УДК 004.056.5

ИССЛЕДОВАНИЕ АЛГОРИТМОВ ОБРАБОТКИ, ДЕТЕКЦИИ И ЗАЩИТЫ ДАННЫХ С ЦЕЛЬЮ МИНИМИЗАЦИИ ВОЗДЕЙСТВИЯ ВРЕДОНОСНОГО ПО И ФИШИНГОВЫХ АТАК НА ПОЛЬЗОВАТЕЛЕЙ ЦИФРОВЫХ ПЛАТФОРМ

Т. С. Волокитина[1] [0000-0002-5493-447X], **М. О. Таныгин**[2] [0000-0002-4099-1414]

[1, 2]*Юго-Западный государственный университет», г. Курск, Россия*

[1]tativolokitina@gmail.com, [2]tanygin@yandex.ru

Аннотация

Статья посвящена разработке научно-методического аппарата повышения эффективности защиты цифровых платформ от киберугроз путем создания алгоритмов обработки и детекции с учетом когнитивных особенностей пользователей. Предложена концептуальная модель трехэтапной системы защиты, интегрирующая технические механизмы безопасности с когнитивными моделями принятия решений. Разработан алгоритм эвристической детекции на основе машинного обучения Random Forest с анализом 47 признаков, включающих технические характеристики URL и когнитивно-семантические характеристики контента. Создана методика динамической интеграции четырех источников данных об угрозах, сокращающая время реагирования с 12–14 ч. до 2 ч. Предложен алгоритм рекурсивного анализа цепочек перенаправлений глубиной до десяти уровней для обнаружения замаскированных угроз. Экспериментальная валидация на эмпирической базе объемом около миллиона записей подтвердила точность детекции 87% при обработке ста тысяч записей в час. Разработанные решения обеспечивают соответствие требованиям ГОСТ Р 57580.1–2017 и российского законодательства в области защиты персональных данных.

Ключевые слова*: эвристическая детекция угроз, машинное обучение, когнитивная безопасность, фишинговые атаки, социальная инженерия, защита данных, интеграция источников угроз.*

ВВЕДЕНИЕ

Стремительное развитие цифровых платформ в Российской Федерации сопровождается критическим ростом киберугроз, эксплуатирующих когнитивные уязвимости пользователей. По данным МВД России, в 2023 г. зарегистрировано свыше 50 тыс. преступлений в сфере информационных технологий, что на 47% превышает показатели предыдущего периода [1, 2]. Особую опасность представляют фишинговые атаки и вредоносное программное обеспечение, распространяемые через социальные сети, где 78% атак осуществляется путем манипуляции восприятием и доверием пользователей [3, 4].

В результате анализа существующих технических решений установлена фундаментальная проблема: традиционные системы защиты, основанные исключительно на черных списках URL и сигнатурном анализе, демонстрируют недостаточную эффективность вследствие игнорирования когнитивно-поведенческих факторов пользователей [5, 6]. Задержки обновления черных списков составляют 12–14 ч. для Google Safe Browsing и 24–48 ч. для реестра Роскомнадзора, в течение которых реализуется 70% кликов пользователей на вредоносные ссылки из-за максимальной когнитивной уязвимости в период новизны угрозы [7].

Когнитивный аспект проблемы усугубляется низкой цифровой грамотностью 60% российских пользователей, которые не способны распознавать признаки социальной инженерии и игнорируют предупреждения систем безопасности вследствие когнитивных искажений восприятия рисков [8–10]. До 70% пользователей игнорируют технические предупреждения из-за когнитивной перегрузки, привычки к игнорированию сообщений и искажений оценки вероятности рисков.

Теоретические основы защиты от фишинга с учетом когнитивных факторов заложены в работах, исследовавших демографические факторы восприимчивости к фишингу и эффективность образовательных интервенций (см., например [11]). Однако эти подходы ориентированы на западную аудиторию и не учитывают специфику когнитивных паттернов российских пользователей социальных сетей, включая эксплуатацию доверия к государственным брендам и омографические атаки с использованием кириллических символов.

В российской науке значительный вклад в развитие методов информационной безопасности внесли исследователи, изучавшие технические аспекты защиты [11–15]. Однако существующие работы недостаточно раскрывают проблемы интеграции технических средств защиты с когнитивно-поведенческими моделями пользователей.

Анализ показывает следующие ограничения: недостаточный учет когнитивных особенностей российских пользователей; слабая интеграция технических средств с когнитивно-поведенческими моделями; отсутствие комплексного анализа многоэтапной структуры защиты; недостаточное внимание к культурной специфике восприятия киберугроз.

Правовые аспекты проблемы регулируются Федеральными законами № 152-ФЗ и № 149-ФЗ, обязывающими операторов цифровых платформ обеспечивать защиту от несанкционированного доступа [15, 16]. Требования ГОСТ Р 57580.1–2017 устанавливают минимальную эффективность защиты не менее 80%, что не достигается большинством платформ при учете когнитивных факторов пользователей [17].

