$OPR = \{opr_v : opr_v = 2^Y \times Z, v = \overline{1,n_{OPR}}\}$ – множество примитивных команд в ДКА, где opr_v – произвольная примитивная команда, n_{OPR} – число примитивных команд. При этом каждая примитивная команда $opr_v \in OPR$ состоит из набора условий выполнения команды и одного примитивного операнда;

$\delta_h (G_h, OPR) = G_h$ – функция переходов ДКА из состояния в состояние, определяемая заданной в системе МБ;

$[\lambda_h]$ – матрица переходов ДКА λ_h, элемент π_{ab} которой на пересечении строки a, и столбца b определяется наличием примитивной команды opr_{ab}, перехода ДКА из состояния g_a^h в состояние g_b^h, при этом $\{g_a^h, g_b^h\} \subseteq G_h$;

$P = \{p\}$ – множество всех возможных путей в ДКА λ_h, описывающем применяемую МБ. Каждый путь $p \in P$ представляет собой последовательность примитивных команд $opr_v \in OPR$, которая в диаграмме переходов ведет из одного состояния в другое;

$p(ab)_l^\omega$ – путь в ДКА длины ω из состояния с литером a (т.е. – g_a^h) в состояние с литером b (т.е. – g_b^h), представляет собой упорядоченную последовательность примитивных команд который представляется упорядоченным произведением примитивных команд и определяется по формуле:

$$p(ab)_l^\omega = opr_{ai_1} \cdot opr_{i_1} opr_{i_2} \cdot \ldots \cdot opr_{i_{\omega-1}b}, \qquad (1)$$

где каждый индекс примитивной команды a, i_1, i_2, ..., $i_{\omega-1}$, b отражает состояние ДКА, из которого (в которое) переводит его примитивная команда, l – номер пути. При этом, если примитивная команда opr_v содержится в пути $p(ab)_l^\omega$, то в рамках работы будем записывать $opr_v \in p(ab)_l^\omega$;

P_{ab}^ω – множество путей длины ω из состояния с литером a (т.е. – g_a^h) в состояние с литером b (т.е. – g_b^h), которое $P_{ab}^\omega \subseteq P$ и представляется неупорядоченной суммой путей:

$$P_{ab}^\omega = \sum_{l=1}^{n_P} p(ab)_l^\omega, \qquad (2)$$

где $p(ab)_l^\omega$ – путь в ДКА, n_P – число путей во множестве P_{ab}^ω, при этом, если путь $p(ab)_l^\omega$ содержится в P_{ab}^ω, то в рамках работы будем записывать $p(ab)_l^\omega \in P_{ab}^\omega$ (рис. 1);

На вербальном уровне задача разработки минимальных сценариев выполнения этапов ЖЦ документа может быть декомпозирована на несколько этапов:

1. Формирование графа диаграммы переходов ДКА, соответствующего применяемой в системе МБ. Дугами графа являются применяемые в ней примитивные команды, вершинами – уникальные наборы элементов (атрибутов) обрабатываемого документа.
2. Определение множества путей в указанном графе, ведущих из начального в допускающее состояние и отражающих сценарии выполнения этапов ЖЦ документа.
3. Определение зависимости показателя затрат каждой примитивной команды в сценарии от предшествующей.
4. Расчет показателей затрат сценариев с учетом указанной зависимости.
5. Определение на множестве сценариев выполнения этапа ЖЦ документа сценария, характеризующегося минимальным показателем затрат.

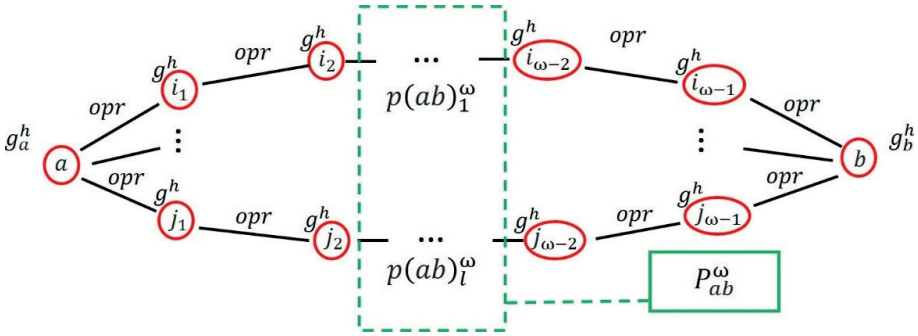


Рис. 1. Схематическое представление порядка формирования пути и множества путей в ДКА.

$t(p)$ – показатель затрат произвольного пути p, выраженный в числе тактов;

$t_y(p)$ – показатель затрат проверки условий срабатывания всех примитивных команд произвольного пути p, выраженный в числе тактов.

На *формальном уровне* постановка задачи имеет вид:

Дано: множество путей P в диаграмме переходов ДКА, описывающего применяемую МБ. Определить множество допустимых решений: множество путей ведущих из начального состояния в допускающее $P_{\alpha\beta}$. На множестве допустимых решений найти: путь p_{min}, характеризующийся минимальными затратами, то есть:

$$f_p(P,f_t(P)) = p_{min} \mid f_t(P) = \min_{p_{min} \in P_{\alpha\beta}} f_t(p_{min}), \quad (3)$$

где $f_p(P,f_t(P)) = p_{min}$ – функция, определяющая на множестве путей, путь – p_{min}, характеризующийся минимальными затратами, $f_t(P) = t(p)$ – функция, ставящая в соответствие пути показатель его затрат.

Методика разработки минимальных сценариев выполнения этапов жизненного цикла документа

Для описания методики введем определения.

Определение 1. Если индексы a, i_1, i_2, ..., $i_{\omega-1}$, b в пути $p(ab)_l^{\omega}$ различны (рис. 1), то его называют элементарным. Путь не являющийся элементарным называется избыточным[6]. Множество элементарных путей представим условной суммой (2) и обозначим – $P'^{\omega}_{ab} \subseteq P^{\omega}_{ab}$, для ДКА λ_h матрицу переходов элементарных путей высшего порядка ω, элементами которой на пересечении строки a и столбца b является P'^{ω}_{ab}, обозначим $[\overline{\lambda'_h}]^{\omega}$.

Исходя из допущений 4, 5 и 6, следует, что при удалении из пути некоторого контура показатель затрат будет снижен по меньшей мере на число примитивных операндов, выполняемых в указанном контуре. Таким образом, для определения минимального пути в ДКА в указанных условиях целесообразно учитывать только элементарные пути. Далее в исследовании при рассмотрении любого пути подразумевается, что он элементарный, т.е. $p(ab)_l^{\omega} \in P'^{\omega}_{ab}$.

Определение 2. Для каждой пары состояний g_a^h, g_b^h существует по меньшей мере один путь $p(ab)_l^{\omega} \in P'^{\omega}_{ab}$ некоторой длины ω такой, что $p(ab)_l^{\omega} \preceq t(p):\forall p \in P'^{\omega'}_{ab}$, $\omega' = \overline{1,n_{G_h}-1}$. Такой путь назовем минимальным с точки зрения затрат (далее – минимальный путь) – p_{min}.

Схема методики разработки минимальных сценариев выполнения этапов ЖЦ документа представлена на рисунке (рис. 2). Остановимся на каждом ее шаге более подробно.

На *первом шаге* инициализируем множество ДКА Λ_U, реализующих все этапы ЖЦ документа в качестве исходных данных методики (блоки 1-2 рис. 2).

Для ДКА каждого этапа $h = \overline{1,n_U}$ ЖЦ документа устанавливаются начальные – нулевые значения минимального пути $f_p(P,f_t(P)) = 0$ и его показателя затрат $f_t(p_{min}) = 0$. Также зададим нулевое значение показателю затрат проверки условий срабатывания примитивных команд первого пути, рассматриваемого в качестве претендента на «минимальность». Такой путь обозначим p_1, т.е. $t_Y(p_1) = 0$ (блоки 3-4 рис. 2).

Затем с применением известных подходов строим матрицу переходов элементарных путей первого порядка, отражающую функцию переходов δ_h рассматриваемого ДКА $\lambda_h \in \Lambda_U$, применяя в качестве дуг диаграммы переходов примитивные команды $opr_v \in OPR$[7].

Прежде всего представим функцию переходов δ_h в виде матрицы переходов (блок 5 рис. 2). В данном случае каждый элемент π_{ab} матрицы $[\lambda_h]$ определяется следующим образом:

$$\pi_{ab} = \begin{cases} opr_{ab}: opr_{ab} \in OPR, \text{если } opr_{ab} \text{ существует;} \\ 0, \text{если } opr_{ab} \text{ не существует.} \end{cases} \quad (4)$$

Преобразуем матрицу переходов в матрицу переходов элементарных путей первого порядка – $f_{\text{ЭП}}([\lambda_h]) = [\overline{\lambda'_h}]$ (блок 6 рис. 2). Исходя из формулы (1) и определения 1, каждый элемент матрицы $[\overline{\lambda'_h}]$ определим путем замены π_{ab} на P'^{1}_{ab} и обнуления диагональных элементов:

$$P'^{1}_{ab} = \begin{cases} \pi_{ab}, \text{если } a \neq b; \\ 0, \text{если } oa = b. \end{cases} \quad (5)$$

Затем для каждого значения порядка $\omega > 1$, вычисляем очередную матрицу переходов элементарных путей (блоки 7-8 рис. 2) по формуле:

$$[\overline{\lambda'_h}]^{\omega} = [\overline{\lambda'_h}]\,[\overline{\lambda'_h}]^{\omega-1}. \quad (6)$$

Данная операция выполняется до тех пока элемент матрицы, отражающий путь из начального состояния в допускающее не примет значение отличное от нуля (блок 9 рис. 2) и повторяется в последующем по мере работы методики до выполнения условий ее останова.

Элементы P'^{ω}_{ab} матриц $[\overline{\lambda'_h}]^{\omega}$ по мере возрастания их порядка $\omega = \overline{2,|G_h|-1}$ последовательно направляются для дальнейшей обработки.

На *втором шаге* для каждого пути из формулы (2) поступившего элемента $p(ab)_l^{\omega} \in P'^{\omega}_{ab}$ (блок 10 рис. 2) осуществляется расчет показателя затрат – $t(p(ab)_l^{\omega})$.

Исходя из допущения 5, каждому примитивному операнду соответствует заранее определенный набор условий его срабатывания, т.е. имеет место функция – $f_Z(Z) = 2^Y$.

Принимая затраты примитивного операнда и проверки каждого условия его срабатывания за один такт согласно допущениям 4 и 5, показатель затрат отдельно взятой примитивной команды определяется по формуле:

$$t(opr_v) = |f_Z(z_v)| + 1. \quad (7)$$

6 Там же. С. 45.

7 Hohn F. E., Seshu S., Aufenkamp D. D. OP. cit. P. 157.

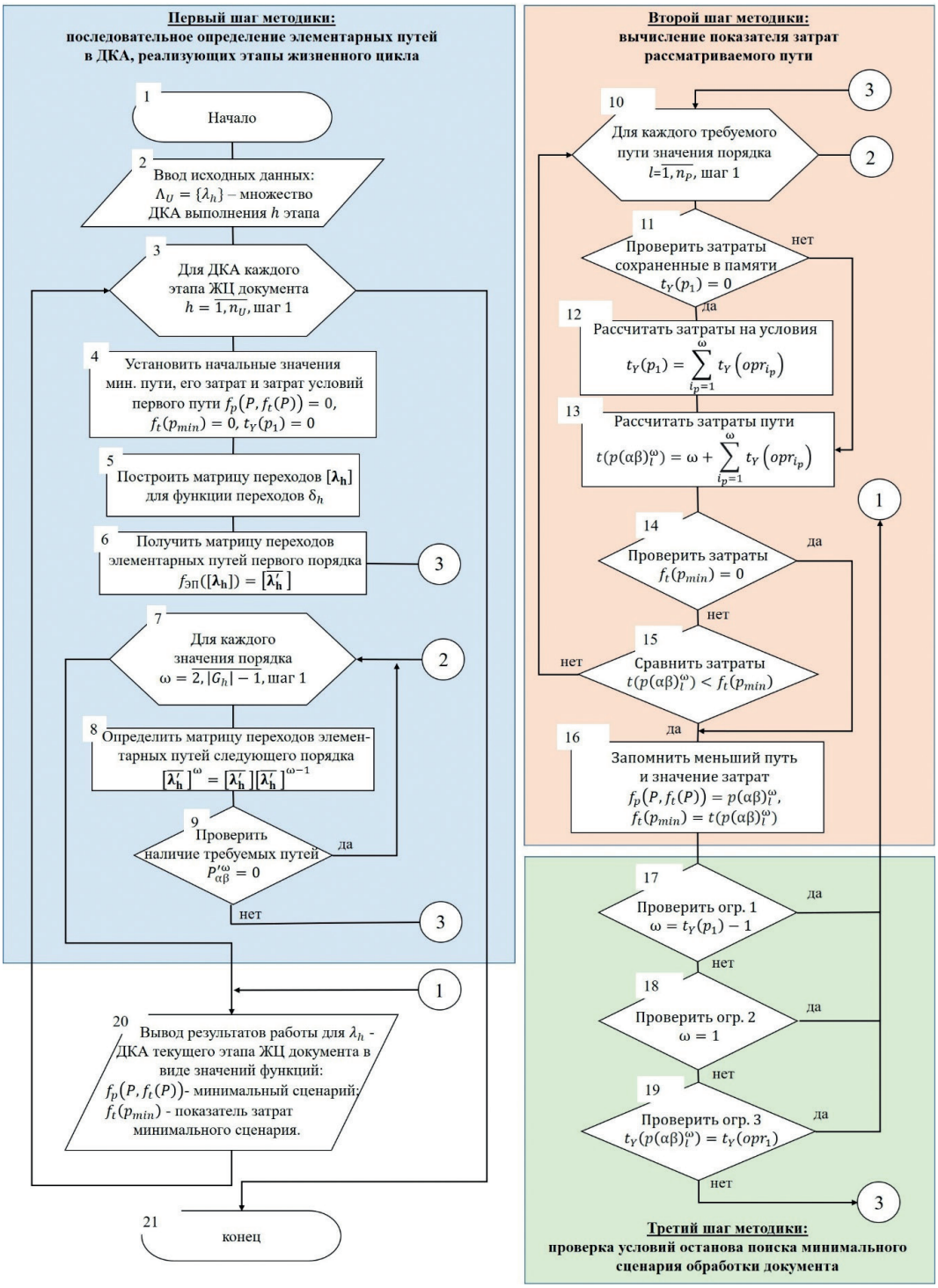


Рис. 2. Схема методики разработки минимальных сценариев выполнения этапов ЖЦ документа

Пронумеруем каждую примитивную команду $opr_v \in p(ab)_l^{\omega}$, как opr_{i_p}, где $i_p = \overline{1,\omega}$ – порядковый номер примитивной команды в пути. Тогда с учетом допущения 6 показатель затрат на проверку условий ее срабатывания будет определяться по формуле:

$$t_Y(opr_{i_p}) = \begin{cases} |f_Z(z_{i_p})|, \text{ если } Y_{i_p} \neq Y_{i_{p-1}}; \\ 0, \text{ если } Y_{i_p} = Y_{i_{p-1}}, \end{cases} \quad (8)$$

где z_{i_p} – примитивный операнд содержащийся в примитивной команде opr_{i_p}, $Y_{i_p} \subseteq Y$ – множество условий срабатывания примитивной команды opr_{i_p}, $Y_{i_{p-1}} \subseteq Y$ – множество условий срабатывания примитивной команды $opr_{i_{p-1}}$. Для наглядности порядок определения значений функции $t_Y(opr_{i_p})$ представлен на рисунке (рис. 3).

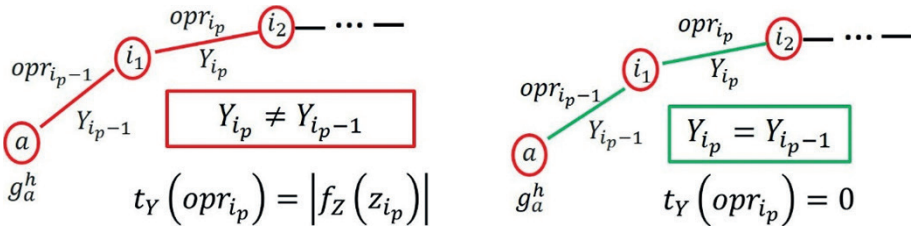


Рис. 3. Порядок определения затрат проверки условий срабатывания примитивной команды в пути

Для первого поступившего пути следует рассчитать и запомнить затраты на проверку условий срабатывания всех примитивных команд (блоки 11, 12 рис. 2), что в дальнейшем потребуется для проверки условий останова. Исходя из (8), данный показатель рассчитывается по формуле:

$$t_Y(p_1) = \sum_{i_p=1}^{\omega} t_Y(opr_{i_p}). \quad (9)$$

С учетом (7) и (9) показатель затрат пути $p(\alpha\beta)_l^{\omega}$ рассчитывается по формуле по формуле (блок 13 рис. 2):

$$p(\alpha\beta)_l^{\omega} = \omega + \sum_{i_p=1}^{\omega} t_Y(opr_{i_p}). \quad (10)$$

После чего первый поступивший путь и его показатель затрат запоминается как минимальный (блоки 14, 16 рис. 2) для чего выполним функции: f_p, f_t, принимая первый путь, как удовлетворяющий условию «минимальности».

Показатели последующих поступающих путей сравниваются с сохраненными в памяти (блок 15 рис. 2). Если выражение $t(p(\alpha\beta)_l^{\omega}) < f_t(p_{min})$ верно, то текущий путь и его показатель затрат запоминается согласно f_p и f_t запоминаются как минимальные.

На *третьем шаге* методики проверяется выполнение условий останова ее работы (блоки 17, 18, 19 рис. 2).

Результатом работы методики являются значения функций $f_p(P,f_t(P))$ и $f_t(p_{min})$, отражающих минимальные сценарии реализации каждого этапа ЖЦ документа и их показатели затрат соответственно (блоки 20-21 рис. 2).

Обоснование условий останова поиска минимального сценария

Каждое условие останова работы методики задается соответствующим ограничением. Рассмотрим каждое из них более подробно.

Ограничение 1.

Если существует путь $p(ab)_l^{\omega}$, то не существует менее затратного пути из состояния g_a^h в состояние g_b^h, длиннее исходного более чем на $\sum_{i_p=1}^{\omega} t_Y(opr_{i_p})$.

Доказательство.

1. Пусть задан путь, имеющий большую длину, но характеризующийся меньшим показателем затрат, т.е. $t(p(ab)_l^{\omega}) > t(p(ab)_l^{\gamma})) : \omega < \gamma$.
2. Тогда с учетом формулы (10) данное выражение будет иметь вид:

$$\omega + \sum_{i_p=1}^{\omega} t_Y(opr_{i_p}) > \gamma + \sum_{i_\gamma=1}^{\gamma} t_Y(opr_{i_\gamma}),$$

где i_p и i_γ – порядковые номера примитивных команд в соответствующих рассматриваемых путях.

3. Представим $\gamma = \omega + \iota$, где ι – величина, на которую длина пути γ превышает ω. Для удобства записи обозначим

$$\sum_{i_p=1}^{\omega} t_Y(opr_{i_p}) = t_Y(p_\omega) \text{ и } \sum_{i_\gamma=1}^{\gamma} t_Y(opr_{i_\gamma}) = t_Y(p_\gamma).$$

Тогда выражение будет иметь вид:

$$\omega + t_Y(p_\omega) > \omega + \iota + t_Y(p_\gamma).$$

Следовательно,

$$t_Y(p_\omega) - t_Y(p_\gamma) > \iota. \quad (11)$$

4. Согласно допущению 5 в крайнем случае выражение (11) будет иметь вид: $t_Y(p_\omega) - 1 > \iota$, т.е. $\sum_{i_p=1}^{\omega} t_Y(opr_{i_p}) - 1 > \iota$. *Ограничение* 1 *доказано*.

Ограничение 2.

Если существует путь $p(ab)_l^{\omega}$, длина которого $\omega = 1$, то он минимальный, т.е. $p(ab)_l^{\omega} = p_{min} : \omega = 1$.

Доказательство.

1. Допустим $p(ab)_l^{\omega} \neq p_{min}$, тогда существует иной минимальный путь длины k, отличный от существующего, т.е. $p(ab)_j^k = p_{min} : p(ab)_j^k \neq p(ab)_l^{\omega}$, $t(p(ab)_j^k) < t(p(ab)_l^{\omega})$.
2. Исходя из допущений 3 и 4, можно сделать вывод, что допускающее состояние отличается от начального единственным атрибутом $x_q \in X$, изменяемым примитивной командой $opr_{ab} \in OPR$.
3. Тогда либо каждый из этих путей состоит из данной примитивной команды и они совпадут $p(ab)_j^k = p(ab)_l^{\omega} = opr_{ab}$, либо путь $p(ab)_j^k$ будет иметь в своем составе примитивные команды создания и последующего удаления атрибутов, не входящих в допускающее состояние.
4. Таким образом, исходя из допущения 3 и формулы (1), такой путь имеет по меньшей мере одну пару совпадающих индексов примитивных команд в своем составе, т.е. по определению 1 является избыточным – $p(ab)_j^k \notin P'^{\omega}_{ab}$.
5. Но по определению 2 – $p_{min} \in P'^{\omega}_{ab}$, следовательно $p(ab)_j^k \neq p_{min}$. *Ограничение* 2 *доказано*.

Ограничение 3.

Если существует кратчайший путь $p(ab)_l^{\omega}$ длины $\omega > 1$ условия срабатывания примитивных команд в котором проверяются только в первой из них, то он минимальный, т.е. $p(ab)_l^{\omega} = p_{min} : t_Y(p(ab)_l^{\omega}) = t_Y(opr_1)$, $\nexists p(ab)_j^k \mid k < \omega$.

Доказательство.

1. Допустим $p(ab)_l^{\omega} \neq p_{min}$, тогда существует иной минимальный путь длины $k \geq \omega$, отличный от существующего, т.е. $p(ab)_j^k = p_{min} : p(ab)_j^k \neq p(ab)_l^{\omega}$, $t(p(ab)_j^k) < t(p(ab)_l^{\omega})$, $k \geq \omega$.
2. Исходя из допущений 3 и 4, можно сделать вывод, что при прохождении пути $p(ab)_l^{\omega}$ изменению подвергается некоторый набор атрибутов $X_b \in X$, каждый из которых изменяется соответствующими примитивными командами $opr_{i_p} \in p(ab)_l^{\omega}$. Т.к. $p(ab)_l^{\omega}$ кратчайший, то в нем не имеется изменений атрибутов, отличных от $x_q \in X_b$ в «промежуточных» состояниях и каждое изменение носит обязательный характер.
3. Тогда: либо $p(ab)_j^k$ состоит из набора примитивных команд, идентичного набору пути $p(ab)_l^{\omega}$, расположенных в ином порядке, что приводит согласно формулы (9) к ложности утверждения $t(p(ab)_j^k) < t(p(ab)_l^{\omega})$; либо путь $p(ab)_j^k$ будет иметь в своем составе примитивные команды создания и последующего удаления атрибутов, не входящих X_b. В данном случае аналогично доказательству ограничения 2 приходим к ложности утверждения $p(ab)_j^k = p_{min}$. *Ограничение* 3 *доказано*.

Вывод

В статье представлена новая методика разработки минимальных сценариев выполнения этапов ЖЦ электронного документа ограниченного доступа с учетом применяемой в системе ПБ и изменчивости показателя затрат каждого отдельно взятого запроса в сценарии.

Обоснованные в работе условия останова функционирования методики обеспечивают минимальность выбранного сценария, исключая необходимость полного перебора множества допустимых решений.

Применение разработанных таким образом сценариев в перспективных МБ позволит: с одной стороны, сократить затраты вычислительного ресурса на обработку документа, с другой стороны – не допустить обработку запросов, нарушающих ЖЦ документа.

Достоверность предлагаемых научных решений подтверждается представленным в статье комплексом строгих доказательств.

Литература

1. Носенко С. В., Королев И. Д., Поддубный М. И. О единой системе электронного документооборота // Военная мысль. 2019. № 3. С. 90–97.
2. Колесник А. В., Кошелев А. В., Поддубный М. И., Васильев В. Д. Актуальность задачи создания единого мультисервисного межведомственного цифрового пространства с повышенным уровнем обеспечения безопасности связи и информации // Состояние и перспективы развития современной науки по направлению «Информационная безопасность». Сборник статей III Всероссийской научно-технической конференции. Анапа, 2021. С. 96–114.
3. Марков А. С. Современные тенденции безопасных информационных технологий // Безопасные информационные технологии. Сборник трудов Двенадцатой международной научно-технической конференции. Москва, 2023. С. 5–10.
4. Зегжда П. Д., Зегжда Д. П., Анисимов В. Г., Анисимов Е. Г., Сауренко Т. Н. Модель формирования программы развития системы обеспечения информационной безопасности организации // Проблемы информационной безопасности. Компьютерные системы. 2021. № 2. С. 109–117.
5. Девянин П. Н. О разработке проекта национального стандарта ГОСТ Р «Защита информации. Формальная модель управления доступом. Часть 3. Рекомендации по разработке» // Труды Института системного программирования РАН. 2024. Т. 36. № 3. С. 63–82. DOI: 10.15514/ISPRAS-2024-36(3)-5.
6. Поддубный М. И. Новый подход к построению моделей безопасности систем электронного документооборота // Инженерный вестник Дона. 2023. № 2 (98). С. 235–245.
7. Поддубный М. И. Разработка концептуальных основ обеспечения безопасности обработки и хранения электронных документов в системе электронного документооборота Вооруженных Сил Российской Федерации // Состояние и перспективы развития современной науки по направлению «IT-технологии». Сборник трудов II Всероссийской научно-технической конференции. Анапа, 2023. С. 266–278.
8. Девянин П. Н., Тележников В. Ю., Хорошилов А. В. Формирование методологии разработки безопасного системного программного обеспечения на примере операционных систем // Труды Института системного программирования РАН. 2021. Т. 33. № 5. С. 25–40. DOI: 10.15514/ISPRAS-2021-33(5)-2.
9. Максимовский А. Ю. О выборе параметров автоматных моделей мониторинга информационной безопасности сетевых объектов // Информация и безопасность. 2020. Т. 23. № 1. С. 31–40.
10. Максимовский А. Ю. О выборе параметров автоматных моделей мониторинга информационной безопасности сетевых объектов (часть 2) // Информация и безопасность. 2020. Т. 23. № 3. С. 327–336.
11. Кузнецова А. Л., Афонин С. А. Автоматная модель проверки корректности атрибутной политики информационной безопасности в системах с конечным числом объектов // Вестник Московского университета. Серия 1: Математика. Механика. 2021. № 5. С. 57–60.
12. Васильева Н. Б. Обзор алгоритмов поиска кратчайших путей в графах // Экспериментальные и теоретические исследования в современной науке. сборник статей по материалам XCVII международной научно-практической конференции. Новосибирск, 2024. С. 17–21.

13. Ходулина Е. А., Шатовкин Р. Р. Анализ алгоритмов поиска пути минимальной стоимости в графе // Радиоэлектроника. Проблемы и перспективы развития. Сборник трудов IX Всероссийской научно-практической конференции с международным участием. Тамбов, 2024. С. 13–15.
14. Азимов Р. Ш., Григорьев С. В. Алгоритм поиска всех путей в графе с заданными контекстно-свободными ограничениями с использованием матриц с множествами промежуточных вершин // Научно-технический вестник информационных технологий, механики и оптики. 2021. Т. 21. № 4. С. 499–505. DOI: 10.17586/2226-1494-2021-21-4-499-505.
15. Кузнецов А. Л. Матричный метод поиска путей на взвешенных ориентированных графах в задачах сетевого планирования при проектировании и эксплуатации морских портов // Вестник государственного университета морского и речного флота им. адмирала С. О. Макарова. 2020. Т. 12. № 2. С. 230–238. DOI: 10.21821/2309-5180-2020-12-2-230-238.
16. Vatutin E. I., Panishchev V. S., Gvozdeva S. N., Titov V. S. Comparison of Decisions Quality of Heuristic Methods Based on Modifying Operations in the Graph Shortest Path Problem // Problems of Information Technology. 2020. № 1. С. 3–15. DOI: 10.25045/jpit.v11.i1.01.

DEVELOPING METHOD OF MINIMUM SCENARIOS OF ELECTRONIC DOCUMENT LIFESPAN STAGES IN RESTRICTED ACCESS

Poddubniy M. I.[8]

Keywords: *finite state machine, computing resource costs, computer system, finding a path in a finite state machine, security policy, document processing scenario, access management in a computer system.*

Relevance: *the features of processing restricted electronic documents in computer systems actualize the issues of formation of minimum scenarios on each stage of the document lifespan fulfilment. Known algorithms in searching for such scenarios do not take into account the variability of the value of indicators of computing spending resource of a single request in the scenario applied by the security policy and cannot be applied.*

The purpose of the study: *is to develop a methodology for developing minimum scenarios of implementation of the stages of the life cycle of an electronic document with limited access processed by the computer system.*

Methods used: *These scenarios are proposed to be considered as ways in the transition diagram of a finite state machine describing the implemented security policy in the computer system. As the weight of edges processing each atomic request in scenario is taken as weight, which allows to apply approaches in building and processing of transformation matrix transition of high-order finite state machine and search for ways in it, based on the works of F. Hohn, S. Seshu, D. Aufenkamp, A. Gill.*

The novelty value: *is the order of calculation of the computing spending resource, taking into account the dependence of successive requests in the scenario between them. The novelty elements should also include the conditions described and justified in the work of stopping the search for minimal scenarios in the considered finite state machine.*

Result: *the developed methodology allows to determine the execution scenario of the lifespan stage of a document processed by the computer system, characterized by minimum computing resources, taking into account the applied security policy, avoiding the method of crude force. The use of such scenarios as a response to a request user of a restricted document processing computer system would eliminate the possibility of the document lifespan failure and minimize the attempts of its processing.*

References

1. Nosenko S. V., Korolev I. D., Poddubnii M. I. O yedinoi sisteme elektronnogo dokumentooborota [About the Unified Electronic Document Management System]. Voennaya misl [Military Thought], 2019, no. 3, pp. 90–97 (in Russian).
2. Kolesnik A. V., Koshelev A. V., Poddubnii M. I., Vasilev V. D. Aktualnost zadachi sozdaniya yedinogo multiservisnogo mezhvedomstvennogo tsifrovogo prostranstva s povishennim urovnem obespecheniya bezopasnosti svyazi i informatsii [The Relevance of the Task of Creating a Single Multiservice Interdepartmental Digital Space With an Increased Ievel of Communication and Information Security]. Sostoyanie i perspektivi razvitiya sovremennoi nauki po napravleniyu «Informatsionnaya bezopasnost». Sbornik statei III Vserossiiskoi nauchno-tekhnicheskoi konferentsii [The State and Prospects of Development of Modern Science in the Field of «Information Security». Collection of Articles of the 3rd All-Russian Scientific and Technical Conference], Anapa, 2021, pp. 96–114 (in Russian).
3. Markov A. S. Sovremennie tendentsii bezopasnikh informatsionnikh tekhnologii [Secure Information Technologies Modern Trends]. Bezopasnie informatsionnie tekhnologii. Sbornik trudov Dvenadtsatoi mezhdunarodnoi nauchno-tekhnicheskoi konferentsii [Secure information technologies. Proceedings of the Twelfth International Scientific and Technical Conference], Moskva, 2023, pp. 5–10 (in Russian).

8 Maxim I. Poddubniy, Ph.D. of Engineering Sciences, doctoral student. Federal State Treasury military educational institution of higher education «Krasnodar High Military Orders of Zhukov and October Revolution Red Banner School named after the army general S.M.Shtemenko» of the Ministry of Defense of the Russian Federation. City of Krasnodar, Russia E–mail: podd.maxim@yandex.ru

4. Zegzhda P. D., Zegzhda D. P., Anisimov V. G., Anisimov Ye. G., Saurenko T. N. Model formirovaniya programmi razvitiya sistemi obespecheniya informatsionnoi bezopasnosti organizatsii [Model for Forming Development Program of Organization's Information Security System]. Problemi informatsionnoi bezopasnosti. Kompyuternie sistemi [Information Security Problems. Computer Systems], 2021, no. 2, pp. 109–117 (in Russian).
5. Devyanin P. N. O razrabotke proekta natsionalnogo standarta GOST R «Zashchita informatsii. Formalnaya model upravleniya dostupom. Chast 3. Rekomendatsii po razrabotke» [on the Development of the Draft Standard Gost R «Information Protection. Formal Access Control Model. Part 3. Recommendations on Development»]. Trudi Instituta sistemnogo programmirovaniya RAN [Proceedings of the Institute for System Programming of the RAS], 2024, vol. 36, no. 3, pp. 63–82 (in Russian), DOI: 10.15514/ISPRAS-2024-36(3)-5.
6. Poddubnii M. I. Novii podkhod k postroeniyu modelei bezopasnosti sistem elektronnogo dokumentooborota [A New Approach to Building Security Models for Electronic Document Management Systems]. Inzhenernii vestnik Dona [Engineering journal of Don], 2023, no. 2 (98), pp. 235–245 (in Russian).
7. Poddubnii M. I. Razrabotka kontseptualnikh osnov obespecheniya bezopasnosti obrabotki i khraneniya elektronnikh dokumentov v sisteme elektronnogo dokumentooborota Vooruzhennikh Sil Rossiiskoi Federatsii [Development of a Conceptual Framework for Ensuring the Security of Electronic Document Processing and Storage in the Electronic Document Management System of the Armed Forces of the Russian Federation]. Sostoyanie i perspektivi razvitiya sovremennoi nauki po napravleniyu «IT-tekhnologii»: Sbornik trudov II Vserossiiskoi nauchno-tekhnicheskoi konferentsii. [The State and Prospects of Development of Modern Science in the Field of « IT technologies». Collection of Articles of the 2rd All-Russian Scientific and Technical Conference], 2023, pp. 266–278 (in Russian).
8. Devyanin P. N., Telezhnikov V. Yu., Khoroshilov A. V. Formirovanie metodologii razrabotki bezopasnogo sistemnogo programmnogo obespecheniya na primere operatsionnikh sistem [Building a Methodology for Secure System Software Development on the Example of Operating Systems]. Trudi Instituta sistemnogo programmirovaniya RAN [Proceedings of the Institute for System Programming of the RAS], 2021, vol. 33, no. 5, pp. 25–40 (in Russian), DOI: 10.15514/ISPRAS-2021-33(5)-2.
9. Maksimovskii A. Yu. O vibore parametrov avtomatnikh modelei monitoringa informatsionnoi bezopasnosti setevikh obektov [About Parameters of Automated Models for Monitoring Information Security of Network Objects]. Informatsiya i bezopasnost [Information and Security], 2020, vol. 23, no. 1, pp. 31–40 (in Russian).
10. Maksimovskii A. Yu. O vibore parametrov avtomatnikh modelei monitoringa informatsionnoi bezopasnosti setevikh obektov (chast 2) [About Parameters of Automated Models for Monitoring Information Security of Network Objects (Part 2)]. Informatsiya i bezopasnost [Information and Security], 2020, vol. 23, no. 3, pp. 327–336 (in Russian).
11. Kuznetsova A. L., Afonin S. A. Avtomatnaya model proverki korrektnosti atributnoi politiki informatsionnoi bezopasnosti v sistemakh s konechnim chislom obektov [Automata Model for Verifying Attibuted-Based Access Control Policy in Systems With a Finite Number of Objects]. Vestnik Moskovskogo universiteta. Seriya 1: Matematika. Mekhanika [Bulletin of the Moscow University. Series 1: Mathematics. Mechanics], 2021, no. 5, pp. 57-60 (in Russian).
12. Vasileva N. B. Obzor algoritmov poiska kratchaishikh putei v grafakh [Review of the Fiding Shortest Paths in Graphs Algorithms]. Eksperimentalnie i teoreticheskie issledovaniya v sovremennoi nauke. sbornik statei po materialam XCVII mezhdunarodnoi nauchno-prakticheskoi konferentsii [Experimental and Theoretical Research in Modern Science. Collection of Articles Based on the Materials of the XCVII International Scientific and Practical Conference], Novosibirsk, 2024, pp. 17–21 (in Russian).
13. Khodulina Ye. A., Shatovkin R. R. Analiz algoritmov poiska puti minimalnoi stoimosti v grafe [Analysis of Algorithms for Finding the Minimum Cost Path in a Graph]. Radioelektronika. Problemi i perspektivi razvitiya. Sbornik trudov IX Vserossiiskoi nauchno-prakticheskoi konferentsii s mezhdunarodnim uchastiem. [Proceedings of the IX All-Russian Scientific and Practical Conference With International Participation], Tambov, 2024, pp. 13–15 (in Russian).
14. Azimov R. Sh., Grigorev S. V. Algoritm poiska vsekh putei v grafe s zadannimi kontekstno-svobodnimi ogranicheniyami s ispolzovaniem matrits s mnozhestvami promezhutochnikh vershin [Context-Free Path Querying with All-Path Semantics Using Matrices with Sets of Intermediate Vertices]. Nauchno-tekhnicheskii vestnik informatsionnikh tekhnologii, mekhaniki i optiki [Scientific and Technical Journal of Information Technologies, Mechanics and Optics], 2021, vol. 21, no. 4, pp. 499–505 (in Russian), DOI: 10.17586/2226-1494-2021-21-4-499-505.
15. Kuznetsov A. L. Matrichnii metod poiska putei na vzveshennikh orientirovannikh grafakh v zadachakh setevogo planirovaniya pri proektirovanii i ekspluatatsii morskikh portov [Matrix Method for Finding the Paths on Weighted Oriented Graphs in the Tasks of Port Net Operational Planning]. Vestnik gosudarstvennogo universiteta morskogo i rechnogo flota im. admirala S. O. Makarova, 2020, vol. 12, no. 2, pp. 230–238 (in Russian), DOI: 10.21821/2309-5180-2020-12-2-230-238.
16. Vatutin E. I., Panishchev V. S., Gvozdeva S. N., Titov V. S. Comparison of Decisions Quality of Heuristic Methods Based on Modifying Operations in the Graph Shortest Path Problem. Problems of Information Technology, 2020, no. 1, pp. 3–15 (in English), DOI: 10.25045/jpit.v11.i1.01.

КОМПЛЕКС МЕТОДОВ ГЕНЕТИЧЕСКОЙ ДЕЭВОЛЮЦИИ ПРЕДСТАВЛЕНИЙ ПРОГРАММЫ

Израилов К. Е.[1]

DOI: 10.21681/2311-3456-2025-4-93-106

Цель исследования: *повышение эффективности нейтрализации уязвимостей программы за счет интеллектуализации ее реверс-инжиниринга с помощью генетических алгоритмов*

Методы исследования: *системный анализ и методы оптимизации, теория графов, функциональный и структурный синтез, общая методология программирования и теория компиляторов.*

Полученные результаты: *синтезирован иерархический трехуровневый комплекс методов, состоящий из метода генетического реверс-инжиниринга программы, метода генетической деэволюции ее соседних представлений (машинного и исходного кода, алгоритмов, архитектуры и т.д.), и группы методов для реализации основополагающих операций генетических алгоритмов.*

Новизна *комплекса методов заключается в их ориентированности на решение задачи реверс-инжиниринга путем прямых преобразований программы в последующие представления, что отличает их от классических, выполняющих обратные преобразования. Также, алгоритмы группы методов комплекса основаны на работе с оригинальной моделью исходного кода, представляющей его как последовательность ген.*

Ключевые слова: *нейтрализация уязвимостей, реверс-инжиниринг, искусственный интеллект, генетические алгоритмы, комплекс методов.*

Введение

Наличие уязвимостей в программном обеспечении является актуальной проблемой современного информационно-технологического мира [1]. Одним из путей разрешения данной проблемы считается непосредственное обнаружение и устранение уязвимостей в различных представлениях программы [2] – машинном, исходном и байткоде, алгоритмах, архитектуре и др. При этом, для достижения высокой эффективности такой нейтрализации уязвимостей необходимо предварительное получение представлений программы, в которых уязвимости были внедрены; каждое же такое преставление после исправления должно преобразовываться в конечную программу. В интересах этого требуется создание нового подхода к реверс-инжинирингу программ, поскольку существующие не удовлетворяют указанным условиям [3] – часть подходов не позволяет работать со всем набором предыдущих представлений программы, а другая часть получает псевдо-представления, не преобразуемые в конечное; так, например, в описании коммерческого продукта для анализа машинного кода IDA Pro изначально заявлено, что он восстанавливает псевдо-исходный код, который в принципе не обязан быть даже компилируемым, не говоря уже о тождественности машинному. Предлагаемый автором подход основан на решении оптимизационной задачи по подбору конструкций предыдущего представления для его полного соответствия заданному, что как раз и позволяет получать высокоуровневые представления, подходящие для поиска уязвимостей, которые затем могут быть «собраны» (например, компиляцией) в выполняемую программу с функциональностью, идентичной первоначальной; такой процесс обратных преобразований между соседними представлениями назван их деэволюцией. Исходя из того, что решение указанной оптимизационной задачи основано на применении генетических алгоритмов (точнее их модифицированной версии), как деэволюция представлений, так и полная их цепочка – т. е. реверс-инжиниринг, были названы генетическими (сокр. ГДЭ и ГРИ, соответственно); при этом, частный случай получения исходного кода программы по ее машинному коду логично был назван генетической декомпиляцией (сокр. ГДК). Проведенные эксперименты показали как работоспособность данной концепции реверс-инжиниринга, так и ее практическую реализуемость и превосходство над аналогами. Описанию же самого подхода ГРИ в виде комплекса методов и посвящена данная статья.

Генетический подход

Прежде чем перейти к описанию предложенного подхода (в виде комплекса методов), рассмотрим основные исторические предпосылки, лежащие в его основе. Также в качестве примеров применения подхода выберем декомпиляцию машинного кода в исходный, как крайне востребованную на сегодняшний день задачу реверс-инжиниринга.

1 Израилов Константин Евгеньевич, кандидат технических наук, доцент, профессор кафедры прикладной математики и безопасности информационных технологий Санкт-Петербургского университета государственной противопожарной службы МЧС России, Санкт-Петербург, ORCHID: http://orcid.org/0000-0002-9412-5693. Scopus Author ID: 56122749800. E-mail: konstantin.izrailov@mail.ru

Существуют различные подходы к проведению реверс-инжиниринга, которые с точки зрения их исторического появления можно упорядочить следующим образом: ручной, алгоритмический, полный перебор (как гипотетический, практически не применимый на практике) и интеллектуальный. При этом у каждого их них есть свои сильные и слабые стороны, которые не позволяют обеспечить качественного повышения эффективности всего процесса нейтрализации уязвимостей.

Рассмотрим более подробно подход, использующий искусственный интеллект (далее – ИИ), поскольку он является современным трендом информационных технологий, востребованным в огромном количестве областей [4].

ИИ активно исследуется и используется при решении относительно частных задач области реверс-инжиниринга, а именно, следующих: глубокое машинное обучение для анализа машинного кода (декомпиляция, восстановление метаданных) [5], декомпиляция машинного кода с применением малых и больших языковых моделей [6, 7], именование функций машинного кода с поддержкой оптимизаций компилятора на базе графовых нейронных сетей [8], восстановление отладочной информации (имен и типов переменных) с использованием машинного обучения [9], идентификация функций в машинном коде на основе рекуррентных нейронных сетей [10]. Впрочем, использование ИИ для деэволюции представлений программы, как правило, ведет к ряду существенных недостатков, таких, как необходимость в датасетах огромного размера, высокая вычислительная стоимость и ресурсоемкость, слабая интерпретируемость получаемых результатов (т. е. отсутствие ответа на вопрос, почему ИИ восстановил по машинному коду именно такой код программы), эффект «галлюцинирования» (когда создается выдуманная информация, не имеющая ничего общего с реальным кодом). Также, ИИ может ошибаться при декомпиляции программ со сложной логикой работы; например, следующий «запутанный» исходный код:

```
int funct(int p) {
  goto aaa2;

aaa1:;
  p++;
  goto aaa3;

aaa2:;
  goto aaa1;

aaa3:;
  return p;
}
```

производит увеличение аргумента «p» функции funct(...) на 1, возвращая тем самым всегда значение «p + 1»; соответствующий ему машинный код для архитектуры x86-64 является следующим:

```
push ebp
mov  ebp, esp

jmp  SHORT $aaa2$3

$aaa1$4:
jmp  SHORT $aaa3$5

$aaa2$3:
mov  eax, DWORD PTR _p$[ebp]
add  eax, 1
mov  DWORD PTR _p$[ebp], eax

jmp  SHORT $aaa1$4

$aaa3$5:
mov  eax, DWORD PTR _p$[ebp]

pop  ebp
ret  0
```

Запрос же на декомпиляцию приведенного машинного кода для генеративного ИИ с моделью GPT-4o mini привел к выдаче следующего результата: «Этот код реализует бесконечный цикл, который увеличивает значение переменной _p$ на 1. Так как после увеличения значения происходит переход обратно к метке $aaa1$4, которая снова переходит к $aaa3$5, где значение снова загружается, цикл не завершится, пока функция не будет принудительно остановлена». Таким образом, данная модель произвела некорректную декомпиляцию (а, точнее, не смогла ее произвести).

