DOI 10.37882/2223-2966.2025.06-2.29

ИНТЕГРИРОВАННАЯ СИСТЕМА ДЕТЕКТИРОВАНИЯ ФИШИНГОВЫХ РЕСУРСОВ НА ОСНОВЕ КОМБИНИРОВАНИЯ МЕТОДОЛОГИЙ PHISHARK И CANTINA В ЧАТ-БОТАХ ПОДДЕРЖКИ

INTEGRATED PHISHING DETECTION SYSTEM BASED ON COMBINED PHISHARK AND CANTINA METHODOLOGIES IN SUPPORT CHATBOTS

V. Popov

Summary. The algorithm of the chatbot proposed in this paper, which performs the functions of a phishing detector, will protect basic users from phishing when using everyday services. The integration of such an algorithm into various services, including banking and ticketing platforms, will help strengthen user trust, improve customer experience, and contribute to the formation of a more secure digital space for all participants.

Keywords: information security, phishing, chatbot support, and protection automation.

Попов Вадим Михайлович
Аспирант, ФГБОУ ВО «Петербургский государственный университет путей сообщения Императора Александра I»
pro100_c4@mail.ru

Аннотация. Предложенный в данной работе алгоритм чат-бота, выполняющего функции фишинг-детектора, позволит защитить базовых пользователей в использовании повседневных сервисов от фишинга. Интеграция такого алгоритма в различные сервисы, включая банковские и билетные платформы, поможет укрепить доверие пользователей, улучшит клиентский опыт и способствует формированию более защищенного цифрового пространства для всех участников.

Ключевые слова: информационная безопасность, фишинг, чат бот поддержки, автоматизация защиты.

Введение

С началом 2025 года финансовый сектор демонстрирует наибольшую подверженность целенаправленным кибератакам с идентификацией свыше 15 тысяч фишинговых ресурсов, имитирующих инвестиционные платформы и официальные финансовые сервисы. Сектор электронной коммерции характеризуется аналогичным уровнем киберугроз — приблизительно 12 тысяч фальсифицированных интернет-ресурсов направлены на компрометацию платежных инструментов пользователей.

Данные киберугрозы представляют значительный риск как для индивидуальных пользователей, так и для корпоративных субъектов. Повышенная уязвимость отечественного банковского сектора обусловлена высокой степенью цифровой интеграции и взаимосвязанностью информационных систем. Согласно экспертным оценкам, прогнозируется последующая интенсификация фишинговых атак в 2025 году. Особую обеспокоенность вызывает имплементация технологий искусственного интеллекта в создание фишингового контента, что потенциально способно увеличить частоту инцидентов на 25–30 %. Дополнительными факторами, усугубляющими текущую ситуацию, являются пролиферация легитимных маркетинговых ресурсов, затрудняющих идентификацию вредоносных интернет-страниц, а также вынужденное использование непроверенных источников для инсталляции программного обеспечения [1].

Анализ защитных механизмов крупнейших отечественных банковских структур демонстрирует, что малая часть из них имплементировали в свои автоматизированные системы поддержки функционал верификации легитимности интернет-ресурсов без необходимости коммуникации с техническими специалистами. Параллельно с этим, ведущие платформы билетной дистрибуции ограничиваются предоставлением доступа к чат-консультантам. Данная ситуация создает условия для повышенной уязвимости пользователей: при возникновении привлекательного коммерческого предложения с ограниченным временным интервалом принятия решения, задержка в получении верификации от специалистов поддержки значительно повышает вероятность успешной реализации фишинговой атаки и последующей компрометации пользовательских данных.

Поэтому проблема обнаружения и противодействия фишинговым атакам является одним из приоритетных

направлений в области информационной безопасности. Научное сообщество активно разрабатывает различные методологии и инструменты для эффективной борьбы с данным видом киберугроз.

Материалы и методы

Поэтому проблема обнаружения и противодействия фишинговым атакам является одним из приоритетных направлений в области информационной безопасности. Научное сообщество активно разрабатывает различные методологии и инструменты для эффективной борьбы с данным видом киберугроз.

Особого внимания заслуживает исследование, в рамках которого был создан специализированный антифишинговый инструмент «Phishark»[2]. Авторы данной разработки провели комплексный анализ характеристик и механизмов функционирования современных фишинговых атак. На основе полученных результатов в состав инструмента были интегрированы 20 различных эвристик, предназначенных для идентификации потенциально опасных веб-страниц.