Целью настоящего исследования была разработка научно-методического аппарата для повышения эффективности защиты цифровых платформ от фишинговых атак и вредоносного ПО путем создания алгоритмов обработки и детекции угроз, учитывающих когнитивные особенности пользователей.

Для достижения цели решались следующие задачи: разработка математической модели многоэтапной системы защиты; создание алгоритма эвристической детекции на основе анализа признаков; разработка методики интеграции источников данных; создание алгоритма анализа цепочек перенаправлений; экспериментальная проверка разработанных методов.

КОНЦЕПТУАЛЬНАЯ МОДЕЛЬ МНОГОЭТАПНОЙ СИСТЕМЫ ЗАЩИТЫ С УЧЕТОМ КОГНИТИВНЫХ ФАКТОРОВ

Предложена концептуальная модель функционирования системы защиты цифровой платформы, интегрирующая технические механизмы безопасности с когнитивными моделями принятия решений пользователями. Модель описывает три последовательных этапа обработки угроз.

Этап публикации представляет первую линию защиты, где автоматические фильтры анализируют контент до его появления в ленте пользователей. Когнитивная оценка угрозы пользователями отсутствует, защита осуществляется исключительно техническими средствами: сигнатурный анализ вредоносных URL по черным спискам, эвристическая детекция на основе анализа признаков контента, проверка цифровых подписей. Эффективность этапа определяется коэффициентом обнаружения.

Этап перехода по ссылке характеризуется когнитивным принятием решения пользователем о клике на подозрительную ссылку. Данный этап критически важен с точки зрения когнитивной безопасности, поскольку здесь проявляются факторы доверия к источнику публикации, эмоциональной вовлеченности в контент, автоматизмов принятия решений. Эффективность определяется долей пользователей, отказавшихся от клика без технических предупреждений.

Этап браузерной защиты активируется при попытке перехода на URL, обнаруженный в черных списках браузера. Пользователю выдается предупреждение о риске с описанием потенциальной угрозы. Эффективность этапа определяется долей правильных реакций на предупреждения, зависящей от когнитивных факторов восприятия технических сообщений.

Концептуальная архитектура системы представлена 7 функциональными блоками, соединенными через системную шину для параллельной обработки данных. Блок аппаратного обеспечения содержит процессор с количеством ядер не менее 28, оперативную память объемом не менее 64 ГБ, устройства энергонезависимой памяти объемом не менее 1 ТБ, сетевые интерфейсы пропускной способностью не менее 10 Гбит/с.

Блок сбора данных выполнен с возможностью извлечения публикаций из API с обработкой исключительно публично доступных данных и анонимизацией персональных идентификаторов посредством SHA-256-хеширования в соответствии с требованиями ФЗ № 152. Блок содержит модуль асинхронных HTTP-запросов для параллельного извлечения не менее 100 тыс. записей в час.

Разработана математическая модель временной динамики когнитивной уязвимости пользователей, описывающая зависимость вероятности реализации клика от времени после публикации угрозы. Модель учитывает три ключевых

фактора: новизну контента в ленте, отсутствие технических предупреждений и эмоциональную вовлеченность.

Временная динамика характеризуется экспоненциальным убыванием вероятности клика. В первые два часа после публикации реализуется 25.2% кликов при максимальной когнитивной уязвимости из-за новизны угрозы и отсутствия информации в черных списках. За 12 ч. происходит 70% кликов, что определяет критический временной интервал. После 24 ч. вероятность клика стабилизируется на низком уровне 5–7% вследствие появления информации об угрозе в черных списках браузеров.

Экспериментально установлено, что задержка реагирования систем защиты критически влияет на эффективность. Традиционные системы с временем обновления черных списков 12–14 ч. перехватывают угрозу только после реализации 70% потенциальных кликов, обеспечивая защиту лишь для 30% пользователей. Сокращение времени реагирования до 2 ч. позволяет перехватить угрозу до реализации 74.8% кликов.

Предложен количественный индекс когнитивной уязвимости пользователей CVI, измеряющий степень подверженности социальной инженерии с учетом базовых цифровых навыков и способности правильно реагировать на предупреждения. Индекс вычисляется как произведение двух компонентов: первый представляет базовую когнитивную уязвимость и равен (1−G), где G является уровнем цифровой грамотности; второй компонент выражает поведенческую уязвимость и равен (1−P_реакция/P_эталон).

Уровень цифровой грамотности G измеряется как доля пользователей, преодолевших пороговое значение функциональной цифровой грамотности в стандартизированном тесте по методике ОЭСР, содержащем не менее 20 заданий по категориям: распознавание фишинга, оценка безопасности URL, понимание технологий защиты, практические действия при угрозах. Пороговое значение установлено на уровне не менее 55% правильных ответов. Для российских пользователей цифровых платформ с преобладанием возрастной группы 45+ значение G составляет 0.40 по данным Росстата 2023 г.