Аналогичный запрос для другого популярного генеративного ИИ с моделью DeepSeek-R1 выдал следующий исходный код:

```
int func(int *p) {
    return ++(*p);
}
```

что, конечно, близко к логике истинного исходного кода, но не соответствует ей, т. к. аргумент функции передается по указателю, а не по значению, а сама некорректно восстановленная функция тем самым обладает эффектом изменения внешних данных.

При этом остается «открытым» вопрос полной корректности восстанавливаемых с помощью ИИ представлений программы – т. е., например, тождественность полученного исходного кода декомпилируемому машинному коду.

Перейдем к описанию самого предложенного подхода ГДЭ, частично лишенного недостатков, свойственных существующим интеллектуальным.

Как указывалось ранее, суть подхода заключается в решении оптимизационной задачи по подбору конструкций предыдущего представления, которые бы были тождественны совокупности конструкций заданного представления; само же решение основано на генетических алгоритмах [11, 12]. Так, если требуется получить исходный код по машинному, то создается случайная популяция экземпляров исходного кода, которые компилируются в машинный и сравниваются с заданным. Из исходных кодов отбираются те, машинный код которых наиболее близок к заданному – производится их селекция, над которыми затем осуществляются операции скрещивания и мутации, получая новое поколение. Данный итеративный процесс завершится, когда будет получен исходный код, дающий после компиляции машинный код, полностью идентичный заданному. При этом существует достаточно большое количество исследований, посвященных оптимизации работы самих генетических алгоритмов [13, 14, 15].

Для работы ГДЭ необходимы синтаксисы соседних представлений программы, которые предварительно преобразуются в соответствующие графы синтаксических правил (далее – ГСП); при этом, сам код программы имеет запись в виде хромосомы, в которой каждый ген определяет выбор пути по данному ГСП. Сам выбор целесообразно делать только в тех узлах ГСП, в которых существуют различные пути продолжения синтаксических правил, т. е. в узлах-альтернативах; так, например, в математических выражениях после переменной может идти целый набор различных бинарных операций (что в ГСП является узлом-альтернативой), выбор каждой из которых и задается геном хромосомы-пути. Таким образом, по сравнению с классическим реверс-инжинирингом, преобразующим текущее представление программы в ее предыдущее, предлагаемый подход имеет противоположную направленность. Особенность подхода и его отличие от альтернативных может быть представлено следующим формальным образом.

Предположим, что представление программы является совокупностью двух неразделимых компонентов: формы и содержания, первый из которых определяет внешнее представление программы, а второй – ее внутреннюю суть или логику; тогда представление программы может быть записано следующим образом:

$$Rep_x = \langle Rep_x^{Form}, Rep_x^{Content} \rangle, \quad (1)$$

где Rep_x – x-ое представление программы, определяемое кортежем из его формы Rep_x^{Form} и содержания $Rep_x^{Content}$. Очевидно, что в процессе разработки логика программы должна сохраняться, а ее форма будет постепенно меняться или эволюционировать (точнее переходить от человеко-ориентированной в машинно-ориентированную), т. е.:

$$\begin{cases} Rep_x^{Form} \neq Rep_{x+1}^{Form} \\ Rep_x^{Content} = Rep_{x+1}^{Content} \end{cases}, \quad (2)$$

где x и $x + 1$ – индексы текущего и последующего представлений программы.

Тогда весь процесс программного инжиниринга может быть записан следующим образом:

$$\begin{cases} Rep_x \longrightarrow Rep_{x+1} \\ Rep_x^{Form} \longrightarrow Rep_{x+1}^{Form} \\ Rep_x^{Content} = Rep_{x+1}^{Content} \end{cases}, \quad (3)$$

где «⟶» – операция прямого преобразования представлений и их компонентов.

Процесс же реверс-инжиниринга в этом случае имеет следующую запись:

$$\begin{cases} Rep_{x-1} \longleftarrow Rep_x \\ Rep_{x-1}^{Form} \longleftarrow Rep_x^{Form} \\ Rep_{x-1}^{Content} = Rep_x^{Content} \end{cases}, \quad (4)$$

где «⟵» – операция обратного преобразования представлений и их компонентов.

Принцип действия классического реверс-инжиниринга (подходы, применяемые в котором уже были упомянуты), можно записать следующим образом:

$$Rep_{x-1}^{Form} = Reverse(Rep_x^{Form}), \quad (5)$$

где $Reverse(\ldots)$ – операция обратного ручного, алгоритмического или интеллектуального преобразования формы представления.

Отличие же подхода ГДЭ от классических в этом случае видно по следующей его записи:

$$\{Rep_{x-1}\}^{n+1} = GenAlgOpt(\{Rep_{x-1}\}^n) : Compile(Rep_{x-1}) = Rep_x, \quad (6)$$

где $\{\ldots\}^n$ и $\{\ldots\}^{n+1}$ – n-ая и $n+1$-ая популяции экземпляров программы в предыдущем представлении (например, исходного кода); $GenOpt(\ldots)$ – итерационная операция решения оптимизационной задачи с помощью генетического алгоритма: $Compile(\ldots)$ – операция получения из программы в предыдущем представлении (например, исходного кода) его текущего представления (например, машинного кода) прямым преобразованием (например, компиляцией).

Предлагаемый подход ГДЭ является некоторым развитием существующих и активно используемых – алгоритмического и интеллектуального (при этом, хотя ручной подход также применяется, однако его использование скорее связано с безысходностью по причине невозможности повеления реверс-инжиниринга остальными, поскольку он крайне время- и ресурсозатратен). Так, хотя алгоритмический подход дает гарантированный результат, тем не менее он основан на заложенных экспертом правилах,

которые не могут покрыть все множество необходимых преобразований машинного кода в исходный. В противоположность этому, хотя интеллектуальный подход способен избегать шаблонных конструкций машинного кода, однако получаемый с помощью него результат не всегда соответствует требуемому – возникают ошибки восстановления исходного кода, который не гарантированно соответствует машинному. ГДК (как частный случай ГДЭ) представляет собой комбинацию наиболее успешных возможностей данных подходов следующим образом. Во-первых, применение средств компиляции при решении оптимизационной задачи позволяет утверждать о тождественности полученного исходного кода заданному машинному – специфика алгоритмического подхода. А, во-вторых, вариативность в работе генетического алгоритма (поскольку, операции скрещивания и мутации построены в том числе на случайных выборах генов) позволяет расширять строгие правила преобразования внешаблонными решениями – специфика интеллектуального подхода.

Для обоснования работоспособности подхода ГДЭ, центрального во всем ГРИ, был создан программный прототип ГДК, производящий декомпиляцию различных экземпляров машинного кода для заданных синтаксисов. При реализации прототипа использовался язык программирования Python версии 3.11, авторский код модифицированного ядра генетических алгоритмов и его основных операций, а также вспомогательные библиотеки: collections, copy, enum, os, random, subprocess и др.

Комплекс методов генетической деэволюции представлений

Для работы всего ГРИ необходим целый комплекс соответствующих методов (далее – Комплекс), связанных иерархически – т. е. каждый вышестоящий метод при выполнении использует результаты работы нижестоящего (или их группу), как параметр. Так, на верхнем уровне расположен метод генетического реверс-инжиниринга программы (далее – Метод-ГРИ), который в процессе работы для обратного преобразования соседних представлений применяет метод их генетической деэволюции (далее – Метод-ГДЭ), который в свою очередь использует группу методов (далее – Группа методов), предназначенных для реализации основополагающих операций генетических алгоритмов, основанных на генетическом представлении особей популяции [16]: методы вычисления метрики близости (двух экземпляров программ в текущем представлении), а также скрещивания и мутации ген особей; пользователем Комплекса является эксперт по информационной безопасности и реверс-инжинирингу (далее – Эксперт). Блок-схемы и псевдокод методов с необходимыми пояснениями приводятся далее.

Метод генетического реверс-инжиниринга программы

Принцип работы Метода-ГРИ заключается в итерационном (т. е. последовательном однотипном) преобразовании представлений программы из текущего (как правило, конечного, т. е. выполняемого, такого, как машинный код) во все более высокоуровневые (например, в исходный код, алгоритмы, архитектуру) с нейтрализацией в них уязвимостей. Затем, требуется пересборка конечного представления программы, которое тем самым становится безопасным. Соответствующая блок-схема Метода-ГРИ приведена на рис. 1; здесь и далее белым фоном обозначены основные шаги метода, зеленым – ввод и вывод данных, желтым – специализированные конструкции (начало, конец, границы цикла), синим – вызовы внешних методов, оранжевым – условный переход, серым – хранилища данных; пунктирные линии обозначают связь по данным.

На блок-схеме (см. рис. 1) присутствуют следующие элементы работы Метода-ГРИ; здесь и далее операция «⟶» означает получение одного объекта из другого, операция «+=» – добавление элемента к набору, а пометка «*» – безопасную версию представления:

1) «Начало» – запуск метод Экспертом;

2) «Ввод конечного (небезопасного) представления программы: Rep_i» – пользователь метода выбирает имеющееся i-е представление программы (Rep_i), реверс-инжиниринг которого с последующим обнаружением и устранением уязвимостей необходимо произвести;

3) «Нейтрализация уязвимостей в конечном представлении: $Rep_i \longrightarrow Rep_i^*$» – внешний вызов процесса нейтрализации уязвимостей в текущем представлении (Rep_i) с получением его безопасной версии (Rep_i^*), поскольку существует определенное количество способов их обнаружения и устранения в конечных представлениях программы (например, статические анализаторы машинного кода);

4) «Идентификация синтаксиса текущего (конечного) представления: $Rep_i^* \longrightarrow Syntax_i$» – получение синтаксиса ($Syntax_i$) текущего представления программы (Rep_i^*), такого, как язык программирования или процессорная архитектура [17], что необходимо для дальнейшей деэволюции представлений;

5) «Модель жизненного цикла программы: $\{Rep_x, x \longrightarrow y\}$» – информационный объект ($\{\ldots\}$), хранящий возможные представления программ (Rep_x) и преобразования между ними ($x \longrightarrow y$) [18, 19];

6) «Существует предыдущее представление: Rep_{i-1}?» – проверка согласно модели жизненного цикла программы наличия предыдущего представления программы (Rep_{i-1}), в котором требуется нейтрализация уязвимостей;

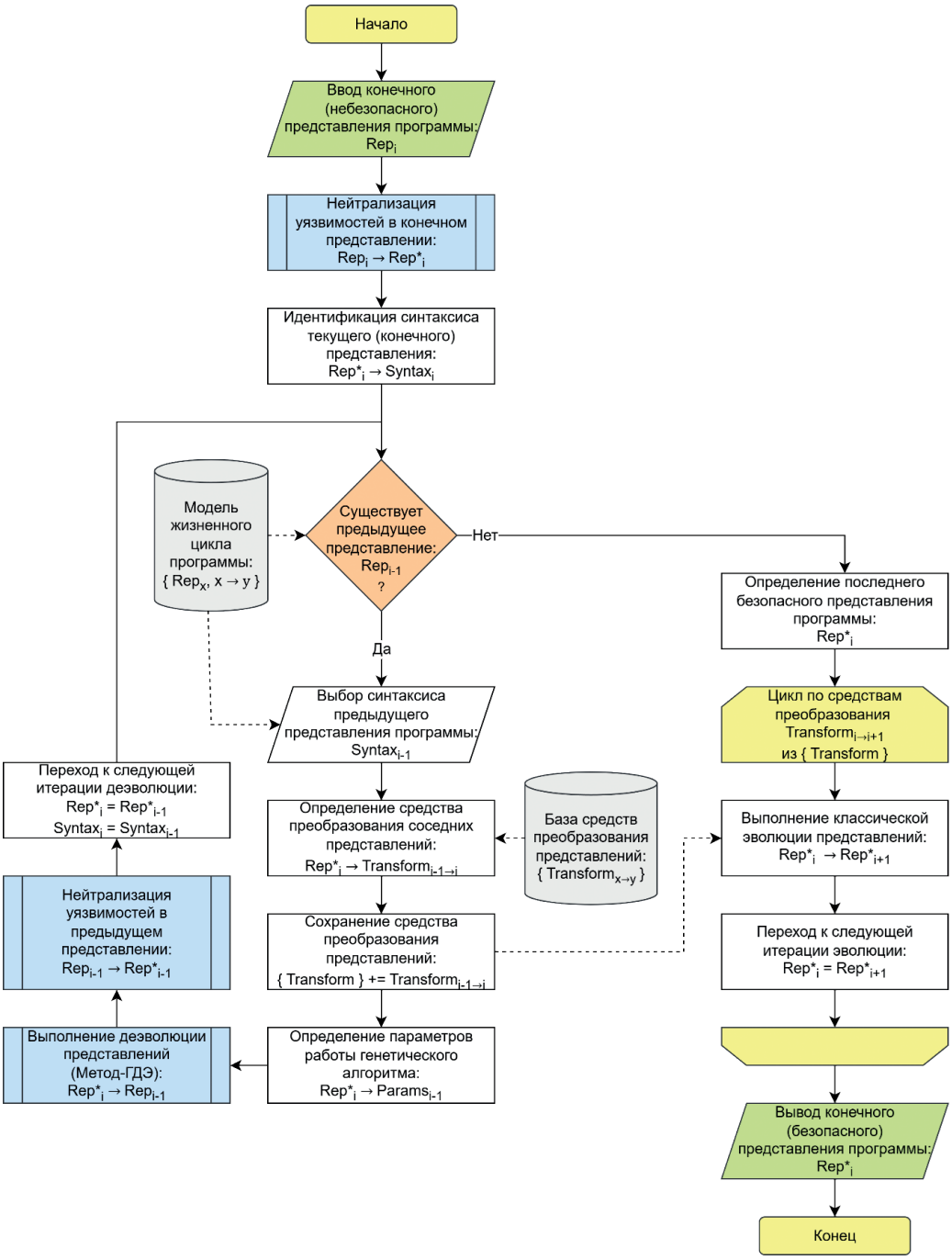


Рис. 1. Блок-схема генетического реверс-инжиниринга программы (с нейтрализацией уязвимостей)

7) «Выбор синтаксиса предыдущего представления программы: $Syntax_{i-1}$» – получение согласно модели жизненного цикла программы синтаксиса ее предыдущего представления ($Syntax_{i-1}$), что необходимо для работы ГДЭ;

8) «База средств преобразования представлений: $\{Transform_{x \to y}\}$» – информационный объект ($\{\ldots\}$) с существующими средствами получения ($Transform$) последующих представлений из текущих ($x \longrightarrow y$), такими, как генераторы исходного кода, средства компиляции, ассемблирования и пр.;

9) «Определение средства преобразования соседних представлений: $Rep_i^* \longrightarrow Transform_{i-1 \to i}$» – определение согласно базе средств преобразования представлений необходимого для получения текущего представления (Rep_i^*) из предыдущего ($i-1 \longrightarrow i$);

10) «Сохранение средства преобразования представлений: $\{Transform\}$ += $Transform_{i-1 \to i}$» – добавление определенного средства преобразования представлений ($Transform_{i-1 \to i}$) в набор ($\{\ldots\}$) таких средств ($Transform$);

11) «Определение параметров работы генетического алгоритма: $Rep_i^* \longrightarrow Params_{i-1}$» – определение по текущему представлению (Rep_i^*) таких параметров генетического алгоритма, работающего с предыдущим представлением ($Params_{i-1}$), как длина хромосомы, алгоритмы операций селекции, скрещивания и мутации, частота выполнения последних операций и др.;

12) «Выполнение деэволюции представлений (Метод-ГДЭ): $Rep_i^* \longrightarrow Rep_{i-1}$» – внешний вызов второго метода в иерархии Комплекса, производящего непосредственное преобразование текущего представления с нейтрализованными уязвимостями (Rep_i^*) в потенциально небезопасное предыдущее (Rep_{i-1}); при этом, в Метод-ГДЭ передаются такие настройки, как программа в текущем представлении (Rep_i^*), синтаксисы текущего ($Syntax_i$) и предыдущего ($Syntax_{i-1}$) представлений, средство преобразования предыдущего представления в текущее ($Transform_{i-1 \to i}$) и параметры генетического алгоритма ($Params_{i-1}$);

13) «Нейтрализация уязвимостей в предыдущем представлении: $Rep_{i-1} \longrightarrow Rep_{i-1}^*$» внешний вызов процесса нейтрализации уязвимостей в предыдущем представлении (Rep_{i-1}) с получением его безопасной версии (Rep_{i-1}^*) по аналогичным с элементом 3 причинам;

14) «Переход к следующей итерации деэволюции: $Rep_i^* = Rep_{i-1}^*$, $Syntax_i = Syntax_{i-1}$» – установление в качестве текущего представления (Rep_i^*) и его синтаксиса ($Syntax_i$) тех, которые ранее являлись предыдущими (Rep_{i-1}^* и $Syntax_{i-1}$), обеспечивая тем самым переход к следующей итерации реверс-инжиниринга программы;

15) «Определение последнего безопасного представления программы: Rep_i^*» – установление в качестве представления для сборки конечной безопасной программы того, которое обрабатывалось на последней итерации реверс-инжиниринга (Rep_i^*);

16) «Цикл по средствам преобразования $Transform_{i \to i+1}$ из $\{Transform_{x \to y}\}$» – осуществление циклического перебора средств трансформации представлений ($Transform_{i \to i+1}$), сохраненных в наборе ранее ($\{Transform_{x \to y}\}$) при осуществлении их деэволюции;

17) «Выполнение классической эволюции представлений: $Rep_i^* \longrightarrow Rep_{i+1}^*$» – получение в цикле каждого последующего представления программы (Rep_{i+1}^*) по текущему (Rep_i^*), например, путем генерации исходного кода, компиляции или ассемблирования;

18) «Переход к следующей итерации эволюции: $Rep_i^* = Rep_{i+1}^*$» – установление в качестве текущего представления (Rep_i^*) того, которое ранее считалось последующим (Rep_{i+1}^*), обеспечивая тем самым переход к следующей итерации получения конечного (безопасного) представления программы;

19) «Вывод конечного (безопасного) представления программы: Rep_i^*» – вывод представления программы, полученного на последней итерации цикла, которое является конечным и безопасным, т. е. результатом ГРИ, в процессе которого была произведена нейтрализация уязвимостей;

20) «Конец» – завершение метода.

Согласно блок-схеме, для работы Метода-ГРИ необходимы методы нейтрализации (т. е. обнаружения и устранения) уязвимостей во всех представлениях программы, а также, второй элемент Комплекс – Метод-ГДЭ.

Метод генетической деэволюции соседних представлений программы

Принцип работы Метода-ГДЭ заключается в итерационном подборе конструкций предыдущего представления программы по мере приближения полученного из них представления к текущему, деэволюцию которого необходимо произвести. При этом, наилучшие (отобранные) экземпляры в предыдущем представлении будут скрещиваться друг с другом и мутировать для получения наиболее близких к искомой программ. Так, например, метод будет подбирать ключевые слова, переменные и константы языка программирования С, компилируя их в машинный код, сравнивая с заданным, исходя из этого отбирая наиболее близкие, «перемешивая» между собой их конструкции и внося случайные изменения. Такой эволюционный (а, точнее, генетический) процесс решения оптимизационной задачи завершится, когда будет найдена программа в предыдущем представлении (например, ее исходный код), которая после преобразования тождественна заданной программе в текущем представлении (например, машинному коду). Соответствующая блок-схема Метода-ГДЭ приведена на рис. 2.

На блок-схеме (см. рис. 2) присутствуют следующие элементы работы Метода-ГДЭ:

1) «Начало» – запуск метод;

2) «Ввод настроек метода: текущее представление программы (Rep_i), синтаксис текущего ($Syntax_i$) и предыдущего представлений ($Syntax_{i-1}$), средство их преобразования ($Transform_{i-1 \to i}$), параметры генетического алгоритма ($Params$)» – в метод передаются параметры, необходимые для получения предыдущего i–1-го представления по i-му текущему (Rep_i) с помощью генетического алгоритма, в операциях которого используются синтаксисы обоих представлений ($Syntax_{i-1}$ и $Syntax_i$) и средство получения текущего представления программы, соответствующего ее некоторому экземпляру в предыдущем ($Transform_{i-1 \to i}$); управление работой

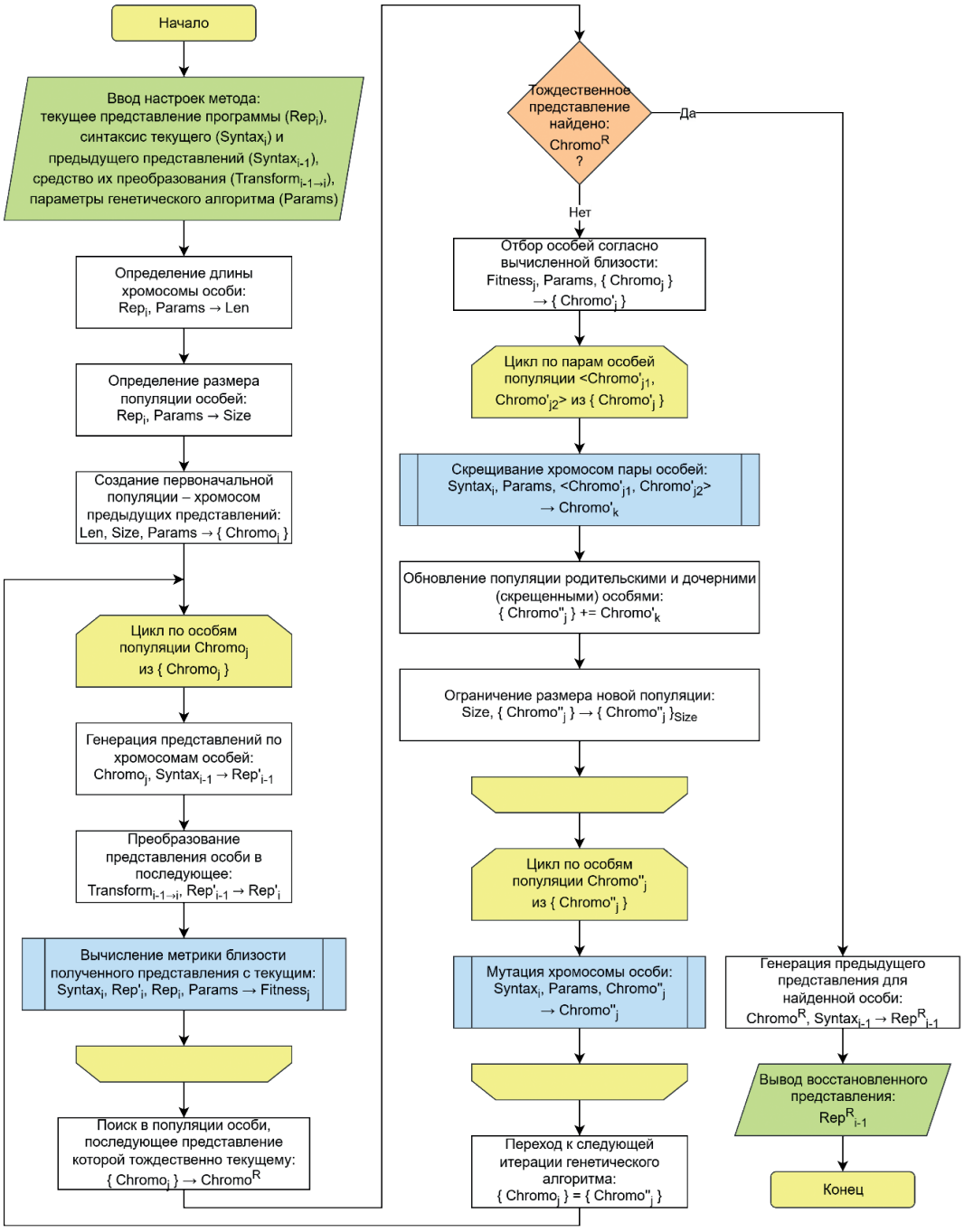


Рис. 2. Блок-схема генетической деэволюции представлений программы

метода осуществляется через переданные в него параметры (*Params*);

3) «Определение длины хромосомы особи: Rep_i, $Params \longrightarrow Len$» – определение близкой к истинной длины (*Len*) хромосомы особи, вычисляемой на основании экземпляра программы в текущем представлении (Rep_i) и заданных настроек метода (*Params*);

4) «Определение размера популяции особей: Rep_i, $Params \longrightarrow Size$» – определение размера популяции особей (*Size*), близкого к оптимальному и вычисляемому на основании экземпляра программы в текущем представлении (Rep_i) и заданных параметров метода (*Params*) [20];

5) «Создание первоначальной популяции – хромосом предыдущих представлений: $Len, Size, Params \longrightarrow \longrightarrow \{Chromo_j\}$» – создание первоначальной популяции особей, как множества хромосом, на основании определенных длины хромосомы (*Len*), размера популяции (*Size*) и заданных параметров метода (*Params*);

6) «Цикл по особям популяции $Chromo_j$ из $\{Chromo_j\}$» – осуществление циклического перебора

хромосом особей ($Chromo_j$) из текущей популяции ($\{Chromo_j\}$);

7) «Генерация представлений по хромосомам особей: $Chromo_j$, $Syntax_{i-1} \longrightarrow Rep'_{i-1}$» – создание предыдущего представления программы (Rep'_{i-1}) на основании ее генетической записи в виде хромосомы ($Chromo_j$) с использованием ГСП, заданного синтаксисом этого представления ($Syntax_{i-1}$);

8) «Преобразование представления особи в последующее: $Transform_{i-1\to i}$, $Rep'_{i-1} \longrightarrow Rep'_i$» – получение программы в текущем представлении (Rep'_i) из ее предыдущего представления (Rep'_{i-1}) с помощью средства преобразования ($Transform_{i-1\to i}$);

9) «Вычисление метрики близости полученного представления с текущим: $Syntax_i$, Rep'_i, Rep_i, $Params \longrightarrow Fitness_j$» – внешний вызов 1-го метода Группы для вычисления численной метрики близости двух программ (которая в теории генетических алгоритмов соответствует приспособленности особи или ее фитнес-функции – $Fitness_j$) в текущем представлении с использованием его синтаксиса ($Syntax_i$), одна из которых задана и требует деэволюции (Rep_i), а другая получена преобразованием из особи популяции (Rep'_{i-1}); управление работой метода осуществляется путем передачи необходимых параметров ($Params$);

10) «Поиск в популяции особи, последующее представление которой тождественно текущему: $\{Chromo_j\} \longrightarrow Chromo^R$» – поиск в текущей популяции особи, метрика близости к которой предельно максимальна (например, равна 1, в случае нормированной к единице метрики), что означает тождественность некоторого экземпляра программы ($Chromo^R$) после преобразования к заданному;

11) «Тождественное представление найдено: $Chromo^R$?» – проверка нахождения в популяции экземпляра программы, тождественного заданному, что означает решение задачи деэволюции ($Chromo^R$) и приводит к последующему завершению метода;

12) «Отбор особей согласно вычисленной близости: $Fitness_j$, $Param$, $\{Chromo_j\} \longrightarrow \{Chromo'_j\}$» – отбор особей из текущей популяции ($\{Chromo_j\}$) в новую популяцию ($\{Chromo'_j\}$) исходя из значения метрики близости каждой из них ($Fitness_j$); в простейшем случае может осуществляться сортировкой по убыванию и выбором первых N-особей;

13) «Цикл по парам особей популяции $\langle Chromo'_{j1}, Chromo'_{j2}\rangle$ из $\{Chromo'_j\}$» – осуществление циклического перебора двух хромосом особей (кортеж из $Chromo'_{j1}$ и $Chromo'_{j2}$) из текущей популяции ($\{Chromo'_j\}$) для «смешивания» их ген;

14) «Скрещивание хромосом пары особей: $Syntax_i$, $Params$, $\langle Chromo'_{j1}, Chromo'_{j2}\rangle \longrightarrow Chromo'_k$» – внешний вызов 2-го метода Группы для выполнения операции скрещивания (в рамках генетического алгоритма) хромосом двух родительских особей ($Chromo'_{j1}$ и $Chromo'_{j2}$), приводящей к получению дочерней особи с генами, содержащими родительские; управление работой метода осуществляется путем передачи синтаксиса текущего представления ($Syntax_i$) и необходимых параметров ($Params$);

15) «Обновление популяции родительскими и дочерними (скрещенными) особями: $\{Chromo''_j\}$ += $Chromo'_k$» – формирование новой текущей популяции ($\{Chromo''_j\}$) из отобранных родительских особей и их дочерних ($Chromo'_k$);

16) «Ограничение размера новой популяции: $Size$, $\{Chromo''_j\} \longrightarrow \{Chromo''_j\}_{Size}$» – ограничение результатов работы операции скрещивания, создающей дочерние особи, для предотвращения превышения текущей популяцией ($\{Chromo''_j\}$) определенного ранее размера ($Size$), гарантируя тем самым постоянное количество особей популяции ($\{Chromo''_j\}_{Size}$);

17) «Цикл по особям популяции $Chromo''_j$ из $\{Chromo''_j\}$» – осуществление циклического перебора хромосом особей ($Chromo''_j$) из текущей популяции ($\{Chromo''_j\}$) для мутации ген их хромосом;

18) «Мутация хромосомы особи: $Syntax_i$, $Params$, $Chromo''_j \longrightarrow Chromo''_j$» – внешний вызов 3-го метода Группы для выполнения операции мутации (в рамках генетического алгоритма) генов хромосомы особи ($Chromo''_j$), приводящей к их случайному изменению; управление работой метода осуществляется путем передачи синтаксиса текущего представления ($Syntax_i$) и необходимых параметров ($Params$);

19) «Переход к следующей итерации генетического алгоритма: $\{Chromo_j\} = \{Chromo''_j\}$» – формальный переход к новой эпохе эволюции, в которой предыдущая популяция ($\{Chromo''_j\}$), полученная операциями селекции, скрещивания и мутации, становится текущей ($\{Chromo_j\}$);

20) «Генерация предыдущего представления для найденной особи: $Chromo^R$, $Syntax_{i-1} \longrightarrow Rep^R_{i-1}$» – создание предыдущего представления программы (Rep^R_{i-1}) на основании особи в популяции с максимально возможной метрикой близости к заданной программе в текущем представлении ($Chromo^R$) с использованием ГСП, заданного синтаксисом этого представления ($Syntax_{i-1}$); полученное таким образом представление программы является искомым результатом работы метода;

21) «Вывод восстановленного представления: Rep^R_{i-1}» – вывод из метода найденного представления программы, которое после преобразования тождественно заданному;

22) «Конец» – завершение метода.

Согласно блок-схеме, для работы Метода-ГДЭ необходима Группа методов, отвечающих за операции

вычисления метрики близости экземпляров программы, а также скрещивания и мутации их хромосом.

Группа методов операций генетических алгоритмов

Группа методов на нижнем уровне в иерархии Комплекса состоит из методов для выполнения операций вычисления метрики близости, скрещивания и мутации, работа которых основана на генетическом представлении программы – как пути по ГСП синтаксиса предыдущего представления. Поскольку блок-схема не даст достаточного понимания принципов и особенностей методов Группы, приведем далее интуитивно-понятный псевдокод алгоритмов работы каждого из них.

Первый метод Группы предназначен для определения близости экземпляра программы, полученного из хромосомы особи с заданной программой (т. е. той, деэволюцию которой необходимо произвести). Работа метода построена на определении признаков каждого из экземпляров, которые являются параметрами модели, соответствующей синтаксису текущего представления программы. Так, например, если представлением является исходный код, а экземпляры имеют текстовую запись программы на соответствующем языке программирования, то параметрами может быть топология дерева абстрактного синтаксиса и свойства его узлов; в ряде случаев, можно использовать более простые модели, такие как списки символьных строк [21]. Затем, используя такие модельные представления экземпляров программ, производится вычисление метрики их близости, целесообразность чего обосновывается использованием единого базиса сравнения (т. е. формальной записью синтаксиса).

Псевдокод алгоритма для метода вычисления метрики близости (ProximityMetric) в Python-подобном стиле приведен ниже; помимо типовых, используются следующие языковые конструкции: «**List<>**» – список элементов, «**double**» – тип для чисел с плавающей точкой.

```
   Name: ProximityMetric()
   Input:
      Syntax - синтаксис представления программы
      Rep1 - представление первой программы
      Rep2 - представление второй программы
      Params - параметры определения метрики близости
   Output:
      Metric - метрика близости двух программ (в диапазоне [0, 1])
   Begin
         // Шаг 1. Получение внутреннего представления
    1:   parser = Params.Parser(Syntax);
    2:   features1 = parser.Run(Rep1);
    3:   features2 = parser.Run(Rep2);

         // Шаг 2. Нормализация внутреннего представления
    4:   normal = Params.Normalizer(Syntax);
    5:   features_normal1 = normal(features1);
    6:   features_normal2 = normal(features2);

         // Шаг 3. Вычисление частных метрик
    7:   List<double> metrics;
    8:   For comparator In Params.Comparators:
    9:      metric = comparator.Eval(features_normal1, features_normal2);
   10:      metrics += metric;
   11:   End For

         // Шаг 4. Вычисление интегральной метрики
   12:   integrator = Params.Integrator();
   13:   Metric = integrator.Combine(metrics);

   14:   Return Metric
   End
```

Метод на вход принимает синтаксис представления программы, два ее представления и параметры; а на выходе возвращает метрику близости этих программ.

Согласно псевдокоду алгоритма ProximityMetric(), он состоит из 4 следующих шагов: 1) экземпляры программ на основе формального синтаксиса преобразуются во внутреннее представление, например, дерево абстрактного синтаксиса [22]; 2) производится нормализация внутреннего представления приведением их к более обобщенному виду, например, сортировкой операндов коммутативных операций или более сложным образом [23]; 3) вычисляются различные частные метрики близости программ, например, с использованием алгоритмов схожести графов [24]; 4) вычисленные частные метрики интегрируются в единую, множество значений которой находится в отрезке от 0 до 1 и которая считается результатом работы метода.

Второй метод Группы скрещивания предназначен для получения дочернего экземпляра программы по хромосомам двух родительских особей-программ в текущей популяции [25]. Метод анализирует два родительских пути и находит в них позиции, которые ведут на один узел-альтернативу в ГСП. Затем, случайным образом выбирается одна из таких позиций, относительно которой происходит взаимный «обмен» частями хромосом родительских узлов для их поддеревьев в ГСП, с получением пары дочерних. В результате, начальная (до общей позиции и поддеревьев ГСП) и конечная части (после поддеревьев ГСП) хромосом дочерних узлов совпадают с аналогичными частями 1-го и 2-го родителя соответственно, а средняя часть хромосом (после общей позиции и в рамках поддерева ГСП) – с противоположными родителями, т.е. с 2-м и 1-м соответственно. Например, если согласно хромосомам родителей их пути

по ГСП проходят через узлы [1 2 9 3 4] и [5 6 9 7 8] (т. е. в которых общий узел 9), то после скрещивания могут получиться две дочерних особи, пути которых проходят через следующие узлы ГСП – [1 2 9 7 8] и [5 6 9 3 4].

Псевдокод алгоритма для метода скрещивания хромосом (ChromosomeCrossover) в Python-подобном стиле приведен ниже; языковые конструкции, указанные для алгоритма ProximityMetric(), дополнены следующими: «**Tuple<>**» – кортеж элементов, «**integer**» – целочисленный тип, «**var_list[p1 : p2]**» – часть списка «**var_list**» с $p1$-го по $p2$-ой элементы (второй не включительно).

```
   Name: ChromosomeCrossover()
   Input:
      Syntax - синтаксис представления программы
      Chromosome1 - хромосома первой особи
      Chromosome2 - хромосома второй особи
      Params - параметры скрещивания хромосом
   Output:
      Chromosome_C1 - хромосома после скрещивания первой особи
      Chromosome_C2 - хромосома после скрещивания второй особи
   Begin:

         // Шаг 1. Инициализация
    1:   random = Random();
    2:   List<Tuple<integer, integer>> points;
    3:   List<Tuple<ChromosomeType, ChromosomeType>> crosses;

         // Шаг 2. Обход генов первой хромосомы
    4:   For i1 In Chromosome1.Gens.Length:
    5:      gen1 = Chromosome1.Gens[i1];

         // Шаг 2.1. Обход генов второй хромосомы
    6:   For i2 In Chromosome2.Gens.Length:
    7:      gen2 = Chromosome2.Gens[i1];

         // Шаг 2.1.1. Поиск генов, связанных с одинаковыми узлами-альтернативами
    8:      If gen1.NodeId == gen2.NodeId Then:

            // Шаг 2.1.2. Получение частей хромосом для обмена между особями
    9:         cross_1 = Chromosome2.Gens[i1 : i1 + gen1.SubTreeLen];
   10:         cross_2 = Chromosome2.Gens[i2 : i2 + gen2.SubTreeLen];

            // Шаг 2.1.3. Проверка отличности частей хромосом
   11:         If cross_1 != cross_2 Then:
            // Шаг 2.1.4. Выбор и сохранение частей хромосом
   12:         If Params.CrossoverRate >= random.float(0, 1) Then
   13:            points += [ i1, i2 ];
   14:            crosses += [ cross_1, cross_2 ];
   15:               End If
   16:            End If
   17:         End If
   18:      End If
   19:   End If

         // Шаг 3. Проверка наличия подходящих хромосом для обмена генами
   20:   If points.Length == 0 Then:
   21:      Return None;
   22:   End If

         // Шаг 4. Выбор части хромосом для обмена генами
   23:   j = random.int(0, points.Length);
   24:   point = points[j];
   25:   cross = crosses[j];

         // Шаг 5. Конструирование дочерних хромосом из родительских
   26:   Chromosome_C1.Gens = Chromosome1.Gens[0 : point.Element1] + cross.Element2 + Chromosome1.Gens[point.Element1 + cross.Element2.Length : Chromosome1.Gens.Length];
   27:   Chromosome_C2.Gens = Chromosome2.Gens[0 : point.Element2] + cross.Element1 + Chromosome2.Gens[point.Element2 + cross.Element1.Length : Chromosome2.Gens.Length];

   28:   Return Chromosome_C1, Chromosome_C2;
   End
```

Метод на вход принимает синтаксис представления программы, хромосомы двух особей и параметры; а на выходе возвращает хромосомы двух дочерних (сгенерированных) особей.

Согласно псевдокоду алгоритма ChromosomeCrossover(), он состоит из 5 следующих шагов: 1) инициализация генератора случайных чисел, создание вспомогательных хранилищ точек пересечения хромосом и самих частей хромосом; 2) обход генов обоих хромосом с целью определения их точек пересечения для обмена при скрещивании; 3) проверка наличия точек пересечения и в ином случае возврат негативного результата скрещивания; 4) случайный выбор частей хромосом для обмена между особями; 5) конструирование пары новых дочерних особей, составленных из начальных и конечных частей хромосом своих родителей с обменом средними частями. При этом, шаг 2 представляет собой два вложенных цикла по генам хромосом особей, в теле последнего из которых выполняются следующие вложенные шаги: 1) поиск генов двух хромосом, соответствующих одному узлу-альтернативе; 2) сохранение точек пересечения и частей хромосом особей, соответствующих поддеревьям ГСП для найденных узлов-альтернатив; 3) обеспечение отличности полученных частей-хромосом для обеих особей (в ином случае, скрещивание не будет иметь эффекта); 4) выбор частей хромосом для скрещивания, используя

случайно сгенерированное число (на отрезке [0,1]) и частоту скрещивания.

Третий метод Группы предназначен для мутации хромосомы отдельной особи [26]. Метод случайным образом (или согласно некоторой логике) меняет гены хромосомы, что приводит к изменению выбора дочерних веток в ГСП в узлах-альтернативах. Так, замена элемента пути на другой случайный, в котором содержатся другие альтернативы, перестроит всю логику обхода ГСП, потенциально увеличив или уменьшив длину хромосомы особи. Естественно, такое произвольное изменение ген может привести к необходимости выбора значений для новых альтернатив, для чего в методе присутствует специальная операция добавления корректного (случайного) окончания пути. Например, если изначально альтернатива вела к сложению с числом, а после мутации – к сложению с функцией, выражение для которой содержит множество пока еще не выбранных альтернатив (например, параметров), то окончание хромосомы будет перегенерировано, а ее длина увеличится.

Псевдокод алгоритма (ChromosomeMutation) для метода мутации хромосомы приведен ниже.

```
Name: ChromosomeMutation()
Input:
   Syntax - синтакс представления программы
   Chromosome - хромосома особи
   Params - параметры мутации хромосом
Output:
   Chromosome_M - хромосома смутировавшей особи
Begin
      // Шаг 1. Инициализация
 1:   random = Random();

      // Шаг 2. Обход генов хросомосы
 2:   For i In Chromosome.Gens.Length:
 3:      gen = Chromosome.Gens[i];

      // Шаг 2.1. Выбор гена для мутации
 4:      If Params.MutationRate >= random.
float(0, 1) Then:

      // Шаг 2.2. Выбор нового значения гена
 5:      j = random.int(0, gen.MaxValue - 1):

      // Шаг 2.3. Проверка отличности нового
гена от имеющегося
 6:      If j >= gen.Value Then:
 7:         j += 1;
 8:      End If

      // Шаг 2.4. Обновление значение гена
 9:      Chromosome.Gens[i].Value = j;
10:      End If

      // Шаг 2.5. Построение корректного пути
по ГСП с учетом нового гена
11:      ChromosomeM = Syntax.CompletePathFrom-
Gen(Chromosome, I, Param);

12:   Return ChromosomeM
End
```

Метод на вход принимает синтаксис представления программы, хромосому и параметры; а на выходе возвращает мутированную хромосому.

Согласно псевдокоду алгоритма ChromosomeMutation(), он состоит из 2 следующих шагов, на которых производится инициализация генератора случайных чисел и обход генов хромосомы для их мутации. Также, второй шаг поделен на следующие вложенные шаги: 1) выбор гена для мутации, используя случайно сгенерированное число (на отрезке [0,1]) и заданную частоту мутации; 2) выбор нового мутированного значения гена, используя случайно сгенерированное число в диапазоне 0 до максимально возможного значения гена (определяемого количеством возможных путей из узла-альтернативы), уменьшенного на 1; 3) обеспечение отличности нового значения гена от имеющегося в хромосоме путем увеличения его значения на 1 в случае равенства или превышения имеющегося значения (для этого было уменьшение диапазона случайных чисел на 1 на предыдущем шаге); 4) обновление имеющегося значения гена хромосомы тем, которое было выбрано случайно; 5) построение корректного окончания генов хромосомы, начиная с мутировавшего, поскольку сделанное таким образом изменение пути по ГСП, как правило, влияет на последующие узлы-альтернативы.

Заключение

В работе описывается авторский подход генетической деэволюции, основным применением которого является получение представлений программы для обнаружения и устранения в них уязвимостей. Подход основан на решении оптимизационной задачи подбора синтаксических конструкций, которые бы соответствовали программе в предыдущем представлении, тождественной имеющейся программе в текущем.

Основным результатом проведенного исследования является комплекс методов генетической деэволюции, состоящих из общего метода проведения реверс-инжиниринга программы, использующего метод последовательной деэволюции соседних представлений программы, который в свою очередь основан на применении генетических алгоритмов с помощью операций, реализованных в виде метода вычисления метрики близости программ в одном представлении, а также методов скрещивания и мутации хромосом их особей.

Значимость полученных результатов заключается как в теоретической интеллектуализации процесса реверс-инжиниринга программ, так и в практическом улучшении результативности нейтрализации уязвимостей программы при неухудшении оперативности и ресурсоэкономности процесса.

Исходя из того, что все методы имеют реализацию в виде программного прототипа, продолжением работы должно стать проведение серии экспериментов, как для подтверждения их работоспособности, так и для оценок границ применимости.