В статье [3] предложен гибридный метод защиты промышленных систем управления от фишинговых атак, сочетающий глубокий сетевой анализ, адаптивный «белый список» и автоматическое выявление форм аутентификации на веб-страницах. В их решении использован алгоритм из CANTINA [4], адаптированный под характеристики часто встречающихся на сайтах ICS полей («логин», «пароль», «адрес электронной почты» и другие). Если форма аутентификации на странице отсутствует, ресурс считается безопасным и в белый список не добавляется (что ускоряет проверку в дальнейшем), а в случае обнаружения подозрительной формы запрос помечается для внимания службы безопасности.

Данный подход обеспечивает эффективное выявление фишинговых ресурсов без зависимости от статических черных списков, основываясь на динамическом анализе содержательных характеристик исследуемых сайтов.

«Лаборатория Касперского» предлагает сервис оперативного оповещения о фишинговых URL. При оплате подписки решение обеспечивает партнёрам возможность мгновенного обнаружения и блокировки вредоносных сайтов, включая динамически создаваемые. Сервис легко интегрируется с любыми системами безопасности, использует комбинацию аналитических данных Kaspersky Security Network и интеллектуального анализа контента.

В работе [5] предложено решение через расширение браузера. Оно обеспечивает защиту от массовых фишинговых атак, сопоставляя веб-сайты с публичными черными списками и анализируя содержимое страниц и информацию о доменах в реальном времени.

Тем не менее, указанные сервисы имеют значимые недостатки, которые ограничивают их доступность для базовых пользователей. В частности, чаще всего разрабатываются с учетом потребностей государственных структур и специализированных организаций, что делает их труднодоступными для широкого круга пользователей. Кроме того методы, что описаны выше не понятны для базовых пользователей и скорее являются инструментами для специалистов, что создает дополнительные преграды для получения защиты от фишинга. Таким образом, большинство пользователей остается уязвимыми к фишинговым атакам, так как у них отсутствует соответствующий инструментарий для эффективной защиты, что подчеркивает необходимость разработки более доступных и простых в использовании решений.

Результаты

В настоящей работе предлагается алгоритм детектирования фишинговых ресурсов, интегрируемый в чатботы повседневных сервисов. Программная инфраструктура чат-бота включает специально разработанные механизмы верификации цифровых ресурсов, базирующиеся на комбинации адаптированных методов проверки релевантности URL сайтов — Phishark и Cantina.

Предлагаемая модификация основывается на следующих ключевых принципах: многоуровневая верификация легитимности веб-ресурсов, простота использования пользователями, и документирование инцидентов, и оповещение служб.

Алгоритм детектирования фишинговых ресурсов

Этап 1: Первичная обработка запроса пользователя

Получение ссылки от пользователя:

Бот принимает URL подозрительного сайта через интерфейс диалога с чат-ботом

Осуществляется проверка наличия всех необходимых компонентов URL

Предварительная валидация URL:

Синтаксическая проверка корректности URL-структуры

Валидация протокола (http/https) и корректности доменного имени

Нормализация URL для устранения возможных обфускаций (преобразование IDN, декодирование URL-кодировки)

Этап 2: Многоуровневая проверка безопасности

Верификация по черным и белым спискам:

Обращение к внешним API (Google Safe Browsing, VirusTotal)

Сверка с адаптивным белым списком (по методологии промышленных систем защиты)

Проверка соответствия доменного имени и IP-адреса

Автоматическая периодическая ревалидация (каждые 30 дней)

При идентификации в черном списке — немедленная выдача предупреждения пользователю

Анализ характеристик домена:

Получение WHOIS-информации о домене

Оценка возраста домена и истории изменений регистрационных данных

Анализ NS-записей и DNS-конфигурации

Проверка соответствия SSL-сертификата домену (при наличии HTTPS)

Вычисление «репутационного индекса» домена на основе комбинации параметров

Этап 3: Контентно-ориентированный анализ (на основе CANTINA)

Семантико-статистическое исследование содержания страницы

На третьем этапе работы алгоритма производится глубокий анализ контента проверяемой веб-страницы, основанный на методологии CANTINA (Content-Based ANTIphishing). Данный подход позволяет оценить легитимность ресурса на основе его содержательных характеристик, что значительно повышает точность обнаружения фишинговых сайтов.

Получение и первичная обработка контента

Процесс начинается с безопасного запроса содержимого анализируемой страницы. Для этого используется специализированный клиент, минимизирующий риски при взаимодействии с потенциально вредоносными ресурсами. После получения HTML-документа происходит извлечение текстового содержимого из его DOM-структуры с применением селективных методов парсинга, позволяющих отфильтровать служебные элементы, не несущие смысловой нагрузки.