Фактическая вероятность правильной реакции измеряется в ходе контролируемых экспериментов с использованием открытых обезличенных данных через

официальный API. Анализ 1625 случаев отображения предупреждений браузеров показал, что только 487 случаев привели к правильной реакции, тогда как 1138 были проигнорированы. Фактическая вероятность составляет 0.30. Эталонная вероятность 0.70 установлена на основе международных исследований [18].

Для российской аудитории индекс CVI вычисляется: первый компонент равен (1−0.40) = 0.60; второй компонент равен (1−0.30/0.70) = 0.57; произведение компонентов дает CVI = 0.34. Значение интерпретируется как средняя когнитивная уязвимость аудитории на верхней границе перехода к высокой уязвимости.

МЕТОДОЛОГИЯ ПРОЕКТИРОВАНИЯ АЛГОРИТМОВ ДЕТЕКЦИИ

Разработан алгоритм эвристической детекции киберугроз на основе Random Forest с комплексным анализом технических и когнитивно-семантических признаков. Алгоритм построен в виде программного модуля, размещенного в оперативной памяти и связанного с процессором через системную шину для обработки потока данных не менее 100 тыс. записей в час.

Модуль извлечения признаков выполняет вычисления на процессоре для извлечения 47 признаков из каждой записи и каждого URL. Признаки разделяются на две категории.

Технические признаки URL включают 32 параметра, характеризующих структурные и сетевые свойства адреса. Длина доменного имени в символах используется для детекции аномально длинных доменов, типичных для фишинга. Наличие IP-адреса вместо буквенного имени домена индицирует попытку скрыть истинного владельца сайта. Возраст домена по данным WHOIS в днях позволяет выявить недавно зарегистрированные домены, характерные для одноразовых фишинговых сайтов. Наличие протокола HTTPS проверяется с учетом того, что его присутствие не гарантирует легитимность сайта.

Количество поддоменов анализируется для обнаружения попыток имитации через размещение названия бренда в поддомене вместо основного домена. Энтропия Шеннона доменного имени вычисляется для детекции случайно сгенерированных доменов. Наличие дефисов в домене проверяется как признак попытки имитации. Наличие символа @ в домене индицирует специфическую атаку с использованием правила парсинга URL браузерами.

Длина пути URL и количество параметров в строке запроса анализируются для выявления аномально сложных адресов с множественными параметрами, используемых для обфускации. Отношение длины домена к общей длине URL вычисляется для детекции URL с избыточно длинным путем. Сходство домена с известными брендами вычисляется через расстояние Левенштейна к списку из 50 брендов российских организаций.

Когнитивно-семантические признаки контента включают 15 параметров, характеризующих психологическое воздействие текста. Наличие ключевых слов социальной инженерии проверяется через список не менее 15 слов: срочно, выигрыш, блокировка, подтверждение, проверка, бесплатно, успей, последний день, ограниченное предложение, кликни, перейди, введи данные, подтверди личность, верни деньги, возврат.

Наличие названий брендов российских государственных организаций и коммерческих компаний проверяется через список не менее 50 брендов: Госуслуги, Сбербанк, ВТБ, Альфа-Банк, Тинькофф, Газпромбанк, Налоговая служба, ФНС, ПФР, МВД, Росреестр, ГИБДД, МФЦ, Почта России, Wildberries, Ozon, Яндекс, ВКонтакте. Анализ выполняется с учетом замены латинских символов на визуально сходные кириллические для детекции омографических атак типа sbe**г**bank.ru с кириллической буквой г вместо латинской r.

Наличие эмоциональных триггеров определяется через список слов, эксплуатирующих страх, жадность, любопытство. Присутствие призывов к немедленному действию проверяется через маркеры временного давления. Длина текстового содержания записи в символах анализируется с учетом того, что фишинговые публикации часто характеризуются краткостью. Количество восклицательных и вопросительных знаков подсчитываются как индикаторы эмоциональной окраски.

Наличие смешения латиницы и кириллицы анализируется для детекции омографических атак, где визуально сходные символы разных алфавитов используются для обмана. Например, в слове сбербанк буква е может быть заменена латинской e, визуально неотличимой, но имеющей другой код.

Модель Random Forest размещается в оперативной памяти объемом около 2 ГБ. Модель представляет собой ансамбль из 300 деревьев решений, каждое

с максимальной глубиной 15 уровней. Обучение модели проводилось на размеченном датасете, содержащем не менее 10 тыс. образцов URL, специфичных для российского сегмента, с балансом вредоносных и легитимных примеров 40:60.

Модуль классификации применяет модель к вектору из 47 признаков и формирует вероятность угрозы в диапазоне от 0 до 1. Классификация выполняется путем голосования деревьев: каждое из 300 деревьев выдает класс 0 для легитимного или 1 для вредоносного, затем вычисляется доля деревьев, проголосовавших за класс 1. Модуль обеспечивает точность классификации не менее 87% при обработке потока данных не менее 100 тыс. записей в час.