Литература

1. Абдуллин Т. И., Баев В. Д., Буйневич М. В. и др. Цифровые технологии и проблемы информационной безопасности / Санкт-Петербург: Санкт-Петербургский государственный экономический университет, 2021. 163 с.
2. Шимчик Н. В., Игнатьев В. Н., Белеванцев А. А. IRBIS: Статический анализатор помеченных данных для поиска уязвимостей в программах на C/C++ // Труды Института системного программирования РАН. 2022. Т. 34. № 6. С. 51–66. DOI: 10.15514/ISPRAS-2022-34(6)-4.
3. David A. Ghidra Software Reverse Engineering for Beginners: Analyze, identify, and avoid malicious code and potential threats in your networks and systems. UK: Packt Publishing Ltd, 2021. 322 p.
4. Жилин В. В., Сафарьян О. А. Искусственный интеллект в системах хранения данных // Вестник Донского государственного технического университета. 2020. Т. 20. № 2. С. 196–200. DOI: 10.23947/1992-5980-2020-20-2-196-200.
5. Artuso F. Deep Learning Based Binary Code Analysis: Ph.D. Program in Engineering in Computer Science / Sapienza University of Rome, 2025. 155 p.
6. Armengol-Estape J., Woodruff J., Cummins C., O'Boyle M. F. SLaDe: A Portable Small Language Model Decompiler for Optimized Assembly // The proceedings of IEEE/ACM International Symposium on Code Generation and Optimization (Edinburgh, United Kingdom, 2–6 March 2024). 2024. PP. 67–80.
7. Tan H., Luo Q., Li J., Zhang Y. LLM4Decompile: Decompiling Binary Code with Large Language Models // The proceedings of Conference on Empirical Methods in Natural Language Processing (USA, Miami, Florida, 12–16 November 2024). 2024. PP. 3473–3487.
8. Zhang X., Xu Z., Yang S., Li Z., Shi Z., Sun L. Enhancing Function Name Prediction using Votes-Based Name Tokenization and Multi-task Learning // The proceedings of ACM on Software Engineering. Vol. 1. No. 75. PP. 1679–1702.
9. He J., Ivanov P., Tsankov P., Raychev V., Vechev M. Debin: Predicting Debug Information in Stripped Binaries // The proceedings of ACM SIGSAC Conference on Computer and Communications Security (Canada, Toronto, 15–19 October 2018). 2018. P. 1667–1680.
10. Shin E. C. R., Song D., Moazzezi R. Recognizing functions in binaries with neural networks // The proceedings of 24th USENIX Conference on Security Symposium (USA, Washington, D.C., 2015 August 12–14). 2015. PP. 611–626.
11. Израилов К. Е. Концепция генетической деэволюции представлений программы. Часть 1 // Вопросы кибербезопасности. 2024. № 1(59). С. 61–66. DOI: 10.21681/2311-3456-2024-1-61-66.
12. Израилов К. Е. Концепция генетической деэволюции представлений программы. Часть 2 // Вопросы кибербезопасности. 2024. № 2(60). С. 81–86. DOI: 10.21681/2311-3456-2024-2-81-86.
13. Силенко Д. И., Лебедев И. Г. Алгоритм глобальной оптимизации, использующий деревья решений для выявления локальных экстремумов // Проблемы информатики. 2023. № 2(59). С. 21–33. DOI: 10.24412/2073-0667-2023-2-21-33.
14. Пикалов М. В., Письмеров А. М. Настройка параметров генетического алгоритма при помощи анализа ландшафта функции приспособленности и машинного обучения // Известия ЮФУ. Технические науки. 2024. № 2(238). С. 221–228. DOI: 10.18522/2311-3103-2024-2-221-228.
15. Петросов Д. А. Анализ и выбор методов представления характеристик состояния популяции генетического алгоритма // Оригинальные исследования. 2023. Т. 13. № 10. С. 235–239.
16. Безгачев Ф. В., Галушин П. В., Рудакова Е. Н. Эффективная реализация инициализации и мутации в генетическом алгоритме псевдо-булевой оптимизации // E-Scio. 2020. № 4(43). С. 224–231.
17. Pan Z., Yan Y., Yu L., Wang T. Identification of binary file compilation information // Proceedings of the IEEE 5th Advanced Information Management, Communicates, Electronic and Automation Control Conference (Chongqing, China, 16–18 December 2022). 2022. PP. 1141–1150. DOI: 10.1109/IMCEC55388.2022.10019958.
18. Израилов К. Е. Моделирование программы с уязвимостями с позиции эволюции ее представлений. Часть 1. Схема жизненного цикла // Труды учебных заведений связи. 2023. Т. 9. № 1. С. 75–93. DOI: 10.31854/1813-324X-2023-9-1-75-93.
19. Израилов К. Е. Моделирование программы с уязвимостями с позиции эволюции ее представлений. Часть 2. Аналитическая модель и эксперимент // Труды учебных заведений связи. 2023. Т. 9. № 2. С. 95–111. DOI: 10.31854/1813-324X-2023-9-2-95-111.
20. Цыганков В. А., Шабалина О. А., Катаев А. В. Исследование воздействия размера популяции на быстродействие генетического алгоритма // Известия ЮФУ. Технические науки. 2024. № 3(239). С. 168–176. DOI: 10.18522/2311-3103-2024-3-168-176.
21. Буйневич М. В., Израилов К. Е. Авторская метрика оценки близости программ: приложение для поиска уязвимостей с помощью генетической деэволюции // Программные продукты и системы. 2025. Т. 38. № 1. С. 89–99. DOI: 10.15827/0236-235X.149.089-099.
22. Грибков Н. А., Овасапян Т. Д., Москвин Д. А. Анализ восстановленного программного кода с использованием абстрактных синтаксических деревьев // Проблемы информационной безопасности. Компьютерные системы. 2023. № 2(54). С. 47–60. DOI: 10.48612/jisp/ruar-u6he-kmd4.
23. Allamanis M., Brockschmidt M., Khademi M. Learning to Represent Programs with Graphs // In proceedings of the 6th International Conference on Learning Representations (Vancouver, Canada, 20 April–3 May 2018). 2018. PP. 1–17. DOI: 10.48550/arXiv.1711.00740.
24. Ормонова Э. М. Определение качества программного продукта на основе теории графов // Наука. Образование. Техника. 2021. № 1(70). С. 37–44.
25. Тотухов К. Е., Романов А. Ю., Лукьянов В. И. Исследование эффективности работы генетических алгоритмов с различными методами скрещивания и отбора // Электронный сетевой политематический журнал «Научные труды КубГТУ». 2022. № 6. С. 98–109.
26. Доманов К. И. Сравнительный анализ эффективности работы генетического алгоритма при модификации оператора мутации в задаче коммивояжера // Политехнический молодежный журнал. 2022. № 1(66). DOI: 10.18698/2541-8009-2022-1-760.

A COMPLEX OF METHODS FOR GENETIC DE-EVOLUTION OF PROGRAM REPRESENTATIONS

Izrailov K. E.[2]

Keywords: *vulnerability neutralization, reverse engineering, artificial intelligence, genetic algorithms, complex of methods.*

The goal of the research: *increasing the efficiency of neutralizing program vulnerabilities by intellectualizing its reverse engineering using genetic algorithms*

Research methods: *system analysis and optimization methods, graph theory, functional and structural synthesis, general programming methodology and compiler theory.*

Results: *a hierarchical three-level set of methods was synthesized, consisting of a genetic reverse-engineering program method, a genetic de-evolution method of its neighboring representations (machine and source code, algorithms, architecture, etc.), and a group of methods for implementing the genetic algorithms fundamental operations.*

The scientific novelty *of the complex methods lies in their focus on solving the reverse engineering problem by direct transformations of the program into subsequent representations, in contrast to classical ones that perform inverse transformations. Also, the algorithms of the methods group of the complex are based on working with the original source code model, representing it as a genes sequence.*

References

1. Abdullin T. I., Baev V. D., Bujnevich M. V. i dr. Cifrovye tehnologii i problemy informacionnoj bezopasnosti / Sankt-Peterburg: Sankt-Peterburgskij gosudarstvennyj jekonomicheskij universitet, 2021. 163 s.
2. Shimchik N. V., Ignat'ev V. N., Belevancev A. A. IRBIS: Staticheskij analizator pomechennyh dannyh dlja poiska ujazvimostej v programmah na C/C++ // Trudy Instituta sistemnogo programmirovanija RAN. 2022. T. 34. № 6. S. 51–66. DOI: 10.15514/ISPRAS-2022-34(6)-4.
3. David A. Ghidra Software Reverse Engineering for Beginners: Analyze, identify, and avoid malicious code and potential threats in your networks and systems. UK: Packt Publishing Ltd, 2021. 322 p.
4. Zhilin V. V., Safar'jan O. A. Iskusstvennyj intellekt v sistemah hranenija dannyh // Vestnik Donskogo gosudarstvennogo tehnicheskogo universiteta. 2020. T. 20. № 2. S. 196–200. DOI: 10.23947/1992-5980-2020-20-2-196-200.
5. Artuso F. Deep Learning Based Binary Code Analysis: Ph.D. Program in Engineering in Computer Science / Sapienza University of Rome, 2025. 155 p.
6. Armengol-Estape J., Woodruff J., Cummins C., O'Boyle M. F. SLaDe: A Portable Small Language Model Decompiler for Optimized Assembly // The proceedings of IEEE/ACM International Symposium on Code Generation and Optimization (Edinburgh, United Kingdom, 2–6 March 2024). 2024. PP. 67–80.
7. Tan H., Luo Q., Li J., Zhang Y. LLM4Decompile: Decompiling Binary Code with Large Language Models // The proceedings of Conference on Empirical Methods in Natural Language Processing (USA, Miami, Florida, 12–16 November 2024). 2024. PP. 3473–3487.
8. Zhang X., Xu Z., Yang S., Li Z., Shi Z., Sun L. Enhancing Function Name Prediction using Votes-Based Name Tokenization and Multi-task Learning // The proceedings of ACM on Software Engineering. Vol. 1. No. 75. PP. 1679–1702.
9. He J., Ivanov P., Tsankov P., Raychev V., Vechev M. Debin: Predicting Debug Information in Stripped Binaries // The proceedings of ACM SIGSAC Conference on Computer and Communications Security (Canada, Toronto, 15–19 October 2018). 2018. P. 1667–1680.
10. Shin E. C. R., Song D., Moazzezi R. Recognizing functions in binaries with neural networks // The proceedings of 24th USENIX Conference on Security Symposium (USA, Washington, D.C., 2015 August 12–14). 2015. PP. 611–626.
11. Izrailov K. E. Koncepcija geneticheskoj dejevoljucii predstavlenij programmy. Chast' 1 // Voprosy kiberbezopasnosti. 2024. № 1(59). S. 61–66. DOI: 10.21681/2311-3456-2024-1-61-66.
12. Izrailov K. E. Koncepcija geneticheskoj dejevoljucii predstavlenij programmy. Chast' 2 // Voprosy kiberbezopasnosti. 2024. № 2(60). S. 81–86. DOI: 10.21681/2311-3456-2024-2-81-86.
13. Silenko D. I., Lebedev I. G. Algoritm global'noj optimizacii, ispol'zujushhij derev'ja reshenij dlja vyjavlenija lokal'nyh jekstremumov // Problemy informatiki. 2023. № 2(59). S. 21–33. DOI: 10.24412/2073-0667-2023-2-21-33.
14. Pikalov M. V., Pis'merov A. M. Nastrojka parametrov geneticheskogo algoritma pri pomoshhi analiza landshafta funkcii prisposoblennosti i mashinnogo obuchenija // Izvestija JuFU. Tehnicheskie nauki. 2024. № 2(238). S. 221–228. DOI: 10.18522/2311-3103-2024-2-221-228.
15. Petrosov D. A. Analiz i vybor metodov predstavlenija harakteristik sostojanija populjacii geneticheskogo algoritma // Original'nye issledovanija. 2023. T. 13. № 10. S. 235–239.
16. Bezgachev F. V., Galushin P. V., Rudakova E. N. Jeffektivnaja realizacija inicializacii i mutacii v geneticheskom algoritme psevdo-bulevoj optimizacii // E-Scio. 2020. № 4(43). S. 224–231.
17. Pan Z., Yan Y., Yu L., Wang T. Identification of binary file compilation information // Proceedings of the IEEE 5th Advanced Information Management, Communicates, Electronic and Automation Control Conference (Chongqing, China, 16–18 December 2022). 2022. PP. 1141–1150. DOI: 10.1109/IMCEC55388.2022.10019958.
18. Izrailov K. E. Modelirovanie programmy s ujazvimostjami s pozicii jevoljucii ee predstavlenij. Chast' 1. Shema zhiznennogo cikla // Trudy uchebnyh zavedenij svjazi. 2023. T. 9. № 1. S. 75–93. DOI: 10.31854/1813-324X-2023-9-1-75-93.

2 Konstantin E. Izrailov, Ph.D., Docent, Professor of the Department of Applied Mathematics and Information Technologies Security of the Saint-Petersburg University of State Fire Service of EMERCOM of Russia, Saint-Petersburg. ORCHID: http://orcid.org/0000-0002-9412-5693. Scopus Author ID: 56123238800. E-mail:konstantin.izrailov@mail.ru

19. Izrailov K. E. Modelirovanie programmy s ujazvimostjami s pozicii jevoljucii ee predstavlenij. Chast' 2. Analiticheskaja model' i jeksperiment // Trudy uchebnyh zavedenij svjazi. 2023. T. 9. № 2. S. 95–111. DOI: 10.31854/1813-324X-2023-9-2-95-111.
20. Cygankov V. A., Shabalina O. A., Kataev A. V. Issledovanie vozdejstvija razmera populjacii na bystrodejstvie geneticheskogo algoritma // Izvestija JuFU. Tehnicheskie nauki. 2024. № 3(239). S. 168–176. DOI: 10.18522/2311-3103-2024-3-168-176.
21. Bujnevich M. V., Izrailov K. E. Avtorskaja metrika ocenki blizosti programm: prilozhenie dlja poiska ujazvimostej s pomoshh'ju geneticheskoj dejevoljucii // Programmnye produkty i sistemy. 2025. T. 38. № 1. S. 89–99. DOI: 10.15827/0236-235X.149.089-099.
22. Gribkov N. A., Ovasapjan T. D., Moskvin D. A. Analiz vosstanovlennogo programmnogo koda s ispol'zovaniem abstraktnyh sintaksicheskih derev'ev // Problemy informacionnoj bezopasnosti. Komp'juternye sistemy. 2023. № 2(54). S. 47–60. DOI: 10.48612/jisp/ruar-u6he-kmd4.
23. Allamanis M., Brockschmidt M., Khademi M. Learning to Represent Programs with Graphs // In proceedings of the 6th International Conference on Learning Representations (Vancouver, Canada, 20 April–3 May 2018). 2018. PP. 1–17. DOI: 10.48550/arXiv.1711.00740.
24. Ormonova Je. M. Opredelenie kachestva programmnogo produkta na osnove teorii grafov // Nauka. Obrazovanie. Tehnika. 2021. № 1(70). S. 37–44.
25. Totuhov K. E., Romanov A. Ju., Luk'janov V. I. Issledovanie jeffektivnosti raboty geneticheskih algoritmov s razlichnymi metodami skreshhivanija i otbora // Jelektronnyj setevoj politematicheskij zhurnal «Nauchnye trudy KubGTU». 2022. № 6. S. 98–109.
26. Domanov K. I. Sravnitel'nyj analiz jeffektivnosti raboty geneticheskogo algoritma pri modifikacii operatora mutacii v zadache kommivojazhera // Politehnicheskij molodezhnyj zhurnal. 2022. № 1(66). DOI: 10.18698/2541-8009-2022-1-760.

МЕРЫ ПРОТИВОДЕЙСТВИЯ ИСПОЛЬЗУЕМЫМ В ХОДЕ ПРОВЕДЕНИЯ КОМПЬЮТЕРНЫХ АТАК СТЕГАНОГРАФИЧЕСКИМ ТЕХНИКАМ

Анисимов Е. С.[1], Крылов Г. О.[2]

DOI: 10.21681/2311-3456-2025-4-107-116

Целью работы *является определение возможных путей противодействия использованию стеганографических техник при проведении компьютерных атак, включая разработку программного средства стеганографического анализа в качестве примера одного из решений.*

Метод исследования: *анализ основных сценариев применения стеганографии при проведении компьютерных атак; анализ основных методов стеганографического анализа и представленных в открытом доступе тестов безопасности; разработка и программная реализация программного средства стеганографического анализа изображений; экспериментальное исследование и оценка разработанного программного средства.*

Результат исследования: *сформулированы основные направления противодействия стеганографическим техникам проведения компьютерных атак. Разработано программное средство комплексного стеганографического анализа изображений с применением нейронной сети для определения наличия в объекте стеганографической LSB-вставки. Проведена оценка полученных результатов работы модели. Сформулированы направления дальнейших исследований, а также обозначены области практического применения результатов работы, в частности, по совершенствованию SIEM решений и DLP систем.*

Научная новизна *заключается в исследовании противодействия стеганографическим механизмам как техникам проведения компьютерных атак. Также предлагается вариант стеганографического анализа, основанный на комплексном применении нескольких методов анализа.*

Ключевые слова: *стеганография, стеганографический анализ, MITRE ATT&CK, SIEM, DLP, компьютерные атаки, нейронные сети, машинное обучение, Cyber Kill Chain.*

Введение

Одним из ориентиров для многих средств защиты информации при определении их функциональных возможностей является MITRE ATT&CK – матрица с описанием техник, применяемых нарушителями в рамках компьютерных атак. В данной матрице отдельные техники представлены идентификаторами с описанием их содержания, что дополнено ссылками на информационные материалы с упоминаниями их реального применения атакующими. Таким образом, для некоторой компьютерной атаки допустимо осуществить декомпозицию на отдельные элементы, каждый из которых является определённой техникой из матрицы. Подобное разделение атак на части несколько упрощает задачи противодействия им, поскольку в отношении одной техники проще выделять факторы её обнаружения (обнаружения последствий её использования нарушителями), а также зачастую легче локализовать фрагмент проведения атаки и принять меры противодействия. При общем взгляде на MITRE ATT&CK в контексте некоторого средства защиты информации допустимой является оценка охвата матрицы данным средством с позиции его назначения. Например, для SIEM системы охват отдельной техники из MITRE ATT&CK означает наличие у решения возможности обнаружить в сети организации активность атакующих, придерживающихся в составе атаки данной техники.

В связи с этим один из возможных путей совершенствования средств защиты информации – их дополнение с целью увеличения покрытия техник MITRE ATT&CK. То есть для некоторого существующего средства проводится оценка его текущих функциональных возможностей, которые соотносятся с техниками MITRE ATT&CK, после чего особое внимание обращается на не входящие в область покрытия техники. В отношении данных техник определяются инструменты их идентификации или противодействия, на основании чего проводится работа по улучшению средства защиты информации.

Ранее в работе [2] отмечалась присутствующая у нарушителей возможность применения стеганографических методов для организации скрытых коммуникаций в рамках проводимых компьютерных атак. Такая возможность представляет особый интерес

1 Анисимов Ефим Сергеевич, магистр, Финансовый университет при Правительстве Российской Федерации, Москва, Россия. E-mail: EAnisimov_Sci@mail.ru, ORCID: 0000-0002-2632-4439.

2 Крылов Григорий Олегович, доктор физико-математических наук, профессор, Военный университет им. кн. Александра Невского Министерства обороны Российской Федерации, Национальный исследовательский ядерный университет «МИФИ», Финансовый университет при Правительстве Российской Федерации, Москва, Россия. E-mail: op50@mail.ru, ORCID: 0000-0001-8145-1994.

по причине высокой сложности идентификации соответствующей злонамеренной активности нарушителей в сети организации, осуществляемой подобным образом. Для описания данных действий сообществом введены идентификаторы T1001.002, T1027.003, T1406.001 в составе техник нарушителей из матрицы MITRE ATT&CK. Обзор некоторых российских средств защиты информации в части покрытия ими матрицы MITRE ATT&CK показал преобладающее отсутствие реализаций, позволяющих полностью покрыть отмеченные «стеганографические» техники: в лучших случаях найдены примеры охвата только техники T1027.003. Данное наблюдение подтверждает актуальность исследования вопросов обнаружения и противодействия использованию нарушителями стеганографии для различных действий. Другая работа [1] также содержит обзор применений стеганографии как элемента компьютерных атак. Соответственно, в рамках настоящей работы авторами преследуется цель оценки использования нарушителями отмеченных техник компьютерных атак и определения возможных мер противодействия.

Стеганографические механизмы как элемент компьютерных атак

Прежде проведём оценку потенциала применения стеганографических алгоритмов для достижения различных целей атакующих. Если рассматривать компьютерные атаки в разрезе известной семиэтапной модели Cyber Kill Chain проведения атак, то основные шаги нарушителей, на которых имеется некоторый обмен информацией между сетью атакуемого объекта и внешним миром:

- доставка вредоносного программного обеспечения – 3 этап;
- организация канала связи и установление соединения с командно-управляющим Сервером (Command and Control, C2 Сервер) – 6 этап;
- осуществление целевых действий (создание утечки защищаемой информации и пр.) – 7 этап.

Соответственно, стоит оценить: возможно ли применять стеганографию для выполнения отмеченных мероприятий в составе атаки. Обзор источников показал, что нарушители действительно применяют стеганографические методы для реализации обозначенных элементов атак.

Например, источники [5, 12, *Staged Malware*[3]] содержат примеры или описание инструментов сокрытия некоторого вредоносного материала в контейнерах-изображениях посредством таких стеганографических методов, как LSB и сокрытие с помощью LSB в коэффициентах DCT (дискретного косинусного преобразования). Данные примеры также доказывают наличие возможности использования стеганографии в целях доставки вредоносного программного обеспечения.

Таким образом, показано наличие возможностей использования стеганографических методов в рамках проведения компьютерных атак. Данные возможности подтверждаются и реальными примерами, составленными специалистами по результатам расследований произошедших атак. Перечислим несколько из подобных примеров в качестве подтверждения представленного тезиса:

- обзор методов работы APT-группировки (APT – Advanced Persistent Threat, целевая продолжительная атака повышенной сложности) Turla показал[4], что потенциальный арсенал для их функционирования включает, среди прочего, техники MITRE ATT&CK T1001.002 и T1027.003;
- атака группировки MuddyWater содержала[5] действия по использованию техники T1027.003: нарушители скрывали обфусцированный исходный код на JavaScript в медиа-файле temp.jpg;
- серия атак с использованием трояна Necro.N, в рамках которых применялись[6] стеганографические механизмы для сокрытия вредоносной нагрузки в изображениях (специалистами соотносится с техникой T1406.001) для осуществления её загрузки с C2 Сервера на заражённое устройство.

В дополнение, без взаимосвязи с реальными случаями данный факт возможности использования стеганографии в рамках компьютерных атак косвенно был подтверждён выше посредством обозначения соответствующих техник в матрице MITRE ATT&CK, которые, в свою очередь, не формализуются и не вносятся в матрицу без веских оснований существования.

Основные подходы противодействия стеганографическим техникам

Представим возможные пути противодействия нарушителям, действующим в обозначенных формах. Основных подходов к решению такой проблемы можно выделить два:

Использование средств комплексного стеганографического анализа, которые будут применяться для

3 Steganography based Staged Malware 001. [Электронный ресурс] URL: https://github.com/Shauryae1337/Steganography-based-Staged-Malware. (дата обращения: 18.04.2025).

4 LunarWeb and LunarMail: The Secret Weapons of the Turla APT. Hiveforce Labs. [Электронный ресурс] URL: https://www.hivepro.com/wp-content/uploads/2024/05/LunarWeb-and-LunarMail-The-Secret-Weapons-of-the-Turla-APT_TA2024191.pdf. (дата обращения: 21.04.2025).

5 Iranian Government-Sponsored Actors Conduct Cyber Operations Against Global Government and Commercial Networks. CISA. [Электронный ресурс] URL:https://www.cisa.gov/news-events/cybersecurity-advisories/aa22-055a. (дата обращения: 21.04.2025).

6 The Mobile Malware Chronicles: Necro.N – Volume 101. Zimperium. [Электронный ресурс] URL: https://zimperium.com/blog/the-necro-n-chronicles-volume-101. (дата обращения: 21.04.2025).

анализа передаваемых файлов на предмет оценки возможного наличия в них стеганографической вставки;

Настройка системы защиты объекта по результатам учений информационной безопасности на основе проведения различных тестов, имитирующих реализацию интересующих с позиции противодействия им техник атакующих (совершенствование системы защиты на основе взаимодействия Red и Blue Team).

Средство стеганографического анализа

Относительно применения средств стеганографического анализа стоит отметить, что они не являются универсальным и безошибочным решением проблемы применения стеганографии в рамках компьютерных атак. Однако их и нельзя игнорировать, напротив, они представляют собой практически полезное дополнение функционирующих на объекте защиты средств в составе систем обеспечения информационной безопасности (в особенности систем предотвращения утечек – DLP-систем – и SIEM решений). Авторами в контексте исследования вопроса противодействия использованию стеганографических методов при проведении компьютерных атак было разработано программное средство, которое предназначено для проведения комплексного стеганографического анализа медиа-файлов с целью идентификации наличия в них стеганографической вставки или её отсутствия. Представим описание структуры данного программного обеспечения с приведением краткой характеристики основных достигнутых параметров работы.

Специалистами разработано множество различных методов стеганографического анализа, и данное направление продолжает развиваться, в частности, с уделением особого внимания применению моделей машинного обучения для данных задач. Подобная тенденция в развитии стеганографического анализа обусловлена среди прочего и тем, что нарушители в последнее время активно используют генеративные модели для сокрытия некоторой вставки в не готовых изображениях, а в генерируемом контейнере. Разработки подобных методов описаны, например, в работах [6, 13]. Подобные изменения требуют более активного развития направлений стеганографического анализа, в особенности с применением искусственного интеллекта. В свою очередь, ряд специалистов отмечает, что научное сообщество стеганографическому анализу уделяет меньшее внимание нежели стеганографическим методам сокрытия [7].

Авторами при создании средства стеганографического анализа была сформирована гипотеза о предпочтительности совместного применения нескольких методов стеганографического анализа с точки зрения повышения точности определения вероятного наличия стеганографической вставки в исследуемом объекте. В связи с этим, учитывая широкое использование метода LSB (Least Significant Bit – стеганографический метод сокрытия данных в младших значащих битах), был проведён обзор основных методов стеганографического анализа изображений на предмет обнаружения в них вставок, осуществлённых именно методом LSB. По итогам анализа содержания источников [3, 8, 9, 10] в качестве основных методов для дальнейшей реализации были выбраны следующие:

- регулярный-сингулярный анализ (Regular or Singular, RS-анализ);
- метод анализа подобия значений пискелей (Pixel Similarity Weights — PSW);
- метод хи-квадрат на основе статистического анализа пар значений.

Отметим ключевые особенности указанных методов стеганографического анализа.

RS-анализ строится на функции гладкости (регулярности) и функции флиппинга. Функция регулярности f однозначно неопределена в рамках метода: она может быть дисперсией значений внутри данного блока изображения (анализируемое изображение разбивается на квадратные блоки G: nxn) или суммой разностей значений пикселей (1):

$$f(G) = \sum_{i=1}^{n-1} |x_{i+1} - x_i|. \quad (1)$$

Флиппинг-функция должна обладать свойством $F(F(x)) = x$, то есть она инволютивна. Основными вариантами функции флиппинга для изображений являются функция инверсии младшего бита значения пикселя (F_1) и такая же инверсия, только с переносом в старший бит (F_{-1}):

$$F_1 = 0 - 1;\ 2-3;\ 4-5;\ \ldots;\ 252-253;\ 254-255,$$
$$F_{-1} = 1 - 2;\ 3-4;\ 5-6;\ \ldots;\ 253-254;\ 255-0.$$

Соответственно, применение к отдельному блоку изображения функции гладкости и флиппинг-функции даст различные результаты, что ложится в основу RS-стеганоанализа. Определённое соотношение результатов функций от данного блока позволяет различать три класса блоков:

- неиспользуемые группы U: это блоки $G \in U \iff$ $\iff f(F(G)) = f(G)$;
- регулярные группы R: блоки $G \in R \iff f(F(G)) =$ $= f(G)$;
- сингулярные группы S: блоки $G \in S \iff f(F(G)) =$ $= f(G)$.

В рамках RS стеганографического анализа аналитик обращает внимание на количество блоков, которые попали в группы R и S, причём отдельно

рассчитывается такое значение для флиппинг-функции F_1 и отдельно для F_{-1}. Соответственно, такие количества блоков в группах при разных флиппинг-функциях можно обозначить как R_{cnt} и S_{cnt}, R_{-cnt} и S_{-cnt}.

RS-метод стеганографического анализа основывается на предположении пренебрежимо малых отличий между количествами блоков в сингулярной и регулярной группах независимо от применённой при расчётах флиппинг-функции в случае «чистого», обычного изображения, то есть: $R_{cnt} \approx R_{-cnt}$ и $S_{cnt} \approx S_{-cnt}$.

Если же для анализируемого изображения данное соотношение не выполняется, использование другой флиппинг-функции вносит значительное расхождение в количество блоков в соответствующих группах, то делается вывод о высокой вероятности факта проведения над изображением стеганографического преобразования методом LSB.

Дополнительно скажем, что RS-метод стеганографического анализа может иметь дополнения и расширения, в частности, с его помощью можно проводить оценки длины стеганографической вставки.

Метод анализа PSW — это пример метода статистического стеганографического анализа, в основе которого лежит вычисление определённым образом доли идентичных пикселей рассматриваемому. Для рассматриваемого пикселя окружающие пиксели, размещённые в соответствующем квадрате, расположенном от текущего пикселя не дальше, чем на 3 (в некоторых работах предлагается до 4 — глубина анализа строго не ограничивается, однако следует соблюдать баланс между объёмом анализа и вычислительной нагрузкой метода), выделяются окружающие его кольца. То есть для рассматриваемого пикселя с координатами (i_0,j_0) в группу пикселей Z_d, где $d \leq 3$, попадут все пиксели с координатами $(i,j): max(|i - i_0|,|j - j_0|) \leq d$. Таким образом, если анализируемый пиксель расположен не вблизи границ изображения, то область Z_1 будет содержать 8 пикселей, область Z_2 – 16 и Z_3 – 24. Наглядно идея такого разделения областей близких пикселей показана на рисунке 1.

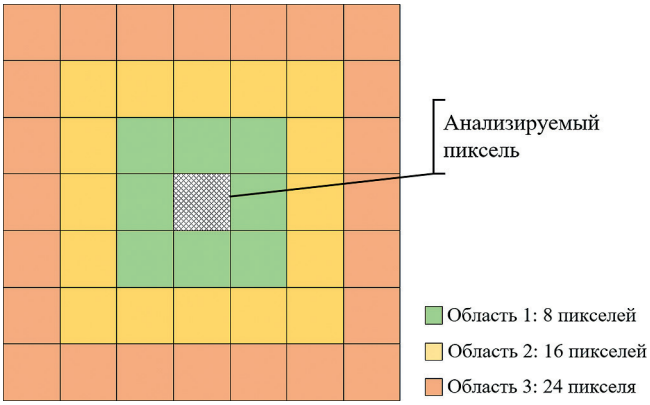


Рис. 1. Области близких пикселей для определения доли идентичных анализируемому пикселей в рамках метода PSW (Источник: составлено автором по материалам[7])

В каждой из таких соседствующих зон подсчитывается количество одинаковых с рассматриваемым пикселей. В соответствии с обозначенной идеей обработки изображения формируются наборы значений для дальнейшего анализа, которые образуют основу для второй и не менее важной части метода стеганографического анализа PSW — метода опорных векторов (Support Vector Machine — SVM). SVM – это алгоритм машинного обучения для решения задач классификации/регрессии. В рамках второй части метода первостепенное значение имеет качественная подготовка обучающих выборок для модели. При хорошем подборе примеров изображений, охватывающих как можно большее число сценариев размещения стеганографической вставки и различных изображений без вставки с соответствующими значениями целевой переменной, такая задача классификации посредством SVM решается достаточно хорошо. Таким образом, метод стеганографического анализа PSW предполагает не только статистический анализ изображений на предмет наличия в них скрытых данных, но и машинное обучение, основывающееся на предварительном анализе. Данный подход даёт большую гибкость и адаптивность метода к различным стеганографическим техникам изначального сокрытия данных в контейнере-изображении.

Рассматривая классический подход LSB, предполагающий изменение только одного, самого младшего бита, можно обратить внимание, что преобразование значений байта соответствующего цветового канала происходит в рамках ограниченного набора пар. Например, из значения $(27)_{10} = (11011)_2$ метод LSB может получить значения $(27)_{10} = (11011)_2$ или $(26)_{10} = (11010)_2$ — других возможных результатов по итогам корректировки наименьшего значащего бита нет. В связи с этим заранее возможно обозначить пары значений пикселей, которые имеют место в рамках LSB преобразования, визуально идея таких пар представлена на рисунке 2.

Соответственно, всего имеется 128 пар значений представленного вида (можно также сказать, что число 128 соответствует количеству компонент сильной связности ориентированного графа, в котором 256 вершин, соответствующих значениям от 0 до 255 включительно; а дуги – отражают возможные результаты изменения наименьшего значащего бита). Данное наблюдение (PoV – Pairs of pixel

7 Chaeikar S., Zamani M., Manaf A. B., Zeki A. M. PSW statistical LSB image steganalysis // Multimedia Tools and Applications. 2018. № 77. pp. 805–835. https://doi.org/10.1007/s11042-016-4273-6.

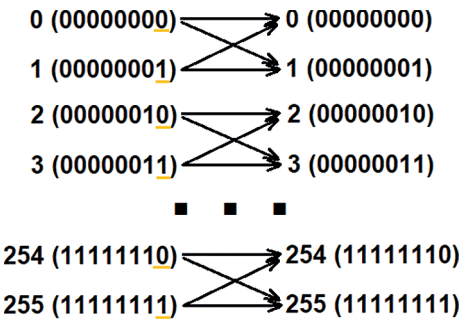


Рис. 2. Возможные пары-изменения значений пикселей в LSB

Values) используется в рамках метода хи-квадрат. Для каждой из 128 пар можно подсчитать количество пикселей, соответствующих конкретной паре. Более формально такое количество c_p для пары $(2p, 2p+1)$, где $p \in [0,1,...,127]$ определяется формулой (2):

$$c_p = \frac{|pixel_{value} \in [2p, 2p+1]|}{2}. \quad (2)$$

Для целей метода стеганографического анализа применяются метрики хи-квадрат, а именно подсчитывает хи-квадрат с $p-1$ степенями свободы (3):

$$\chi^2_{p-1} = \sum_{i=1}^{p} \frac{(c_i - c_i^*)^2}{c_i^*}. \quad (3)$$

Теперь готовы все значения для расчёта вероятности P наличия скрытой вставки (4):

$$P = 1 - \frac{1}{2^{\frac{p-1}{2}}\Gamma\frac{p-1}{2}} \int_0^{\chi^2_{(p-1)}} e^{x/2}\, x^{\frac{p-1}{2}-1}\, dx. \quad (4)$$

От полученной вероятности отталкиваются вердикты данного хи-квадрат метода стеганографического анализа.

В соответствии с содержанием методов была осуществлена их программная реализация на языке программирования Python. Каждый из методов составил основу отдельных модулей программного средства, возвращаемым значением которых является некоторое значение вероятности наличия в переданном на анализ в метод изображении LSB-вставки. По предположению, объект анализируется каждым из трёх методов, в результате чего формируется вектор из трёх значений, который служит для формирования матрицы признаков, соответствующей используемому датасету из изображений. По готовности такой матрицы из трёх признаков и соответствующего ей вектора со значениями целевой переменной (0 – стеганографическая вставка отсутствует, 1 – имеется LSB вставка) имеется датасет для обучения и тестирования нейронной сети (для реализации выбрана из пакета Keras). Данный подход к построению программного средства определил его архитектуру, в общем виде которая показана на рисунке 3.

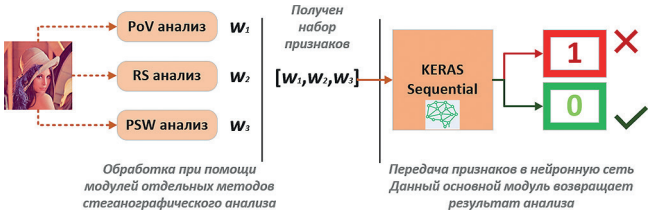


Рис. 3. Архитектура разработанного программного средства

Исходный датасет изображений, использованный для подготовки описанной матрицы признаков с результатами стеганографического анализа каждым из методов в отдельности, выбран из источника[8]. Описание данного датасета представлено далее по тексту в таблице 3. По результатам его обработки реализованными модулями PoV (метод хи-квадрат), RS и PSW анализа получена матрица размерности 10000x3 вместе с вектор-столбцом значений булевой целевой переменной. В целом, полученные элементы матрицы содержат подтверждения гипотезы о возможности различных результатов стеганографического анализа одного объекта разными методами. Вычисленные значения свидетельствуют об этом, в связи с чем потенциально возможно достижение синергетического эффекта в сценарии совместного применения различных методов. По причине необходимости получения однозначного результата для отдельного изображения имеющиеся три соответствующих значения-признака определённым образом требовалось свернуть в булево значение целевой переменной. Для этих целей выбрана нейронная сеть с последовательным расположением слоёв из пакета Keras (Keras Sequential). Конечно, пересчёт можно было сделать математически на основе введения коэффициентов и расчёта интегрального показателя, однако при выборе финальной основной модели авторами принималась во внимание возможность дальнейшего её масштабирования (в том числе посредством добавления других методов стеганографического анализа), для чего более применимы нейросетевые решения.

Прокомментируем достигнутые результаты созданного программного средства. До момента окончательного выбора моделей из пакета Keras в качестве основного решения проводилось их сравнение с наиболее сильными представителями не-нейросетевых решений, а именно LightGBM и XGBoost – модели градиентного бустинга над решающими деревьями. В таблице 1 содержатся значения метрик Precision и F1, полученные по результатам тестирования каждой из сравниваемых моделей на тестовой выборке после их обучения.

8 Stego-Images-Dataset. Kaggle. [Электронный ресурс] URL: https://www.kaggle.com/datasets/marcozuppelli/stegoimagesdataset. (дата обращения: 21.04.2025).

Таблица 1.

Показатели работы основного модуля, формирующего результат комплексного стеганографического анализа изображения

Модель	Время обучения, с	Время теста, с	Точность (precision)	F1 мера
LightGBM	0,053773	0,006964	0,676190	0,633333
XGBoost	0,065580	0,004790	0,676190	0,633333
Keras Sequential	73,067876	0,365499	0,762100	0,686217

Представленные значения дают подтверждение предпочтительности выбора нейросетевых решений в качестве основного модуля программного средства, возвращающего некоторое интегральное значение стеганографического анализа на основе результатов отдельных методов, причём не только в качестве отвечающих задачам возможного масштабирования модели, но и по достигнутым в ходе работы показателям.

Приведённое описание позволяет зафиксировать несколько основных выводов о результатах разработки программного средства комплексного стеганографического анализа. Начальная созданная модель характеризуется относительно позитивными результатами работы. Показатель Precision на уровне 76,21% для исследуемой нетривиальной задачи идентификации в анализируемом изображении стеганографической LSB-вставки является весьма хорошим.

Таблица 2.

Существующие тесты безопасности для имитации элементов компьютерных атак с использованием стеганографических методов

Техника	Источники тестов безопасности и описание
T1001.002	Тест[9] имитирует сокрытие tar архива в изображении. Тест[10] имитирует сокрытие в изображении скрипта PowerShell с помощью стеганографических методов. Тест[11] имитирует исполнение встроенного в изображение (при помощи стеганографии) Shell скрипта с предварительным его кодированием в BASE64.
T1027.003	Точечный тест[12], его можно считать примером сигнатурного стеганографического анализа, который направлен на обнаружение применения программного обеспечения steghide для сокрытия информации. Авторами данный тест аудита отмечен как соответствующий технике атаки T1027.003, однако область применения представляется практически более широкой, относящийся в то же время и к двум другим рассматриваемым стеганографическим техникам. [11] В данной научной статье авторами также проведена работа, идейно близкая к сигнатурному стеганографическому анализу. Данные материалы не являются тестом, готовым к прямому применению в практических системах, но в то же время представленные в работе результаты формируют ориентир для элементов объектов проверки, на которые следует обращать внимание при анализе. Хотя данное исследование и проводилось применительно к задачам компьютерной криминалистики (форензики), его материалы представляют практическую значимость для задач стеганографического анализа в ходе процессов обнаружения и противодействия компьютерным атакам.
T1406.001	Не найдены тесты безопасности, предназначенные для отработки техники матрицы MITRE ATT&CK с явно указанным идентификатором.

9 T1001.002 – Data Obfuscation via Steganography. Atomic Test #1 – Steganographic Tarball Embedding. Atomic Red Team. [Электронный ресурс] URL: https://www.atomicredteam.io/atomic-red-team/atomics/T1001.002#atomic-test-1---steganographic-tarball-embedding (дата обращения: 21.04.2025).

10 T1001.002 – Data Obfuscation via Steganography. Atomic Test #2 – Embedded Script in Image Execution via Extract-Invoke-PSImage. Atomic Red Team. [Электронный ресурс] URL: https://www.atomicredteam.io/atomic-red-team/atomics/T1001.002#atomic-test-2---embedded-script-in-image-execution-via-extract-invoke-psimage (дата обращения: 21.04.2025).

11 T1001.002 – Data Obfuscation via Steganography. Atomic Test #3 – Execute Embedded Script in Image via Steganography. Atomic Red Team. [Электронный ресурс] URL: https://www.atomicredteam.io/atomic-red-team/atomics/T1001.002#atomic-test-3---execute-embedded-script-in-image-via-steganography (дата обращения: 21.04.2025).

12 Sigma – Generic Signature Format for SIEM Systems. lnx_auditd_steghide_embed_steganography.yml. [Электронный ресурс. (дата обращения: 21.04.2025).] URL: https://github.com/SigmaHQ/sigma/blob/master/rules/linux/auditd/lnx_auditd_steghide_embed_steganography.yml

Применение тестов безопасности

По направлению организации проведения тестов безопасности, имитирующих конкретные техники атакующих, стоит отметить наличие примеров конкретных тестов безопасности, которые подготовлены различными крупными командами. В таблице 2 собраны ссылки на тесты безопасности, которые могут использоваться Red Team организации для имитации соответствующих «стеганографических» техник из матрицы MITRE ATT&CK.

Результаты обзора источников различных категорий свидетельствуют о недостаточном охвате различных сценариев применения рассматриваемых стеганографических техник MITRE ATT&CK. В этой связи видится практически полезным накопление материалов, а также в целом использование готовых датасетов для формирования разнообразных примеров медиа-файлов, которые будут в таком случае являться подобным тестом безопасности. Решением для обработки этих материалов может являться модуль в виде модели машинного обучения или нейронной сети, который будет обучаться на подготовленных материалах, результатом же такого обучения являются обновления моделей, способные учитывать ситуации, аналогичные проведённым тестам.

Примерами доступных в открытом доступе датасетов подобного рода, подходящих для тестов, соответствующих названным в настоящей работе техникам матрицы MITRE ATT&CK с основой в виде стеганографических методов, являются следующие наборы данных, собранные в виде таблицы 3.

Выше представлен, конечно, не исчерпывающий перечень допускающих использование для целей настроек моделей датасетов, однако достаточный для получения начального материала работы по рассматриваемому направлению. В то же время не стоит забывать о возможности самостоятельной подготовки обучающих материалов аналогичного

Таблица 3.

Материалы для обучения моделей, предназначенных для идентификации медиа-файлов, содержащих стеганографическую вставку

Наименование	Описание датасета
Stego-Images-Dataset. N. Cassavia, L. Caviglione, M. Guarascio, G. Manco, M. Zuppelli	Содержит 44000 изображений, содержащих стеганографические вставки (по содержанию: ссылки URL, вредоносный исходный код, адреса контрактов блокчейн-платформы Ethereum и прочие), осуществлённые при помощи алгоритма LSB[13].
JPEG StegoChecker dataset. Mikołaj Płachta, Marek Krzemień, Krzysztof Szczypiorski, Artur Janicki.	Содержит 10000 чёрно-белых изображений, содержащих стеганографические вставки случайного содержания, осуществлённые при помощи алгоритмов J-Uniward, nsF5 и UERD[14].
Stego-Favicons-Dataset. M. Guarascio, M. Zuppelli, N. Cassavia, L. Caviglione, G. Manco.	Репозиторий состоит из двух датасетов: – первый содержит 360000 изображений, из которых 60000 – исходные изображения без вставок, 60000 – изображения со стеганографическими вставками вредоносного PHP кода, 240000 изображений со стеганографическими вставками URL; – второй содержит 90000 изображений, из которых 15000 – чистые исходные изображения, 15000 – изображения со стеганографическими вставками PHP кода в формате BASE64, 60000 – изображения со стеганографическими вставками URL. Стеганографические вставки сделаны при помощи стеганографического алгоритма LSB[15].
Прочие датасеты	Собрание ссылок на прочие датасеты также составлены в работах других авторов, например, примеры имеются в статье [4]

13 Stego-Images-Dataset. Kaggle. [Электронный ресурс (дата обращения: 21.04.2025).] URL: https://www.kaggle.com/datasets/marcozuppelli/stegoimagesdataset (дата обращения: 21.04.2025).

14 JPEG StegoChecker dataset. Kaggle. [Электронный ресурс (дата обращения: 21.04.2025)] URL: https://www.kaggle.com/datasets/h2020simargl/jpeg-stegochecker-dataset?select=gfr_cover.csv

15 Stego-Favicons-Dataset. GitHub. [Электронный ресурс] URL: https://github.com/Ocram95/Stego-Favicons-Dataset (дата обращения: 21.04.2025).

характера, которые в зависимости от постановки задач могут и превосходить указанные по вкладу, вносимому в конечное качество работы моделей. Например, улучшение может заключаться в использовании различных стеганографических алгоритмов для сокрытия вставок; формировании пула разнообразных по содержанию и характеристикам изображений, используемых в качестве контейнера, а также составлении наборов вставок вариативного содержания.

Заключение

В рамках настоящей статьи отмечена актуальность совершенствования систем обеспечения информационной безопасности в направлении учёта ими угроз, связанных с применением стеганографических методов сокрытия информации. Подобные техники имеют собственные идентификаторы в матрице MITRE ATT&CK, и они могут быть использованы для различных действий, в частности, для доставки вредоносного программного обеспечения, осуществления взаимодействия с C2 Сервером, выполнения целевых действий, организации утечки защищаемой информации. Латентная природа действий по передаче данных стеганографическими методами служит хорошим мотивом для взятия таких техник на вооружение нарушителями, поэтому осуществление указанного круга действий является актуальной угрозой, связанный с реализацией которой риск требует определённой обработки.

Теоретическая значимость работы заключается в исследовании направления стеганографического анализа, в отношении которого отмечают меньшее внимание со стороны сообщества. Полученные результаты вносят вклад и в прикладную науку, состоящий в определении возможностей противодействия стеганографическим техникам, применяемым при проведении компьютерных атак.