Разработана методика интеграции четырех внешних источников черных списков URL для повышения охвата обнаружения угроз и сокращения времени реагирования. Методика реализована в виде блока динамической интеграции источников данных с возможностью асинхронного опроса через сетевые интерфейсы.

Четыре внешних источника данных выбраны на основании анализа их характеристик. Google Safe Browsing обеспечивает наибольший охват и точность 85%, но характеризуется задержкой обновления 12–14 ч. PhishTank представляет краудсорсинговую базу с охватом 70% и задержкой 6–8 ч. OpenPhish обеспечивает охват 65% с задержкой 4–6 ч. Реестр Роскомнадзора охватывает специфичные для России угрозы с охватом 50% и задержкой 24–48 ч.

Весовые коэффициенты источников установлены на основе эмпирической калибровки: Google Safe Browsing – 0.35; PhishTank – 0.25; OpenPhish – 0.20; реестр Роскомнадзора – 0.20. Сумма весовых коэффициентов равна 1. Коэффициенты отражают компромисс между точностью и оперативностью источников.

Модули интерфейсов реализуют асинхронные HTTP-запросы с использованием библиотеки aiohttp для Python. Асинхронность обеспечивает параллельное выполнение четырех запросов одновременно, сокращая общее время проверки URL с потенциальных 8–12 с. до 2–3 с. Каждый модуль выполняет опрос своего источника с интервалом не более 30 мин.

Модуль кеширования результатов записывает результаты проверок в энергонезависимую память. Для каждого URL сохраняется запись в формате: хеш SHA-

256 от URL, временная метка проверки в формате Unix timestamp, 4 бинарных флага результатов от источников. Время жизни записи в кеше составляет 1 ч.

Модуль агрегирования вычисляет агрегированную оценку угрозы как взвешенную сумму бинарных откликов 4 источников по формуле:

$$0.35 \times \text{отклик\_GSB} + 0.25 \times \text{отклик\_PhishTank} +$$
$$+0.20 \times \text{отклик\_OpenPhish} + 0.20 \times \text{отклик\_РКН},$$

где каждый отклик принимает значение 1, если URL обнаружен, или 0, если не обнаружен.

Методика обеспечивает охват обнаружения угроз не менее 90% против 60–70% у отдельных источников за счет покрытия различных сегментов пространства угроз. Сокращение времени реагирования до 2 ч. достигается за счет интервала опроса 30 мин. и интеграции источников с различными задержками обновления.

Разработан алгоритм рекурсивного анализа цепочек перенаправлений для обнаружения замаскированных вредоносных ссылок, скрытых за несколькими уровнями сокращенных URL и HTTP-редиректов. Алгоритм учитывает когнитивные особенности восприятия пользователями сокращенных URL.

Модуль обнаружения систем сокращения URL проверяет каждый URL на принадлежность к известным сервисам путем сравнения доменного имени со списком не менее 50 доменов-сокращателей. Список включает глобальные сервисы: bit.ly, goo.gl, tinyurl.com, ow.ly; специфичные для России: vk.cc, clck.ru; сокращатели социальных платформ: okl.lt для Одноклассников, vk.link для ВКонтакте.

Модуль HTTP HEAD-запросов выполняет запрос типа HEAD к URL для получения финального адреса без загрузки полного контента. HEAD-запрос возвращает только HTTP-заголовки без тела ответа, что экономит пропускную способность. Параметр allow_redirects = True включает автоматическое следование редиректам для получения конечного URL в цепочке.

Процесс раскрытия ограничен тайм-аутом не более 5 с. на 1 URL для предотвращения атак типа бесконечный редирект. Установлена максимальная глубина цепочки (не более 10 перенаправлений) на основе анализа легитимных сокращателей. Превышение лимита глубины индицирует попытку обфускации и классифицируется как подозрительное поведение.

Буфер хранения цепочек размещается в оперативной памяти и фиксирует полную последовательность URL в цепочке от исходного до финального. Для каждого уровня сохраняется URL, HTTP-код ответа и заголовок Location. Зафиксированная цепочка передается через системную шину в блок эвристической детекции, который применяет модель Random Forest к финальному URL в цепочке для классификации угрозы.

Алгоритм обеспечивает эффективность обнаружения замаскированных угроз не менее 78% на тестовой выборке из 500 сокращенных URL с цепочками перенаправлений от 2 до 10 уровней. Сравнительный анализ показал, что системы защиты без анализа цепочек пропускают до 50% замаскированных угроз.

РЕАЛИЗАЦИЯ И ЭКСПЕРИМЕНТАЛЬНАЯ ОЦЕНКА

Создан программный комплекс для экспериментальной проверки разработанных алгоритмов на реальных данных цифровой платформы Одноклассники. Программная составляющая зарегистрирована как программа для ЭВМ № 2025683166 от 02.09.2025 под названием OKPHISH, что подтверждает воспроизводимость технического решения.