Исследование отражает несколько возможных путей противодействия стеганографическим техникам: проведение стеганографического анализа файлов и моделирование действий нарушителей, применяющих подобные техники, на основе которого проводится донастройка средств защиты информации, по итогам чего допустима отработка основных сценариев активности нарушителей, включённых в использованные тесты безопасности.

В рамках первой группы мер приведён пример разработки программного средства стеганографического анализа с обзором результатов его работы.

В части проведения тестов безопасности в работе авторами собраны примеры ссылок на применимые для этого материалы. Отметим, что в открытом доступе представлено незначительное количество тестов рассматриваемого класса, поэтому в этой части оптимальной может стать собственная подготовка материалов.

По мнению авторов, предложенные направления мер защиты и, в особенности, разработанное в рамках исследования программное средство стеганографического анализа имеют возможность получить практическое применение в рамках совершенствования как систем обеспечения информационной безопасности организаций в целом, так и применительно к улучшению отдельных классов средств защиты информации, среди которых в данном случае основными являются DLP решения и SIEM системы. Упомянем, что обзор возможностей ряда ведущих российских SIEM решений в контексте охвата ими техник матрицы MITRE ATT&CK отразил неполное покрытие ими рассматриваемых в настоящем исследовании техник T1001.002, T1027.003, T1406.001, что одновременно дополнительно подтверждает актуальность вопроса и служит подтверждением потенциала практической применимости полученных результатов. В свою очередь, положительный вклад в развитие обозначенных классов средств защиты информации, которые занимают далеко не последнее место в структуре продуктов сферы, будет способствовать улучшению состояния информационной безопасности Российской Федерации.

Литература

1. Клишин Д. В., Федосенко М. Ю. Применение стеганографии при осуществлении компьютерных атак на информационную инфраструктуру предприятий // Экономика и качество систем связи. 2024. № 2(32). с. 158–166.
2. Ревенков П. В., Анисимов Е. С. Утечка информации: классификация каналов и влияние на типичные банковские риски // В центре экономики. 2025. № 1. Т. 6. с. 1–6.
3. Apau R., Hayfron-Acquah J. B., Asante M., Twum F. A multilayered secure image steganography technique for resisting regular-singular steganalysis attacks using elliptic curve cryptography and genetic algorithm // International conference on ICT for sustainable development. Singapore: Springer Nature Singapore, 2023. с. 427–439. https://doi.org/10.1371/journal.pone.0308807.
4. Badar L. T., Carminati B., Ferrari E. A comprehensive survey on stegomalware detection in digital media, research challenges and future directions // Signal Processing. 2025. С. 109888. https://doi.org/10.1016/j.sigpro.2025.109888.
5. Chaganti R., Ravi V., Alazab M., Pham T. D. Stegomalware: A Systematic Survey of MalwareHiding and Detection in Images, Machine LearningModels and Research Challenges // arXiv preprint arXiv:2110.02504. 2021. https://doi.org/10.48550/arXiv.2110.02504.
6. Huo L., Chen R., Wei J., Huang L. A high-capacity and high-security image steganography network based on chaotic mapping and generative adversarial networks // Applied Sciences. 2024. 14 (3). с. 1225. https://doi.org/10.3390/app14031225.
7. Kombrink M. H., Geradts Z. J. M. H., Worring M. Image steganography approaches and their detection strategies: A survey // ACM Computing Surveys. 2024. № 57 (2). с. 1–40. https://doi.org/10.1145/3694965.

8. Lin W. B., Lai T. H., Chou C. L. Chi-square-based steganalysis method against modified pixel-value differencing steganography // Arabian Journal for Science and Engineering. 2021. Т. 46. №. 9. с. 8525–8533. https://doi.org/10.1007/s13369-021-05554-2.
9. Shankar D. D., Azhakath A. S. Random embedded calibrated statistical blind steganalysis using cross validated support vector machine and support vector machine with particle swarm optimization // Scientific Reports. 2023. № 13(1). с. 2359. https://doi.org/10.1038/s41598-023-29453-8.
10. Shehab D. A., Alhaddad M. J. Comprehensive survey of multimedia steganalysis: Techniques, evaluations, and trends in future research // Symmetry. 2022. № 14(1). с. 117. https://doi.org/10.3390/sym14010117.
11. Stefan Kiltz, Jana Dittmann, Fabian Loewe, Christian Heidecke, Max John, Jonas Mädel and Fabian Preißler. 2024. Forensic Image Trace Map for Image-Stego-Malware Analysis: Validation of the Effectiveness with Structured Image Sets. In Proceedings of 2024 ACM 12th ACM Workshop on Information Hiding and Multimedia Security (ACM IH&MMSEC'24), June 24–26, 2024, Baiona, Spain. ACM, New York, NY, USA, с. 6. https://doi.org/10.1145/3658664.3659659.
12. Strachanski F., Petrov D., Schmidbauer T., Wendzel S. A Comprehensive Pattern-based Overview of Stegomalware // Proceedings of the 19th International Conference on Availability, Reliability and Security. 2024. с. 1–10. https://doi.org/10.1145/3664476.3670886.
13. Volkhonskiy D., Nazarov I., Burnaev E. Steganographic generative adversarial networks // Twelfth international conference on machine vision (ICMV 2019). SPIE, 2020. Т. 11433. с. 991–1005. https://doi.org/10.48550/arXiv.1703.05502.

COUNTERMEASURES APPLICABLE FOR CYBERATTACK STEGANOGRAPHIC TECHNIQUES

Anisimov E.S.[16], Krylov G. O.[17]

Keywords: *steganography, steganalysis, MITRE ATT&CK, SIEM, DLP, cyberattack, neural networks, machine learning, Cyber Kill Chain.*

The purpose of the article *is defining of main countermeasures for steganographic techniques usage in cyberattacks, comprising the development of a steganalysis tool as an example of one of the ways.*

Research methods: *steganography in cyberattacks usage scenario analysis; steganalysis methods and available security tests review; image steganalysis software tool development; experimental evaluation of the developed tool.*

The result obtained: *main directions of countermeasures for cyberattack steganographic techniques were formulated in the article. Image complex steganalysis software tool using a neural network to LSB insertion detection in an object has been developed. The results of the developed tool were evaluated. The study reveals directions for further researches and main applications of the study's results, in particular, for improving DLP and SIEM solutions.*

The scientific novelty of the article is countermeasures for steganographic mechanisms as cyberattack techniques research. In the article proposed a steganalysis scenario based on several analysis methods combined usage.

References

1. Klishin D., Fedosenko M. The use of steganography in the implementation of computer attacks on the information infrastructure of enterprises. Economy and quality of communication systems. 2024; 2 (32). pp. 158–166.
2. Revenkov P. V., Anisimov E. S. Information Leak: Classification of Channels and Impact on Typical Banking Risks. In the Center of Economy. 2025; 1 (6). pp. 1–6.
3. Apau R., Hayfron-Acquah J. B., Asante M., Twum F. A multilayered secure image steganography technique for resisting regular-singular steganalysis attacks using elliptic curve cryptography and genetic algorithm // International conference on ICT for sustainable development. Singapore: Springer Nature Singapore, 2023. pp. 427–439. https://doi.org/10.1371/journal.pone.0308807.
4. Badar L. T., Carminati B., Ferrari E. A comprehensive survey on stegomalware detection in digital media, research challenges and future directions // Signal Processing. 2025. p. 109888. https://doi.org/10.1016/j.sigpro.2025.109888.
5. Chaganti R., Ravi V., Alazab M., Pham T. D. Stegomalware: A Systematic Survey of MalwareHiding and Detection in Images, Machine LearningModels and Research Challenges // arXiv preprint arXiv:2110.02504. 2021. https://doi.org/10.48550/arXiv.2110.02504.
6. Huo L., Chen R., Wei J., Huang L. A high-capacity and high-security image steganography network based on chaotic mapping and generative adversarial networks // Applied Sciences. 2024. 14(3). p. 1225. https://doi.org/10.3390/app14031225.
7. Kombrink M. H., Geradts Z. J. M. H., Worring M. Image steganography approaches and their detection strategies: A survey // ACM Computing Surveys. 2024. № 57(2). pp. 1–40. https://doi.org/10.1145/3694965.
8. Lin W. B., Lai T. H., Chou C. L. Chi-square-based steganalysis method against modified pixel-value differencing steganography // Arabian Journal for Science and Engineering. 2021. V. 46. №. 9. pp. 8525–8533. https://doi.org/10.1007/s13369-021-05554-2.
9. Shankar D. D., Azhakath A. S. Random embedded calibrated statistical blind steganalysis using cross validated support vector machine and support vector machine with particle swarm optimization // Scientific Reports. 2023. 13 (1). pp. 2359. https://doi.org/10.1038/s41598-023-29453-8.

16 Efim S. Anisimov, master's degree, Financial University under the Government of the Russian Federation, Moscow, Russia. E-mail: EAnisimov_Sci@mail.ru, ORCID: 0000-0002-2632-4439.
17 Grigorii O. Krylov, Dr. Sc. (Phys.-Math.), Professor, Prince Alexander Nevsky Military University of the Ministry of Defense of the Russian Federation, National Research Nuclear University MEPhI, Financial University under the Government of the Russian Federation, Moscow, Russia. E-mail: op50@mail.ru, ORCID: 0000-0001-8145-1994.

10. Shehab D. A., Alhaddad M. J. Comprehensive survey of multimedia steganalysis: Techniques, evaluations, and trends in future research // Symmetry. 2022; 14 (1). p. 117. https://doi.org/10.3390/sym14010117.
11. Stefan Kiltz, Jana Dittmann, Fabian Loewe, Christian Heidecke, Max John, Jonas Mädel and Fabian Preißler. 2024. Forensic Image Trace Map for Image-Stego-Malware Analysis: Validation of the Effectiveness with Structured Image Sets. In Proceedings of 2024 ACM 12th ACM Workshop on Information Hiding and Multimedia Security (ACM IH&MMSEC'24), June 24–26, 2024, Baiona, Spain. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3658664.3659659.
12. Strachanski F., Petrov D., Schmidbauer T., Wendzel S. A Comprehensive Pattern-based Overview of Stegomalware // Proceedings of the 19th International Conference on Availability, Reliability and Security. 2024. pp. 1–10. https://doi.org/10.1145/3664476.3670886.
13. Volkhonskiy D., Nazarov I., Burnaev E. Steganographic generative adversarial networks // Twelfth international conference on machine vision (ICMV 2019). SPIE, 2020. V. 11433. pp. 991–1005. https://doi.org/10.48550/arXiv.1703.05502.

АНАЛИЗ ПРОБЛЕМЫ ФОРМИРОВАНИЯ НАБОРА СРЕДСТВ ЗАЩИТЫ ИНФОРМАЦИИ В РАДИОКАНАЛАХ РОБОТОТЕХНИЧЕСКИХ КОМПЛЕКСОВ

Головской В. А.[1]

DOI: 10.21681/2311-3456-2025-4-117-126

Цель работы *– провести анализ проблематики автоматизированного оценивания достаточности некриптографических средств защиты информации в радиоканалах радиосистем передачи данных робототехнических комплексов.*

Методы исследования: *общенаучные методы – анализ, дедуктивный вывод, методы системного анализа и теории алгоритмов, применение сопутствующих абстракций потенциальной реализуемости и актуальной бесконечности.*

Результат исследования: *предложен подход к формализации проблем в области информационной безопасности в виде конструктивных объектов, применение которого позволило сформировать соответствующие массовые проблемы защиты информации в радиоканалах радиосистем передачи данных робототехнических комплексов и оценить их на предмет алгоритмической разрешимости. Предложено использовать описание средств защиты информации через совокупность нетривиальных семантических свойств алгоритмов, ими управляющих. Данный подход обеспечивает возможность абстрагироваться от особенностей реализации средств защиты информации и использовать такие описания в составе конструктивных объектов при работе алгоритмов оценивания достаточности и выбора оптимального набора средств защиты информации в радиоканалах. Предложена гипотеза о взаимосвязи указанных массовых проблем, для проверки которой сформулирована и доказана теорема.*

Научная значимость: *представленные результаты формируют основания для исследования вычислительных аспектов задачи построения эффективного алгоритма формирования набора средств защиты информации в радиоканалах робототехнических комплексов. Предложенное описание средств защиты информации совокупностью нетривиальных семантических свойств управляющих ими алгоритмов обеспечивает возможность адекватного учета их сущностного содержания и значимых для решения задачи особенностей без необходимости рассмотрения их программной или аппаратно-программной реализации.*

Ключевые слова: *Алгоритм, алгоритмическая проблема, конфиденциальность, криптографическая защита информации, массовая проблема, машина Тьюринга, моделирование, угроза, проблема эквивалентности.*

Введение

В настоящее время требования практики к сокращению сроков разработки и поставки потребителю робототехнических комплексов (РТК) обострились критически. При этом, несмотря не существующие успехи в развитии теории и практики испытаний, наблюдаемые тенденции к увеличению функционала и интеллектуализации РТК обуславливают усложнение этапов проверки их заявленных свойств [1, 2]. Обусловленная указанными факторами проблема построения объяснимого искусственного интеллекта актуальна и для исследований вопросов защиты информации (ЗИ) [3]. Наличие подсистем криптографической ЗИ (КЗИ) в радиоканалах радиосистем передачи данных (РС) РТК обуславливает как снижение эффективной скорости передачи информации, так и увеличение массогабаритных характеристик робототехнических средств (РТС), сроков разработки таких РТК, как надсистемы, за счет необходимости проведения ряда обязательных исследований[2] и, соответственно, стоимости РТК. Также немаловажным моментом является необходимость выполнения ряда организационно-технических мероприятий при подготовке к применению РТК, содержащих подсистемы КЗИ, что в критических условиях является осложняющим фактором [4]. Указанные аспекты в совокупности с повышенным вниманием к безопасности информации в РТК [5, 6] актуализируют известную научно-техническую проблему [7–9] формирования оптимального набора средств ЗИ (СЗИ) в условиях ограниченности ресурсов. Дополнительную остроту этой проблеме придает существенное влияние человеческого фактора при оценивании угроз безопасности информации и принятии соответствующих решений на основании полученных оценок [10]. Так, методический документ[3] декларирует, что «независимо от результата формирования экспертной группы при оценке угроз безопасности информации существуют субъективные факторы,

1 Головской Василий Андреевич, кандидат технических наук, доцент, Краснодарское высшее военное училище имени генерала армии С.М. Штеменко, г. Краснодар, Россия. E-mail: golovskoy_va@mail.ru

2 Положение о разработке, производстве, реализации и эксплуатации шифровальных (криптографических) средств защиты информации (Положение ПКЗ-2005). Утверждено приказом ФСБ РФ от 9 февраля 2005 г. N 66.

3 Методический документ «Методика оценки угроз безопасности информации». Утвержден ФСТЭК России 5 февраля 2021 г.

связанные с психологией принятия решений. Это также может приводить как к занижению (ослаблению), так и к завышению (усилению) экспертами прогнозов и предположений при оценке угроз безопасности информации, что в свою очередь может привести к пропуску отдельных угроз безопасности информации или к неоправданным затратам на нейтрализацию неактуальных угроз».

Ввиду указанных выше аспектов вопрос о возможности обеспечения безопасности информации в радиоканалах не криптографическими СЗИ (НКСЗИ) инициирует исследования по анализу возможностей и синтезу НКСЗИ [3, 11–16], а также – по оцениванию их достаточности для обеспечения защищенности информации [9, 17–19].

Вопросы отнесения различных не криптографических средств обработки и преобразования информации и сигналов, ее переносящих, предназначенных для обеспечения конфиденциальности, целостности и доступности информации, к классу СЗИ являются дискуссионными [7, 17] ввиду важности дефиниции терминологического аппарата ЗИ, как и вопросы оценивания эффективности таких средств относительно ставших для ряда задач ЗИ традиционными криптографических методов [11]. При этом ввиду высокой практической значимости предметом многих исследований являются различные подходы, призванные обеспечить отдельные составляющие защищенности информации – конфиденциальность, целостность, доступность – в том числе, при передаче ее в радиоканалах [9, 11–20]. Необходимо отметить, что открытость разделяемого радиосистемами ресурса – радиочастотного спектра, который становится одновременно все более доступным для информационно-технических воздействий ввиду развития средств информационного противоборства, как интенсивного, так и экстенсивного, и все более загруженным [21], обостряет проблему защиты передаваемой по радиоканалам информации [7] в условиях антагонистического информационного конфликта.

Предлагаемая статья является развитием предложенной ранее [8] идеи построения алгоритма оценивания достаточности СЗИ и описания проблем ЗИ как конструктивных объектов со структурой, обусловленной предложенным для решения поставленной задачи теоретико-алгоритмическим подходом. Данный подход имеет достаточную историю[4] эффективного применения при исследованиях проблем информационной безопасности и обеспечивает возможность абстрагироваться от способа реализации анализируемых СЗИ.

4 Cohen F. Computational aspects of computer viruses // Computers & Security. – 1989. – Vol. 8, No. 4. P. 297–298.

1. Постановка задачи

Задача исследования возможностей создания конструктивного подхода и программы оценивания достаточности набора СЗИ в радиоканале, позволяющих минимизировать влияние субъективности экспертов, была предложена в [8]. Однако рассмотрение СЗИ при этом осуществлялось с позиций анализа лишь классов методов, составляющих функциональное наполнение СЗИ, а также остались нерассмотренными вопросы проверки гипотезы, доказательства необходимости и достаточности наборов входных данных и ряд других. В работе [21] была предложена модель антагонистического информационного конфликта и рассмотрено ее наполнение, в основном, данными о возможностях средств радиомониторинга и анализа его результатов, однако не был представлен анализ проблематики в широкой постановке и не определен перечень необходимых данных о РС РТК. Настоящая статья является промежуточным результатом исследования, имеющего цель – построение научно-методического аппарата формирования набора СЗИ, передаваемой по радиоканалам РС РТК. Цель статьи – провести анализ проблематики автоматизированного оценивания достаточности НКСЗИ в радиоканалах РС РТК. Объект исследования – функционирующая в условиях антагонистического информационного конфликта РС РТК, по радиоканалам которой передается подлежащая защите информация ограниченного распространения. Предмет исследования – обеспечение конфиденциальности информации, передаваемой в РС РТК по радиоканалам в условиях антагонистического информационного конфликта. Ограничение вопросов ЗИ только обеспечением ее конфиденциальности обусловлено сформировавшимися на практике актуальными угрозами безопасности информации, передаваемой по радиоканалам РС РТК определенного типа [4, 22].

Под РТК далее понимается автоматизированная система, являющаяся надсистемой для входящих в ее состав следующих подсистем [4, 21]: группа РТС, РС, сопряженная с системой ЗИ, включающей подсистему КЗИ, а также пункт управления. Рис. 1 призван проиллюстрировать принятую с учетом сформулированных объекта и предмета исследования декомпозицию РТК на подсистемы, осуществленную на логическом уровне.

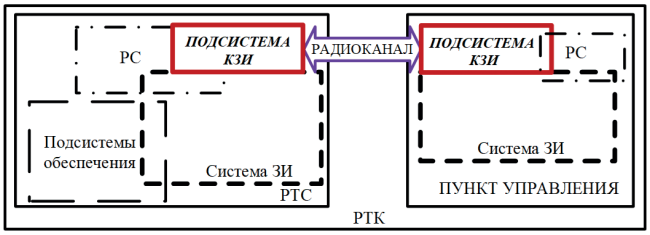


Рис. 1. Декомпозиция РТК на подсистемы

С учетом приведенных формальных элементов исследования сформулированы следующие ограничения:

- рассматривается применение средств КЗИ (СКЗИ) непосредственно и только для обеспечения конфиденциальности передаваемой в радиоканале информации, и не рассматриваются вопросы использования криптографических протоколов при аутентификации абонентов и прочие, формирующие актуальные уязвимости [4, 9, 23] в системе ЗИ;
- рассматриваются только вопросы разрешимости массовых проблем без анализа класса сложности разрешающего их алгоритма;
- по радиоканалам рассматриваемой РС передается только один вид информации, с одной меткой конфиденциальности;
- рассматривается наличие только внешнего нарушителя;
- не рассматриваются характеристики РТК, однако анализируемая проблематика наиболее актуальна для беспилотных летательных аппаратов.

К СКЗИ будем относить аппаратные, программные и аппаратно-программные средства, реализующие алгоритмы криптографического преобразования информации[5].

2. Формализация проблемы защиты информации в радиоканалах

Рассмотрим место проблемы обеспечения конфиденциальности информации среди множества традиционно рассматривающихся задач ЗИ.

С учетом цели работы используем методологию системного анализа, обеспечивающего адекватность предложений [8] по наполнению элементами проблематики ЗИ следующего известного[6] формализма, позволяющего построить модель M интересующего явления ISP:

$$M = \langle ISP, SM, T, IM, \Lambda \rangle, \quad (1)$$

где ISP – множество неформализованных проблем ЗИ, выступающих в качестве объекта-оригинала; SM – субъект моделирования, т.е. осуществляющая автоматизированное оценивание сущность, нуждающаяся в формальном описании проблемы ЗИ [8]; T – цель моделирования – обеспечение SM информацией, необходимой для формирования требуемого набора СЗИ [8]; IM – инфраструктура моделирования; Λ – язык описания, представляющий собой искомое отображение объекта в модель

$$\begin{cases} \Lambda: \Pi_k \longrightarrow P_k, \\ A_{SM}(P_k) = answer, \end{cases}$$

где $\Pi_k \in ISP$ – k-я неформализованная проблема ЗИ; $P_k = code(\Pi_k)$ – запись Π_k, обеспечивающая ее эффективную программную обработку за счет согласованности применяемых формализмов $code(\bullet)$ и A_{SM}; A_{SM} – алгоритм, используемый SM для вычисления формального ответа $answer$ о характеристиках требуемого набора СЗИ $W_h = [w_j, \ldots, w_p]$, обеспечивающего защищенность информации в P_k, $h = \overline{1, N_H}$, N_H – количество рассматриваемых СЗИ, $|W_h| \geq 1$.

Необходимо отметить, что цель моделирования T может иметь различные формулировки. Первоначальный запрос от практики был в обеспечении достаточности набора СЗИ [8], что будет соответствовать цели T_1. Однако первым автором работы [5] при обсуждении доклада [8] предложена иная формулировка цели, предполагающая формирование оптимального набора СЗИ, которой будет соответствовать T_2. При оптимизации набора СЗИ могут быть рассмотрены такие параметры СЗИ как пропускная способность, величина «накладных расходов» при шифровании, вычислительная эффективность, массогабаритные и температурные характеристики, стоимость и т.д. Далее будут последовательно рассмотрены проблемы, соответствующие обеим представленным формулировкам цели из (1).

В докладе [8] было предложено для анализа проблем ЗИ Π_k при конструировании Λ использовать теоретико-алгоритмический подход, характеризующийся формализацией любой массовой проблемы P_k конструктивным объектом – языком этой проблемы L_{P_k}, что позволило сформировать гипотезу об адекватности оценивания достаточности СЗИ согласно T_1 с позиций алгоритмической разрешимости массовых проблем P_k, соответствующих Π_k.

Под массовой или алгоритмической проблемой далее понимается класс однотипных задач, для которых необходимо найти единый разрешающий их алгоритм [2, 8]. Проблема P считается разрешимой в алгоритмическом смысле при существовании разрешающего ее язык $L_P = \{(data_j, decision_m)_k\}$ алгоритма $A_P(code(data_j)) = decision_m$, где $code(data_j) \in \Sigma^*$ обозначает непустое слово, кодирующее данные $data_j$ в алфавите Σ рассматриваемой машины Тьюринга (МТ) Θ_P, формализующей алгоритм A_P, или эквивалентного ей формализма. При этом A_P вычисляет слово $decision_m$, получив на вход $code(data_j)$.

Использование принятого теоретико-алгоритмического подхода позволяет с привлечением характерных для него абстракций при рассмотрении

5 МР 26.2.006-2021. Методические рекомендации «Информационная технология. Криптографическая защита информации. Термины и определения». Технический комитет по стандартизации ТК 26 «Криптографическая защита информации», М.: 2021 г. – 87 с.

6 Волкова В. Н., Козлов В. Н., Магер В. Е., Черненькая Л. В. Классификация методов и моделей в системном анализе // Сборник докладов XX Международной конференции по мягким вычислениям и измерениям. – СПб.: СПбГЭТУ(ЛЭТИ). 2017. – С. 223–226.

работающих по алгоритмам СЗИ пренебречь особенностями их реализации – аппаратная, программная или аппаратно-программная. Это обеспечивает адекватность представления СЗИ w_p как соответствующего набора нетривиальных семантических свойств алгоритма A_{w_p}, по которому функционирует w_p, т.е. отображение из (1) имеет вид $w_p \xrightarrow{\Lambda} \{s_r\}_p$, $r = \overline{1,N_R^p}$, где N_R^p – количество нетривиальных семантических свойств s_r, исчерпывающе описывающих алгоритм A_{w_p} функционирования СЗИ w_p. В общем случае предполагается, что для функционально различных СЗИ w_p, w_f и w_j мощности соответствующих множеств s_r не равны, т.е. $N_R^p \neq N_R^f \neq N_R^f$. Принято также допущение, что известны нетривиальные семантические свойства, соответствующие всем характеристикам анализируемых СЗИ и их наборам.

Использование предложенного отображения $w_p \xrightarrow{\Lambda} \{s_r\}_p$ позволяет повысить уровень конструктивности моделирования относительно [8] и сформулировать массовую проблему проверки задач, разрешаемых СКЗИ, на разрешимость их и НКСЗИ. С учетом сформулированного объекта исследования и содержания антагонистического информационного конфликта [21] массовую проблему P_S оценивания обеспечения защищенности информации в радиоканалах РС РТК опишем следующим языком:

$$L_S = \{(code(RS_i),code(D_j),code(ENW_q))_k, \\ code(W_h^l),dec_m\},\ n = \overline{1,N_K}, \quad (2)$$

где RS, $i = \overline{1,N_I}$ – модель i-й РС, реализующей передачу информации, подлежащей защите, между РТС по радиоканалу в условиях среды ENW_q, $q = \overline{1,\infty}$; D_j – достаточное описание угроз безопасности передаваемой по радиоканалу информации, инициируемых антагонистической стороной информационного конфликта, $j = \overline{1,N_D}$; W_h^l – набор используемых в RS_i СЗИ l-го класса, $l \in \{1,2\}$; $dec_m \in \{0,1\}$ – решение об эффективности набора СЗИ W_h^l в RS_i, $m = \overline{1,2}$. Рассматриваемые классы СЗИ – СКЗИ (К) при $l = 1$ и НКСЗИ (Ф) при $l = 2$ – не пересекаются при формировании набора СЗИ W_h^l, т.е. $\mathrm{K} \cap \Phi = \varnothing$.

Разрешимость массовой проблемы P_S с языком (2) означает потенциальное существование разрешающего ее алгоритма A_{L_S}, который для каждого набора $((code(RS_i),code(D_j),code(ENW_q))_k,code(W_h^l))$ дает бинарный ответ dec_m на вопрос об эффективности ЗИ, передаваемой в RS_i, с помощью набора СЗИ W_h^l при информационном конфликте с D_j в условиях ENW_q. Такая интерпретация P_S соответствует цели T_1. С учетом привлечения абстракций, принятых в теории алгоритмов, разрешающий L_S алгоритм описывается таким выражением:

$$A_{L_S}((code(RS_i),code(D_j),code(ENW_q))_k, \\ code(W_h^l)) = dec_m.$$

Представляется логичным связать с возможностью построения программы оценивания эффективности ЗИ в РС РТК следующее высказывание:

$$S^{IS} \Longleftrightarrow \exists A_{L_S}: A_{L_S}((code(RS_i),code(D_j),code(ENW_q))_k, \\ code(W_h^l)) = dec_m. \quad (3)$$

С учетом цели работы выделим из множества проблем (2), разрешаемых СКЗИ, т.е. при $l = 1$, собственное подмножество – класс проблем P_C обеспечения конфиденциальности передаваемой в RS_i информации, формализуемый языком L_C, что иллюстрирует рис. 2. Также выделен из $\mathrm{K} \cup \Phi$ подкласс СЗИ C, обеспечивающих ее конфиденциальность, т.е. $C \subseteq \{\mathrm{K},\Phi\}$. С учетом принятых при формировании (2) обозначений выделенное собственное подмножество $L_C \subseteq L_S$, представляющее особый интерес для настоящей работы, формализуемо языком следующей структуры:

$$L_C = \{(code(RS_i),code(D_j),code(ENW_q))_k, \\ code(C),dec_m\}. \quad (4)$$

$$L_C = \{(RS_i, D_j, ENW_q)_n, C, dec\}$$

$$L_S = \{(RS_i, D_j, ENW_q)_k, W_h^l, dec_m\}$$

Рис. 2. Вложенность массовых проблем ЗИ

С учетом вложенности $C \subseteq \{\mathrm{K},\Phi\}$ и $L_C \subseteq L_S$ очевидно, что при разрешимости проблемы L_S будет разрешимой и проблема L_C.

По аналогии с (3) свяжем с возможностью построения программы оценивания эффективности обеспечения применяемыми СЗИ конфиденциальности информации в радиоканале RS_i разрешимость массовой проблемы (4):

$$S^C \Longleftrightarrow \exists A_{L_C}: A_{L_C}((code(RS_i),code(D_j),code(ENW_q))_k, \\ code(C)) = dec_m, \quad (5)$$

где A_{L_C} – разрешающий L_C алгоритм.

Теория и практика показывают, что у множества L_C существует подмножество L_K, для которого конфиденциальность передаваемой информации обеспечивается СКЗИ. Тогда при разрешимости проблемы L_C будет разрешимой и ее подпроблема, формализуемая языком

$$L_K = \{(code(RS_i),code(D_j),code(ENW_q))_k, \\ code(K),1\}. \quad (6)$$

Выделим из $L_C \subseteq L_S$ подмножество проблем $L_\Phi \subseteq L_C$, разрешаемых НКСЗИ, т.е. содержащее такие наборы $(code(RS_i),code(D_j),code(ENW_q))_k = \alpha_k$, для которых

конфиденциальность передаваемой по радиоканалам RS_i информации обеспечивается при использовании набора СЗИ $W_h^{\Phi} \in \Phi$:

$$L_{\Phi} = \{\alpha_k, code(W_h^{\Phi}), 1\}, \quad (7)$$

где $j = \overline{1, N_{D_C}}$, N_{D_C} – количество детализированных моделей нарушителя, способного нарушить конфиденциальность передаваемой в радиоканале информации, $N_{D_C} < N_D$.

В работе [7] было показано и практика подтверждает, что существует подмножество проблем $P_{\text{КФ}} \subseteq P_C$, характеризующееся тем, что конфиденциальность передаваемой по радиоканалам RS_i информации обеспечивается, как СКЗИ, так и НКСЗИ, формализуемое

$$L_{\text{КФ}} = L_{\text{К}} \cap L_{\Phi}. \quad (8)$$

Рис. 3 иллюстрирует предлагаемое выделение подпроблем $L_{\text{К}}$, L_{Φ} и $L_{\text{КФ}}$ из массовой проблемы $L_C \subseteq L_S$. При разрешимости проблемы L_C будут разрешимыми проблемы $L_{\text{К}}$ и $L_{\text{КФ}}$ ввиду их вложенности в L_C. Однако разрешимость $L_{\text{КФ}}$ не может рассматриваться даже как необходимое условие для разрешимости L_C.

Теперь для рассмотрения возможности автоматизации оценивания достаточности СЗИ и выбора их подходящего набора $W_h^l \in \Phi$ проведем анализ проблемы, формализуемой языком (8), на предмет определения ее алгоритмической разрешимости. Известно, что при использовании дедуктивного подхода для автоматического построения решения задачи необходимым является доказательство существования ее решения.

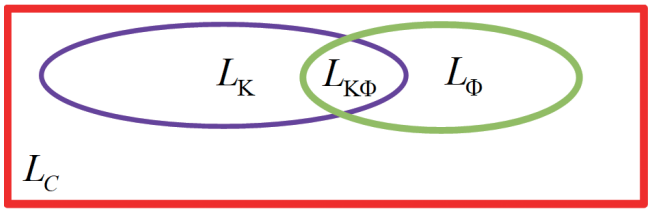


Рис. 3. Выделение подпроблем из массовой проблемы L_C

Разрешимость $L_{\text{КФ}}$ означает, что для любого набора α_k существует алгоритм $A_{L_{\text{КФ}}}(\alpha_k, \beta_h) = dec_m$, где $\beta_h = code(W_h^l)$, dec_m – ответ на вопрос об обеспечении конфиденциальности информации как с использованием СКЗИ ($W_h^{\text{К}}$), так и с использованием НКСЗИ (W_h^{Φ}). По аналогии с (3) и (5) с разрешимостью свяжем высказывание

$$S^{\text{КФ}} \Longleftrightarrow \exists A_{L_{\text{КФ}}} : A_{L_{\text{КФ}}}(\alpha_k, \beta_h) = dec_m. \quad (9)$$

С учетом алгоритмической природы β_h, как входа для предложенного $A_{L_{\text{КФ}}}$, очевидно, что истинность высказывания (9) согласована с разрешимостью вариации алгоритмической проблемы распознавания эквивалентности МТ. Необходимо отметить, что несмотря на известную неразрешимость ставшей классической массовой проблемы распознавания функциональной эквивалентности МТ, известны [24] варианты указанной проблемы, разрешимые за счет конкретизации вычислительной модели, содержания программы МТ и ее поведения.

Одним из направлений по конструированию разрешимых подпроблем неразрешимых массовых проблем является разрабатываемая в рамках генерического подхода [25] методология формирования такого подмножества множества входов, на элементах которого алгоритм остановится всегда. При этом за счет снижения требования массовости разрешающего алгоритма возможно обеспечение его приемлемой сложности. Однако сильная конкретизация делает проблему менее массовой, что обуславливает снижение ее интереса для практики. Отсюда возникают две задачи: определения условий баланса «массовость/разрешимость» и выбор наилучшей вычислительной модели. С учетом приведенных аргументов будем считать, что для массовой и интересной для практики версии (8) существует $A_{L_{\text{КФ}}}$, обеспечивающий истинность высказывания (9).

Рассмотрим алгоритмическую проблему, соответствующую цели T_2 из (1). В результате проведенного анализа сформулирована следующая **Гипотеза**: существование $A_{L_{\text{КФ}}}$ является необходимым условием существования алгоритма $A_{\Phi_C}(\alpha_k, B_c) = \beta_h^O$, разрешающего массовую проблему P_{Φ_C} определения оптимального набора СЗИ, формализуемую языком

$$L_{\Phi_C} = \{(\alpha_k, B_c)_y, \beta_h^O\},$$

где $B_c = \beta_0 \# \beta_1 \# \ldots \# \beta_{N_H - 1} \# \beta_{N_H}$, $\beta_h^O = code(\tilde{W}_h^l)$, $\tilde{W}_h^l \in \{W_h^l\}_{N_H}^{h=1}$ – оптимальный в определенном смысле набор СЗИ, # – служебный символ, предназначенный для разделения слов на ленте МТ. Свяжем с разрешимостью L_{Φ_C} высказывание

$$S^{\Phi_C} \Longleftrightarrow \exists A_{\Phi_C} : A_{\Phi_C}(\alpha_k, B_c) = \beta_h^O. \quad (10)$$

Для проверки гипотезы сформулирована **Теорема**: Взаимосвязь массовых проблем P_{Φ_C} и $P_{\text{КФ}}$ описывается высказыванием $S^{\Phi_C} \longrightarrow S^{\text{КФ}}$.

Доказательство. Для доказательства будем использовать широко применяемую в исследованиях массовых проблем [2, 8] технику сведения известной проблемы к исследуемой. Построены представленные на рис. 4а и 4б соответственно МТ $\Theta_{\text{КФ}}$, формализующая описанный выше алгоритм $A_{L_{\text{КФ}}}$, и ее модификация $\Theta_{\text{КФ}}^M$, заключающаяся в том, что $\Theta_{\text{КФ}}^M$ печатает только слова β_h, на которых $\Theta_{\text{КФ}}$ дает положительный ответ. Также построена представленная на рис. 4, в МТ Θ_{sort}, осуществляющая сортировку последовательности слов β_h в надслове

B_c по заданному параметру, соответствующему конкретной задаче оптимизации при цели T_2.

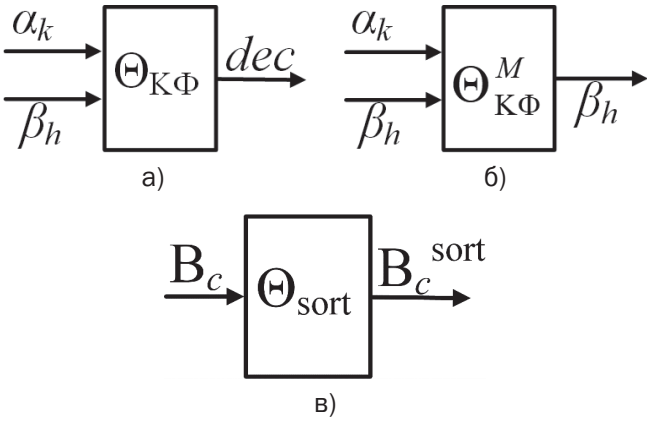


Рис. 4. Схемы машины Тьюринга, разрешающей проблему $L_{КФ}$ (а), ее модификация (б) и машины Тьюринга, разрешающей проблему сортировки (в)

Для доказательства теоремы построена схема сведения массовой проблемы $P_{КФ}$, приведенная на рис. 5, где между описанными МТ $\Theta^M_{КФ}$ и Θ_{sort} помещен буфер в виде ленты классической МТ, накапливающий все выходы $\Theta^M_{КФ}$. МТ Θ_{sort} начинает считывать с ленты надслово B_C только после останова $\Theta^M_{КФ}$, и, переработав B_C, печатает только слово β^O_h, соответствующее оптимальному набору СЗИ $\tilde{W}^l_h$. Очевидно, что ввиду известной разрешимости массовой проблемы сортировки из существования алгоритма $\Theta^M_{КФ}$ следует существование алгоритма A_{Φ_C} $(\alpha_k,B_c) = \beta^O_h$. Теорема доказана.

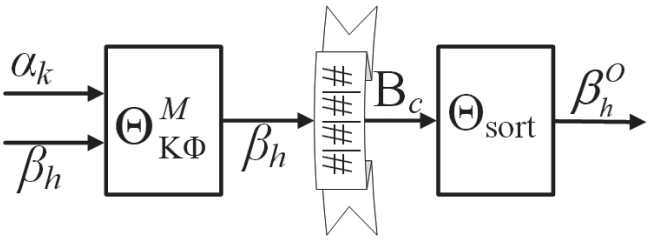


Рис. 5. Схема сведения $P_{КФ}$ к проблеме P_{Φ_C}

3. Обсуждение результатов

Алгоритм A_{Φ_C} позволит осуществлять интерпретируемое и объективно обоснованное формирование набора W^l_h СЗИ, обрабатываемой в РС РТК в условиях антагонистического информационного конфликта. Массовость в этом контексте будет означать инвариантность A_{Φ_C} к конкретному содержанию алгоритма функционирования СЗИ [2] и особенностям его реализации. В результате анализа содержания L_{Φ_C} сформулирована гипотеза об адекватности представления соответствующей проблемы P_{Φ_C} в терминах задачи о рюкзаке. Также важным фактором является правовая плоскость, и для практической разрешимости проблемы L_{Φ_C} представляется целесообразным нормативно определить новый тип информации, учитывающий зависимость от времени ее ценности и ограничения в распространении.

Сложность задачи построения эффективного алгоритма $A_{L_{КФ}}$, разрешающего интересную в практическом смысле соответствующую массовую проблему $P_{КФ}$, обуславливается необходимостью формирования требований к содержанию входа для $A_{L_{КФ}}$, поскольку эффективность любого алгоритма неразрывно связана с характеристиками входных для него данных [2]. При формировании требований к конкретному содержанию (α_k,β_h) должна быть обеспечена согласованность описаний RS_i, D_j, ENW_q и W^l_h еще до применения к ним операции $code(\bullet)$.

В работе [21] был предложен подход к наполнению D_j, ENW_q, однако наиболее сложным представляется вопрос конструирования Λ, обеспечивающего при описании W^l_h учет алгоритмом $A_{L_{КФ}}$ сущностного содержания НКСЗИ и криптографических СЗИ, имеющих подчас принципиальные различия, в том числе и на физическом уровне [7, 18]. Предложенный в настоящей работе подход к описанию СЗИ w_p набором нетривиальных семантических свойств $\{s_r\}_p$ алгоритма A_{w_p} позволит использовать в вычислениях в качестве исходных данных и криптоалгоритмы, как показано в работе[7], что полностью согласуется с возможностью построения алгоритмов над алгоритмами, обеспечиваемой формализмом МТ. Однако такой подход из-за принятой абстракции не будет учитывать некоторые особенности, обусловленные именно реализацией криптоалгоритмов в конкретных СКЗИ, что является при верификации криптографических протоколов недостатком [23], однако на данном этапе исследования считается приемлемым для решаемой задачи.

Рассмотрим некоторые предложенные ранее [8] требования к содержанию RS_i, которые призваны обеспечить согласованность с D_j и возможность построения интересной для практики применения РТК [4, 22] версии соответствующей массовой проблемы $P_{КФ}$. Раздельное описание РС РТК и обрабатываемой в ней информации, обусловленное необходимостью защиты именно информации (inf) в РС (rs), с учетом представленной детализации может быть формализовано на уровне грамматики как $RS_i = (rs\#inf)_i$. Малое, относительно других этапов жизненного цикла, время «активного существования» РТК в совокупности с высокой изменчивостью

7 Goldwasser S., Kalai Y. T., Popa R. A., Vaikuntanathan V., Zeldovich N. How to Run Turing Machines on Encrypted Data // Advances in Cryptology – CRYPTO 2013. CRYPTO 2013. Lecture Notes in Computer Science, 2013. Vol. 8043. pp. 536–553, Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-40084-1_30

условий функционирования РТК обуславливают ограниченность времени [4], в течение которого информация в радиоканалах РС РТК может снять имеющуюся неопределенность и обеспечить эффективность выполнения задачи надсистемой, включающей РТК. Таким образом подтверждается тезис [8] о необходимости учета ценности информации в ее описании (inf), что может быть формализовано как $inf = object_v\#atr_f\#\gamma_l$, где $object_v$ – обозначение v-го объекта, данные о котором содержатся в inf, $v = \overline{1,N_V}$; atr_f – данные об атрибутах v-го объекта; γ_l – значения ценности информации inf, передаваемой по радиоканалам RS_i .

Необходимо отметить, что понятие ценности информации, введенное академиком А. А. Харкевичем[8] вместе с соответствующей мерой, активно используется в настоящее время при исследованиях различных научно-технических проблем [26, 27]. Ценность информации не является постоянной величиной в общем случае. Это косвенно подтверждает, как наличие процедур снижения грифа секретности и рассекречивания сведений, так и отнесение данных, считавшихся открытыми, к конфиденциальным.

С учетом приведенных выше аргументов для решения поставленной задачи необходимо построение вычислимой n-местной функции ценности информации $f^{\Gamma}(arg_1,\ldots,arg_n) = \gamma_l$, одним из аргументов которой является время t_2. Смещая акцент из области вычислимости функций в теоретико-алгоритмическую плоскость, последний тезис будет преобразован в задачу построения алгоритма $A_{\Gamma}(\alpha_k,t_z,\{arg_n\}) = \gamma_l$, разрешающего массовую проблему определения ценности информации в РС РТК, формализуемую языком $L_{\Gamma} = \{(\alpha_k,t_z,\{arg_n\})_x,\gamma_l\}$, $x = \overline{1,\infty}$, $n = \overline{1,N_N}$. Представленные предложения по учету ценности информации согласуется с положениями риск-ориентированного подхода [6].

8 Харкевич А. А. О ценности информации // Проблемы кибернетики, 1960. № 4. С. 53–57.

Выводы

Рассматриваемая проблема оценивания достаточности НКСЗИ обусловлена регуляторной политикой. В результате анализа проблемы осуществлена декомпозиция предварительно формализованной массовой проблемы оценивания достаточности СЗИ, передаваемой в радиоканалах РС РТК в условиях антагонистического информационного конфликта. Предложена и подтверждена гипотеза, имеющая как теоретическое, так и практическое значение.

Предложенное описание СЗИ совокупностью нетривиальных семантических свойств соответствующего алгоритма представляется адекватным для решаемой задачи, однако требует дальнейшего исследования.

Развитие результатов работы представляется целесообразным в следующих направлениях: исследование предложенного описания отображения $w_p \xrightarrow{\Lambda} \{s_r\}_p$ для корректной обработки результатов его применения; постановка задачи для T_2 в виде графовой модели, естественной для задачи о рюкзаке; определение необходимых и достаточных условий разрешимости L_{Φ_C} при разрешимости $L_{\text{КФ}}$; формирование интересной формулировки разрешимого L_{Γ} и конструирование эффективного алгоритма A_{Γ}.