Программный комплекс реализован на Python 3.9 с использованием библиотек: pandas 1.3 для обработки структурированных данных, scikit-learn 1.0 для реализации алгоритмов машинного обучения, aiohttp 3.8 для асинхронных HTTP-запросов, numpy 1.21 для численных вычислений, nltk 3.6 для лингвистического анализа текста.

Архитектура программного комплекса соответствует блочной структуре устройства. Модуль сбора данных реализован с использованием асинхронного программирования через библиотеку asyncio. Модуль подключается к API Одноклассников по протоколу HTTPS с аутентификацией OAuth 2.0, извлекая публично доступные записи с соблюдением ограничений скорости не более 300 запросов в минуту.

Модуль анонимизации применяет хеширование SHA-256 к персональным идентификаторам с добавлением соли длиной 32 байта в соответствии с требованиями ФЗ № 152-ФЗ. Соль генерируется единожды при инициализации системы и сохраняется в защищенном хранилище ключей операционной системы.

Модуль извлечения признаков реализован с использованием векторизованных операций numpy для обеспечения производительности. Технические признаки URL извлекаются через парсинг компонентов адреса библиотекой urllib.parse. Когнитивно-семантические признаки контента извлекаются через токенизацию текста библиотекой nltk с последующим поиском ключевых слов.

Модель Random Forest загружается из сериализованного файла формата pickle при инициализации для размещения в оперативной памяти. Обучение модели проводилось на выделенном вычислительном кластере с использованием процедуры кросс-валидации по 5 блокам. Финальная модель демонстрирует точность 87% на независимой тестовой выборке.

Экспериментальная валидация разработанных алгоритмов проводилась на эмпирической базе, полученной из открытых обезличенных данных пользовательских записей социальной сети Одноклассники через официальный API. Платформа выбрана как репрезентативный пример российской социальной сети с аудиторией 20 млн активных пользователей в месяц.

Период наблюдения составил с 01.02.2024 по 10.10.2025, общей продолжительностью 20 месяцев. Объем эмпирической выборки составил 1 млн записей, включающих 500 тыс. публикаций, 300 тыс. комментариев, 200 тыс. кликов по ссылкам с временными метками взаимодействий.

Обработка осуществлялась исключительно с публично доступными записями, размещенными пользователями в открытом доступе без ограничений видимости. Персональные идентификаторы необратимо хешировались алгоритмом SHA-256 с добавлением соли перед обработкой, что исключает возможность деанонимизации.

Идентификация вредоносных URL осуществлялась через комбинацию методов. Первичная верификация проводилась через проверку в черных списках Google Safe Browsing, PhishTank, OpenPhish и реестре Роскомнадзора. Дополнительная верификация выполнялась с помощью ручного экспертного анализа выборки из 500 URL специалистами в области информационной безопасности. Из массива 1 млн записей идентифицировано 5 тыс. записей с подтвержденными вредоносными URL.

Экспериментальная валидация подтвердила достижение заявленных технических характеристик. Точность детекции угроз блоком эвристической детекции составила 87% на независимой тестовой выборке объемом 5 тыс. URL, включающей 2500 вредоносных и 2500 легитимных адресов. Полнота обнаружения составила 85%, что означает корректную идентификацию 2125 вредоносных URL из 2500.

Сравнительный анализ с четырьмя распространенными системами защиты на той же тестовой выборке показал превосходство разработанного алгоритма. Google Safe Browsing продемонстрировал точность 85%, PhishTank 79%, Yandex Safe Browsing 83%, Kaspersky URL Advisor 84%. Разработанный алгоритм превосходит ближайшего конкурента на 2% по точности и на 5% по полноте обнаружения.

Статистическая значимость различий с системой Google Safe Browsing проверена при помощи теста χ^2. Значение статистики составило 12.4 с. p-значением 0.002, что подтверждает статистически значимое превосходство разработанного алгоритма на уровне достоверности, равном 98%.

Разработанный алгоритм имеет явное преимущество для угроз, специфичных для российского сегмента. На подвыборке из 500 URL с имитацией российских брендов точность составила 92% против 78% у Google Safe Browsing. Это объясняется включением признаков анализа кириллических омографических атак типа sbe**r**bank.ru.

Время реагирования на новые угрозы составило 2 ч. благодаря интеграции источников с асинхронным опросом каждые 30 мин. и блока эвристической детекции, работающего независимо от черных списков. Для сравнения: аналоги демонстрируют время реагирования 12–14 ч. для Google Safe Browsing и 24–48 ч. для реестра Роскомнадзора.

Охват обнаружения угроз составил 90% за счет интеграции четырех источников данных против 60–70% у отдельных источников. Эффективность обнаружения замаскированных угроз блоком анализа цепочек перенаправлений составила 78% на выборке из 500 сокращенных URL с цепочками от 2 до 10 уровней.