Литература

1. Абросимов В. К. Принципы испытаний образцов вооружения, военной и специальной техники с реализацией технологии машинного обучения (полемические заметки) // Вооружение и экономика. 2024. № 2(68). С. 23–32.
2. Головской В. А. Анализ проблематики прогнозирования поведения когнитивных радиосистем // Радиотехника. 2024. Т. 88, № 12. С. 134–145.
3. Capuano N., Fenza G., Loia V., Stanzione C. Explainable Artificial Intelligence in CyberSecurity: A Survey // IEEE Access. 2022. vol. 10, pp. 93575–93600. DOI: 10.1109/ACCESS.2022.3204171.
4. Головской В. А., Винокуров А. В. Модель подсистемы выработки криптографических ключей системы защиты информации киберфизической системы // Известия ЮФУ. Технические науки. 2025. № 2 (244). С. 202–211. DOI: 10.18522/2311-3103-2025-2-202-211.
5. Pavlenko E. Y., Vasileva K. V., Lavrova D. S., Zegzhda D. P. Counteraction the cybersecurity threats of the in-vehicle local network // Journal of Computer Virology and Hacking Techniques. 2023. Vol. 19, No. 3. P. 399–408. DOI: 10.1007/s11416-022-00451-0.
6. Anagnostis I., Kotzanikolaou P., Douligeris C. Understanding and Securing the Risks of Uncrewed Aerial Vehicle Services // IEEE Access. 2025. vol. 13. pp. 47955–47995. DOI: 10.1109/ACCESS.2025.3549861.
7. Махов Д. С. Анализ некриптографических методов защиты информации в радиоканалах информационных систем // Вопросы кибербезопасности. 2024. № 1(59). С. 82–88. DOI: 10.21681/2311-3456-2024-1-82-88.
8. Головской В. А. Алгоритмические аспекты проблемы оценивания достаточности средств защиты информации // Перспективы безопасности – 2024: сборник материалов II НТК, посвященной информационной безопасности, Санкт-Петербург, 19–20 июня 2024 года. – Санкт-Петербург: ООО «Специальный Технологический Центр», 2024. С. 17–22.
9. Лэ В. Х., Комаров И. И., Привалов А. А., Пыркин А. А. Модель обеспечения непрерывности безопасного функционирования системы прослеживаемости качества продукции в условиях неустойчивой коммуникации // Научно-технический вестник информационных технологий, механики и оптики. 2024. Т. 24. № 6. С. 949–961. DOI: 10.17586/2226-1494-2024-24-6-949-961.
10. Варенница В. В., Марков А. С., Савченко В. В., Цирлов В. Л. Практические аспекты выявления уязвимостей при проведении сертификационных испытаний программных средств защиты информации // Вопросы кибербезопасности. 2021. № 5(45). С. 36–44. DOI: 10.21681/2311-3456-2021-5-36-44.

11. Pandey G. K., Gurjar D. S., Nguyen H. H, Yadav S. Security Threats and Mitigation Techniques in UAV Communications: A Comprehensive Survey // IEEE Access, 2022, vol. 10, pp. 112858–112897. DOI: 10.1109/ACCESS.2022.3215975.
12. Кукушкин С. С., Рубан Д. А., Козлов Е. В. Математическая модель и алгоритм формирования помехозащищённого сигнала синхронизации на основе использования составных кодовых конструкций псевдослучайных последовательностей // Двойные технологии. 2022. № 1(98). С. 40–45.
13. Глобин Ю. О., Финько О. А. Способ обеспечения имитоустойчивой передачи информации по каналам связи // Наукоемкие технологии в космических исследованиях Земли. 2020. Т. 12. № 2. С. 30–43. DOI: 10.36724/2409-5419-2020-12-2-30-43.
14. Басан Е. С., Прошкин Н. А., Силин О. И. Повышение защищенности беспроводных каналов связи для беспилотных летательных аппаратов за счет создания ложных информационных полей // Сибирский аэрокосмический журнал. 2022. Т. 23. № 4. С. 657–670. DOI: 10.31772/2712-8970-2022-23-4-657-670.
15. Tikhonov V., Taher A., Tikhonov S., Shulakova K., Hluschenko V., Chaika A. Turing Machine Development for High-Secure Data Link Encoding in the Internet of Things Channel // Proceedings of the 12th International Conference on Applied Innovations in IT (ICAIIT), 2024, Vol. 12, Iss. 1, pp. 1–10. DOI: 10.25673/1156354.
16. Gvozdeva I. G., Gromov A. S., Gvozdeva O. M. Development and implementation of the digital steganography method based on the embedding of pseudoinformation // Proceedings of the Institute for System Programming of the RAS. 2023. Vol. 35, No. 3. P. 63–70.
17. Белокопытов М. Л., Бянкин А. А., Алехин С. А. Способ защиты телеметрической информации при передаче в радиолиниях комплексов вооружения и военной техники // Вопросы оборонной техники. Серия 16: Технические средства противодействия терроризму. 2023. № 7-8(181-182). С. 81–87. DOI: 10.53816/23061456_2023_7-8_81.
18. Мальцев Г. Н., Матвеев С. А. Исследование защищенности системы командного радиоуправления подвижным объектом с использованием марковской модели преодоления нарушителем многоуровневой системы защиты информации // Труды Военно-космической академии имени А. Ф. Можайского. 2021. № 677. С. 153–163.
19. Ватрухин Е. М. Комплексная защита информации в каналах «земля-борт» // Вестник Концерна ВКО «Алмаз-Антей». 2020. № 4. С. 6–14. DOI: 10.38013/2542-0542-2020-4-6-14.
20. Манаенко С. С., Дворников С. В., Пшеничников А. В. Теоретические аспекты формирования сигнальных конструкций сложной структуры // Информатика и автоматизация. 2022. Т. 21, № 1. С. 68–94. DOI: 10.15622/ia.2022.21.3.
21. Головской В. А. Модель сложного информационного конфликта для робототехнических комплексов // Вопросы кибербезопасности. 2025. № 1 (65). С. 86–95. DOI: 10.21681/2311-3456-2025-1-86-95.
22. Бирюков П. А., Тимохин А. А., Макаренко С. И. Бригады сухопутных войск, вооруженные беспилотными летательными аппаратами: обоснование создания, предложения по их структуре, способам боевого применения и техническому обеспечению с учетом опыта специальной военной операции на Украине // Системы управления, связи и безопасности. 2024. № 2. С. 43–70. DOI: 10.24412/2410-9916-2024-2-043-070.
23. Бабенко Л. К., Писарев И. А. Язык PDA для динамического анализа криптографических протоколов // Вопросы кибербезопасности. 2020. № 5(39). С. 19–29. DOI: 10.21681/2311-3456-2020-05-19-29.
24. Zakharov V. A. Efficient Equivalence Checking Technique for Some Classes of Finite-State Machines // Automatic Control and Computer Sciences. 2021. Vol. 55, pp. 670–701. DOI: 10.3103/S014641162107018X.
25. Рыбалов А. Н. Генерически неразрешимые и трудноразрешимые проблемы // Прикладная дискретная математика. 2024. № 63. С. 109–116. DOI: 10.17223/20710410/63/7.
26. Чечин И. В., Маринин А. А., Новиков П. А., Диченко С. А., Самойленко Д. В. Комбинационное кодирование данных с учетом анализа ценности содержащейся информации // Проблемы информационной безопасности. Компьютерные системы. 2023. № 4(57). С. 31–41. DOI: 10.48612/jisp/mvrb-h5xa-xx1r.
27. Копкин Е. В., Деев В. В. Алгоритм построения оптимальной диагностической процедуры по показателю ценности информации на основе принципа максимума Понтрягина // Информация и космос. 2024. № 1. С. 65–72.

ANALYSIS OF THE PROBLEM OF FORMING A SET OF INFORMATION SECURITY TOOLS IN THE RADIO CHANNELS OF ROBOTIC COMPLEXES

Golovskoy V. A.[9]

Keywords: *algorithm, algorithmic problem, confidentiality, cryptographic protection of information, mass problems, Turing machine, simulation, threat, equivalence problems.*

The purpose of the work *is to analyze the problems of automated assessment of the sufficiency of non-cryptographic information security tools in radio channels of radio data transmission systems of robotic complexes.*

Research methods*: general scientific methods – analysis, deductive inference, methods of system analysis and theory of algorithms, application of related abstractions of potential realizability and actual infinity.*

The result of the study: *an approach to formalizing problems in the field of information security in the form of constructive objects is proposed, the application of which made it possible to form the corresponding massive problems*

9 Golovskoy Vasiliy Andreevich, Ph.D. (in Engineering sciences), Associate Professor, Krasnodar Higher Military School named after army general S. M. Shtemenko, Krasnodar, Russia. E-mail: golovskoy_va@mail.ru

of information protection in radio channels of radio data transmission systems of robotic complexes and evaluate them for algorithmic solvability. It is proposed to use the description of information security tools through a set of non-trivial semantic properties of the algorithms that control them. This approach provides an opportunity to abstract from the specifics of the implementation of information security tools and use such descriptions as part of constructive objects when using algorithms for assessing sufficiency and choosing the optimal set of information protection tools in radio channels. A hypothesis about the interrelation of these mass problems is proposed, for which a theorem is formulated and proved.

Scientific significance: *the presented results form the basis for the study of computational aspects of the task of constructing an effective algorithm for the formation of a set of information security tools in the radio channels of robotic complexes. The proposed description of information security tools by a set of non-trivial semantic properties of the algorithms controlling them provides the possibility of adequate consideration of their essential content and features significant for solving the problem without the need to consider their software or hardware-software implementation.*

References

1. Abrosimov V. K. Principy ispytanij obrazcov vooruzhenija, voennoj i special'noj tehniki s realizaciej tehnologii mashinnogo obuchenija (polemicheskie zametki) // Vooruzhenie i jekonomika. 2024. № 2(68). S. 23–32.
2. Golovskoj V. A. Analiz problematiki prognozirovanija povedenija kognitivnyh radiosistem // Radiotehnika. 2024. T. 88, № 12. S. 134–145.
3. Capuano N., Fenza G., Loia V., Stanzione C. Explainable Artificial Intelligence in CyberSecurity: A Survey // IEEE Access. 2022. vol. 10, pp. 93575–93600. DOI: 10.1109/ACCESS.2022.3204171.
4. Golovskoj V. A., Vinokurov A. V. Model' podsistemy vyrabotki kriptograficheskih kljuchej sistemy zashhity informacii kiberfizicheskoj sistemy // Izvestija JuFU. Tehnicheskie nauki. 2025. № 2 (244). S. 202–211. DOI: 10.18522/2311-3103-2025-2-202-211.
5. Pavlenko E. Y., Vasileva K. V., Lavrova D. S., Zegzhda D. P. Counteraction the cybersecurity threats of the in-vehicle local network // Journal of Computer Virology and Hacking Techniques. 2023. Vol. 19, No. 3. P. 399–408. DOI: 10.1007/s11416-022-00451-0.
6. Anagnostis I., Kotzanikolaou P., Douligeris C. Understanding and Securing the Risks of Uncrewed Aerial Vehicle Services // IEEE Access. 2025. vol. 13. pp. 47955–47995. DOI: 10.1109/ACCESS.2025.3549861.
7. Mahov D. S. Analiz nekriptograficheskih metodov zashhity informacii v radiokanalah informacionnyh sistem // Voprosy kiberbezopasnosti. 2024. № 1(59). S. 82–88. DOI: 10.21681/2311-3456-2024-1-82-88.
8. Golovskoj V. A. Algoritmicheskie aspekty problemy ocenivanija dostatochnosti sredstv zashhity informacii // Perspektivy bezopasnosti – 2024: sbornik materialov II NTK, posvjashhennoj informacionnoj bezopasnosti, Sankt-Peterburg, 19–20 ijunja 2024 goda. – Sankt-Peterburg: OOO «Special'nyj Tehnologicheskij Centr», 2024. S. 17–22.
9. Lje V. H., Komarov I. I., Privalov A. A., Pyrkin A. A. Model' obespechenija nepreryvnosti bezopasnogo funkcionirovanija sistemy proslezhivaemosti kachestva produkcii v uslovijah neustojchivoj kommunikacii // Nauchno-tehnicheskij vestnik informacionnyh tehnologij, mehaniki i optiki. 2024. T. 24. № 6. S. 949–961. DOI: 10.17586/2226-1494-2024-24-6-949-961.
10. Varenica V. V., Markov A. S., Savchenko V. V., Cirlov V. L. Prakticheskie aspekty vyjavlenija ujazvimostej pri provedenii sertifikacionnyh ispytanij programmnyh sredstv zashhity informacii // Voprosy kiberbezopasnosti. 2021. № 5(45). S. 36–44. DOI: 10.21681/2311-3456-2021-5-36-44.
11. Pandey G. K., Gurjar D. S., Nguyen H. H, Yadav S. Security Threats and Mitigation Techniques in UAV Communications: A Comprehensive Survey // IEEE Access, 2022, vol. 10, pp. 112858–112897. DOI: 10.1109/ACCESS.2022.3215975.
12. Kukushkin S. S., Ruban D. A., Kozlov E. V. Matematicheskaja model' i algoritm formirovanija pomehozashhishhjonnogo signala sinhronizacii na osnove ispol'zovanija sostavnyh kodovyh konstrukcij psevdosluchajnyh posledovatel'nostej // Dvojnye tehnologii. 2022. № 1(98). S. 40–45.
13. Globin Ju. O., Fin'ko O. A. Sposob obespechenija imitoustojchivoj peredachi informacii po kanalam svjazi // Naukoemkie tehnologii v kosmicheskih issledovanijah Zemli. 2020. T. 12. № 2. S. 30–43. DOI: 10.36724/2409-5419-2020-12-2-30-43.
14. Basan E. S., Proshkin N. A., Silin O. I. Povyshenie zashhishhennosti besprovodnyh kanalov svjazi dlja bespilotnyh letatel'nyh apparatov za schet sozdanija lozhnyh informacionnyh polej // Sibirskij ajerokosmicheskij zhurnal. 2022. T. 23. № 4. S. 657–670. DOI: 10.31772/2712-8970-2022-23-4-657-670.
15. Tikhonov V., Taher A., Tikhonov S., Shulakova K., Hluschenko V., Chaika A. Turing Machine Development for High-Secure Data Link Encoding in the Internet of Things Channel // Proceedings of the 12th International Conference on Applied Innovations in IT (ICAIIT), 2024, Vol. 12, Iss. 1, pp. 1–10. DOI: 10.25673/1156354.
16. Gvozdeva I. G., Gromov A. S., Gvozdeva O. M. Development and implementation of the digital steganography method based on the embedding of pseudoinformation // Proceedings of the Institute for System Programming of the RAS. 2023. Vol. 35, No. 3. P. 63–70.
17. Belokopytov M. L., Bjankin A. A., Alehin S. A. Sposob zashhity telemetricheskoj informacii pri peredache v radiolinijah kompleksov vooruzhenija i voennoj tehniki // Voprosy oboronnoj tehniki. Serija 16: Tehnicheskie sredstva protivodejstvija terrorizmu. 2023. № 7-8(181-182). S. 81–87. DOI: 10.53816/23061456_2023_7-8_81.
18. Mal'cev G. N., Matveev S. A. Issledovanie zashhishhennosti sistemy komandnogo radioupravlenija podvizhnym ob#ektom s ispol'zovaniem markovskoj modeli preodolenija narushitelem mnogourovnevoj sistemy zashhity informacii // Trudy Voenno-kosmicheskoj akademii imeni A. F. Mozhajskogo. 2021. № 677. S. 153–163.
19. Vatruhin E. M. Kompleksnaja zashhita informacii v kanalah «zemlja-bort» // Vestnik Koncerna VKO «Almaz-Antej». 2020. № 4. S. 6–14. DOI: 10.38013/2542-0542-2020-4-6-14.
20. Manaenko S. S., Dvornikov S. V., Pshenichnikov A. V. Teoreticheskie aspekty formirovanija signal'nyh konstrukcij slozhnoj struktury // Informatika i avtomatizacija. 2022. T. 21, № 1. S. 68–94. DOI: 10.15622/ia.2022.21.3.
21. Golovskoj V. A. Model' slozhnogo informacionnogo konflikta dlja robototehnicheskih kompleksov // Voprosy kiberbezopasnosti. 2025. № 1 (65). S. 86–95. DOI: 10.21681/2311-3456-2025-1-86-95.

22. Birjukov P. A., Timohin A. A., Makarenko S. I. Brigady suhoputnyh vojsk, vooruzhennye bespilotnymi letatel'nymi apparatami: obosnovanie sozdanija, predlozhenija po ih strukture, sposobam boevogo primenenija i tehnicheskomu obespecheniju s uchetom opyta special'noj voennoj operacii na Ukraine // Sistemy upravlenija, svjazi i bezopasnosti. 2024. № 2. S. 43–70. DOI: 10.24412/2410-9916-2024-2-043-070.
23. Babenko L. K., Pisarev I. A. Jazyk PDA dlja dinamicheskogo analiza kriptograficheskih protokolov // Voprosy kiberbezopasnosti. 2020. № 5(39). S. 19–29. DOI: 10.21681/2311-3456-2020-05-19-29.
24. Zakharov V. A. Efficient Equivalence Checking Technique for Some Classes of Finite-State Machines // Automatic Control and Computer Sciences. 2021. Vol. 55, pp. 670–701. DOI: 10.3103/S014641162107018X.
25. Rybalov A. N. Generic heski nerazreshimye i trudnorazreshimye problemy // Prikladnaja diskretnaja matematika. 2024. № 63. S. 109–116. DOI: 10.17223/20710410/63/7.
26. Chechin I. V., Marinin A. A., Novikov P. A., Dichenko S. A., Samojlenko D. V. Kombinacionnoe kodirovanie dannyh s uchetom analiza cennosti soderzhashhejsja informacii // Problemy informacionnoj bezopasnosti. Komp'juternye sistemy. 2023. № 4(57). S. 31–41. DOI: 10.48612/jisp/mvrb-h5xa-xx1r.
27. Kopkin E. V., Deev V. V. Algoritm postroenija optimal'noj diagnosticheskoj procedury po pokazatelju cennosti informacii na osnove principa maksimuma Pontrjagina // Informacija i kosmos. 2024. № 1. S. 65–72.

ПОДХОДЫ КВАНТОВОГО ОТЖИГА К ВЗЛОМУ ШИФРОВАНИЯ RSA

Холодов Я. А.[1], Саллум Х.[2], Агапова Н. А.[3]

DOI: 10.21681/2311-3456-2025-4-127-133

Цель исследования: *изучение трансформационного потенциала квантового отжига в решении проблемы простой факторизации.*

Метод(ы) исследования: *наш подход включает в себя всесторонний обзор последних экспериментальных прорывов и теоретических инноваций. В частности, мы анализируем такие методики, как память поверхностного кода, формулировки HUBO и QUBO, алгоритмы гамильтониана, зависящие от диапазона, модульные локально-структурированные методы встраивания и модифицированный метод таблицы умножения. Кроме того, для подтверждения наших выводов представлены предварительные эксперименты по генерации случайных чисел.*

Результат(ы) исследования: *исследование оценивает применение квантового отжига для факторизации простых чисел, показывая, что продвинутые техники отображения задач, такие как формулировки HUBO и QUBO, значительно повышают эффективность представления сложных задач факторизации на квантовом оборудовании. Примечательно, что включение памяти поверхностных кодов повышает стабильность состояний кубитов во время отжига, снижая количество ошибок и повышая точность вычислений. Исследование также демонстрирует, что алгоритмы гамильтониана, зависящие от диапазона, и модульные локально-структурированные методы встраивания способствуют оптимизации взаимодействия кубитов, обеспечивая более точное выполнение процесса факторизации. Представлен модифицированный метод таблицы умножения, обеспечивающий оптимизированную вычислительную стратегию, особенно эффективную для больших составных чисел. Предварительные эксперименты со случайными числами подтверждают теоретические выводы, указывая на то, что эти интегрированные методы позволяют повысить производительность по сравнению с традиционными подходами. В совокупности полученные результаты подчеркивают потенциал квантового отжига как надежной основы для решения сложных криптографических задач и закладывают основу для будущих исследований масштабируемых квантовых алгоритмов и аппаратных реализаций.*

Научная новизна: *эта работа объединяет несколько передовых методов квантового отжига для факторизации простых чисел, соединяя экспериментальные инновации с теоретическими разработками, чтобы предложить новую структуру, которая повышает эффективность криптографических вычислений.*

Ключевые слова: *Quantum Annealing, RSA, QUBO, Prime Factorization.*

Введение

Квантовые вычисления стали преобразующей областью, обладающей потенциалом превзойти классические вычислительные возможности в решении определенных классов задач. Недавние достижения в области квантовой коррекции ошибок, такие как продемонстрированные Google Quantum AI на их сверхпроводниковом процессоре Willow, свидетельствуют о значительном прогрессе в направлении практических квантовых вычислений. В их реализации использовались два запоминающих устройства на основе поверхностного кода, включая 101-кубитный код с расстоянием 7 и код с расстоянием 5 с декодированием в реальном времени. Эти реализации позволили достичь ключевых результатов, таких как подавление логических ошибок в 2,14 раза, увеличение времени жизни по сравнению с лучшим физическим кубитом и задержка декодирования в реальном времени, превышающая миллион циклов. [1]

В качестве теста производительности процессора Willow использовался эталонный тест Random Circuit Sampling (RCS). RCS, изначально разработанный Google, стал стандартом для оценки квантовых вычислительных возможностей. Однако, несмотря на установление нового рубежа в квантовом превосходстве, практическая полезность RCS остается ограниченной, что подчеркивает необходимость разработки значимых, прикладных задач для квантовых вычислений. [2]

В отличие от этого, недавние разработки в области квантового отжига продемонстрировали превосходство в решении практических задач. 1 марта 2024 года компания D-Wave объявила о достижении вычислительного превосходства в квантовом моделировании с использованием квантового отжига. Они продемонстрировали, что сверхпроводниковые квантовые отжигатели могут быстро генерировать выборки, соответствующие решениям уравнения

1 Холодов Ярослав Александрович, д.ф.-м.н., профессор, главный научный сотрудник Научного центра информационных технологий и искусственного интеллекта АНОО ВО «Университет «Сириус», Федеральная территория «Сириус», Россия. ORCID: https://orcid.org/0000-0003-2466-1594. Scopus Author ID: 6602420821. E-mail: kholodov.ya@talantiuspeh.ru

2 Саллум Хади, АНО ВО «Университет Иннополис», г. Иннополис, научный центр информационных технологий и искусственного интеллекта АНОО ВО «Университет «Сириус», Федеральная территория «Сириус», Россия. E-mail: h.salloum@innopolis.ru. ORCID: https://orcid.org/ 0009-0005-6068-0532.

3 Агапова Наталья Аркадьевна, АНО ВО «Университет Иннополис», г. Иннополис, E-mail: agapnatalya004@mail.ru

Шрёдингера. В частности, было продемонстрировано масштабирование закона площади запутанности при резких изменениях двух-, трех и бесконечномерных спин-стекол. Кроме того, анализ приближенных классических методов, основанных на тензорных сетях и нейронных сетях, показал, что ни один из известных классических подходов не достигает той же точности, что и квантовый отжигатель, в разумные сроки. Эти результаты подчеркивают способность квантового отжига решать практически значимые задачи, остающиеся недоступными для классических вычислительных методов. [3]

Основным принципом компании D-Wave всегда было создание квантовых вычислительных инструментов для решения сложных задач, а не немедленное стремление к универсальным квантовым вычислениям. Этот подход остается актуальным. Хотя алгоритм Шора — известный квантовый алгоритм для разложения целых чисел на множители — пока не смог успешно разложить число с использованием квантовых компьютеров на логических вентилях, квантовый отжиг уже позволил факторизовать числа до определенного масштаба. Это естественным образом поднимает важный вопрос: можно ли использовать квантовый отжиг для взлома шифрования RSA?

В данной работе представлен краткий обзор квантового отжига и различных подходов к решению задачи разложения на простые множители [4,5,6,7], которая лежит в основе шифрования RSA.

Квантовый отжиг: краткий обзор

Квантовый отжиг — это парадигма квантовых вычислений, основанная на оптимизации, которая использует квантовые флуктуации для нахождения основного состояния заданного гамильтониана. В отличие от квантовых вычислений на логических вентилях, которые опираются на унитарные операции, квантовый отжиг следует процессу адиабатической эволюции, постепенно преобразуя начальный тривиальный гамильтониан в гамильтониан, специфичный для решаемой задачи. Финальное состояние системы соответствует оптимальному решению исходной задачи оптимизации [8,9].

Основной принцип работы квантового отжига основан на **адиабатической теореме квантовой механики**, которая утверждает, что квантовая система, изначально находящаяся в основном состоянии, останется в этом состоянии, если гамильтониан изменяется достаточно медленно. Это свойство позволяет квантовому отжигу эффективно решать задачи комбинаторной оптимизации, включая разложение целых чисел на простые множители, путем кодирования задачи в энергетический ландшафт, где основное состояние представляет правильное решение факторизации.

Чтобы формализовать этот процесс, система начинает с начального гамильтониана H_0, где кубиты подготавливаются в состоянии суперпозиции:

$$H_0 = -\sum_i \sigma_x^{(i)}, \quad (1)$$

где $\sigma_x^{(i)}$ – оператор Паули-X, действующий на i-й кубит. Это гарантирует, что каждый кубит находится в равновероятной суперпозиции состояний $|0\rangle$ и $|1\rangle$, что означает отсутствие закодированной информации в системе на начальном этапе.

По мере продвижения процесса отжига гамильтониан системы постепенно трансформируется в гамильтониан, специфичный для данной задачи H_f:

$$H(t) = [1 - a(t)]H_0 + a(t)H_p, \quad (2)$$

где $a(t)$ – функция, изменяющаяся от 0 до 1 во времени, а H_p – конечный гамильтониан, соответствующий решаемой задаче. Обычно H_p представляется в форме модели Изинга:

$$H_p = -\sum_i h_i \sigma_z^{(i)} - \sum_{i<j} J_{i,j} \sigma_z^{(i)} \sigma_z^{(j)}, \quad (3)$$

где $\sigma_z^{(i)}$ – операторы Паули-Z, h_i – локальные магнитные поля, а $J_{i,j}$ – коэффициенты взаимодействия между кубитами.

В ходе процесса отжига квантовые флуктуации вызывают туннельные эффекты, которые помогают системе избегать локальных минимумов, в конечном итоге направляя её к глобальному минимуму H_p. По мере приближения $a(t)$ к 1 квантовые флуктуации подавляются, и система достигает классического представления модели Изинга, соответствующего оптимальному решению.

Подходы к решению задачи разложения на простые множители с помощью квантового отжига

Модель Изинга — это фундаментальная математическая модель, используемая в статистической механике для описания ферромагнетизма. Энергетический гамильтониан (или функция стоимости) формулируется следующим образом:

$$H(\sigma) = -\sum_{i=1} nh_i\sigma_i - \sum_{i<j} J_{i,j}\sigma_i\sigma_j, (4)$$

где $\sigma = (\sigma_1,...,\sigma_n)^T$, при этом $\sigma_i \in \{+1,-1\}$. Здесь σ_i представляет спин i-го кубита, а h_i и $J_{i,j}$ — коэффициенты, отвечающие за спины кубитов и их связи соответственно [?].

Альтернативно, задачу можно сформулировать как задачу QUBO (Quadratic Unconstrained Binary Optimization — квадратичная безусловная бинарная оптимизация). В этом представлении функция стоимости f определяется в n-мерном бинарном пространстве B^n следующим образом:

$$f(q) = q^T Q q, \quad (5)$$

где Q – верхнетреугольная матрица, а $q = (q_1,...,q_n)^T$ — бинарный вектор. Поскольку для бинарных переменных выполняется $q_i^2 = q_i$, функция стоимости может быть эквивалентно записана так:

$$f(q) = \sum_{i=1} nQ_{i,i}q_i + \sum_{i<j} Q_{i,i}q_iq_j. \quad (6)$$

Неизвестные переменные в модели Изинга (σ) и в модели QUBO (q) связаны между собой следующим образом:

$$\sigma = 2q - 1 \text{ или } q = \frac{1}{2}(\sigma + 1). \quad (7)$$

Предположим, что целое число N является произведением двух простых чисел p и q, рассмотрим следующую задачу наименьших квадратов:

$$\underset{p,q}{arg\,min}\ (pq - N)^2, \quad (8)$$

которая достигает минимального значения 0, когда $pq = N$.

Для удобства вычислений применим 2-норму:

$$\|pq - N\|^2 = p^2q^2 - 2pqN + N^2. \quad (9)$$

Модель HUBO

В формулировке HUBO (Higher-order Unconstrained Binary Optimization) для бинарной задачи наименьших квадратов числа p и q представлены в виде комбинаций кубитов $q_l \in \{0,1\}$. Их представления в системе счисления с основанием 2 записываются следующим образом:

$$p = \sum_{l=0}^{n-1} 2^l q_l,\ q = \sum_{l=0}^{n-1} 2^l q_{n+l}. \quad (10)$$

Это представление позволяет подставить эти выражения в формулировку задачи наименьших квадратов, тем самым генерируя суммируемые члены для функции стоимости. Например, первый член в уравнении (6) принимает вид:

$$(\sum_{l=0}^{n-1} 2^l\, q_l)^2\, (\sum_{l=0}^{n-1} 2^l\, q_{n+l})^2, \quad (11)$$

которое затем расширяется и упрощается с учетом того, что $q_l^2 = q_l$.

Модель QUBO

Поскольку модель HUBO для разложения на простые множители содержит квадратичные, кубические и квартные (четвертой степени) члены, необходимо преобразовать не квадратичные (высшей степени) полиномы в формулировку QUBO.

Члены вида $cxyz$ (где c - оэффициент) заменяются на квадратичные члены путем введения вспомогательного кубита w. В частности, для всех $x,y,z \in \{0,1\}$ можно преобразовать $cxyz$ в комбинацию линейных и квадратичных членов, тем самым упростив его интеграцию в QUBO-модель.

Аналогично, квартные члены можно свести к более простым выражениям, вводя дополнительные переменные (например, для каждого квартного члена требуется ввести семь новых кубитов $x_1,x_2,...,x_7$).

Модель HUBO с алгоритмом гамильтониана, зависящего от диапазона

Недавно был предложен алгоритм гамильтониана, зависящего от диапазона (range-dependent Hamiltonian algorithm) [10]. Этот алгоритм делит область на подрегионов, которые могут быть представлены желаемым количеством кубитов. Применяя этот алгоритм, p и q могут быть выражены следующим образом:

$$p \approx \sum_{l=0}^{n-1} 2^l q_l + S_i, \text{ и } q \approx \sum_{l=0}^{n-1} 2^l q_{n+l} + S_j, \quad (12)$$

где S_i и S_j – это целые числа, регулирующие представление.

Чтобы вывести модель HUBO, подставим уравнение (1) в функцию стоимости наименьших квадратов:

$$p^2q^2 - S_i^2S_j^2 = (\sum_{l=0}^{n-1} 2^l q_l + S_i)^2 (\sum_{l=0}^{n-1} 2^l q_{n+l} + S_j)^2 - S_i^2S_j^2. \quad (13)$$

После раскрытия скобок получаем ряд суммирующихся членов, соответствующих линейным, квадратичным, кубическим и квартным взаимодействиям. Например, линейные члены выглядят так:

$$\sum_{l=0}^{n-1} [(2^{2l} + 2^{l+1}S_i)S_j^2q_l + (2^{2l} + 2^{l+1}S_i)S_j^2q_{n+l}], \quad (14)$$

И аналогичные расширения для квадратичных и более высоких порядков.

Полная модель HUBO получается путем комбинирования этих расширений со вторым членом выражения наименьших квадратов:

$$-2pqN = -2N(\sum_{l=0}^{n-1} 2^l q_l + S_i)(\sum_{l=0}^{n-1} 2^l q_{n+l} + S_j) + 2NS_iS_j. \quad (15)$$

Таким образом, глобальная минимальная энергия, которую нужно получить, равна:

$$-N^2 - S_i^2S_j^2 + 2NS_iS_j. \quad (16)$$

Метод модульного локально-структурированного встраивания

Подведем итог концепциям из [4,7]. Задача разложения на простые множители (PF) для числа N может быть решена с помощью решателей SAT путем кодирования умножителя размером $n \times m$ в булевую формулу и фиксации значений выходных битов, чтобы представить N. В [7] была представлена модульная инкапсуляция бинарного умножителя в архитектуру Pegasus QA, основанная на локально-структурированном встраивании задач SAT.

Цепочка умножителя представлена как конъюнкция логических функций Controlled Full-Adder (CFA), связанных эквивалентностями между переменными. Каждый CFA встраивается в 8-кубитный модуль с эквивалентностями переменных, реализуемыми через цепочки. Каждый CFA $F(x)$ кодируется через штрафную функцию:

$$P_F(z - x,a|\theta)\theta_0 + \sum_{z_i \in V} \theta_i z_i + \sum_{(z_i,z_j)\in E, i<j} \theta_{ij} z_i z_j, \quad (17)$$

где $z_i \in \{-1,1\}$, и при этом выполняется ограничение:

$$\begin{cases} PF(x,a|\theta) = 0 \text{ если } F(x) = \top, \\ PF(x,a|\theta) \geq g_{min} \text{ если } F(x) = \bot. \end{cases} \quad (18)$$

Здесь булевы переменные x и вспомогательные переменные a отображаются на подмножество $z \subset V$ кубитов в топологическом графе (V,E), где значения кубитов $\{1,-1\}$ соответствуют значениям истинности $\{\top,\bot\}$, соответственно. Параметры θ_0, θ_i, θ_{ij} и g_{min} — это сдвиг, смещения, связи и зазор соответственно. Стоит отметить, что смещения и связи имеют ограниченные диапазоны (например, смещения в $[-4,+4]$ и связи в $[-2,+1]$), в то время как сдвиг не ограничен. Вспомогательные переменные a включаются для решения задач с избыточным кодированием. Штрафная функция для всего умножителя строится как сумма штрафных функций для отдельных CFA, а также дополнительные члены, такие как $(2 - 2zz')$ для каждой цепочки $\langle z,z' \rangle$.

Итоговая штрафная функция подается в отжигатель, при этом выходные кубиты фиксируются, чтобы представить число N, с соответствующей инициализацией (например, принудительное значение для кубита carry-in самого правого CFA в каждой строке и кубита in2 для CFA в первой строке, равное -1). Если отжигатель находит основное состояние, для которого штрафная функция равна нулю, то значения кубитов представляют собой валидное решение задачи разложения на простые множители.

Метод модифицированной таблицы умножения

Этот подход основывается на модифицированной таблице умножения, которая уменьшает диапазон значений параметров Изинга, используемых как коэффициенты для локальных полей и взаимодействий. Этот метод также минимизирует количество переменных переноса, устраняя необходимость в обширной предварительной обработке. Метод модифицированной таблицы умножения выполняет локальные минимизации по произведению отдельных бинарных подстрок, представляющих числа p и q. Таблица умножения делится на несколько блоков, каждый из которых можно оптимизировать независимо. Размер блока можно выбрать таким образом, чтобы сбалансировать желаемый диапазон параметров и количество переменных.

Например, рассмотрим иллюстративный случай, где $N = 143$, $p = 13$ и $q = 11$. В предыдущих подходах система уравнений строилась из каждого столбца (или частичных столбцов) таблицы умножения, при этом каждое уравнение учитывало один или несколько битов переноса. В нашем подходе таблица умножения делится на блоки, требующие переносов только между блоками. Это значительно сокращает общее количество переносов и соответствующее количество переменных.

Как показано в Таблице 1 для $N = 143$, вводятся два набора битов переноса, обозначаемых $c_i \in \{0,1\}$. Двухбитовые числа $(c_2c_1)_2 = 2c_2 + c_1$ и $(c_4c_3)_2 = 2c_4 + c_3$ представляют биты переноса для каждого блока. В этой формулировке суммы вычисляются по четырёхбитным числам, при этом сложение внутри каждого блока выполняется по двухбитным числам. Полученная система уравнений, выведенная из таблицы умножения, записывается следующим образом:

$$\begin{aligned} (p_2 + p_1q_1 + q_2) \times 2 + (p_1 + q_1) &= c_2 \times 2^3 + c_1 \times 2^2 + (11)^2 \\ &= c_2 \times 8 + c_1 \times 4 + 3 \\ (q_1 + p_2q_2 + p_1 + c_2) \times 2 + (1 + p_2q_1 + p_1q_2 + 1 + c_1) &= c_4 \times 2^3 + c_3 \times 2^2 + (01)_2 \\ &= c_4 \times 8 + c_3 \times 4 + 1 \\ (1 + c_4) \times 2 + (q_2 + p_2 + c_3) &= (100)_2 \\ &= 4. \end{aligned}$$

Метод модифицированной таблицы умножения устраняет необходимость в бите переноса в каждом столбце, вычисляя переносы только внутри блоков, что значительно снижает общую вычислительную сложность. В предельных случаях восстанавливается обычная таблица умножения при использовании одного столбца на блок, а прямой метод восстанавливается при использовании одного уравнения. Вместо того, чтобы заставлять сумму каждого столбца совпадать с каждым битом числа, которое нужно разложить на множители (как в обычных методах), модифицированный подход заставляет каждый блок таблицы умножения равняться соответствующему блоку числа N. Это приводит к положительной штрафной функции вида:

$$\begin{aligned} f(p,q,c) = {} & (2p_2 + 2p_1q_1 + 2q_2 - 8c_2 - 4c_1 + p_1 + q_1 - 3)^2 \\ & + (2q_1 + 2p_2q_2 + 2p_1 + 2c_2 - 8c_4 - 4c_3 + p_2q_1 + p_1q_2 + c_1 + 1)^2 \\ & + (q_2 + p_2 + c_3 + 2c_4 - 2)^2. \end{aligned}$$

После расширения и упрощения, используя свойство $x^2 = x$ для $x \in \{0,1\}$, получаются кубические и более высокие порядки, которые затем сводятся к квадратичной форме через введение вспомогательных переменных. Например, квадратирование отрицательных членов выполняется аналогично позиционным членам, как подробно описано в дополнительном материале. Для $N = 143$ эта процедура в конечном итоге приводит к соответствующим параметрам для гамильтониана Изинга.

Экспериментальные результаты для случайных чисел

Для проверки описанных методик мы провели серию экспериментов на случайных двуделимых числах. Экспериментальный протокол включал следующие шаги:

1. Генерация случайных двуделимых чисел в заранее определённом диапазоне.
2. Формулировка соответствующей задачи разложения на простые множители как для моделей HUBO, так и для моделей QUBO.
3. Встраивание полученной задачи в квантовый отжигатель с использованием трёх различных методов:
   - Алгоритм зависимого от диапазона гамильтониана;
   - Модифицированный метод таблицы умножения;
   - Модульный метод локально-структурного встраивания.
4. Проведение процесса отжига на доступном оборудовании и сравнение результатов с классическими алгоритмами разложения на множители.

Квантовый отжигатель успешно определил простые множители для следующих двуделимых чисел, используя все три метода:

Таблица 1.

Результаты разложения для случайно выбранных двуделимых чисел с использованием всех трёх методов

Двупростое число	Простые множители	Гамильтонов метод	Табличный метод	Модульный метод
323	17*19	✓	✓	✓
437	19*23	✓	✓	✓
667	23*29	✓	✓	✓
899	29*31	✓	✓	✓
1081	31*37	✓	✓	✓
1619	37*43	✓	✓	✓

Эти результаты подтверждают эффективность подхода квантового отжига для разложения двуделимых чисел в пределах тестируемого диапазона. Успех всех трёх методов подчеркивает их согласованность в правильной идентификации простых множителей. Кроме того, эксперименты выявляют компромисс между количеством введённых переменных (например, вспомогательных кубитов) и точностью значений коэффициентов в модели Изинга. Модульный метод локально-структурного встраивания демонстрирует улучшенную масштабируемость и эффективность встраивания, что делает его перспективным кандидатом для более крупных задач. Эти результаты поддерживают дальнейшее масштабирование подхода по мере развития квантового оборудования.

Выводы

В этой статье представлен подробный обзор квантового отжига и его применения к разложению на простые множители. Мы рассмотрели несколько методов, включая формулировки HUBO и QUBO, алгоритм зависимого от диапазона гамильтониана, методы модульного локально-структурного встраивания и модифицированный метод таблицы умножения. Эти подходы предлагают перспективные пути для использования квантового отжига в решении задач, которые классически являются неразрешимыми, таких как взлом шифрования RSA. Первоначальные эксперименты с случайными двуделимыми числами дополнительно подтверждают осуществимость этих методов, с многообещающими результатами, полученными на текущем квантовом отжигательном оборудовании. В дальнейшем работа будет направлена на масштабирование этих техник для более крупных чисел, улучшение эффективности уменьшения вспомогательных переменных и усовершенствование стратегий встраивания для дальнейшего повышения производительности квантовых отжигателей.

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. Google Quantum AI and Collaborators. (2024). Quantum error correction below the surface code threshold. Nature. https://doi.org/10.1038/s41586-024-08449-y.
2. Coenen, C., Grinbaum, A., Grunwald, A., Milburn, C., & Vermaas, P. (2022). Quantum technologies and society: Towards a different spin. NanoEthics, 16, 1–6. https://doi.org/10.1007/s11569-021-00409-4.
3. King, A. D., Nocera, A., Rams, M. M., Dziarmaga, J., Wiersema, R., Bernoudy, W., Raymond, J., Kaushal, N., Heinsdorf, N., Harris, R., Boothby, K., Altomare, F., Berkley, A. J., Boschnak, M., Chern, K., Christiani, H., Cibere, S., Connor, J., Dehn, M. H., ... Amin, M. H. (2024, March 1). Computational supremacy in quantum simulation [Preprint]. arXiv. https://doi.org/10.48550/arXiv:2403.00910v1.
4. Ding, J., Spallitta, G., & Sebastiani, R. (2024). Experimenting with D-Wave quantum annealers on prime factorization problems. Frontiers in Computer Science, 6. https://doi.org/10.3389/fcomp.2024.1335369.
5. Jun, K., & Lee, H. (2023). HUBO and QUBO models for prime factorization. Scientific Reports, 13, 10080. https://doi.org/10.1038/s41598-023-36813-x.

6. Jiang, S., Britt, K. A., McCaskey, A. J., Humble, T. S., & Kais, S. (2018). Quantum annealing for prime factorization. Scientific Reports, 8, 17667. https://doi.org/10.1038/s41598-018-36058-z.
7. Ding, J., Spallitta, G., & Sebastiani, R. (2024). Effective prime factorization via quantum annealing by modular locally-structured embedding. Scientific Reports, 14, 3518. https://doi.org/10.1038/s41598-024-53708-7.
8. Salloum, H., Sabbagh, K., Savchuk, V., Lukin, R., Orabi, O., & Isangulov, M. (2025). Performance of quantum annealing machine learning classification models on ADMET datasets. IEEE Access, 13, 16263–16287. https://doi.org/10.1109/ACCESS.2025.3531391.
9. Neukart, F., Compostella, G., Seidel, C., von Dollen, D., Yarkoni, S., & Parney, B. (2017). Traffic flow optimization using a quantum annealer. Frontiers in ICT, 4, 29. https://doi.org/10.3389/fict.2017.00029.
10. Lee, H., & Jun, K. (2022, February 15). Range dependent Hamiltonian Algorithm for numerical QUBO formulation [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2202.07692v1.

QUANTUM ANNEALING APPROACHES TO BREAKING RSA ENCRYPTION

Kholodov Y. A.[4], Salloum H.[5], Agap N. A.[6]

Keywords: *Quantum Annealing, RSA, QUBO, Prime Factorization.*

Objective: *to study the transformational potential of quantum annealing in solving the problem of simple factorization.*

Research method(s): *the approach includes a comprehensive review of recent experimental breakthroughs and theoretical innovations. In particular, we analyze techniques such as surface code memory, HUBO and QUBO formulations, range-dependent Hamiltonian algorithms, modular locally structured embedding methods, and a modified multiplication table method random number.*

Research Output(s): *the study evaluates the application of quantum annealing to factorization of primes, showing that advanced problem mapping techniques, such as the HUBO and QUBO formulations, significantly improve the efficiency of representing complex factorization problems on quantum hardware. Notably, the inclusion of surface code memory increases the stability of qubit states during annealing, reducing errors and improving computational accuracy It also demonstrates that Hamiltonian's range-dependent algorithms and modular, locally structured embedding methods help optimize qubit interaction, enabling a more accurate factorization process. A modified multiplication table method is presented, providing an optimized computational strategy, especially effective for large composite numbers. Preliminary experiments with random numbers confirm the theoretical conclusions, indicating that these integrated methods allow for better performance than traditional approaches. Taken together, the results highlight the potential of quantum annealing as a solid foundation for solving complex cryptographic problems and lay the foundation for future research into scalable quantum algorithms and hardware implementations.*

Scientific novelty: *the work combines several advanced quantum annealing techniques for factorization of prime numbers, combining experimental innovations with theoretical developments to propose a new framework that improves the efficiency of cryptographic computing.*

References

1. Google Quantum AI and Collaborators. (2024). Quantum error correction below the surface code threshold. Nature. https://doi.org/10.1038/s41586-024-08449-y.
2. Coenen, C., Grinbaum, A., Grunwald, A., Milburn, C., & Vermaas, P. (2022). Quantum technologies and society: Towards a different spin. NanoEthics, 16, 1–6. https://doi.org/10.1007/s11569-021-00409-4.
3. King, A. D., Nocera, A., Rams, M. M., Dziarmaga, J., Wiersema, R., Bernoudy, W., Raymond, J., Kaushal, N., Heinsdorf, N., Harris, R., Boothby, K., Altomare, F., Berkley, A. J., Boschnak, M., Chern, K., Christiani, H., Cibere, S., Connor, J., Dehn, M. H., ... Amin, M. H. (2024, March 1). Computational supremacy in quantum simulation [Preprint]. arXiv. https://doi.org/10.48550/arXiv:2403.00910v1.
4. Ding, J., Spallitta, G., & Sebastiani, R. (2024). Experimenting with D-Wave quantum annealers on prime factorization problems. Frontiers in Computer Science, 6. https://doi.org/10.3389/fcomp.2024.1335369.
5. Jun, K., & Lee, H. (2023). HUBO and QUBO models for prime factorization. Scientific Reports, 13, 10080. https://doi.org/10.1038/s41598-023-36813-x.
6. Jiang, S., Britt, K. A., McCaskey, A. J., Humble, T. S., & Kais, S. (2018). Quantum annealing for prime factorization. Scientific Reports, 8, 17667. https://doi.org/10.1038/s41598-018-36058-z.
7. Ding, J., Spallitta, G., & Sebastiani, R. (2024). Effective prime factorization via quantum annealing by modular locally-structured embedding. Scientific Reports, 14, 3518. https://doi.org/10.1038/s41598-024-53708-7.