Анализ паттернов взаимодействия пользователей с вредоносным контентом выявил критическую роль когнитивных факторов. Из 5 тыс. выявленных угроз

системы фильтрации заблокировали 2 тыс. угроз, обеспечив коэффициент обнаружения 40%. Оставшиеся 3 тыс. угроз прошли фильтры и появились в лентах пользователей.

1350 угроз привели к реализованным кликам пользователей, преодолевших все уровни технической защиты. Анализ показал доминирующую роль доверия к источнику: 743 клика (55%) произошли через контент от пользователей в списке друзей, 135 кликов (10%) от участников тех же социальных групп, 472 клика (35%) от незнакомых пользователей.

Фактор воспринимаемой срочности играл роль в 405 кликах (30% от общего числа). Публикации с явными маркерами временного давления демонстрировали на 40% более высокую эффективность. Социальное подтверждение влияло на 270 кликов (20%). Критическим порогом оказалось наличие минимум 15–20 позитивных реакций.

Анализ реакции на предупреждения браузеров выявил критически низкий уровень эффективности защитных сообщений. Из 1200 угроз, обнаруженных браузерами и сопровожденных предупреждениями, только 360 пользователей (30%) прекратили попытку перехода. Остальные 840 пользователей (70%) проигнорировали предупреждения.

Демографический анализ показал значительную вариацию уязвимости. Пользователи старше 50 лет составляли 540 жертв (40%) при доле в общей аудитории 30%. Жители населенных пунктов с населением менее 100 тыс. человек составляли 877 жертв (65%) при доле в общей аудитории 45%.

ЗАКЛЮЧЕНИЕ

Решена важная научно-техническая задача повышения эффективности защиты цифровых платформ от фишинговых атак и вредоносного ПО путем разработки алгоритмов обработки и детекции угроз с учетом когнитивных особенностей пользователей и обеспечения соответствия требованиям российского законодательства.

Основные научные результаты включают разработку концептуальной модели трехэтапной системы защиты, интегрирующей технические механизмы безопасности с когнитивными моделями принятия решений пользователями. Модель учитывает временные характеристики когнитивной обработки информации.

Предложен алгоритм эвристической детекции на основе Random Forest с комплексным анализом 47 технических и когнитивно-семантических признаков. Алгоритм обеспечивает точность классификации 87% при обработке 100 тыс. записей в час. Показана высокая эффективность алгоритма для обнаружения угроз, специфичных для российского сегмента, установлена точность 92% против 78% у глобальных систем защиты.

Разработана методика динамической интеграции 4 разнородных источников данных об угрозах, позволяющая увеличить охват обнаружения до 90% и сократить время реагирования с 12–14 ч. до 2 ч. Сокращение времени реагирования критически важно с учетом временных характеристик когнитивной уязвимости: 70% кликов происходит в первые 12 ч. после публикации угрозы.

Создан алгоритм рекурсивного анализа цепочек перенаправлений глубиной до 10 уровней с эффективностью 78%, учитывающий когнитивные особенности восприятия сокращенных URL. Алгоритм обнаруживает замаскированные угрозы, которые пропускаются системами без анализа цепочек в 50% случаев.

Обоснована система показателей эффективности защиты, интегрирующая технические метрики с когнитивно-поведенческими индикаторами. Система адаптирована к требованиям российского законодательства ФЗ № 152-ФЗ, ФЗ № 149-ФЗ и ГОСТ Р 57580.1–2017.

Практическая значимость результатов подтверждена экспериментальной проверкой на реальных данных социальной сети Одноклассники объемом 1 млн записей за период 20 месяцев. Внедрение разработанных алгоритмов позволяет повысить эффективность защиты с 67% базового уровня до 80%, что соответствует требованиям ГОСТ Р 57580.1–2017.

Когнитивный анализ 1350 успешных кибератак выявил доминирующую роль доверия к источнику информации, определяющего 55% инцидентов. Критически низкая реакция пользователей на предупреждения браузеров на уровне

30% указывает на проблему привыкания к предупреждениям и необходимость разработки адаптивных форматов коммуникации рисков.

Направления дальнейших исследований включают адаптацию разработанных алгоритмов к другим российским социальным платформам с учетом специфики их аудиторий. Перспективным является исследование влияния культурных и возрастных факторов на когнитивную уязвимость для разработки более точных моделей сегментации пользователей.

СПИСОК ЛИТЕРАТУРЫ

1. *Селиверстов В.В., Корчагин С.А.* Анализ актуальности и состояния современных фишинг-атак на объекты критической информационной инфраструктуры // Инженерный вестник Дона. 2024. № 6 (114). С. 17.

2. *Group-IB.* Отчет о киберугрозах в России за 2023 год: анализ трендов и прогнозы. М.: Group-IB, 2024. 89 с.