4 Yaroslav A. Kholodov, Doctor of physico-mathematical sciences, Professor, Chief Researcher of Scientific Center of Information Technologies and Artificial Intelligence of Sirius University of Science and Technology, Sirius Federal Territory Sirius University of Science and Technology. ORCID: https://orcid.org/0000-0003-2466-1594. Scopus Author ID: 6602420821. E-mail: kholodov.ya@talantiuspeh.ru

5 Sallum Hadi, Innopolis University, Innopolis, Scientific Center for Information Technologies and Artificial Intelligence, Sirius University, Sirius Federal Territory, Russia. E-mail: h.salloum@innopolis.ru. ORCID: https://orcid.org/ 0009-0005-6068-0532.

6 Natalia A. Agapova, Innopolis University, Innopolis, Sirius Federal Territory, Russia. E-mail: agapnatalya004@mail.ru

8. Salloum, H., Sabbagh, K., Savchuk, V., Lukin, R., Orabi, O., & Isangulov, M. (2025). Performance of quantum annealing machine learning classification models on ADMET datasets. IEEE Access, 13, 16263–16287. https://doi.org/10.1109/ACCESS.2025.3531391.
9. Neukart, F., Compostella, G., Seidel, C., von Dollen, D., Yarkoni, S., & Parney, B. (2017). Traffic flow optimization using a quantum annealer. Frontiers in ICT, 4, 29. https://doi.org/10.3389/fict.2017.00029.
10. Lee, H., & Jun, K. (2022, February 15). Range dependent Hamiltonian Algorithm for numerical QUBO formulation [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2202.07692v1.

ОБНАРУЖЕНИЕ ФИШИНГОВЫХ ЭЛЕКТРОННЫХ ПИСЕМ С ПОМОЩЬЮ РЕКУРРЕНТНЫХ НЕЙРОННЫХ СЕТЕЙ

Болдырихин Н. В.[1], Ядрец Э. А.[2]

DOI: 10.21681/2311-3456-2025-4-134-141

Цель исследования: *рассмотреть особенности применения рекуррентных нейронных сетей при решении задачи обнаружения фишинговых писем.*

Метод(ы) исследования: *сравнение и сопоставление, математическое и программное моделирование, системный анализ.*

Результат(ы) исследования: *рассмотрены понятие и виды фишинговых атак. Проведен анализ современных публикаций по теме использования рекуррентных нейронных сетей в задачах обнаружения фишинга, который показал, что использование рекуррентных сетей даёт возможность с большой вероятностью обнаруживать фишинговые письма. Проанализированы датасеты, имеющиеся в открытом доступе: большинство датасетов ориентированы на обнаружение фишинговых URL – адресов. Немногочисленные датасеты, ориентированные на текст электронного письма, в подавляющем большинстве являются англоязычными, качественные русскоязычные датасеты в открытом доступе отсутствуют, поэтому был собран собственный датасет из русскоязычных электронных писем. Проведено также математическое и программное моделирование различных рекуррентных нейронных сетей для выявления фишинговых писем: RNN, LSTM, BiLSTM и проведен сравнительный анализ их характеристик. Выявлены зависимости характеристик потерь и точности от числа эпох. Сравнительный анализ рекуррентных сетей показал, что наиболее эффективной в решении задач обнаружения фишинга в рамках исследований оказалась сеть BiLSTM, которая обнаружила 91,43 % фишинговых писем. Худшие характеристики показала сеть RNN, которая обнаружила только 50,71 % фишинговых писем из тестовой выборки. Следует отметить, что данные результаты получены для сетей, обучаемых на датасетах малого объёма (300 писем).*

Научная новизна: *результаты исследований позволяют аргументировано заключить, что из рассмотренных рекуррентных нейронных сетей именно BiLSTM наилучшим образом справляются с задачами выявления фишинговых писем при небольших объемах обучающего датасета.*

Ключевые слова: *кибермошенничество, защита от фишинга, рекуррентные нейронные сети RNN, LSTM, BiLSTM.*

Введение

В настоящее время задача обеспечения информационной безопасности становится очень актуальной как для государственных учреждений, коммерческих организаций, так и для физических лиц. Повсеместное внедрение цифровых технологий стало причиной небывалого всплеска киберпреступной деятельности [1–16]. Большое количество таких преступлений совершается с использованием фишинговых технологий [4–9].

Фишинг – вид кибермошенничества, нацеленный на получение доступа к конфиденциальной информации посредством побуждения пользователя перейти по ссылке на интернет-ресурс, содержащий вредоносный код. Если пользователь переходит на данный ресурс, то ему на устройство загружается вредоносный код или пользователь сам вводит логин и пароль на поддельном сайте, имитирующем, например, ресурс банка [4–9].

Фишинговые ссылки могут распространяться при помощи электронных писем, СМС, через мессенджеры и т.д.

Главная проблема состоит в том, что мошенники постоянно подстраиваются под текущую информационную повестку, модифицируя и усложняя схемы фишинговых атак.

Самой распространённой разновидностью фишинга является «массовый» или «слепой» фишинг. На электронную почту жертв, вслепую, рассылаются шаблонные варианты сообщений, ведущие на заражённые или поддельные сайты. Письма маскируются под официальные, а темы могут быть разнообразными: сообщения от службы поддержки, почтового сервиса, уведомления о закрытии, открытии, блокировке банковских счетов, извещения из государственных органов или популярных услуг и сервисов. В таких письмах обычно пользователя просят обновить свои данные или войти в аккаунт для дальнейших манипуляций. Главная особенность такого метода в том, что атакующий заранее не знает кого именно он атакует.

Самой опасной формой является целевой фишинг, направленный на конкретного человека. Этот тип фишинга представляет собой одну из ключевых

1 Болдырихин Николай Вячеславович, кандидат технических наук, доцент кафедры «Кибербезопасность информационных систем» Донского государственного технического университета, г. Ростов-на-Дону, Россия, E-mail: boldyrikhin@mail

2 Ядрец Эдуард Александрович, магистрант кафедры «Кибербезопасность информационных систем» Донского государственного технического университета, г. Ростов-на-Дону, Россия, E-mail: xperia1058@gmail.com

тактик, применяемых в практике кибершпионажа. Его отличительная черта – целенаправленность и индивидуальный подход к реализации. Изучается сам человек и его окружение. Собирается различная персональная информация: фотографии, стилистика переписки, информация о местах отдыха, предпочтениях, увлечениях и т.д. Подобные атаки нацелены на известных медийных персон, бизнесменов и политических деятелей.

Процесс распознавания фишингового письма весьма затруднительный. Для этого применяются антивирусные программы, антифишинговые расширения браузеров, спам-фильтры, системы обнаружения вторжений, DLP-системы (Data loss prevention) и т.д., которые используют как сигнатурные методы, так и методы, основанные на применении искусственного интеллекта [4–14].

Искусственный интеллект на основе нейросетей в настоящее время широко применяется в решении задач информационной безопасности. Это связано с тем, что системы, основанные на машинном обучении, могут выявлять не только известные атаки, но и атаки «нулевого дня», при этом показывая хорошие результаты. Хотя конечно такие системы не являются панацеей, они, тем не менее, набирают большую популярность. Поэтому исследования в данном направлении представляют теоретический и практический интерес.

В рамках данной статьи проведено исследование особенностей применения рекуррентных нейронных сетей при обнаружении фишинговых электронных писем.

Особенности применения рекуррентных нейронных сетей различных типов при решении задач обнаружения фишинга

В работах [4–9] подробно рассмотрены самые разнообразные методы решения задач обнаружения фишинга, в том числе применение нейронных сетей. На основе анализа данных работ выявлено, что наиболее часто для таких задач применяются рекуррентные (Recurrent neural network, RNN) и свёрточные нейронные сети (Convolutional neural network, CNN).

В работе [5] авторы Корнюхина С. П., Лапонина О. Р. подробно рассмотрели решение задачи обнаружения фишинга различными нейронными сетями, в том числе CNN и LSTM (Long short-term memory). Экспериментальные результаты показали, что LSTM в случаях обнаружения фишинговых URL-адресов, демонстрирует точность свыше 99 %, а при обнаружении фишинговых писем – 98 %. Сеть CNN показывает максимальные результаты при обнаружении фишинговых URL-адресов – свыше 98 %.

В научной статье [6] авторами Suleiman Y. Yerima и Mohammed K. Alzaylaee рассмотрен подход применения CNN для высокоточной классификации фишинговых сайтов. Основываясь на результатах обширных экспериментов, авторы пришли к выводу, что модели на основе CNN оказались очень эффективными в обнаружении неизвестных фишинговых сайтов. Более того, подход на основе CNN показал лучшие результаты, чем традиционные классификаторы машинного обучения, достигнув 98,2 % обнаружения фишинга.

В работе [7] авторами Weiping Wang, Feng Zhang, Xi Luo и Shigeng Zhang представлен быстрый подход к обнаружению фишинговых сайтов под названием Precise Phishing Detection with Recurrent Convolutional Neural Networks (PDRCNN), который опирается только на URL сайта. PDRCNN не нужно извлекать содержимое целевого сайта и использовать сторонние сервисы. Представленный метод кодирует информацию об URL-адресе в двумерный тензор и передаёт этот тензор в нейронную сеть глубокого обучения для классификации исходного URL-адреса. Авторы работы [7] используют двунаправленную LSTM-сеть (Bidirectional Long Short-Term Memory, BiLSTM) для извлечения глобальных признаков из построенного тензора и передают всю строковую информацию каждому символу в URL. После этого используют CNN для автоматического определения того, какие символы играют ключевую роль, захватывают ключевые компоненты URL и сжимают извлечённые характеристики в векторное пространство фиксированной длины. Комбинируя эти два типа сетей, PDRCNN достигает более высокой производительности, чем при использовании сетей по отдельности.

В работе [8] автором Cagatay Catal и др., приведён систематический обзор использования алгоритмов глубокого обучения для обнаружения фишинга.

Авторы работы [8] провели систематический обзор литературы, с целью обобщить результаты применения подходов глубокого обучения для обнаружения фишинга, представленные в отобранных научных публикациях. Были рассмотрены девять исследовательских вопросов и представлен обзор того, как применялись алгоритмы глубокого обучения для обнаружения фишинга. Во всех моделях применялись алгоритмы глубокого обучения с супервизией. В качестве источников использовались наборы данных, связанные с URL, информацией о третьих лицах на сайте, содержимое сайта и электронная почта.

Наиболее часто применялись различные алгоритмы глубокого обучения. Среди них выделяются нейронные сети, известные как глубокие нейронные сети (Deep Neural Network, DNN). DNN и гибридные алгоритмы глубокого обучения обеспечили наилучшую производительность среди других алгоритмов глубокого обучения. В 72 % исследований для построения модели прогнозирования не применялся какой-либо алгоритм отбора признаков. PhishTank был наиболее используемым набором данных среди

других наборов данных. Несмотря на то, что Keras и Tensorflow были наиболее предпочтительными фреймворками для глубокого обучения, в 46 % статей, исследуемых авторами, не упоминался ни один фреймворк [8].

Таким образом, можно утверждать, что рекуррентные сети очень популярный вид сетей при решении задач обнаружения фишинга.

В рамках данного исследования рассмотрены особенности применения различных типов рекуррентных нейронных сетей для выявления фишинговых писем: RNN, LSTM, BiLSTM.

Описание наборов данных

Задача подготовки наборов данных является в целом нетривиальной, поскольку за качественными наборами данных (датасетами) идёт настоящая охота. Компании-разработчики крайне редко раскрывают свои качественные датасеты, хотя алгоритмы публикуют довольно охотно. Это связано с тем, что сбор данных зачастую является сложным, трудоемким и долгим. Так же при подготовке датасетов следует учитывать необходимость очистки данных от выбросов и пропусков, необходимость нормализации и отбора признаков [5], что также сопряжено с существенными затратами.

При исследовании готовых датасетов, находящихся в открытом доступе на порталах VirusTotal, PhishTank OpenPhish, оказалось, они в основном содержат данные о URL-адресах.

Поскольку изначально ставилась задача определения фишинга не только по URL-адресам, содержащимся в письмах, но и по тексту самого письма, то указанные датасеты не подошли для проведения исследования.

На портале GitHub и Kaggle размещены датасеты фишинговых электронных писем, однако они англоязычные и по этой причине тоже не подошли, ввиду того, в рамках данного исследования интерес представляли именно русскоязычные фишинговые письма, поскольку анализировались не только ссылки, содержащиеся в письме, но и сам текст письма.

Для создания обучающего набора данных был разработан скрипт взаимодействия с сервером электронной почты по протоколу IMAP. Скрипт способен работать с различными кодировками текста, такими как UTF-8 и Windows-1252. Скрипт подключается к почтовому серверу перебирает электронные письма на сервере, извлекая из них данные и сохраняет их. Перед сохранением происходит декодирование заголовков и тела письма электронной почты, уделяя особое внимание теме письма и текстовому содержимому тела. Это важная деталь для выявления потенциальных фишинговых писем, поскольку эти поля часто содержат вводящее в заблуждение или вредоносное содержимое. В результате работы данного скрипта, из личного электронного почтового ящика был выполнен сбор данных, ставший основой для формирования обучающего набора данных, включающего в себя 300 писем, 120 из которых не относились фишингу, и 180 писем являлись фишинговыми. Особенностью собранного набора данных является его структурирование, где каждое письмо сохранено в виде отдельного текстового файла, обеспечивая таким образом удобство для последующего анализа и обработки информации.

Дополнительно была создана проверочная выборка из 140 фишинговых писем, которая не участвовала в обучении и валидации модели, но имеет схожие стилистические паттерны текста и ключевые фразы для проверки работоспособности выбранных нейросетевых моделей.

Эффективность LSTM, как и других нейронных сетей в задачах обработки текста в значительной степени зависит от качества входных данных, оптимальная подготовка которых, существенно повышает способность сети к обучению и обобщению. В рамках исследования для подготовки данных использовалась SpaCy. Библиотека SpaCy — является мощным инструментом, позволяющим решать такие задачи подготовки данных, как лемматизация, токенизация, анализ зависимостей и множество других задач.

В качестве предобработки применялась токенизация [15], а затем лемматизация [16] слов, на основе средней русскоязычной модели SpaCy «ru_core_news_md» для выделения лемм алфавитных токенов.

Модель «ru_core_news_md» в SpaCy обеспечивает высокую точность токенизации и лемматизации. Эта модель обучена на большом объёме текстовых данных, позволяя ей эффективно справляться с разнообразными текстовыми задачами и даёт преимущество в точности и надёжности предобработки данных. На рисунке 1 приведены свойства языковой модели, представленные на официальном сайте https://spacy.io/.

Рис. 1. Свойства языковой модели

```
nlp = spacy.load("ru_core_news_md")

def tokenize_text(text):
    doc = nlp(text)
    return [token.lemma_ for token in doc if token.is_alpha]

def vectorize_text(tokens):
    vectors = [nlp.vocab[token].vector for token in tokens if nlp.vocab[token].has_vector]
    if vectors:
        vectors = np.array(vectors)
        return torch.tensor(vectors, dtype=torch.float32)
    else:
        return torch.empty((0, nlp.vocab.vectors_length), dtype=torch.float32)
```

Рис. 2. Код для предобработки текстовых данных

Токенизация – это процесс разделения текста на составляющие его элементы, называемые токенами. В контексте текста токены обычно представляют собой слова, но также могут включать пунктуацию и другие символы. Цель токенизации – преобразовать непрерывный текст в управляемый набор данных, которые далее могут быть обработаны и проанализированы.

Лемматизация – это процесс приведения слова к его базовой форме (лемме). Лемматизация учитывает морфологический анализ слов, чтобы вернуть правильную базовую форму слова с учётом его использования в тексте.

Реализация функций предварительной обработки текста с помощью библиотеки SpaCy представлена на рисунке 2.

Лемматизированные токены затем векторизуются с использованием предварительно обученных векторов слов, доступных в модели SpaCy. Учитываются только те токены, для которых в словаре SpaCy имеются векторы, каждый токен преобразуется в вектор с использованием соответствующего вложения. Затем векторы преобразуются в тензоры, обеспечивая совместимость с моделью LSTM или любыми другими.

Подготовленные векторы и соответствующие метки сохраняются, затем данные в рамках своего класса случайным образом перемешиваются и разделяются на обучающую и тестовую выборки в соответствии с заданным соотношением: 80 % на обучение и 20 % для валидации модели. В результате предобработки получились хорошо структурированные и нормализованные входные данные для обучения нейросети.

Характеристики обучения выбранных топологий нейросетей

Следует отметить, что задача обнаружения фишинга относится к задаче бинарной классификации [6]. В этом случае матрица ошибок имеет следующий вид (см. рис. 3).

В качестве показателей качества работы нейронной сети использовались бинарная кросс-энтропийная функция потерь (Loss)

Реальное событие / Результат работы нейросети	**Фишинг присутствует (Positive)**	**Фишинг отсутствует (Negative)**
Фишинг присутствует (Positive)	True Positive (TP) — правильное определение фишинга	FP (False Positive) — ошибочное определение фишинга, хотя фактически его не было
Фишинг отсутствует (Negative)	FN (False Negative) — когда письмо определено как безопасное, хотя фактически оно фишинговое	True Negative (TN) — правильное определение отсутствия фишинга

Рис. 3. Матрица ошибок

$$\text{Loss} = P_p * \ln(P_{np}) + (1 - P_p) * \ln(1 - P_{np}), \quad (1)$$

и точность (Accuracy),

$$\text{Accuracy} = \frac{TP + TR}{TN + FP + FN + TP}, \quad (2)$$

где P_p – вероятность наличия фишинга, P_{np} – вероятность отсутствия фишинга; TP, TR, TN, FP, FN – переменные, обозначенные на рисунке 3.

Далее приведены характеристики обучения для различных топологий нейросетей.

Стандартная рекуррентная нейронная сеть (RNN) обладает простой архитектурой. На рисунке 4 представлены основные параметры модели, а также проиллюстрирован процесс её обучения.

```
Модель:  RNNClassifier(
  (rnn): RNN(300, 256, num_layers=2, batch_first=True)
  (fc): Linear(in_features=256, out_features=2, bias=True)
)
Начало обучения
Epoch 1, Train Loss: 0.6933, Train Acc: 60.00%, Val Loss: 0.6325, Val Acc: 63.33%
Epoch 2, Train Loss: 0.6753, Train Acc: 60.37%, Val Loss: 0.6279, Val Acc: 56.67%
Epoch 3, Train Loss: 0.6593, Train Acc: 62.22%, Val Loss: 0.7825, Val Acc: 56.67%
Epoch 4, Train Loss: 0.6483, Train Acc: 61.85%, Val Loss: 0.6949, Val Acc: 56.67%
Epoch 5, Train Loss: 0.6498, Train Acc: 61.48%, Val Loss: 0.8026, Val Acc: 56.67%
Epoch 6, Train Loss: 0.6861, Train Acc: 61.11%, Val Loss: 0.7081, Val Acc: 56.67%
Epoch 7, Train Loss: 0.6617, Train Acc: 61.11%, Val Loss: 0.7368, Val Acc: 56.67%
Epoch 8, Train Loss: 0.6508, Train Acc: 61.85%, Val Loss: 0.6735, Val Acc: 56.67%
Epoch 9, Train Loss: 0.6833, Train Acc: 61.11%, Val Loss: 0.6809, Val Acc: 56.67%
Epoch 10, Train Loss: 0.6613, Train Acc: 61.85%, Val Loss: 0.6067, Val Acc: 63.33%
```

Рис. 4. Демонстрация процесса обучения стандартной RNN

На рисунке 5 приведена динамика изменения потерь на обучающей и валидационной выборках в ходе обучения модели. График наглядно демонстрирует, как изменяются значения потерь по мере увеличения числа эпох.

Рис. 5. Изменение потерь в процессе обучения RNN

Рисунок 6 отображает изменение точности модели с каждой новой эпохой, иллюстрируя эти изменения как для обучающих, так и для валидационных данных.

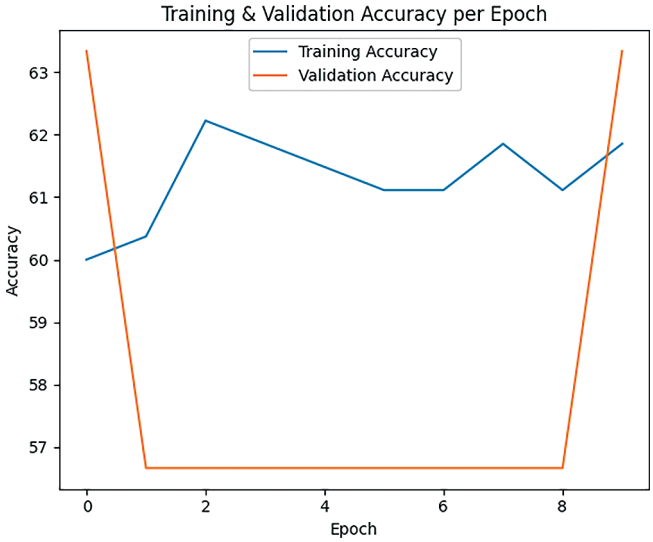


Рис. 6. Изменение точности RNN в процессе обучения

Валидационная точность остаётся на уровне 56,67 % на протяжении большинства эпох, она ниже, чем точность на обучающей выборке, такое поведение модели вероятно является следствием проблемы с обобщающей способностью.

Долгосрочная кратковременная память (LSTM) отличается более сложной архитектурой по сравнению с стандартной рекуррентной сетью (RNN). На рисунке 7 демонстрируются ключевые параметры модели и показан процесс её обучения.

На рисунке 8 показана динамика изменения потерь на обучающей и валидационной выборках

```
Модель:  LSTMClassifier(
  (lstm): LSTM(300, 256, num_layers=2, batch_first=True)
  (fc): Linear(in_features=256, out_features=2, bias=True)
)
Начало обучения
Epoch 1, Train Loss: 0.6991, Train Acc: 47.08%, Val Loss: 0.6799, Val Acc: 63.33%
Epoch 2, Train Loss: 0.6737, Train Acc: 61.25%, Val Loss: 0.6679, Val Acc: 63.33%
Epoch 3, Train Loss: 0.6639, Train Acc: 61.67%, Val Loss: 0.6604, Val Acc: 63.33%
Epoch 4, Train Loss: 0.6646, Train Acc: 61.67%, Val Loss: 0.6751, Val Acc: 61.67%
Epoch 5, Train Loss: 0.6600, Train Acc: 62.08%, Val Loss: 0.6686, Val Acc: 61.67%
Epoch 6, Train Loss: 0.6547, Train Acc: 62.50%, Val Loss: 0.6972, Val Acc: 61.67%
Epoch 7, Train Loss: 0.6760, Train Acc: 59.58%, Val Loss: 0.6741, Val Acc: 61.67%
Epoch 8, Train Loss: 0.6468, Train Acc: 63.33%, Val Loss: 0.6718, Val Acc: 61.67%
Epoch 9, Train Loss: 0.6521, Train Acc: 61.67%, Val Loss: 0.6685, Val Acc: 61.67%
Epoch 10, Train Loss: 0.6508, Train Acc: 62.08%, Val Loss: 0.6569, Val Acc: 61.67%
```

Рис. 7. Демонстрация хода обучения LSTM

в ходе обучения модели LSTM. Важно отметить, что кривая потерь на обучающей выборке показывает тенденцию к снижению, поскольку модель постепенно адаптируется к данным, минимизируя ошибку предсказания.

Результаты в виде графика точности модели изображены на рисунке 9. Модель LSTM показывает

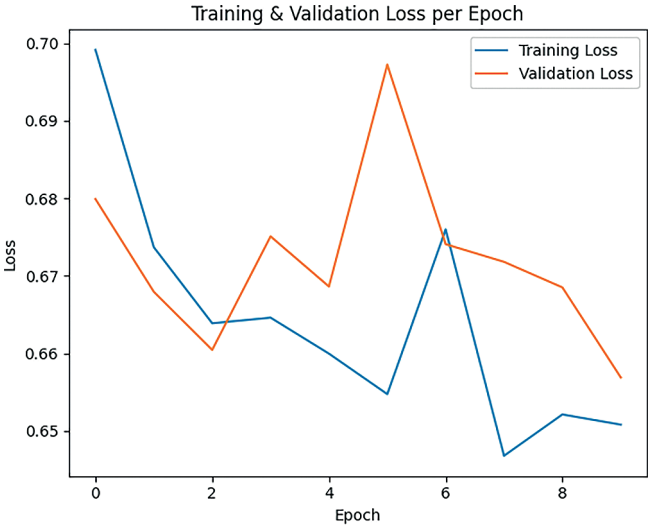


Рис. 8. Потери LSTM при обучении

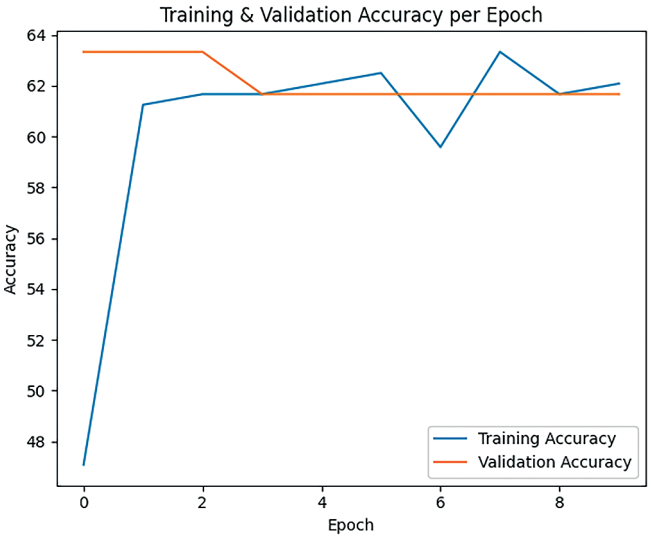


Рис. 9. Точность LSTM по пройденным эпохам

```
Модель:  BiLSTMClassifier(
  (lstm): LSTM(300, 256, num_layers=2, batch_first=True, bidirectional=True)
  (fc): Linear(in_features=512, out_features=2, bias=True)
)
Начало обучения
Epoch 1, Train Loss: 0.6611, Train Acc: 62.22%, Val Loss: 0.6444, Val Acc: 60.00%
Epoch 2, Train Loss: 0.5659, Train Acc: 71.85%, Val Loss: 0.5580, Val Acc: 66.67%
Epoch 3, Train Loss: 0.4113, Train Acc: 80.37%, Val Loss: 0.4133, Val Acc: 73.33%
Epoch 4, Train Loss: 0.2391, Train Acc: 90.74%, Val Loss: 0.3616, Val Acc: 83.33%
Epoch 5, Train Loss: 0.1738, Train Acc: 94.44%, Val Loss: 0.2044, Val Acc: 90.00%
Epoch 6, Train Loss: 0.1346, Train Acc: 94.81%, Val Loss: 0.1653, Val Acc: 96.67%
Epoch 7, Train Loss: 0.0636, Train Acc: 98.89%, Val Loss: 0.3152, Val Acc: 90.00%
Epoch 8, Train Loss: 0.0728, Train Acc: 97.04%, Val Loss: 0.1137, Val Acc: 96.67%
Epoch 9, Train Loss: 0.0463, Train Acc: 98.52%, Val Loss: 0.1826, Val Acc: 90.00%
Epoch 10, Train Loss: 0.0326, Train Acc: 99.12%, Val Loss: 0.1126, Val Acc: 96.67%
```

Рис. 10. Параметры BiLSTM модели и процесс обучения

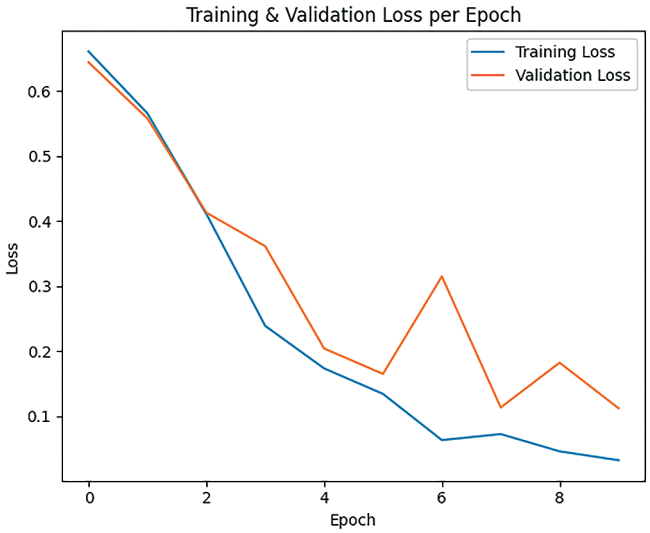


Рис. 11. Потери BiLSTM при обучении

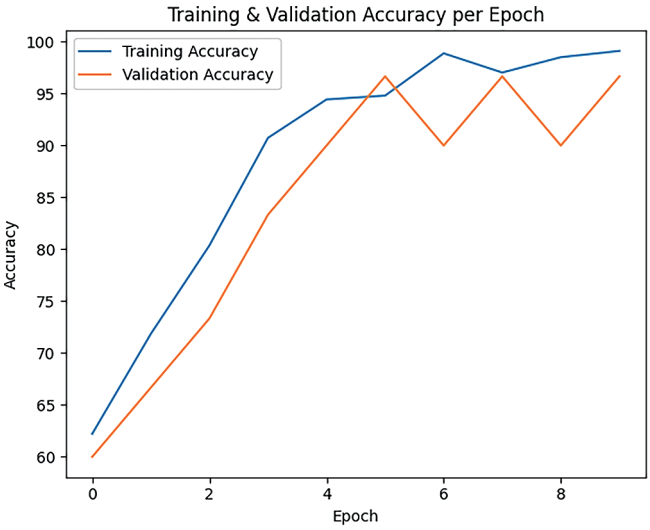


Рис. 12. Точность BiLSTM модели при обучении

средние результаты точности на тренировочных данных и на валидационном наборе она остаётся практически неизменной (около 62 %). Модель явно сталкивается с трудностями в обобщении информации и эффективной обработке длинных текстов, так как важные зависимости между словами могут находиться далеко друг от друга.

Двунаправленная модель LSTM или Bidirectional LSTM (BiLSTM) отличается от стандартной LSTM своей архитектурой, которая позволяет анализировать последовательность данных в двух направлениях вперёд и назад. На рисунке 10 представлены ключевые параметры модели BiLSTM, а также отображён процесс её обучения.

График потерь модели при обучении представлен на рисунке 11.

График точности BiLSTM модели изображён на рисунке 12.

Изображения графиков, доказывают высокую эффективность модели BiLSTM в обработке и понимании обучающего набора данных, поскольку обработка данных проводится в двух направлениях, как было сказано ранее. Достоинство модели состоит в понимании контекста и более точной обработке входных данных, в отличие от простой LSTM.

Тестирование нейросетей на основе дополнительной выборки

На рисунке 13 приведён результат обработки 140 фишинговых писем RNN моделью. Из них модель неверно определила 69 писем как легитимные, что соответствует 49 % от общего количества писем.

```
Файл: ph_9_8.txt - Результат: Легитимное - Вероятности: 0.70, 0.30
Файл: ph_9_9.txt - Результат: Легитимное - Вероятности: 0.71, 0.29
RNNClassifier(
  (rnn): RNN(300, 256, num_layers=2, batch_first=True)
  (fc): Linear(in_features=256, out_features=2, bias=True)
)
Общее количество фишинговых писем: 140
Определены как Легитимные письма: 69 49.29%
Определены как Фишинговые письма: 71 50.71%
```

Рис. 13. Результат обработки RNN

Результат обработки LSTM моделью представлен на рисунке 14.

```
Файл: ph_9_8.txt - Результат: Легитимное - Вероятности: 0.64, 0.36
Файл: ph_9_9.txt - Результат: Легитимное - Вероятности: 0.62, 0.38
LSTMClassifier(
  (lstm): LSTM(300, 256, num_layers=2, batch_first=True)
  (fc): Linear(in_features=256, out_features=2, bias=True)
)
Общее количество фишинговых писем: 140
Определены как Легитимные письма: 36 25.71%
Определены как Фишинговые письма: 104 74.29%
```

Рис. 14. Результаты обработки LSTM

Исходя из результата обработки, можно заметить, что 36 писем были неверно классифицированы, а это 25% от общего их количества. Данный результат намного лучше, чем у RNN сети, но все же далёк от идеала.

Результаты обработки для модели BiLSTM приведены на рисунке 15.

```
Файл: ph_9_8.txt - Результат: Фишинг - Вероятности: 0.00, 1.00
Файл: ph_9_9.txt - Результат: Фишинг - Вероятности: 0.10, 0.90
BiLSTMClassifier(
  (lstm): LSTM(300, 256, num_layers=2, batch_first=True, bidirectional=True)
  (fc): Linear(in_features=512, out_features=2, bias=True)
)
Общее количество фишинговых писем: 140
Определены как Легитимные письма: 12 8.57%
Определены как Фишинговые письма: 128 91.43%
```

Рис.15. Результаты обработки для BiLSTM

Здесь результаты выглядят заметно лучше, по сравнению с ранее рассматриваемыми сетями: всего 12 писем из 140 были определены как легитимные, что составило приблизительно 8 % от их общего количества.

Выводы

В результате проведённого исследования была достигнута основная цель данной статьи: рассмотрены особенности применения рекуррентных нейронных сетей при решении задачи обнаружения фишинговых писем.

В процессе работы проведен анализ предметной области, который показал, что применение рекуррентных нейронных сетей значительно повышает точность обнаружения фишинга.

В рамках данного исследования авторами произведено формирование собственного датасета на основе собранных фишинговых электронных писем на русском языке. Это связано с отсутствием в открытом доступе качественных русскоязычных датасетов, содержащих тексты фишинговых писем.

В рамках разработки нейросетей были рассмотрены этапы предобработки входных данных, а также выполнен полный цикл обучения и валидации.

В результате сравнения характеристик обучения выбранных рекуррентных нейронных сетей, можно аргументировано заключить, что наиболее эффективной для решения задачи выявления фишинговых писем оказалась BiLSTM. Созданная сеть BiLSTM продемонстрировала достаточно высокую точность в обнаружении фишинговых сообщений: 91,43 % несмотря на то, что объем используемого датасета небольшой. Остальные рекуррентные сети в этих условиях показали худший результат.

Следует отметить, что результаты, полученные в рамках данного исследования, являются в основном практическими и справедливы только в конкретных условиях данного программного моделирования. При больших объемах датасета различие в точности обнаружения фишинга между BiLSTM и другими рекуррентными сетями скорее всего будет не столь значительным.

Литература

1. Кострикина А. О., Лазунин К. А. Информационная безопасность в критической информационной безопасности // Проблемы научной мысли. 2024. Т. 4. № 1. С. 82–85.
2. Чапис М. А. Информационная безопасность государства как правовой порядок обеспечения национальной безопасности в информационной сфере // Наукосфера. 2024. № 6(1). С. 551–557. DOI: 10.5281/zenodo.11638587.
3. Добродеев А. Ю. Показатели информационной безопасности как характеристика (мера) соответствия сетей и организаций связи требованиям информационной безопасности // Труды ЦНИИС. Санкт-Петербургский филиал. 2020. Т. 2. № 10. С. 50–78.
4. Лукманова К. А., Картак В. М. Разработка системы защиты от фишинговых атак с использованием программно-аппаратной реализации методов машинного обучения // Моделирование, оптимизация и информационные технологии. 2024. 12(4). DOI: 10.26102/2310-6018/2024.47.4.033.
5. Корнюхина С. П., Лапонина О. Р. Исследование возможностей алгоритмов глубокого обучения для защиты от фишинговых атак // International Journal of Open Information Technologies. 2023. Т. 11. № 6. С. 163–174.
6. Yerima S. Y., Alzaylaee M. K. High accuracy phishing detection based on convolutional neural networks // 2020 3rd International Conference on Computer Applications & Information Security (ICCAIS). IEEE, 2020. С. 1–6. DOI:10.1109/ICCAIS48893.2020.9096869.
7. Wang W. et al. PDRCNN: Precise phishing detection with recurrent convolutional neural networks //Security and Communication Networks. 2019. Т. 2019. С. 1–15. DOI:10.1155/2019/2595794.
8. Catal C. et al. Applications of deep learning for phishing detection: a systematic literature review // Knowledge and Information Systems. 2022. Т. 64. №. 6. С. 1457–1500. DOI:10.1007/s10115-022-01672-x.
9. Dhanavanthini P., Chakkravarthy S. S. Phish-armour: phishing detection using deep recurrent neural networks. Soft Comput. 2023. DOI: 10.1007/s00500-023-07962-y.
10. Филимонов А. В., Осипов А. В., Плешакова Е. С., Гатауллин С. Т. Нейросетевые методы распознавания эмоций речи для противодействия мошенничеству в телекоммуникационных системах // Вопросы кибербезопасности. 2022. № 6(52). С. 83–92. DOI:10.21681/2311-3456-2022-6-83-92.
11. Технологии искусственного интеллекта и кибербезопасность: монография / А. Б. Менисов. – М: Ай Пи Ар Медиа, 2022. 133 с.
12. Применение искусственного интеллекта для решения задачи обеспечения безопасности информации, передаваемой в сетях / В. И. Юхнов, А. И. Сосновский, Н. В. Болдырихин, И. А. Сосновский // Труды Северо-Кавказского филиала Московского технического университета связи и информатики. 2023. № 2. С. 26–28.
13. Бимолдина Ж. А. Как искусственный интеллект меняет правила игры в кибербезопасности // Форум. Серия: Роль науки и образования в современном информационном обществе. 2024. № S2(32). С. 235–240.
14. Букин А. В., Самонов А. В., Тихонов Э. И. Обнаружение инцидентов информационной безопасности на основе технологии нейронных сетей // Вопросы кибербезопасности. 2022. № 5(51). С. 61–73. DOI: 10.21681/2311-3456-2022-5-61-73.
15. Карпенко М. П. Токенизация как метод количественного измерения информации и знаний в учебных текстах профессионального образования // Инновации в образовании. 2025. № 3. С. 40–50.
16. Обработка естественного языка в действии / Л. Хобсон, Х. Ханнес, Х. Коул / перевод с английского И. Пальти, С. Черников. – СПб: Питер, 2020. 575 с.

DETECTION OF PHISHING EMAILS USING RECURRENT NEURAL NETWORKS

Boldyrikhin N. V.[3], Yadrets E. A.[4]

Keywords: *cyberbullying, phishing protection, recurrent neural networks RNN, LSTM, BiLSTM.*

Purpose of the study: *to consider the features of the use of recurrent neural networks in solving the problem of detecting phishing emails.*

Methods of research: *comparison, mathematical and software modeling, system analysis.*

Result(s): *The concept and types of phishing attacks are considered. The analysis of modern publications on the use of recurrent neural networks in phishing detection tasks has been carried out, which has shown that the use of recurrent networks makes it possible to detect phishing emails with high probability. Publicly available datasets have been analyzed: most datasets are focused on detecting phishing URLs. The few datasets focused on the text of an email are overwhelmingly in English, and high-quality Russian-language datasets are not publicly available, so our own dataset of Russian-language emails was compiled. Mathematical and software modeling of various recurrent neural networks for detecting phishing emails has also been carried out: RNN, LSTM, BiLSTM and a comparative analysis of their characteristics has been carried out. The dependences of loss characteristics and accuracy on the number of epochs are revealed. A comparative analysis of recurrent networks has shown that the BiLSTM network, which detected 91.43 % of phishing emails, was the most effective in solving phishing detection problems in the framework of research. The RNN network showed the worst characteristics, which detected only 50.71 % of phishing emails from the test sample. It should be noted that these results were obtained for networks trained on small-volume datasets (300 emails).*

Scientific novelty: *the research results allow us to reasonably conclude that of the considered recurrent neural networks, BiLSTM is the one that best copes with the tasks of detecting phishing emails with small amounts of training dataset.*

References

1. Kostrikina A. O., Lazunin K.A. Informacionnaya bezopasnost' v kriticheskoj informacionnoj bezopasnosti // Problemy nauchnoj mysli. 2024. Vol. 4. № 1. pp. 82–85.
2. Chapis M. A. Informacionnaya bezopasnost' gosudarstva kak pravovoj poryadok obespecheniya nacional'noj bezopasnosti v informacionnoj sfere // Naukosfera. 2024. № 6(1). pp. 551–557. DOI: 10.5281/zenodo.11638587.
3. Dobrodeev A. Yu. Pokazateli informacionnoj bezopasnosti kak xarakteristika (mera) sootvetstviya setej i organizacij svyazi trebovaniyam informacionnoj bezopasnosti // Trudy CNIIS. Sankt-Peterburgskij filial. 2020. Vol. 2. № 10. pp. 50–78.
4. Lukmanova K. A., Kartak V. M. Razrabotka sistemy zashhity ot fishingovyx atak s ispol'zovaniem programmno-apparatnoj realizacii metodov mashinnogo obucheniya // Modelirovanie, optimizaciya i informacionnye texnologii. 2024. 12(4). DOI: 10.26102/2310-6018/2024.47.4.033.
5. Kornyuxina S. P., Laponina O. R. Issledovanie vozmozhnostej algoritmov glubokogo obucheniya dlya zashhity ot fishingovyx atak // International Journal of Open Information Technologies. 2023. Vol. 11. № 6. pp. 163–174.
6. Yerima S. Y., Alzaylaee M. K. High accuracy phishing detection based on convolutional neural networks // 2020 3rd International Conference on Computer Applications & Information Security (ICCAIS). IEEE, 2020. pp. 1–6. DOI:10.1109/ICCAIS48893.2020.9096869.
7. Wang W. et al. PDRCNN: Precise phishing detection with recurrent convolutional neural networks // Security and Communication Networks. 2019. Vol. 2019. pp. 1–15. DOI:10.1155/2019/2595794.
8. Catal C. et al. Applications of deep learning for phishing detection: a systematic literature review // Knowledge and Information Systems. 2022. Vol. 64. № 6. pp. 1457–1500. DOI:10.1007/s10115-022-01672-x.
9. Dhanavanthini P., Chakkravarthy S. S. Phish-armour: phishing detection using deep recurrent neural networks. Soft Comput (2023). DOI: 10.1007/s00500-023-07962-y.
10. Filimonov A. V., Osipov A. V., Pleshakova E. S., Gataullin S. T. Nejrosetevye metody raspoznavaniya emocij rechi dlya protivodejstviya moshennichestvu v telekommunikacionnyx sistemax // Voprosy kiberbezopasnosti [Cybersecurity issues]. 2022. № 6(52). pp. 83–92. DOI:10.21681/2311-3456-2022-6-83-92.
11. Texnologii iskusstvennogo intellekta i kiberbezopasnost': monografiya / A. B. Menisov. – M: Aj Pi Ar Media, 2022. 133 p.
12. Primenenie iskusstvennogo intellekta dlya resheniya zadachi obespecheniya bezopasnosti informacii, peredavaemoj v setyax / V. I. Yuxnov, A. I. Sosnovskij, N. V. Boldyrixin, I. A. Sosnovskij // Trudy Severo-Kavkazskogo filiala Moskovskogo texnicheskogo universiteta svyazi i informatiki. 2023. № 2. pp. 26–28.
13. Bimoldina Zh. A. Kak iskusstvennyj intellekt menyaet pravila igry v kiberbezopasnosti // Forum. Seriya: Rol' nauki i obrazovaniya v sovremennom informacionnom obshhestve. 2024. № S2(32). pp. 235–240.
14. Bukin A. V., Samonov A. V., Tixonov E. I. Obnaruzhenie incidentov informacionnoj bezopasnosti na osnove texnologii nejronnyx setej // Voprosy kiberbezopasnosti [Cybersecurity issues]. 2022. № 5(51). pp. 61–73. DOI: 10.21681/2311-3456-2022-5-61-73.
15. Karpenko M. P. Tokenizaciya kak metod kolichestvennogo izmereniya informacii i znanij v uchebnyx tekstax professional'nogo obrazovaniya // Innovacii v obrazovanii. 2025. № 3. pp. 40–50.
16. Obrabotka estestvennogo yazyka v dejstvii / L. Xobson, X. Xannes, X. Koul. SPb: Piter, 2020. 575 p.