3. *Kaspersky Lab.* Развитие киберугроз в 2023 году: статистика и аналитика инцидентов информационной безопасности. М.: Лаборатория Касперского, 2024. 156 с.

4. *Русских Е.И.* Прошлое, настоящее и будущее фишинговых атак // ББК 1 Н 34. С. 6015.

5. *Назаров А.К.* Некоторые современные средства защиты от киберугроз // редакционно-издательским советом Краснодарского университета МВД России. С. 76.

6. *Брюханов В.А., Грызунов В.В., Шестаков А.В.* Выявление проблем информационной безопасности методом систематического обзора литературы. 2024.

7. *Токолов А.В.* Социальная инженерия в вопросах обеспечения информационной безопасности // Криминологический журнал. 2024. № 4. С. 175–182.

8. *Горбунова Е.А., Сайкинов В.Е.* Российская Федерация. Проблема фишинга в использовании информационных систем на основе облачных технологий // И74 Информационное общество: современное состояние и перспективы развития: сборник материалов XI международного студенческого форума. Краснодар: КубГАУ, 2018. С. 103.

9. *Сергеев А.Ю., Широкова О.В.* Мошенничество в цифровом обществе в условиях социальных изменений // Цифровая социология. 2023. Т. 6, № 1. С. 59–71.

10.*Мрочко В.Л., Рощина Т.М., Тарасов М.Д.* Обеспечение безопасности в сети Интернет: психолого-педагогические аспекты // Экономические и социально-гуманитарные исследования. 2024. № 3 (43). С. 196–204.

11.*Серік А.С.* Правовые основы предотвращения кибермошенничества: состояние и перспективы развития. 2022.

12.*Швецова Е.Э.* Виды мошенничества в сфере дистанционного банковского обслуживания и способы борьбы с ними // Сборник материалов Всероссийской научной конференции молодых исследователей с международным участием ИНТЕКС-2024. 2024. С. 269–272.

13.*Уваров А.А.* Информационная безопасность граждан России: современное состояние // Lex russica. 2024. Т. 77, № 1 (206). С. 133–143.

14.*Харисова З.И.* Генезис преступности в сфере компьютерной информации и ее детерминанты // Общество, право, государственность: ретроспектива и перспектива. 2025. № 1 (21). С. 57–65.

15.*Битюкова А.Ф.* Направления развития банковских электронных услуг и способы обеспечения их безопасности. 2019.

16.*ГОСТ Р 57580.1-2017.* Безопасность финансовых (банковских) операций. Требования к организации и проведению работ по обеспечению безопасности. М.: Стандартинформ, 2017. 26 с.

17.*Федеральный закон от 27.07.2006 № 152-ФЗ «О персональных данных»* (ред. от 14.07.2022). Доступ из справочно-правовой системы «КонсультантПлюс».

18.*Федеральный закон от 27.07.2006 № 149-ФЗ* «Об информации, информационных технологиях и о защите информации» (ред. от 14.07.2022). Доступ из справочно-правовой системы «КонсультантПлюс».

19.*Sheng S., Holbrook M., Kumaraguru P., Cranor L.F., Downs J.* Who falls for phish? A demographic analysis of phishing susceptibility and effectiveness of interventions // Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Atlanta, GA, USA, 2010. P. 373-382. https://doi.org/10.1145/1753326.1753383

20.*Guarino N.* Formal ontology, conceptual analysis and knowledge representation // Int. J. of Human Computer Studies. 1995. Vol. 43 (5/6). P. 625–640.

RESEARCH OF DATA PROCESSING, DETECTION AND PROTECTION ALGORITHMS TO MINIMIZE THE IMPACT OF MALWARE AND PHISHING ATTACKS ON USERS OF DIGITAL PLATFORMS

T. S. Volokitina[1] [0000-0002-5493-447X], M. O. Tanygin[2] [0000-0002-4099-1414]

[1,2]Southwest State University, Kursk, Russia

[1]tativolokitina@gmail.com, [2]tanygin@yandex.ru

Abstract

The article is devoted to the development of a scientific and methodological apparatus for improving the effectiveness of protecting digital platforms from cyber threats by creating processing and detection algorithms that take into account the cognitive characteristics of users. A conceptual model of a three-stage protection system is proposed, integrating technical security mechanisms with cognitive decision-making models. A heuristic detection algorithm based on Random Forest machine learning with analysis of 47 features, including technical URL characteristics and cognitive-semantic content characteristics, has been developed. A methodology for dynamic integration of four threat data sources has been created, reducing response time from 12–14 hours to two hours. An algorithm for recursive analysis of redirection chains up to ten levels deep to detect masked threats is proposed. Experimental validation on an empirical base of approximately one million records confirmed detection accuracy of 87% when processing one hundred thousand records per hour. The developed solutions ensure compliance with the requirements of GOST R 57580.1-2017 and Russian legislation in the field of personal data protection.

Keywords: *heuristic threat detection, machine learning, cognitive security, phishing attacks, social engineering, data protection, threat source integration.*

REFERENCES

1. *Seliverstov V.V., Korchagin S.A.* Analysis of the relevance and state of modern phishing attacks on critical information infrastructure objects // Engineering Bulletin of the Don. 2024. No. 6 (114). P. 17.

2. *Group-IB.* Report on cyber threats in Russia for 2023: analysis of trends and forecasts. Moscow: Group-IB, 2024. 89 p.

3. Kaspersky Lab. Development of cyber threats in 2023: statistics and analytics of information security incidents. Moscow: Kaspersky Laboratory, 2024. 156 p.

4. *Russkikh E.I.* Past, present and future of phishing attacks // BBK 1 N 34. P. 6015.

5. *Nazarov A.K.* Some modern means of protection against cyber threats // Editorial and publishing council of the Krasnodar University of the Ministry of Internal Affairs of Russia. P. 76.

6. Bryukhanov V.A., Gryzunov V.V., Shestakov A.V. Identification of information security problems by the method of systematic literature review. 2024.

7. *Tokolov A.V.* Social engineering in information security issues // Criminological Journal. 2024. No. 4. P. 175–182.

8. *Gorbunova E.A., Saykinov V.E.* Russian Federation The problem of phishing in the use of information systems based on cloud technologies // I74 Information Society: current state and development prospects: collection of materials of the XI international student forum. Krasnodar: KubSAU, 2018. P. 103.

9. *Sergeev A.Yu., Shirokova O.V.* Fraud in digital society under conditions of social change // Digital Sociology. 2023. Vol. 6, No. 1. P. 59–71.

10. *Mrochko V.L., Roschina T.M., Tarasov M.D.* Ensuring security on the Internet: psychological and pedagogical aspects // Economic and socio-humanitarian research. 2024. No. 3 (43). P. 196–204.

11. *Serik A.S.* Legal foundations for preventing cybercrime: state and development prospects. 2022.

12. *Shvetsova E.E*. Types of fraud in the field of remote banking and methods of combating them // Collection of materials of the All-Russian scientific conference of young researchers with international participation INTEX-2024. 2024. P. 269–272.

13. Uvarov A.A. Information security of Russian citizens: current state // Lex russica. 2024. Vol. 77, No. 1 (206). P. 133–143.

14. *Kharisova Z.I.* Genesis of crime in the field of computer information and its determinants // Society, law, statehood: retrospective and perspective. 2025. No. 1 (21). P. 57–65.

15. *Bityukova A.F.* Directions for the development of banking electronic services and methods of ensuring their security. 2019.

16. *GOST R 57580.1-2017.* Security of financial (banking) operations. Requirements for the organization and conduct of security work. Moscow: Standartinform, 2017. 26 p.

17. *Federal Law No. 152-FZ of July 27, 2006 "On Personal Data"* (as amended on July 14, 2022). Access from the reference legal system "ConsultantPlus".

18. Federal Law No. 149-FZ of July 27, 2006 "On Information, Information Technologies and Information Protection" (as amended on July 14, 2022). Access from the reference legal system "ConsultantPlus".

19. Sheng S., Holbrook M., Kumaraguru P., Cranor L.F., Downs J. Who falls for phish? A demographic analysis of phishing susceptibility and effectiveness of interventions // Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Atlanta, GA, USA, 2010. P. 373–382. https://doi.org/10.1145/1753326.1753383

20. Guarino N. Formal ontology, conceptual analysis and knowledge representation // Int. J. of Human Computer Studies. 1995. Vol. 43 (5/6). P. 625–640.

СВЕДЕНИЯ ОБ АВТОРАХ

ВОЛОКИТИНА Татьяна Сергеевна, окончила Юго-Западный государственный университет в 2021 г. Аспирант кафедры информационной безопасности Юго-Западного государственного университета. В списке научных трудов более 50 работ в области кибербезопасности и защиты информации.

Tatiana Sergeevna VOLOKITINA graduated from South-Western State University in 2021. She is currently a postgraduate student at the Department of Information Security of South-Western State University. She has authored more than 50 scientific publications in the fields of cybersecurity and information protection.

email: tativolokitina@gmail.com

ORCID: 0000-0002-5493-447X

ТАНЫГИН Максим Олегович, окончил Курский государственный технический университет в 2001 г., д. т. н. (2022). Доцент кафедры информационной безопасности Юго-Западного государственного университета. В списке научных трудов более 100 работ в области информационной безопасности и анализа данных.

Maxim Olegovich TANYGIN graduated from Kursk State Technical University in 2001, Doctor of Technical Sciences (2022). He is an Associate Professorat the Department of Information Security of South-Western State University. He has authored more than 100 scientific publications in the fields of information security and data analysis.

email: tanygin@yandex.ru

ORCID: 0000-0002-4099-1414

Материал проступил в редакцию 12 декабря 2025 года