3 Nikolay N. Boldyrikhin, Ph.D. (in Engineering sciences), Associate Professor of the Department of Cyber security of Information Systems at the Don State Technical University, Rostov-on-Don, Russia. E-mail: boldyrikhin@mail.ru

4 Eduard A. Yadrets, master's student of the Department of Cybersecurity of Information Systems at the Don State Technical University. Rostov-on-Don, Russia, E-mail: xperia1058@gmail.com

ПОДХОД К ОБЪЯСНИМОМУ ОБНАРУЖЕНИЮ АНОМАЛИЙ В ПОТОКЕ ДАННЫХ ОТ ТЕХНОЛОГИЧЕСКИХ ПРОЦЕССОВ

Новикова Е. С.[1], Бухтияров М. А.[2], Котенко И. В.[3], Саенко И. Б.[4], Федорченко Е. В.[5]

DOI: 10.21681/2311-3456-2025-4-142-151

Цель исследования: *разработка подхода к выявлению аномалий в данных технологических процессов на основе объяснимого машинного обучения в целях дальнейшего выбора контрмер с учетом возможных источников аномалий.*

Методы исследования: *статистический анализ, методы машинного обучения, методы генерации объяснений к прогнозам модели машинного обучения.*

Полученные результаты: *предложен подход к объяснимому обнаружению аномалий в потоке данных от технологических процессов, и представлены его основные этапы, в основе которых лежит преобразование входного вектора данных в матрицу и выявление аномалий с помощью сверточной нейронной сети; разработана методика трансформации вектора данных в матрицу, и оценено влияние алгоритма преобразования данных на эффективность решения задачи выявления аномалий; разработана методика тестирования точности генерируемых объяснений и выполнена экспериментальная оценка методов SHAP, Grad-CAM и Guided Grad-CAM.*

Научная новизна: *предложенный подход к выявлению аномалий в данных технологического процесса отличается от существующих использованием разработанной методики преобразования вектора входных данных в матрицу, что позволяет применить сверточную нейронную сеть в качестве аналитической модели выявления аномалий и методы генерации объяснений, разработанные специально для нейронных сетей данной архитектуры.*

Вклад: *Новикова Е. С. – разработка методики преобразования входного потока данных; Бухтияров М. А. – экспериментальное исследование предложенного подхода; Котенко И. В. – разработка общего подхода к объяснимому обнаружению аномалий в рамках концепции динамического оценивания защищенности информационных систем в условиях неопределенности исходных данных; Котенко И. В., Саенко И. Б. и Федорченко Е. В. – анализ современных исследований по выявлению аномалий в технологических процессах и формированию объяснений к прогнозам моделей машинного обучения.*

Ключевые слова: *обнаружение кибератак и аномалий, промышленные киберфизические системы, генерация аномалий, оценка точности объяснений.*

Введение

Цифровая трансформация производственных систем связана с внедрением технологий Интернета вещей, которые позволяют усовершенствовать производственные процессы, повысить эффективность их управления, осуществлять мониторинг состояния оборудования [1]. Однако интеграция сетевых технологий, обеспечивающих в том числе удаленное подключение к корпоративным информационным системам, приводит к тому, что промышленные системы управления сталкиваются с повышенными рисками информационной безопасности [2]. Обеспечение безопасности таких систем является критически важной задачей, поскольку последствия реализаций информационных угроз могут нанести серьезный экономический и экологический ущерб. Например, нарушение процессов водоочистки или водоподготовки воды могут привести не только к значительным сбоям в работе этих систем, но и к загрязнению источников водоснабжения и потенциальной опасности для здоровья [3].

Для своевременного обнаружения аномалий в технологических процессах предложены разнообразные методы как на основе статистического анализа данных, так и на основе машинного обучения, в том числе глубокого обучения [4, 5]. Методы на основе глубокого обучения показали высокую эффективность решения данной задачи, однако их применение значительно усложняет анализ первопричин аномалий несмотря на то, что эта задача важна для промышленных киберфизических систем (КФС) [6]. Определение источника аномалий входит в процедуры оценки рисков, в частности, на основе этих

1 Новикова Евгения Сергеевна, кандидат технических наук, старший научный сотрудник, ФГБУН «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН), г. Санкт-Петербург, Россия. E-mail: novikova@comsec.spb.ru

2 Бухтияров Марат Андреевич, программист, ФГБУН «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН), г. Санкт-Петербург, Россия. E-mail: buhtiarov.marat@gmail.com

3 Котенко Игорь Витальевич, заслуженный деятель науки РФ, доктор технических наук, профессор, главный научный сотрудник и руководитель лаборатории проблем компьютерной безопасности, ФГБУН «Санкт Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН), г. Санкт-Петербург, Россия. E-mail: ivkote@comsec.spb.ru

4 Саенко Игорь Борисович, доктор технических наук, профессор, главный научный сотрудник, ФГБУН «Санкт Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН), г. Санкт-Петербург, Россия. E-mail: ibsaen@comsec.spb.ru

5 Федорченко Елена Владимировна, кандидат технических наук, старший научный сотрудник, ФГБУН «Санкт Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН), г. Санкт-Петербург, Россия. E-mail: doynikova@comsec.spb.ru

данных вычисляются оценки киберрисков для активов организации и формируются возможные контрмеры [7, 8].

В настоящей работе предлагается подход к объяснимому выявлению аномалий, который включает этапы предобработки входных данных, выявления аномалий методами машинного обучения и генерации объяснений к прогнозам обученной модели. Отличительной особенностью подхода является преобразование входного вектора данных от технологических процессов в матрицу, что позволяет применить сверточные нейронные сети и методы генерации объяснений, разработанные для нейронных сетей с данной архитектурой. В работе исследуется точность различных методов генерации объяснений, в частности рассмотрены метод SHAP, который не зависит от архитектуры нейронной сети, и методы Grad-CAM (Gradient-weighted Class Activation Mapping, градиентно-взвешенное отображение активации класса) и Guided Grad-CAM (управляемое градиентно-взвешенное отображение активации класса), разработанные специально для сверточных нейронных сетей. Таким образом, основным вкладом авторов являются: разработка общего подхода к объяснимому выявлению аномалий в потоке данных от технологических процессов; методика выявления аномалий на основе преобразования входного вектора данных в матрицу и сверточной нейронной сети; сравнительный анализ методов генерации объяснений в задачах выявления аномалий в многомерных временных рядах.

Работа построена следующим образом. В разделе 2 обсуждаются исследования в области обнаружения аномалий в технологических процессах и методы генерации объяснений. В разделе 3 представлен разработанный подход к объяснимому обнаружению аномалий, описаны основные его этапы. В разделе 4 рассмотрен сценарий эксперимента, и обсуждаются полученные результаты эффективности обнаружения аномалий и точности генерируемых объяснений. В заключении представлены основные результаты и определяются дальнейшие направления работ.

Анализ релевантных работ

Основным типом данных от КФС являются многомерные временные ряды. При их анализе необходимо учитывать взаимосвязи между различными атрибутами, которые могут быть нелинейными и динамически развивающимися во времени, что обеспечивает выявление аномалий традиционными методами. В последнее время для выявления аномалий предложены методы на основе глубоких нейронных сетей [9, 10].

В частности, в [11] для выявления аномалий во временных рядах представлено решение OmniAnomaly, в основе которого лежит стохастическая рекуррентная нейронная сеть и вариационный автокодировщик для извлечения временных зависимостей между атрибутами. В [12] предложен подход, основанный на применении сверточной нейронной сети и двух автокодировщиков со слоями долгой короткосрочной памяти. Автокодировщики используются для обнаружения аномалий и редких событий путем выявления краткосрочных и долгосрочных отклонений фактических значений датчиков от прогнозируемых значений. Похожее решение представлено в [13], однако сверточные слои нейронной сети здесь дополнены механизмом внимания, что позволяет сфокусировать акцент сети на наиболее важных извлекаемых признаках.

Джао и др. [14] адаптировали генеративные состязательные сети для решения задачи обнаружения аномалий в условиях несбалансированных наборов данных, причем при генерации синтетических данных реализован принцип полного ассоциативного отображения, то есть нормальные данные используются для генерации аномальных данных и наоборот.

Применение методов глубокого обучения для обнаружения аномалий во временных рядах усложняет определение источника аномалий. В зависимости от подхода и целей существует две основные группы методов, которые могут объяснить предсказания модели машинного обучения: методы, учитывающие специфику архитектуры модели, и методы, не зависящие от модели.

Методы, учитывающие специфику конкретной модели, учитывают встроенные свойства модели для формирования объяснимости. К таким методам относятся методы на основе построения карт значимости (Class Activation Maps, CAM, карты активации классов), разработанных для сверточных нейронных сетей и рассчитываемых на основе оценки градиентов; методы послойного распространения релевантности (Layer-wise Relevance Propagation, LRP), применяемые для анализа прогнозов рекуррентных и сверточных нейронных сетей; методы на основе анализа механизмов внимания.

Методы, не зависящие от модели, позволяют объяснить предсказания моделей машинного обучения, не опираясь на специфические свойства этих моделей, и могут применяться к любой модели, независимо от используемых алгоритмов обучения. Методы этой группы, как правило, работают после обучения основной модели (post-hoc методы) и не влияют на процесс обучения и генерацию предсказаний. К таким методам относятся метод аддитивных объяснений на основе вектора Шэпли (SHAP) [15] и метод модельно-независимых локальных объяснений (LIME) [16], которые широко используются

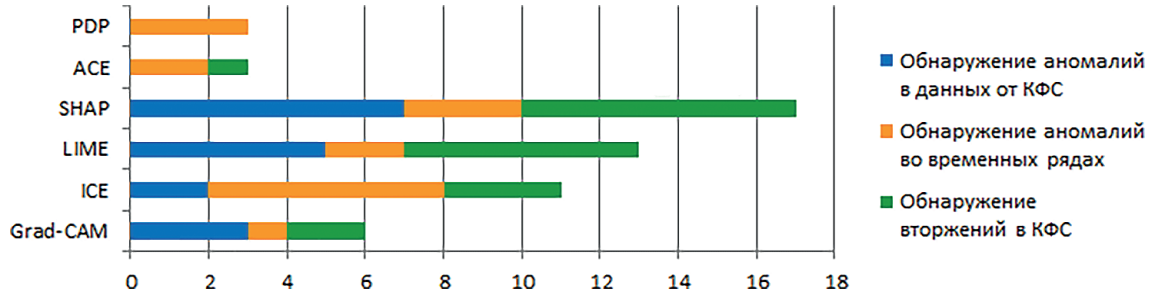


Рис. 1. Частота использования методов объяснения в различных задачах

для объяснения выхода моделей обнаружения аномалий. Следует отметить, что приведенные выше методы относятся к так называемым локальным методам объяснения, то есть позволяют объяснить отдельные предсказания для каждого экземпляра данных, но не поведение модели в целом, как это делают глобальные методы объяснения. В общем случае результатом локального метода объяснения является подмножество признаков данных, значения которых наиболее сильно повлияли на предсказание, сделанное используемой моделью машинного обучения.

Для генерации объяснений к выявленным аномалиям во временных рядах применяются методы, которые, в основном, были разработаны для текстов и изображений. В частности, в [9, 17–19] применяются методы LIME и SHAP. Амели и др. [20] предложили использовать методы на основе карт значимости. Для того, чтобы понять какие методы генерации объяснений наиболее часто используются для формирования объяснений аномалий во временных рядах, было проанализировано более 300 научных статей, извлеченных из электронной базы данных Elsevier Science Direct, опубликованных в интервале с 2021 по 2024 год по модели открытого доступа. Для поиска статей использовались следующие ключевые слова: «anomaly detection in CPS» (обнаружение аномалий в КФС), «anomaly detection in time series» (обнаружение аномалий во временных рядах), «intrusion detection» (обнаружение вторжений). На рис. 1 показана частота использования различных методов генерации объяснений в выбранных статьях (по оси абсцисс показано количество статей): GRAD-CAM, ICE (Individual Conditional expectation, индивидуальное условное ожидание), LIME, SHAP, Shapley values (значения Шепли), ALE (Accumulated Local Effects, накопленные локальные эффекты), PDP (Partial Dependence Plot, график частичной зависимости).

Подход к объяснимому обнаружению аномалий в потоке данных от технологического процесса

Предлагаемый подход к объяснимому обнаружению аномалий в потоке данных от технологического процесса состоит из следующих шагов (рис. 2):

1) предобработка входного потока данных, включающая преобразование входного вектора в матрицу;
2) выявление аномалий на основе сверточной нейронной сети;
3) генерация объяснений к прогнозам модели в виде вектора датчиков и актуаторов технологического процесса.

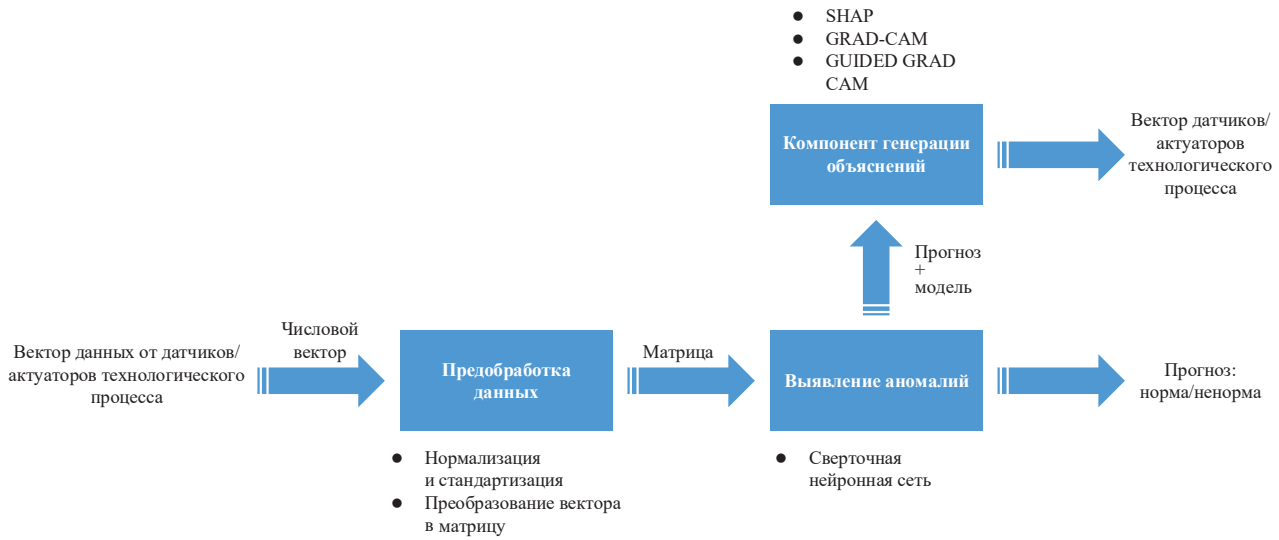


Рис. 2. Основные этапы объяснимого обнаружения аномалий в потоке данных от технологического процесса

Особенностью предлагаемого подхода является преобразование вектора данных в матрицу и использование сверточных нейронных сетей для выявления аномалий, что позволяет использовать методы генерации объяснений, специально разработанные для нейронных сетей данной архитектуры.

Преобразование входного вектора в матрицу выполняется в несколько шагов. Сначала значения входного вектора нормализуются, а затем компонуются в матрицу $n \times n$. Определение размеров матрицы осуществляется по формуле, предложенной в [3]:

$$n = \text{ceil}\left((K + N)/2\right),$$

где K – число количественных атрибутов, а N – число значений, которые могут принимать все категориальные атрибуты, ceil – функция округления в большую сторону до ближайшего целого числа.

Компоновка атрибутов в матрицу может быть осуществлено двумя способами. Первый способ основан только на последовательности атрибутов во входном векторе и не учитывает их подобия друг с другом. Строки матрицы заполняются последовательно, а неиспользованные элементы заполняются нулями. Такой способ часто называется прямой компоновкой данных [3]. В основе второго способа лежит идея, что атрибуты, которые подобны друг другу, должны располагаться в матрице ближе друг к другу, и в этом случае генерируемая матрица отражает пространственные закономерности в данных [21–23].

Подобие между атрибутами чаще всего вычисляется на основе оценки попарного сходства атрибутов, которая может быть представлена косинусным расстоянием, евклидовым расстоянием [24], коэффициентом корреляции Пирсона [25, 26] и т.д. Матрица попарного расстояния служит основой для упорядочивания атрибутов в исходной матрице [23].

В [21, 22] предложен другой подход к упорядочиванию атрибутов в матрице – подобие атрибутов устанавливается на основе построения их проекции в двумерном пространстве. Для нахождения проекции атрибутов сначала выполняется транспонирование обучающей выборки данных, в этом случае каждый атрибут описывается многомерным вектором, а далее могут быть применены как линейные, так и нелинейные алгоритмы снижения размерности. В данной работе анализируются два разных способа построения изображения: прямое преобразование и нелинейное преобразование DeepInsight на основе алгоритма t-SNE.

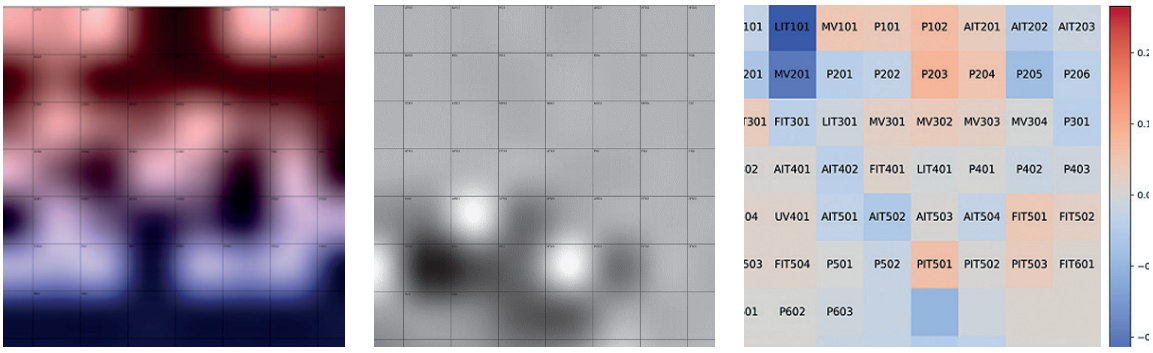
Формат входных данных в виде матрицы позволяет выбирать архитектуры нейронных сетей, в которых сверточные слои используются для извлечения анализируемых признаков. В настоящей работе в качестве аналитической модели предложено использовать простую двуслойную сверточную сеть.

Для генерации объяснений прогнозов сверточной сети предложено исследовать несколько подходов: метод SHAP и методы, разработанные специально для сверточных нейронных сетей Grad-CAM, Guided Grad-CAM.

Метод SHAP не зависит от архитектуры анализируемой модели и применим как для табличных данных, так и для изображений. В его основе лежит теория кооперативных игр, что позволяет оценить вклад каждого признака в конечное решение модели. В контексте решаемой задачи SHAP может быть использован для выявления пикселей, которые наиболее значимы для принятия решения.

Методы Grad-CAM и Guided Grad-CAM основаны на вычислении градиентов выхода модели относительно карт признаков последнего сверточного слоя. Метод Grad-CAM позволяет получить тепловую карту, подсвечивающую важные области входного изображения. Метод Guided Grad-CAM строит более детализированные и точные объяснения, благодаря комбинированию методов Grad-CAM и метода управляемого обратного распространения ошибки (guided backpropagation).

На рис. 3 представлены примеры объяснений, генерируемых разными способами.



а) б) в)

Рис. 3. Примеры генерируемых тепловых карт разными методами: а) методом Grad-CAM, б) методом Guided Grad-CAM, в) методом SHAP

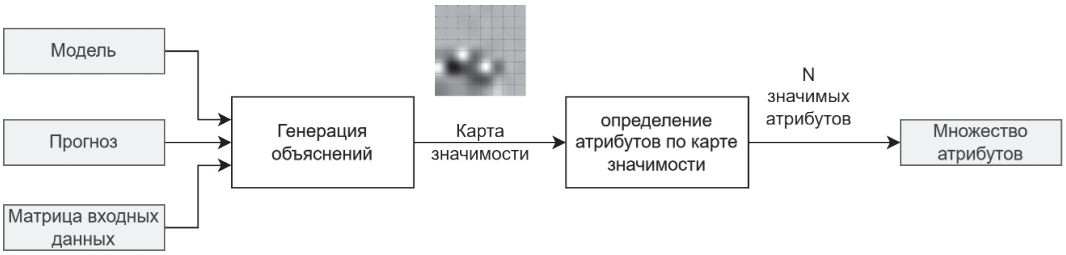


Рис. 4. Общая схема генерации объяснений различными методами

Для метода Grad-CAM значимость признака определяется по цвету: красный – признаки с высокой значимостью, синий – с низкой. В Guided Grad-CAM выполняется выделение областей значимых областей – чем они темнее, тем больший вклад они оказывают на выполненное аналитической моделью решение. В методе SHAP «красные» признаки указывают на то, что признак увеличивает вероятность предсказания, в то время как признаки, отмеченные синим цветом, наоборот, уменьшают эту вероятность. Таким образом, для получения списка потенциально аномальных данных необходимо выполнить обратную операцию нахождения атрибута по матрице. Общая схема генерации объяснений для каждого из метода представлена на рис. 4.

Для выполнения экспериментальной оценки была разработана библиотека DatasetToImageTransformer[6] на языке Python, которая выполняет преобразование числовых данных в матрицы (изображений) разными способами. Для каждого способа построения матрицы формируется словарь позиций атрибутов, в котором каждому элементу матрицы ставится в соответствие название атрибутов. Такое решение позволяет использовать данную библиотеку для анализа потока данных, а также выполнять обратное преобразование – по координатам элемента матрицы получить название атрибута. Для размеченных данных библиотека создает обучающие наборы данных, сгруппированные по каталогам, которые могут быть использованы для обучения различных аналитических моделей.

Экспериментальная оценка предложенного подхода

Целью эксперимента являлось определение эффективности предложенного подхода к выявлению аномалий, который состоит из одинаково значимых задач – обнаружения аномалий и их объяснений. Сценарий эксперимента был разработан таким образом, чтобы оценить эффективность каждого компонента подхода. В первой части эксперимента выполнялась оценка эффективности обнаружения аномальных данных от технологического процесса, на второй части эксперимента – оценка эффективности методов генерации объяснений.

6 https://github.com/Kotolow/FTIConverter.git

В обоих случаях использовался набор данных Secure Water Treatment (SWaT) версии 2015 [27], созданный с помощью программно-аппаратного стенда, представляющего собой уменьшенную копию водоочистных сооружений. Данный набор отражает 11 дней функционирования системы, 7 дней из которых соответствуют норме, а 4 дня содержат 36 атак разной длительности. В таблице 1 представлены характеристики набора SWaT версии 2015 года для 4-х аномальных дней, включая число датчиков, значения которых были изменены.

Таблица 1.
Характеристика атак в наборе данных SWaT

Тип записи в наборе	Число датчиков, измененных в результате атак	Число записей
Норма	0	399157
Аномалия	1	43213
	2	7789
	3	2452

Атаки отличаются числом атакуемых датчиков. В частности, было выполнено 6 атак, целью которых была подмена значений двух и трех датчиков, а также 7 атак на датчики, принадлежащие разным технологическим подпроцессам. Кроме того, отличительной особенностью данного набора является наличие текстовых объяснений, какие вредоносные воздействия проводились со стендом, и какие датчики и актуаторы были изменены, что делает его пригодным для оценки эффективности методов генерации объяснений.

На первом этапе в качестве метрик эффективности обнаружения аномалий были использованы показатели точности (precision), полноты (recall) и F1-мера, которые вычисляются на основе матрицы ошибок. Результаты экспериментов для разных способов генерации матрицы представлены в таблице 2. Из нее следует, что способ построения входной матрицы не влияет на точность обнаружения аномалий в потоке данных. Применение достаточно простой сверточной нейронной сети дает высокую точность решения задачи.

Таблица 2.

Оценка эффективности обнаружения аномалий

Способ генерации матрицы	Прямое преобразование			Нелинейное преобразование DeepInsight на основе t-SNE			Число записей
Класс	Точность (precision)	Полнота (recall)	F1-мера	Точность (precision)	Полнота (recall)	F1-мера	
Аномалия (атака)	0,99	0,94	0,96	0,98	0,93	0,96	54621
Норма	0,99	1,00	0,99	0,99	1,00	0,99	395298
Макро среднее	0,99	0,97	0,98	0,99	0,96	0,98	449919
Микро среднее	0,99	0,99	0,99	0,99	0,99	0,99	449919

На втором этапе эксперимента была выполнена оценка эффективности компонента генерации объяснений, которая определялась как точность объяснений. В данной работе точность объяснений предлагается оценивать на основе сравнения множества датчиков/актуаторов, которые были реально изменены в ходе деструктивных воздействий на систему, со множеством датчиков/актуаторов, которое было получено в результате применения методов генерации объяснений для каждого прогноза модели. В этих целях предложена метрика AHR (Any Hit Rate)[7], которая вычисляется следующим образом.

Пусть E_i есть множество датчиков/актуаторов, которые демонстрируют аномальное поведение в i-й момент времени, т.е. определены для i-ой точки данных, а E_i^* – множество датчиков/актуаторов, которые были получены в результате применения метода генерации к i-му прогнозу. Тогда

$$AHR = \frac{\sum_{i=1}^{N} any\_hit(i, max\_overlap)}{N},$$

$$\text{где } any\_hit(i) = \begin{cases} 1, E_i \cap E_i^* \neq \emptyset; \\ 0, E_i \cap E_i^* = \emptyset \end{cases},$$

где функция $any\_hit(i, max\_overlap)$ возвращает 1, если число совпадающих датчиков между объяснением и реальными данными больше или равно порогового значения $max\_overlap$, и 0 – в противном случае. Метрика AHR обладает высокой практической значимостью для промышленных систем, так как даже частичная локализация аномального поведения позволит специалистам выявить причину аномалии и своевременно принять необходимые контрмеры.

Очевидно, что для практического применения данной метрики необходимо выполнить преобразование исходного набора данных, дополнив его данными от аномальных сенсоров. В таблице 3 представлен пример измененного набора данных.

Следует также отметить, что из анализа были исключены записи, для которых не были указаны аномальные датчики; примером такой записи служит последняя строка в таблице 3, в которой «[]» обозначают пустой массив аномальных датчиков.

Также исходное множество было разбито на три подмножества:

- TP – строки из подмножества, для которых реально определено атакующее воздействие, которое было верно определено бинарным классификатором;
- $TP + FN$ – все строки из подмножества, для которых реально определено атакующее воздействие;
- $TP + FN + FP$ – все строки из подмножества, для которых реально определено атакующее воздействие, и строки, для которых детектор аномалий ошибочно предсказал состояние «аномалия».

Таблица 3.

Фрагмент измененного набора данных SWaT

Временная метка	FIT101	LIT101	MV101	...	Датчики
2015-12-28 10:28:14	2.494	817.674	2	...	[MV101]
2015-12-28 10:28:15	2.536	817.974	2	...	[MV101, P205]
...	...	...	...	...	...
2016-01-01 10:28:14	2.420	573.522	2	...	[]

7 https://www.researchgate.net/publication/360076778_Unsupervised_Multi-Sensor_Anomaly_Localization_with_Explainable_AI

Таблица 4.

Точность сформированных объяснений различными методами генерации объяснений

Метод трансформации вектора данных	Методы генерации объяснений	*AHR* для множества *TP*	*AHR* для множества *TP + FN*	*AHR* для множества *TP + FN + FP*
Нелинейное преобразование DeepInsight на основе t-SNE	Grad-CAM	0,2120	0,0154	0,0049
	Guided Grad-CAM	0,0372	0,0904	0,0009
	SHAP	0,0263	0,0034	0,0002
Прямое преобразование	Grad-CAM	0,0678	0,0466	0,0007
	Guided Grad-CAM	0,1296	0,0718	0,0014
	SHAP	0,0667	0,5973	0,0005

Полученные результаты представлены в таблице 4.

Очевидно, что все методы дают крайне низкую точность, неприемлемую для случаев практического использования. Так, например, метод Grad-CAM достигает максимальной точности для нелинейного метода построения матрицы DeepInsight на множестве верно выявленных аномалий и составляет 0,21. Метод Guided Grad-CAM достигает максимальной точности на множестве верно выявленных аномалий для прямого преобразования. Метод SHAP демонстрирует максимальную точность на множестве $TP + FN$, т.е. на множестве, на котором определены реальные атакующие воздействия. Его точность вначале составляет 0,5973, но резко падает до 0,0005 на множестве векторов, которые включают вектора, которые классификатор относит к аномальным. Это делает неприемлемым использование и данного метода на практике. Возможной причиной такой низкой точности является природа как самих анализируемых данных – временные ряды, так и самих аномалий, которые имеют длительность и могут выражаться различной степенью изменения атрибутов. Предложенные преобразования над входными данными и сама модель учитывают только пространственные связи между данными, при этом временные зависимости между ними не учитываются. Между тем, упомянутые выше методы основаны на предположении, что атрибуты между собой независимы, а данные не зависят друг от друга, что неверно для набора SWaT. Таким образом, выявлена острая необходимость в разработке методов генерации объяснений для моделей машинного обучения, предназначенных для временных рядов.

Заключение

В работе представлен подход к объяснимому выявлению аномалий в потоке данных от технологических процессов. Отличительной особенностью предложенного подхода является использование преобразования входного вектора данных в матрицу, что позволяет применять сверточные нейронные слои для извлечения анализируемых признаков. Выявление аномалий осуществляется при помощи двухслойной сверточной сети, которая показала высокую точность обнаружения аномалий для тестируемого набора данных SWAT, описывающего функционирование системы водоочистных сооружений.

Для реализации компонента формирования объяснений было использовано несколько методов генерации объяснений: методы Grad-CAM, Guided Grad-CAM и SHAP. Было показано, что на текущий момент точность формируемых объяснений низкая, что делает невозможным применение данных методов на практике. Возможной причиной низкой точности являются сами исследуемые данные – многомерные временные ряды.

Дальнейшие направления исследований по этой задаче связаны с разработкой новых моделей выявления аномалий, которые учитывают не только пространственные связи между атрибутами, но и временные. Возможным решением служит использование графовых нейронных сетей. Кроме того, планируется исследование и разработка методов генерации объяснений, которые учитывают особенности многомерных временных рядов, а именно наличие связей между атрибутами как во времени, так и между собой.

Благодарность. Исследование выполнено при поддержке гранта Российского научного фонда №23-11-20024, https://rscf.ru/project/23-11-20024/, и Санкт-Петербургского научного фонда в СПб ФИЦ РАН.

Рецензент: Лаута Олег Сергеевич, доктор технических наук, профессор кафедры комплексного обеспечения информационной безопасности Государственного университета морского и речного флота имени адмирала С. О. Макарова, Санкт-Петербург, Россия. E-mail: laos-82@yandex.ru

Литература

1. Левшун Д. А., Левшун Д. С., Котенко И. В. Обнаружение и объяснение аномалий в индустриальных системах Интернета вещей на основе автокодировщика // Онтология проектирования. 2025. Т.15, № 1(55). С.96–113. DOI:10.18287/2223-9537-2025-15-1-96-113.
2. Котенко И. В., Федорченко Е. В., Новикова Е. С., Саенко И. Б., Данилов А. С. Методология сбора данных для анализа безопасности промышленных киберфизических систем // Вопросы кибербезопасности. 2023. № 5(57). С. 69–79. https://doi.org/10.21681/2311-3456-2023-5-69-79.
3. Novikova E. S., Fedorchenko E. V., Bukhtiyarov M. A., Saenko I. B. Anomaly detection in wastewater treatment process for cyber resilience risks evaluation // Journal of Mining Institute. 2024. Vol. 267. P. 488–500.
4. Dong H., Kotenko I. Cybersecurity in the AI era: analyzing the impact of machine learning on intrusion detection // Knowledge and Information Systems, 2025, 67(5), P. 3915–3966, 102748. DOI: 10.1007/s10115-025-02366-w.
5. Kotenko I. V., Levshun D. A. Machine Learning Methods of Intelligent System Event Analysis for Multistep Cyberattack Detection // Scientific and Technical Information Processing, 2024, Vol. 51, No. 5, P.372–381. Allerton Press, Inc., 2024. Springer Nature. ISSN 0147-6882. DOI: 10.3103/S0147688224700254.
6. Dong H., Kotenko I., Levshun D. Next-Generation IIoT Security: Comprehensive Comparative Analysis of CNN-based Approaches // Knowledge Based Systems, Vol.316, 12 May 2025, 113337. https://doi.org/10.1016/j.knosys.2025.113337.
7. Doynikova E., Novikova E., Murenin I., Kolomeec M., Gaifulina D., Tushkanova O., Levshun D., Meleshko A., Kotenko I. Security Measuring System for IoT Devices // Lecture Notes in Computer Science. 2022. Vol. 13106. P. 256–275.
8. Ning X., Jiang J. Design, Analysis and Implementation of a Security Assessment/Enhancement Platform for Cyber-Physical Systems // IEEE Transactions on Industrial Informatics. 2022. Vol. 18. No. 2. P. 1154–1164.
9. Wang C., Wang B., Liu H., Qu H. Anomaly detection for industrial control system based on autoencoder neural network // Wirel. Commun. Mob. Comput. 2020. P. 8897926–1889792610.
10. Rodríguez M., Tobón D., Múnera D. A framework for anomaly classification in Industrial Internet of Things systems // Internet of Things. 2025. Vol. 29. Article 101446. https://doi.org/10.1016/j.iot.2024.101446.
11. Su Y., Zhao Y., Niu C., Liu R., Sun W., Pei D. Robust anomaly detection for multivariate time series through stochastic recurrent neural network // Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '19). ACM, New York, NY, USA, 2019, pp. 2828–2837. https://doi.org/10.1145/3292500.3330672.
12. Nizam H., Zafar S., Lv Z., Wang F., Hu X. Real-Time Deep Anomaly Detection Framework for Multivariate Time-Series Data in Industrial IoT // IEEE Sensors Journal. 2022. Vol. 22. No. 23. P. 22836–22849, doi: 10.1109/JSEN.2022.3211874.
13. Liu Y. et al. Deep Anomaly Detection for Time-Series Data in Industrial IoT: A Communication-Efficient On-Device Federated Learning Approach // IEEE Internet of Things Journal. 2021. Vol. 8. No. 8. P. 6348–6358. doi: 10.1109/JIOT.2020.3011726.
14. Zhao P., Ding Z., Li Y., Zhang X., Zhao Y., Wang H, Yang Y. SGAD-GAN: Simultaneous Generation and Anomaly Detection for time-series sensor data with Generative Adversarial Networks // Mechanical Systems and Signal Processing. 2024. Vol. 210. Article 111141. https://doi.org/10.1016/j.ymssp.2024.111141.
15. Lundberg S. M., Lee S.-I. A unified approach to interpreting model predictions // Advances in neural information processing systems (NIPS'17), 2017, pp. 4768–4777.
16. Ribeiro M. T., Singh S., Guestrin C. Why Should I Trust You?: Explaining the Predictions of Any Classifier // Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'16). ACM, NY, USA, 2016, pp. 1135–1144.
17. Neshenko N., Bou-Harb E., Furht B. A behavioral-based forensic investigation approach for analyzing attacks on water plants using GANs // Forensic Science International: Digital Investigation. 2021. Vol. 37. Article 301198.
18. Antwarg L., Miller R. M., Shapira B., Rokach L. Explaining anomalies detected by autoencoders using SHAP. arXiv preprint arXiv:1903.02407. 2019.
19. Oliveira D., Vismari L. F., Nascimento A. M., de Almeida J. R., Cugnasca P. S., Camargo J. B., Almeida L., Gripp R., Neves M. A new interpretable unsupervised anomaly detection method based on residual explanation // IEEE Access. 2021. Vol. 10, pp. 1401–1409.
20. Ameli M., Becker P. A., Lankers K., van Ackeren M., Bähring H., Maaß W. Explainable unsupervised multi-sensor industrial anomaly detection and categorization // 21st IEEE International Conference on Machine Learning and Applications (ICMLA), 2022, pp. 1468–1475.
21. Sharma A., Vans E., Shigemizu D., Boroevich K. A., Tsunoda T. DeepInsight: A methodology to transform a non-image data to an image for convolution neural network architecture. Sci. Rep. 2019. Vol. 9. Article 11399. https://doi.org/10.1038/s41598-019-47765-6.
22. Bazgir O., Zhang R., Dhruba S. R., Rahman R., Ghosh S., Pal R. Representation of features as images with neighborhood dependencies for compatibility with convolutional neural networks. Nat. Commun. 2020. Vol. 11. Article 4391. https://doi.org/10.1038/s41467-020-18197-y.
23. Zhu Y., Brettin T., Xia F., Partin A., Shukla M., Yoo H., Evrard Y. A., Doroshow J. H., Stevens R. L. Converting tabular data into images for deep learning with convolutional neural networks. Sci. Rep. 2021. Vol. 11. Article 11325. https://doi.org/10.1038/s41598-021-90923-y.
24. Zhou Q., Chen J., Liu H., He S., Meng W. Detecting Multivariate Time Series Anomalies with Zero Known Label. 2022. arXiv.org/abs/2208.02108.
25. Xie Y., Zhang H., Babar M. A. Multivariate Time Series Anomaly Detection by Capturing Coarse-Grained Intra- and Inter-Variate Dependencies. 2025. arXiv.org/abs/2501.16364.
26. Kamarthi H., Kong L., Rodriguez A., Zhang C., Prakash B. A. Learning Graph Structures and Uncertainty for Accurate and Calibrated Time-series Forecasting. 2024. arXiv.org/abs/2407.02641.
27. Goh J., Adepu S., Junejo K., Mathur A. A Dataset to Support Research in the Design of Secure Water Treatment Systems // Critical Information Infrastructures Security. CRITIS 2016. Lecture Notes in Computer Science. Vol. 10242. Springer, Cham. https://doi.org/10.1007/978-3-319-71368-7_8.

AN APPROACH TO EXPLAINABLE ANOMALY DETECTION IN DATA STREAMS FROM TECHNOLOGICAL SYSTEMS

Novikova E. S.[8], Bukhtiarov M. A.[9], Kotenko I. V.[10], Saenko I. B.[11], Fedorchenko E. V.[12]

Keywords: *cyber attack and anomaly detection, industrial cyberphysical systems anomaly generation, evaluation of explanation accuracy.*

The purpose of the study: *development of an approach to identify anomalies in process data based on explainable machine learning in order to further select countermeasures taking into account possible sources of anomalies.*

Research methods: *statistical analysis, machine learning methods, methods of generating explanations for machine learning model predictions.*

Results obtained: *an approach to explainable anomaly detection in the flow of data from technological processes is proposed, its main stages are presented, which is based on the transformation of the input data vector into a matrix, and the detection of anomalies using a convolutional neural network; the method of transformation of the data vector into a matrix is developed and the influence of the data transformation algorithm on the efficiency of solving the problem of anomaly detection is evaluated; the method of testing the accuracy of the generated explanations is developed and the experimental evaluation is carried out.*

Scientific novelty: *the proposed approach to the identification of anomalies in process data differs from the existing ones by using the technique of transforming the input data vector into a matrix, which allows us to apply a convolutional neural network as an analytical model of anomaly detection and methods of generating explanations developed specifically for neural networks of this architecture.*

Contributions: *Evgenia Novikova – development of a method for converting the input data flow; Marat Bukhtiyarov – experimental study of the proposed approach; Igor Kotenko – development of a general approach to explainable detection of anomalies of the concept of dynamic assessment of the security of information systems in conditions of uncertainty of the initial data; Igor Kotenko, Igor Saenko and Elena Fedorchenko – analysis of the state of arts in identifying anomalies in technological processes and forming explanations for forecasts of machine learning models.*

References

1. Levshun D. A., Levshun D. S., Kotenko I. V. Detecting and explaining anomalies in industrial Internet of things systems using an autoencoder // Ontology of designing. 2025. Vol.15, No.1(55). P.96-113. DOI:10.18287/2223-9537-2025-15-1-96-113.
2. Kotenko I. V., Fedorchenko E. V., Novikova E. S., Saenko I. B., Danilov A. S. Methodology of data collection for security analysis of industrial cyber-physical systems // Cybersecurity Issues. 2023. No. 5 (57). P. 69-79. https://doi.org/10.21681/2311-3456-2023-5-69-79.
3. Novikova E. S., Fedorchenko E. V., Bukhtiyarov M. A., Saenko I. B. Anomaly detection in wastewater treatment process for cyber resilience risks evaluation // Journal of Mining Institute. 2024. Vol. 267. P. 488–500.
4. Dong H., Kotenko I. Cybersecurity in the AI era: analyzing the impact of machine learning on intrusion detection // Knowledge and Information Systems, 2025, 67(5), P. 3915–3966, 102748. DOI: 10.1007/s10115-025-02366-w.
5. Kotenko I. V., Levshun D. A. Machine Learning Methods of Intelligent System Event Analysis for Multistep Cyberattack Detection // Scientific and Technical Information Processing, 2024, Vol. 51, No. 5, P.372–381. Allerton Press Inc., 2024. Springer Nature. ISSN 0147-6882. DOI: 10.3103/S0147688224700254
6. Dong H., Kotenko I., Levshun D. Next-Generation IIoT Security: Comprehensive Comparative Analysis of CNN-based Approaches // Knowledge Based Systems, Vol.316, 12 May 2025, 113337. https://doi.org/10.1016/j.knosys.2025.113337.
7. Doynikova E., Novikova E., Murenin I., Kolomeec M., Gaifulina D., Tushkanova O., Levshun D., Meleshko A., Kotenko I. Security Measuring System for IoT Devices // Lecture Notes in Computer Science. 2022. Vol. 13106. P. 256–275.
8. Ning X., Jiang J. Design, Analysis and Implementation of a Security Assessment/Enhancement Platform for Cyber-Physical Systems // IEEE Transactions on Industrial Informatics. 2022. Vol. 18. No. 2. P. 1154–1164.
9. Wang C., Wang B., Liu H., Qu H. Anomaly detection for industrial control system based on autoencoder neural network // Wirel. Commun. Mob. Comput. 2020. P. 8897926–1889792610.
10. Rodríguez M., Tobón D., Múnera D. A framework for anomaly classification in Industrial Internet of Things systems // Internet of Things. 2025. Vol. 29. Article 101446. https://doi.org/10.1016/j.iot.2024.101446.

8 Evgenia S. Novikova, Ph.D. of Technical Sciences, Senior researcher of Laboratory of Computer Security Problems at St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS), St. Petersburg, Russia. E-mail: novikova@comsec.spb.ru

9 Marat A. Bukhtiarov, Software Developer of Laboratory of Computer Security Problems at St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS), St. Petersburg, Russia. E-mail: buhtiarov.marat@gmail.com

10 Igor V. Kotenko, Honored Worker of Science of the Russian Federation, Dr.Sc. of Technical Sciences, Professor, Chief Scientist and Head of Laboratory of Computer Security Problems at St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS), St. Petersburg, Russia. E-mail: ivkote@comsec.spb.ru

11 Igor B. Saenko, Dr.Sc. of Technical Sciences, Professor, Leading researcher of Laboratory of Computer Security Problems at St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS), St. Petersburg, Russia. E-mail: ibsaen@comsec.spb.ru

12 Elena V. Fedorchenko, Ph.D. of Technical Sciences, Senior researcher of Laboratory of Computer Security Problems at St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS), St. Petersburg, Russia. E-mail: doynikova@comsec.spb.ru

11. Su Y., Zhao Y., Niu C., Liu R., Sun W., Pei D. Robust anomaly detection for multivariate time series through stochastic recurrent neural network // Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '19). ACM, New York, NY, USA, 2019, pp. 2828–2837. https://doi.org/10.1145/3292500.3330672.
12. Nizam H., Zafar S., Lv Z., Wang F., Hu X. Real-Time Deep Anomaly Detection Framework for Multivariate Time-Series Data in Industrial IoT // IEEE Sensors Journal. 2022. Vol. 22. No. 23. P. 22836–22849, doi: 10.1109/JSEN.2022.3211874.
13. Liu Y. et al. Deep Anomaly Detection for Time-Series Data in Industrial IoT: A Communication-Efficient On-Device Federated Learning Approach // IEEE Internet of Things Journal. 2021. Vol. 8. No. 8. P. 6348–6358. doi: 10.1109/JIOT.2020.3011726.
14. Zhao P., Ding Z., Li Y., Zhang X., Zhao Y., Wang H, Yang Y. SGAD-GAN: Simultaneous Generation and Anomaly Detection for time-series sensor data with Generative Adversarial Networks // Mechanical Systems and Signal Processing. 2024. Vol. 210. Article 111141. https://doi.org/10.1016/j.ymssp.2024.111141.
15. Lundberg S. M., Lee S. -I. A unified approach to interpreting model predictions // Advances in neural information processing systems (NIPS'17), 2017, pp. 4768–4777.
16. Ribeiro M. T., Singh S., Guestrin C. Why Should I Trust You?: Explaining the Predictions of Any Classifier // Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'16). ACM, NY, USA, 2016, pp. 1135–1144.
17. Neshenko N., Bou-Harb E., Furht B. A behavioral-based forensic investigation approach for analyzing attacks on water plants using GANs // Forensic Science International: Digital Investigation. 2021. Vol. 37. Article 301198.
18. Antwarg L., Miller R. M., Shapira B., Rokach L. Explaining anomalies detected by autoencoders using SHAP. arXiv preprint arXiv:1903.02407. 2019.
19. Oliveira D., Vismari L. F., Nascimento A. M., de Almeida J. R., Cugnasca P. S., Camargo J. B., Almeida L., Gripp R., Neves M. A new interpretable unsupervised anomaly detection method based on residual explanation // IEEE Access. 2021. Vol. 10, pp. 1401–1409.
20. Ameli M., Becker P. A., Lankers K., van Ackeren M., Bähring H., Maaß W. Explainable unsupervised multi-sensor industrial anomaly detection and categorization // 21st IEEE International Conference on Machine Learning and Applications (ICMLA), 2022, pp. 1468–1475.
21. Sharma A., Vans E., Shigemizu D., Boroevich K. A., Tsunoda T. DeepInsight: A methodology to transform a non-image data to an image for convolution neural network architecture. Sci. Rep. 2019. Vol. 9. Article 11399. https://doi.org/10.1038/s41598-019-47765-6.
22. Bazgir O., Zhang R., Dhruba S. R., Rahman R., Ghosh S., Pal R. Representation of features as images with neighborhood dependencies for compatibility with convolutional neural networks. Nat. Commun. 2020. Vol. 11. Article 4391. https://doi.org/10.1038/s41467-020-18197-y.
23. Zhu Y., Brettin T., Xia F., Partin A., Shukla M., Yoo H., Evrard Y. A., Doroshow J. H., Stevens R. L. Converting tabular data into images for deep learning with convolutional neural networks. Sci. Rep. 2021. Vol. 11. Article 11325. https://doi.org/10.1038/s41598-021-90923-y.
24. Zhou Q., Chen J., Liu H., He S., Meng W. Detecting Multivariate Time Series Anomalies with Zero Known Label. 2022. arXiv.org/abs/2208.02108.
25. Xie Y., Zhang H., Babar M. A. Multivariate Time Series Anomaly Detection by Capturing Coarse-Grained Intra- and Inter-Variate Dependencies. 2025. arXiv.org/abs/2501.16364.
26. Kamarthi H., Kong L., Rodriguez A., Zhang C., Prakash B. A. Learning Graph Structures and Uncertainty for Accurate and Calibrated Time-series Forecasting. 2024. arXiv.org/abs/2407.02641.
27. Goh J., Adepu S., Junejo K., Mathur A. A Dataset to Support Research in the Design of Secure Water Treatment Systems // Critical Information Infrastructures Security. CRITIS 2016. Lecture Notes in Computer Science. Vol. 10242. Springer, Cham. https://doi.org/10.1007/978-3-319-71368-7_8.

АНАЛИЗ РАЗМЕЩАЕМЫХ В СЕТИ ОТКРЫТЫХ ДАННЫХ В ЦЕЛЯХ ПОЛУЧЕНИЯ ИНФОРМАЦИИ О КРИМИНОГЕННОЙ ОБСТАНОВКЕ

Жарова А. К.[1], Елин В. М.[2], Атласов И. В.[3]

DOI: 10.21681/2311-3456-2025-4-152-159

Цель статьи: *предложить методику формирования цифрового профиля человека, который может быть использован для анализа и прогнозирования криминогенной обстановки.*

Метод исследования: *использованы логико-математические методы, такие как типологическая модель, детерминистская модель и имитационное моделирование. Кроме того, используется метод анализа математических моделей, включая стохастическую модель, что позволяет получить более точную и детализированную картину. Входными данными для анализа являлись данные, оставляемые человеком в процессе своей деятельности в интернете*

Результат: *системы анализа данных могут быть применены для извлечения, анализа, преобразования и представления информации, имеющей существенное значение при проведении оперативно-розыскных и следственных мероприятий. Авторы, ссылаясь на имеющуюся судебную практику, раскрывают значение коммуникационных данных для получения цифрового профиля человека, формально не относящихся к персональным данным как категории информации ограниченного доступа. Экспериментальная часть статьи представляет собой математическое моделирование криминогенной обстановки на основании анализа независимых цифровых данных, оставленных пользователем социальной Сети. Таким образом, в результате проведенного исследования выявлены закономерности, позволяющие в дальнейшем предсказывать поведение групп людей, осуществляющих передачу вредоносной информации в сети, либо размещение информации указанной категории.*

Практическая ценность: *на основании проведенного теоретического эксперимента сделан вывод о возможности применения математических методов в криминологическом анализе преступности*

Ключевые слова: *информационные технологии, коммуникационные данные, анализ цифровых теней и цифровых следов, персональные данные, математическое моделирование.*

Введение

Активность пользователей Интернета фиксируется и отражается в тех или иных интернет-данных, которые в дальнейшем могут быть проанализированы в целях получения цифрового профиля человека. Каждый фрагмент интернет-контента окружён множеством элементов коммуникационных данных. Даже в случае, когда контент зашифрован и системы анализа данных не могут получить информацию о содержании персональных данных (какую-либо личную информацию об отправителе или получателе), существующие системы анализа коммуникационных данных (в том числе, связанных с зашифрованным контентом), могут содержать обширный массив личной информации, включая личность, его друзей, географические координаты отправителя и получателя, IP устройства передачи сообщения, включая его полные технические характеристики. Анализ связанных коммуникационных данных усиливает возможности получения полной информации о человеке[4].

Европейский суд по правам человека (ЕСПЧ) обратил внимание на проблему, связанную с тем, что системы анализа данных могут формировать достоверный цифровой профиль человека на основе остаточных цифровых следов человека, оставленных им в интернете. В своём постановлении ЕСПЧ пришёл к выводу, что в интернете содержится гораздо больше коммуникационных данных, чем самого контента[5].

В России, согласно Федеральному закону «О персональных данных»[6], коммуникационные данные сами по себе не являются персональными. Однако, когда они собираются вместе и обрабатываются с помощью специальных систем анализа данных, они могут быть преобразованы в персональные

1 Жарова Анна Константиновна, доктор юридических наук, профессор Финансового университета при Правительстве Российской Федерации, Москва. E-mail: anna_jarova@mail.ru

2 Елин Владимир Михайлович, кандидат педагогических наук, доцент кафедры информационной безопасности Московского университета МВД России имени В.Я.Кикиотя, доцент кафедры информационной безопасности Финансового университета при Правительстве Российской Федерации, Москва. E-mail: elin_vm@mail.ru

3 Атласов Игорь Викторович, доктор физико-математических наук, профессор Московского университета МВД России имени В.Я.Кикиотя, Москва. E-mail: atlasov.igor.777@gmail.com

4 Big Brother Watch and Others v. the United Kingdom, Application Nos. 58170/13, 62322/14 and 24960/15 (ECtHR May 25, 2021).

5 Big Brother Watch and Others v. the United Kingdom.

6 Федеральный закон от 27.07.2006 N 152-ФЗ «О персональных данных»//СЗ РФ. 2006. № 31 (1 ч.) Ст. 3451.

данные [1, 2]. Более того, в нашей стране уже сложилась судебная практика по делам, в которых оспаривается правомерность применения подобных систем анализа данных[7,8].

Системы анализа больших данных способны как обрабатывать коммуникационные данные без необходимости идентификации пользователя сети [3, 4], так и идентифицировать конкретного человека. Например, в целях выявления лица, совершившего правонарушение с использованием информационно-коммуникационных технологий (ИКТ).

Можно ли скрыть данные для систем анализа данных?

Возможность обеспечения конфиденциальности информации ограниченного доступа зависит от используемых методов обработки информации. Для технологий анализа интернет-данных основным является факт их размещения в интернете. Сочетание технологий анализа больших данных и искусственного интеллекта (ИИ) позволяет улучшить прогнозирование, принятие решений, оптимизировать процессы и автоматизировать решение задач, например, провести анализ любой информации, в том числе, данных, которые так или иначе можно отнести к персональным данным [5, 6.]. Например, для решения задачи анализа информации, размещенной на странице социальной сети, в отношении которой ее обладатель поставил ограничения по доступу к ней только определенной группы людей, например, друзей.

Но и даже в том случае, если информация не размещена в Сети, технологии анализа данных могут получить ее самостоятельно на основе анализа цифровых теней и цифровых следов [7]. Например, швейцарские учёные провели эксперимент, в целях подтверждения гипотезы – могут ли большие языковые модели (LLM) собирать и раскрывать личную информацию пользователей. В качестве примера ученые взяли 1,5 тысячи случайных профилей с площадки Reddit и проанализировали их активность с помощью LLM. LLM смогли точно определить место рождения и жительства, а также уровень дохода людей по вторичной информации, оставленной пользователями Сети. GPT-4 идентифицировал с точностью 85 %, а LlaMA-2-7b с точностью 51 %[9].

Пока наиболее действенным методом противодействия получению доступа к содержанию информации является метод криптографического шифрования данных. Но, со временем, как только будет создан квантовый компьютер, все зашифрованные данные станут доступны для изучения [8].

В связи с этим проблема защиты права человека на неприкосновенность частной жизни в связи с массовым использованием различных технологий анализа данных [9, 10], со временем не только не теряет своей остроты, но и становится всё более значимой.

Для охвата всех возможных цифровых данных, оставленных пользователями в Сети, например, в 2022–2024 годах в странах Евросоюза (Германия[10] и Англия[11]) внесены изменения в законодательство, разрешающие массовый перехват сообщений. Тем самым в европейских странах требования о массовом перехвате данных стали законным инструментом для обработки больших данных, применяемым в целях получения разведывательной информации и выявления новых угроз, которые могут исходить как от известных, так и от неизвестных источников.

Рассмотрим возможность получения представления о причастности лица к тем или иным аспектам криминогенной деятельности с помощью анализа открытых данных из сетей с применением математических и логико-математических методов. Также оценим интернет-активность людей. Это позволит нам получить количественные характеристики данных процессов, выявить их закономерности, оптимизировать и прогнозировать поведение человека.

Использование в криминологической деятельности математических методов и моделей

В работе правоохранительных органов широко применяются специальные математические методы. В правовой сфере на смену приближенным качественным оценкам все чаще приходят точные количественные оценки. В правоохранительной деятельности актуально понимание возможностей математического моделирования, анализа, поддержки принятия решений, причинно-следственного анализа и вывода. Так, в целях математической оценки криминогенной обстановки[12] может проводиться оценка процессов и параметров математическими методами. На основании применения этих методов, строятся различные модели оценки криминогенной обстановки. Анализируемыми процессами в этом случае выступают действия пользователей в социальных сетях, а параметры меняются в зависимости от применяемой модели.

7 Определение Конституционного Суда РФ от 02.10.2003 N 345-О «Об отказе в принятии к рассмотрению запроса Советского районного суда города Липецка о проверке конституционности части четвертой статьи 32 Федерального закона от 16 февраля 1995 года «О связи» // Вестник Конституционного Суда РФ», N 1, 2004.

8 Определение Верховного Суда РФ от 29 января 2018 г. N 305-КГ17-21291 // СПС «КонсультантПлюс».

9 Нейросети раскрыли личные данные пользователей соцсетей // https://lenta.ru/news/2023/11/01/llm/

10 Gesetz zur Beschränkung des Brief-, Post- und Fernmeldegeheimnisses (Artikel 10-Gesetz – G 10) // https://www.gesetze-im-internet.de/g10_2001/BJNR125410001.html

11 Interception of communications code of practice 2022 // https://www.gov.uk/government/publications/interception-of-communications-code-of-practice-2022/interception-of-communications-code-of-practice-2022-accessible

12 Криминогенная обстановка определяется как совокупность процессов и их параметров, влияющих на состояние и динамику преступности.

Например, ***стохастическая модель динамики преступности*** базируется на том, что преступность — это случайный процесс, который зависит от множества факторов. Поэтому основным математическим аппаратом, применяемым в этой модели, является теория вероятностей и случайных процессов. Наличие множества возможных состояний, а также связей между ними и с окружающей средой позволяет считать структуру моделируемой системы заданной [11].

Если в исследуемых негативных процессах можно определить два состояния k и l, то можно построить граф, который будет отображать эти состояния и возможные переходы между ними в течение небольшого периода времени Δt. Этот граф представляет собой развитие модели Марковского процесса, где узлы отражают состояния моделируемого объекта, а дуги — вероятность перехода из одного состояния в другое.

Вероятно, модель динамики преступности можно отнести к непрерывной цепи Маркова, где система может менять свое состояние в любой момент времени (q-схема).

Граф изменения состояний в стохастической модели преступности представлен на рис. 1.

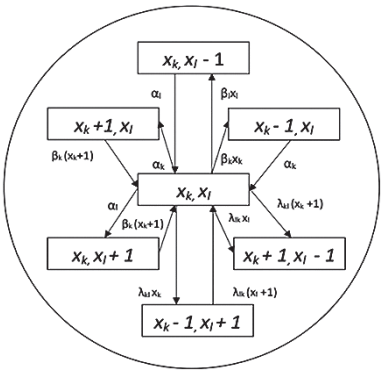


Рис. 1. Граф изменения состояний в стохастической модели динамики преступности

Детерминистская модель анализа динамики преступности позволяет оценить негативные процессы, происходящие в интернете. Она помогает анализировать особенности и закономерности преступной деятельности в Сети. Хотя эта математическая модель не может точно воспроизвести уникальность каждого отдельного преступления, она позволяет выявить факторы, влияющие на динамику процессов.

Поскольку в отношении каждого фактора достаточно четко определены количественные и качественные характеристики, параметры оценки могут быть положены в основу моделируемой системы.

Благодаря этому можно представить общую криминологическую картину и объяснить многие эмпирические наблюдения в этой области. Кроме того, создание математической модели динамики преступности может служить методологической основой для разработки системы криминологических гипотез.

При создании этих моделей индивид выступает в роли элемента системы, а группа индивидов формирует определённое множество процессов. Между ними существует взаимосвязь, которая определяется цепочкой переходов.

Модель устанавливает статистическую зависимость между следующими параметрами:

- Количество индивидов, находящихся в конкретном состоянии в определённый момент времени.
- Параметры, описывающие переходы между состояниями.
- Потоки индивидов, входящих в систему и покидающих её (рис. 2), где блоки – это люди, а дуги графа – это отношения между ними.

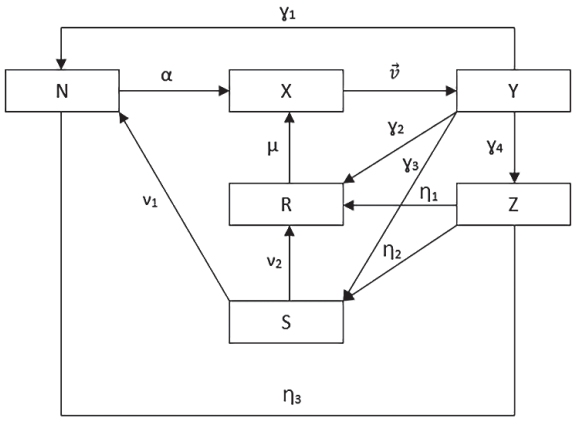


Рис. 2. Граф детерминистской модели динамики преступности.

Имитационная модель анализа динамики преступности является следующим этапом в развитии моделей, применяемых для анализа процессов и представляет собой универсальный программный комплекс, который позволяет воссоздать функционирование сложного процесса. Суть имитационного моделирования заключается в проведении эксперимента, позволяющего определить во времени варианты поведения людей при изменении внешних воздействующих факторов.

В настоящее время в криминологической практике имитационные модели стали эффективным инструментом для анализа, прогнозирования и управления в сфере уголовной юстиции. Они дают возможность комплексно исследовать преступность, учитывая влияние изменяющихся факторов. Так, например, российскими исследователями описана имитационная модель получения вероятностно-временных оценок длительности действий органов внутренних дел

при возникновении чрезвычайных обстоятельств криминального характера на примере массовых беспорядков [12], а также произведено использование модели на примере одного из видов массовых беспорядков с учетом фактора латентности [13].

Математическое моделирование криминогенной обстановки на основании независимых цифровых данных, оставленных пользователем социальной Сети

В данном разделе представлено решение задачи получения достаточно достоверных сведений о физическом лице на основании обработки данных, оставленных человеком в Сети, которые в дальнейшем будут служить входными данными для моделирования криминогенной обстановки, криминогенных наклонностях или интереса некоторой социальной группы к противоправной деятельности, вредоносному материалу, информации вредоносной направленности.

Для этого выберем такие натуральные числа n, s, l, при которых будет справедливо равенство $n = sl$. Здесь считаем $n = 1000$ и $l = 100$.

Далее рассмотрим n человек в социальных сетях. Выделяем тех людей, которые достаточно часто (например, 10 раз) размещают информацию, связанную со своими интересами, в том числе информацию криминальной направленности, которую можно определить как вредоносную информацию.

Для каждого i:$1 \le i \le n$ человека построим случайную величину ξ_i, принимающую два значения 1 и 0 в зависимости от того, было ли 10 упоминаний в социальных сетях об определенной вредоносной категории информации, лежащих в некотором промежутке, необходимой нам для исследования.

$$\xi_i = \begin{pmatrix} 1, & \text{были упоминания с вероятностью} \quad p = P(\xi_i = 1) \\ 0, & \text{не было упоминаний с вероятностью} \quad q = P(\xi_i = 0) \end{pmatrix}, \quad (1)$$

где $p + q = 1$ и значения этих величин неизвестны. Предположим, что случайные величины $\{\xi_i\}_{i=1}^n$ независимы в совокупности, то есть справедливо равенство

$$P(\xi_1 < t_1, \ldots, \xi_n < t_n) = P(\xi_i < t_1) \ldots P(\xi_n < t_n),$$

справедливое для всех действительных чисел $\{t_i\}_{i=1}^n$. Это условие говорит о том, что люди в социальных сетях, размещающие информацию о своих интересах, практически не знают друг друга. Они могут размещать значимую информацию, в том числе одобрять или не одобрять какую-либо информацию.

Нам понадобится еще одно определение.

Определение 1. *Пусть задана последовательность функций* $\{f_n(x)\}_{n=1}^{\infty}$ *и функция* $f(x)$, *со значениями в множестве действительных чисел, заданных для всех* $x \in D$, *где* D – *замыкание некоторого открытого множества на прямой*.

Скажем, что последовательность функций $\{f_n(x)\}_{n=1}^{\infty}$ равномерно сходится к функции $f(x)$, если для некоторого $\varepsilon > 1$ существует такое натуральное n_0, что для всех натуральных $n > n_0$ справедливо равенство

$$f_n(x) - f(x) \vee \varepsilon$$

для всех $x \in D$. Обозначим этот факт через $f_n(x) \longrightarrow f(x)$.

Заметим, что случайная величина ξ_i обладает математическим ожиданием $M(\xi_i) = 1 \cdot p + 0 \cdot q = p$. Квадрат случайной величины ξ_i^2 также обладает математическим ожиданием $M(\xi_i^2) = 1^2 \cdot p + 0^2 \cdot q = p$. Поэтому случайная величина ξ_i обладает дисперсией $D(\xi_i) = M(\xi_i^2) - M^2(\xi_i) = p - p^2 = pq$.

Согласно центральной предельной теореме, для одинаково распределенных случайных величин, имеющих дисперсию [14], справедливо утверждение

$$P\left(\frac{1}{\sqrt{lD(\xi_i)}}\sum_{k=1}^{l}\left(\xi_{l\cdot(j-1)+k} - M(\xi_i)\right) < t\right) =$$

$$= P\left(\frac{1}{\sqrt{lpq}}\sum_{k=1}^{l}\left(\xi_{l\cdot(j-1)+k} - p\right) < t\right) \longrightarrow \frac{1}{2\pi}\int_{-\infty}^{t} e^{\frac{-x^2}{2}}dx,$$

которое означает, что для некоторого $\varepsilon > 0$ существует натуральное n_0, такое, что для всех натуральных $l > n_0$ справедливо равенство

$$\left|P\left(\frac{1}{\sqrt{lpq}}\sum_{k=1}^{l}\left(\xi_{l\cdot(j-1)+k} - p\right) < t\right) \longrightarrow \frac{1}{2\pi}\int_{-\infty}^{t} e^{\frac{-x^2}{2}}dx\right| < \varepsilon,$$

$$\left|P\left(\frac{1}{l}\sum_{k=1}^{l}\xi_{l\cdot(j-1)+k} < t\sqrt{\frac{pq}{l}} + p\right) - \frac{1}{2\pi}\int_{-\infty}^{t} e^{\frac{-x^2}{2}}dx\right| < \varepsilon.$$

Обозначим $y = t\sqrt{\frac{pq}{l}} + p$, $t = (y - p)\sqrt{\frac{l}{pq}}$.

В этом случае имеем

$$\left|P\left(\frac{1}{l}\sum_{k=1}^{l}\xi_{l\cdot(j-1)+k} < y\right) - \frac{1}{2\pi}\int_{-\infty}^{(y-p)\sqrt{\frac{l}{pq}}} e^{\frac{-x^2}{2}}dx\right| < \varepsilon.$$

Упростим интеграл

$$\frac{1}{2\pi}\int_{-\infty}^{(y-p)\sqrt{\frac{l}{pq}}} e^{\frac{-x^2}{2}}dx = \left|\begin{array}{l} x = (z-p)\sqrt{\frac{l}{pq}}\ dx = \sqrt{\frac{l}{pq}}\ dz \\ x = (z-p)\sqrt{\frac{l}{pq}}\ z = y \\ x = -\infty,\ z = -\infty \end{array}\right| =$$

$$= \frac{1}{2\pi}\sqrt{\frac{l}{pq}}\int_{-\infty}^{y} e^{\frac{-(z-p)^2}{l}}dz.$$

Окончательно имеем

$$\left|P\left(\frac{1}{l}\sum_{k=1}^{l}\varepsilon_{l(j-1)+k} < y\right) - \frac{1}{2\pi}\sqrt{\frac{l}{pq}}\int_{-\infty}^{y} e^{\frac{-(z-p)^2}{2\left(\sqrt{\frac{pq}{l}}\right)^2}}dz\right| < \varepsilon$$

Таким образом мы доказали, что для всех возможных, $1 \le j \le \frac{n}{l}$ функция распределения случайной величины $\xi_i = \frac{1}{l}\sum_{k=1}^{l}\varepsilon_{l(j-1)+k}$ равномерно сходится к нормальной случайной величине с математическим ожиданием $m = p$ и дисперсией $\sigma^2 = \frac{pq}{l}$. То есть можно считать, что случайная величина $\xi_i = \frac{1}{l}\sum_{k=1}^{l}\varepsilon_{l(j-1)+k}$ является нормальной случайной величиной с математическим ожиданием p и дисперсией $\sigma^2 = \frac{pq}{l}$.

В дальнейшем необходимо найти некоторое среднее значение (математического ожидания) для группы случайных величин, рассмотренной выше, с определенной степенью достоверности γ, под которой нам следует понимать уровень уверенности в результатах исследования и выводах. (Так, при высоком уровне достоверности, если мы примем $\gamma = 0.95$, можно сделать вывод о том, что при повторении эксперимента независимой группой исследователей мы получим тот же результат в 95 случаев из 100).

Итак, если нам удастся оценить число m, то мы сможем сказать, что с достоверностью γ (в примере: 95 случаев из 100) каждые ml человек из l совершают действия в описанных нами промежутках, или sml из n осуществляют деятельность (интересуются вредоносной информацией) в описанных нами промежутках.

Далее наша задача заключается в оценке математического ожидания нормальной случайной величины τ_i. Для решения указанной задачи авторами предлагается использовать распределение Стьюдента, отвечающее признакам определения 2.

Определение 2. *Распределение τ с плотностью вероятности*

$$f_\tau(t) = \frac{1}{\sqrt{\pi n}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{\frac{-n+1}{2}},$$

называется распределением Стьюдента с n степенями свободы. Символом $\Gamma(x)$ обозначена гамма-функция

$$\Gamma(k) = \int_0^\infty t^{k-1}\, exp(-t)dt,$$

определенная для всех $k > 0$.

Далее, вычислим величины

$$\bar{\zeta} = \frac{1}{l} \Sigma_{k=1}^{l} \xi_k 2, \tag{2}$$

$$\bar{\theta} = \frac{1}{l} \Sigma_{k=1}^{l} \xi_k 2 - (\bar{\zeta})^2. \tag{3}$$

Случайная величина

$$\tau = \frac{\sqrt{l}\,(\bar{\zeta} - m)}{\sqrt{\frac{n\bar{\theta}^2}{l-1}}} = \sqrt{l-1}\,\frac{\bar{\zeta} - m}{\bar{\theta}} \tag{4}$$

имеет распределение Стьюдента с $l - 1$ степенями свободы, или

$$P\left(t_1 < \sqrt{l-1}\,\frac{\bar{\zeta} - m}{\bar{\theta}} < t_2\right) = \int_{t_1}^{t_2} f_\tau(x)dx.$$

Или то же самое утверждение – для плотности вероятности f_τ случайной величины τ с $l - 1$ степенями свободы справедливо равенство

$$P\left(\bar{\zeta} - \frac{\bar{\theta} t_2}{\sqrt{n-1}} < m < \bar{\zeta} - \frac{\bar{\theta} t_1}{\sqrt{n-1}}\right) = \int_{t_1}^{t_2} f_\tau(x)dx =$$

$$= \int_0^{t_2} f_\tau(x)dx - \int_0^{t_1} f_\tau(x)dx.$$

То есть за счет выбора n для любых значений t_1 и t_2 можно сделать величину $\gamma = \int_{t_1}^{t_2} f_\tau(x)dx$ сколь угодно близкой к единице. Далее за счет выбора t_1 и t_2 сделать интервал

$$\left[\bar{\zeta} - \frac{\bar{\theta} t_2}{\sqrt{n-1}} < m < \bar{\zeta} - \frac{\bar{\theta} t_1}{\sqrt{n-1}}\right]$$

сколь угодно малой длины и окончательно сделаем вывод – с вероятностью γ величина m лежит на отрезке

$$\left[\bar{\zeta} - \frac{\bar{\theta} t_2}{\sqrt{n-1}} < m < \bar{\zeta} - \frac{\bar{\theta} t_1}{\sqrt{n-1}}\right].$$

Таким образом, поставленная перед нами задача определения достаточно достоверных необходимых сведений для идентификации физического лица в результате обработки данных, формирующих косвенные признаки на основании определения группы людей, проявляющих интерес к вредоносной информации, имеет следующее решение:

Пусть изучаются 1000 человек на используемую ими вредоносную информацию. В результате, согласно формуле (1), получим 1000 независимых дискретных случайных величин ξ_i.

Далее, разобьем все испытания на непересекающиеся классы по 100 элементов в каждом и посмотрим сколько раз в каждом классе было более 10 упоминаний об информации, отнесенной к вредоносной. Далее, разобьем все испытания на непересекающиеся классы по 100 элементов в каждом и посмотрим сколько раз в каждом классе было более 10 обращений к вредоносной информации, что отобразим в таблице 1.

Будет получена таблица.

Таблица 1.

Количество обращений к вредоносной информации для каждого из классов.

классы	**1**	**2**	**3**	**4**	**5**
обращения	53	48	74	26	35
классы	**6**	**7**	**8**	**9**	**10**
обращения	66	79	81	95	18

Согласно таблице 1, в первом классе 53 раза было более 10 обращений к вредоносной информации, в втором классе 48 раз было более 10 обращений, и так далее.

То есть в соответствии с (1) случайные величины ε_i, i = 1, ..., 100 приняли значение 1 ровно 53 раза (т.е. 53 раза было упоминание указанной категории информации и значение 0 ровно 47 раз (не было упоминаний информации (100 – 53 = 47)).

Также, применяя (1), получаем, что случайные величины ε_i, i = 101, ..., 200 приняли значение 1 ровно 48 раз и значение 0 ровно 52 раза и так далее.

Как доказано в работе, случайные величины $\zeta_j = \frac{1}{100}\sum_{k=1}^{100}\varepsilon_{100*(j-1)+k}$ можно считать нормальными случайными величинами. В итоге получим таблицу 2 нормальных случайных величин для каждого из классов.

Таблица 2.

Значения нормальных случайных величин

случайная величина	ζ_1	ζ_2	ζ_3	ζ_4	ζ_5
значение сл. вел.	0,53	0,48	0,74	0,26	0,35
случайная величина	ζ_6	ζ_7	ζ_8	ζ_9	ζ_{10}
значение сл. вел.	0,66	0,79	0,81	0,95	0,18

Подставляя значения случайных величин в формулы (2) и (3), получим:

$$\bar{\zeta} = \frac{1}{10}\sum_{k=1}^{10}\xi_k = \frac{1}{10}(0{,}53+0{,}48+0{,}74+0{,}26+0{,}35+ +0{,}66+0{,}79+0{,}81+0{,}95+0{,}18) = 0{,}575 \quad (5)$$

$$\bar{\theta}^2 = \frac{1}{10}\sum_{k=1}^{10}\xi^2 - (\bar{\zeta}) = = \frac{1}{10}(0{,}53^2+0{,}48^2+0{,}74^2+0{,}26^2+0{,}35^2+0{,}66^2+ +0{,}79^2+0{,}81^2+0{,}95^2+0{,}18^2) - 0{,}575^2 = 0{,}059345 \quad (6)$$

Подставляя полученные значения (5) и (6) в формулу (4), получаем, что значение случайной величины

$$\tau = \sqrt{l-1}\,\frac{\bar{\zeta}-m}{\bar{\theta}} = \sqrt{10-1}\,\frac{0{,}575-m}{\sqrt{0{,}059345}} = 12{,}31 \quad (7)$$

имеет распределение Стьюдента с 9 степенями свободы, или,

$$P(0{,}575 - 0{,}0812t_2 < m < 0{,}575 - 0{,}0812t_1) = = \int_{t_1}^{t_2} f_\tau(x)dx.$$

По таблицам для $\gamma = 0{,}95$ из неравенства $\int_{t_1}^{t_2} f_\tau(x)dx > \gamma$ выбираем $t_1 = 0{,}12$ и $t_2 = 0{,}83$.

$P(0{,}508 < m < 0{,}584) > 0{,}95$, примем усредненное $m = 0{,}54$.

В итоге получим, что в 95 случаях из 100 аналогичных случаях исследований, в среднем $mls = 0{,}54 * 100 * 10 = 540$ человек осуществляют обращение вредоносной информации.

Заключение

Таким образом, представленные в настоящей статье вычисления доказывают возможность применения математических методов в криминологическом анализе преступности и позволяют осуществлять цифровое профилирование человека на основании данных, размещаемых в социальных сетях и формально не относящихся к персональным данным, как категории информации ограниченного доступа.

Для достижения заявленных целей сформулирована задача получения с достаточной достоверностью криминологически значимой информации в результате обработки данных, оставленных в Сети группой пользователей, и на основе косвенных признаков, на основании анализа которых и полученных закономерностей можно сделать предположение о криминогенных наклонностях или интересе некоторой социальной группы к противоправной деятельности, вредоносному материалу, информации вредоносной направленности.

Таким образом, поставленная перед нами задача об определении достаточно достоверных необходимых сведений для идентификации физического лица в результате обработки данных, формирующих косвенные признаки на основании поведения группы незнакомых друг с другом людей, проявляющих интерес к вредоносной информации, либо же размещающих значимую информацию о своих интересах, решена.

С помощью проведенного исследования представлена модель оценки с высокой достоверностью γ (в примере: 95 случаев из 100) о том, что ml человек из l совершают действия в описанных нами промежутках, или sml из n осуществляют деятельность (интересуются вредоносной информацией) в описанных нами промежутках.

В результате проведенного исследования установлены закономерности, позволяющие в дальнейшем предсказывать поведение групп людей, осуществляющих обращение вредоносной информации в сети, либо предоставляющих информацию указанной категории.

Литература

1. Редкоус, В. М. О совершенствовании правовой основы деятельности органов внутренних дел по объявлению официальных предостережений / В. М. Редкоус // Закон и право. – 2020. – № 9. – С. 157–159. – DOI 10.24411/2073-3313-2020-10453.
2. Степанов, О. А. О правовых особенностях и рисках реализации цифрового профилирования / О. А. Степанов, Д. А. Басангов // Российская юстиция. – 2024. – № 1. – С. 59–69. – DOI 10.52433/01316761_2024_01_59.
3. «Цифровой поворот» в правовых исследованиях / И. Р. Бегишев, А. К. Жарова, Е. А. Громова [и др.] // Journal of Digital Technologies and Law. – 2024. – Т. 2, № 1. – С. 7–13. – DOI 10.21202/jdtl.2024.1.
4. Жарова, А. К. Система организационно-правового выявления лиц, разместивших информацию в интернете о намерении совершить преступление / А. К. Жарова // Пробелы в российском законодательстве. – 2024. – Т. 17, № 1. – С. 122–130. – DOI 10.33693/2072-3164-2024-17-1-122-130. – EDN OHAZYD.

5. Шутова, А. А. Обеспечение цифровой безопасности системы здравоохранения уголовно-правовыми средствами / А. А. Шутова // Russian Journal of Economics and Law. – 2024. – Т. 18, № 4. – С. 936–953. – DOI 10.21202/2782-2923.2024.4.936-953.
6. Дейнеко, А. Г. Публичное право в киберпространстве: публично-правовое регулирование информационных отношений / А. Г. Дейнеко. – Москва: Общество с ограниченной ответственностью «Проспект», 2025. – 248 с.
7. Жарова, А. К. Парадигма цифрового профилирования деятельности человека: риски, угрозы, преступления / А. К. Жарова, В. М. Елин, А. В. Минбалеев. – Москва: Общество с ограниченной ответственностью «Русайнс», 2022. – 240 с.
8. Жарова, А. К. Обзор нормативных требований, обеспечивающих национальную безопасность США в сфере квантовых технологий / А. К. Жарова // Информационное общество. – 2023. – № 3. – С. 69–77. – DOI 10.52605/16059921_2023_03_69.
9. Залоило, М. В. Циклично-волновая модель интерпретации истории права на основе теории технологических укладов / М. В. Залоило // Историко-правовой ежегодник – 2023. – Москва: Infotropic Media, 2024. – С. 48–72.
10. Твердова, Т. В. § 3. Риски правового регулирования отношений, возникающих по поводу искусственного интеллекта / Т. В. Твердова // Теоретико-правовая парадигма существования кибернетической (информационной) цивилизации : монография. – Москва : Межрегиональная общественная организация «Межрегиональная ассоциация теоретиков государства и права», 2022. – С. 244–273. – EDN ZXOJCC.
11. Максимов, С. В. Стохастическая модель репрессивно-превентивного воздействия на преступность: от интуиции к расчетам / С. В. Максимов, Ю. Г. Васин, К. А. Утаров // Всероссийский криминологический журнал. – 2021. – Т. 15, № 6. – С. 665–680. – DOI 10.17150/2500-4255.2021.15(6).665-680.
12. Моделирование процессов принятия решения в правоохранительной деятельности / О. Ю. Данилова, А. В. Меньших, В. В. Меньших [и др.]. – Воронеж : Воронежский институт Министерства внутренних дел Российской Федерации, 2021. – 103 с. – ISBN 978-5-88591-856-5. – EDN FENSWM.
13. Минаев В. А. Моделирование динамики преступности с учетом фактора латентности// Криминологический журнал. Естественные науки. Компьютерные науки и информатика. 2022. № 2. С. 67–78.
14. Малахова, В. В. Анализ статистических данных с использованием математического аппарата искусственного интеллекта / В. В. Малахова, О. В. Малахов // Вестник Луганского государственного университета имени Владимира Даля. – 2023. – № 11. – С. 177–179. – EDN EADYTE.

ANALYSIS OF OPEN DATA POSTED ON THE NETWORK IN ORDER TO OBTAIN INFORMATION ABOUT THE CRIME SITUATION

Zharova A. K.[13], Elin V. M.[14], Atlasov I. V.[15]

Keywords: *information technology, communication data, analysis of digital shadows and digital footprints, personal data, mathematical modeling.*

The purpose of the article *is to propose a method for forming a digital profile of a person, which can be used to analyze and predict the criminogenic situation.*

Research method: *logical and mathematical methods, such as typological model, deterministic model and simulation modeling, are used. In addition, the method of analysis of mathematical models, including the stochastic model, is used, which allows to obtain a more accurate and detailed picture.*

Result: *data analysis systems can be used to extract, analyze, transform and present information that is essential for operational-search and investigative activities. The authors, referring to the existing judicial practice, reveal the importance of communication data for obtaining a digital profile of a person, which formally do not belong to personal data as a category of restricted information. is a mathematical modeling of the criminogenic situation based on the analysis of independent digital data left by the user of the social network. Thus, as a result of the study, patterns have been identified that make it possible to predict the behavior of groups of people who transmit harmful information on the network, or the placement of information of this category.*

Practical value: *on the basis of the theoretical experiment, a conclusion is made about the possibility of using mathematical methods in the criminological analysis of crime.*

References

1. Redkous, V. M. O sovershenstvovanii pravovoj osnovy deyatel'nosti organov vnutrennih del po ob"yavleniyu oficial'nyh predosterezhenij / V. M. Redkous // Zakon i pravo. – 2020. – № 9. – S. 157–159. – DOI 10.24411/2073-3313-2020-10453. – EDN EIBWAV.
2. Stepanov, O. A. O pravovyh osobennostyah i riskah realizacii cifrovogo profilirovaniya / O. A. Stepanov, D. A. Basangov // Rossijskaya yusticiya. – 2024. – № 1. – S. 59–69. – DOI 10.52433/01316761_2024_01_59. – EDN JJSSAU.

13 Anna K. Zharova, Dr.Sc. (of Law), Professor, Financial University under the Government of the Russian Federation, Moscow. E-mail: anna_jarova@mail.ru
14 Vladimir M. Elin, Ph.D., Associate Professor of the Department of Information Security of the Moscow University of the Ministry of Internal Affairs of Russia named after V. Y. Kikiot, Associate Professor of the Department of Information Security of the Financial University under the Government of the Russian Federation, Moscow. E-mail: elin_vm@mail.ru
15 Igor V. Atlasov, Dr.Sc. (of Physical and Mathematical), Professor of the Moscow University of the Ministry of Internal Affairs of Russia named after V. Y. Kikiot, Moscow. E-mail: atlasov.igor.777@gmail.com

3. Begishev I. R., Zharova A. K., Gromova E. A., Zaloilo M. V., Filipova I. A., Shutova A. A. «Cifrovoj povorot» v pravovyh issledovaniyah // Journal of Digital Technologies and Law. 2024. № 2(1). EDN: IWWUBP.
4. Zharova, A. K. Sistema organizacionno-pravovogo vyyavleniya lic, razmestivshih informaciyu v internete o namerenii sovershit' prestuplenie // Probely v rossijskom zakonodatel'stve. – 2024. – T. 17, № 1. – S. 122–130. – DOI 10.33693/2072-3164-2024-17-1-122-130. – EDN OHAZYD.
5. Shutova, A. A. Obespechenie cifrovoj bezopasnosti sistemy zdravoohraneniya ugolovno-pravovymi sredstvami / A. A. Shutova // Russian Journal of Economics and Law. – 2024. – T. 18, № 4 – S. 936–953. – DOI 10.21202/2782-2923.2024.4.936-953. – EDN SHZTFY.
6. Dejneko, A. G. Publichnoe pravo v kiberprostranstve: publichno-pravovoe regulirovanie informacionnyh otnoshenij / A. G. Dejneko. – Moskva : Obshchestvo s ogranichennoj otvetstvennost'yu \«Prospekt»\, 2025. – 248 s. – ISBN 978-5-392-42996-7. – EDN SBSONL.
7. Zharova, A. K. Paradigma cifrovogo profilirovaniya deyatel'nosti cheloveka: riski, ugrozy, prestupleniya / A. K. Zharova, V. M. Elin, A. V. Minbaleev. – Moskva : Obshchestvo s ogranichennoj otvetstvennost'yu \"Rusajns\", 2022. – 240 s. – ISBN 978-5-466-00766-4. – EDN DNKVPR.
8. Zharova, A. K. Obzor normativnyh trebovanij, obespechivayushchih nacional'nuyu bezopasnost' SShA v sfere kvantovyh tekhnologij / A. K. Zharova // Informacionnoe obshchestvo. – 2023. – № 3. – S. 69–77. – DOI 10.52605/16059921_2023_03_69. – EDN CCHNJY.
9. Zaloilo, M. V. Ciklichno-volnovaya model' interpretacii istorii prava na osnove teorii tekhnologicheskih ukladov / M. V. Zaloilo // Istoriko-pravovoj ezhegodnik – 2023. – Moskva : Infotropic Media, 2024. – S. 48–72. – EDN IETWNM.
10. Tverdova, T. V. § 3. Riski pravovogo regulirovaniya otnoshenij, voznikayushchih po povodu iskusstvennogo intellekta / T. V. Tverdova // Teoretiko-pravovaya paradigma sushchestvovaniya kiberneticheskoj (informacionnoj) civilizacii : monografiya. – Moskva : Mezhregional'naya obshchestvennaya organizaciya \«Mezhregional'naya associaciya teoretikov gosudarstva i prava»\, 2022. – S. 244-273. – EDN ZXOJCC.
11. Maksimov S. V. Stohasticheskaya model' repressivnopreventivnogo vozdejstviya na prestupnost': ot intuicii k raschetam / S. V. Maksimov, Yu. G. Vasin, K. A. Utarov. — DOI 10.17150/2500-4255.2021.15(6).665-680 // Vserossijskij kriminologicheskij zhurnal. – 2021. – T. 15, № 6. – S. 665–680.
12. Modelirovanie processov prinyatiya resheniya v pravoohranitel'noj deyatel'nosti / O. Yu. Danilova, A. V. Men'shih, V. V. Men'shih [i dr.]. – Voronezh : Voronezhskij institut Ministerstva vnutrennih del Rossijskoj Federacii, 2021. – 103 s. – ISBN 978-5-88591-856-5. – EDN FENSWM.
13. Minaev V. A. Modelirovanie dinamiki prestupnosti s uchetom faktora latentnosti// Kriminologicheskij zhurnal. Estestvennye nauki. Konp'yuternye nauki i informatika. 2022. № 2. S. 67–78.
14. Malahova, V. V. Analiz statisticheskih dannyh s ispol'zovaniem matematicheskogo apparata iskusstvennogo intellekta / V. V. Malahova, O. V. Malahov // Vestnik Luganskogo gosudarstvennogo universiteta imeni Vladimira Dalya. – 2023. – № 11. – S. 177–179. – EDN EADYTE.

CYBERSECURITY ISSUES

SCIENTIFIC PEER-REVIEWED JOURNAL

2025, № 4 (68)

Cybersecurity Issues is a research periodical scientific and practical publication specializing in information security.
Published six times a year

https://cyberrus.info

The journal is being published from 2013
(Registration Certificate PI No. FS 77-75239).
CrossRef number (DOI): 10.21681/2311-3456

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

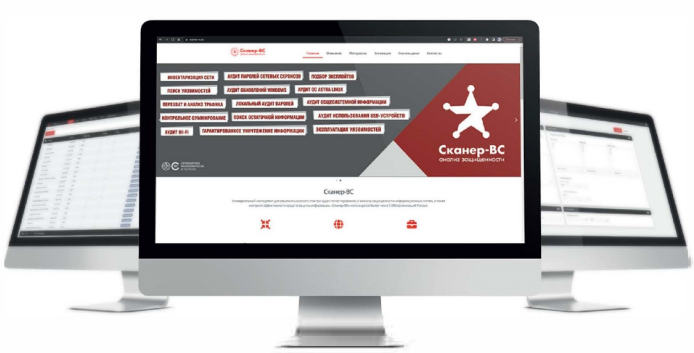
Founder and publisher
JSC «NPO «Echelon»
Postal address: Elektrozavodskaya str., 24, bld. 1, 107023, Moscow, Russia
E-mail: editor@cyberrus.info

CONTENTS

Эшелон
комплексная безопасность

Сканер-ВС
анализ защищенности

СКАНИРОВАНИЕ НА УЯЗВИМОСТИ НИКОГДА НЕ БЫЛО ТАКИМ БЫСТРЫМ!



ГК «Эшелон» представляет новый релиз системы управления уязвимостями Сканер-ВС 6. Сканер-ВС используется более чем в 5 000 организаций в России и позволяет как проводить периодическое сканирование на поиск уязвимостей, так и организовать непрерывный контроль защищенности.

Решение является ключевым компонентом, позволяющим внедрить эффективный процесс управления уязвимостями.

Скачать демо-версию «Сканер-ВС 6» (количество IP: 16, пробный период: 2 месяца) можно на сайте продукта: https://scaner-vs.ru/.

Получить техническую консультацию в группе продукта в телеграм: https://t.me/scanervs

Высокая скорость поиска

Сканер-ВС 6 обладает высокой скоростью поиска уязвимостей благодаря технологии «без скриптов»

Актуальная база уязвимостей

Ежедневно обновляемая база данных уязвимостей позволяет держать руку на пульсе последних изменений

Комплексный подход

Комплексное тестирование защищенности позволяет выявлять максимальное количество нарушений ИБ

Работа в защищенной среде

Работа в среде защищенной операционной системы Astra Linux 1.7

Отчетность

Единая среда для проведения тестирования и формирования отчетов, содержащих различную информацию в зависимости от степени детализации

Исполнение

Наличие исполнений в виде дистрибутива под Astra Linux 1.7 и LiveUSB с предустановленной ОС и с поддержкой режима сохранения изменений.

CYBERSECURITY ISSUES
VOPROSY KIBERBEZOPASNOSTI

№ 4 2025

DOI: 10.21681/2311-3456

Methodology for the synthesis of quantum-resistant blockchain platforms

Critical Information Infrastructure Security Assessment

Detecting phishing emails using neural networks

www.cyberrus.info
editor@cyberrus.info