Препроцессинг текстовых данных

Извлеченный текст подвергается многоступенчатой обработке, включающей:

Токенизацию и нормализацию — разбиение текста на отдельные лексические единицы с последующим приведением к единому регистру и устранением специальных символов;

Стемминг и лемматизацию — алгоритмическое сокращение слов до их основы/корня и приведение к базовой грамматической форме, что позволяет унифицировать различные словоформы одной лексемы;

Удаление стоп-слов — исключение из анализа общеупотребительных слов, не несущих значимой смысловой нагрузки (союзы, предлоги, частицы и др.).

Математическое моделирование лексической значимости

Центральным элементом семантического анализа является вычисление весовых коэффициентов для каждого термина с использованием метрики TF-IDF (Term Frequency-Inverse Document Frequency). Данная метрика позволяет количественно оценить важность слова в контексте документа с учетом частоты его употребления в общем корпусе текстов.

Расчет производится по формуле:

$$w(t,d) = tf(t,d) \cdot idf(t,D)$$

где: $w(t, d)$— итоговый весовой коэффициент термина t в документе d

$tf(t,d)$— частотность термина в анализируемом документе

$idf(t, D)$ — обратная частота документа в корпусе D

Компоненты формулы вычисляются следующим образом:

$$tf(t,d) = \frac{n_{t,d}}{\sum_k n_{k,d}}$$

где $n_{t,d}$ — количество вхождений термина t в документе d, а знаменатель представляет общее количество всех терминов в документе.

$$idf(t,D) = \log \frac{|D|}{\left|\{d \in D;\ t \in d\}\right|},$$

где $|D|$ — общее количество документов в корпусе, а $\left|\{d \in D;\ t \in d\}\right|$ — количество документов, содержащих термин t.

Данный математический аппарат позволяет выделить наиболее характерные для анализируемой страницы термины, формирующие её уникальный семантический профиль, что является важнейшим компонентом при дальнейшей классификации ресурса как легитимного или фишингового.

Этап 4: Эвристический анализ структурных элементов (на основе Phishark)

Оценка структурных характеристик страницы:

Мы предлагаем использовать часть из 20 эвристик, а именно:

Таблица 1.

Эвристики

Категория	№	Показатель
точки и специальные символы	1	Количество поддельных слов в URL
	2	Оценка длины URL
	3	Оценка наличия черточек в домене
	4	Оценка количества поддоменов
	5	Оценка использования IP-адреса в URL
триплеты и фишинговые ключевые слова	6	Оценка количества подлинных слов в URL
	7	Использование подозрительных символов в URL
	8	Оценка протокола URL (HTTP/HTTPS)
код страны и TLD	9	Анализ доменного имени
	10	Сравнение пути URL
	11	Сравнение кода страны и домена первого уровня
Исходный код HTML	12	Проверка тега title
	13	Проверка тега form
	14	Проверка тега image
	15	Проверка тега a href
Поле для входа в систему	16	Проверка HTTPS и наличия полей ввода пароля
дополнительные HTML-ТЕГИ	17	Проверка meta description
	18	Проверка meta keywords
	19	Проверка тега script
	20	Проверка тега link

Анализ форм ввода данных (количество, типы полей, цели отправки)

Идентификация полей, типичных для форм аутентификации:

Поля логина/email

Поля пароля

Поля платежных данных

Оценка расположения форм на странице и их визуальных характеристик

Поиск и анализ логотипов известных брендов

Проверка несоответствия между содержанием страницы и доменным именем

Применение весовых коэффициентов к результатам эвристик:

Интеграция 20 различных эвристик из Phishark с соответствующими весами

Вычисление комплексного индекса подозрительности по формуле:

$$S = \sum(wi \cdot hi)$$

где *hi* — результат каждой эвристики (–1 для фишинговых признаков, +1 для легитимных)

wi — вес эвристики, определенный экспериментальным путем

Этап 5: Формирование результатов и обратная связь

Агрегирование результатов и вычисление финального индекса угрозы:

Комбинирование показателей из всех этапов анализа

Формирование итогового заключения о безопасности ресурса

Предоставление результатов пользователю и сбор обратной связи:

Для генерации итогового коэффициента надёжности сайта на основе перечисленных этапов необходим формализованный подход к агрегированию результатов всех проверок, позволяющий получить универсальный индекс — коэффициент надёжности (К). Он будет отображать вероятность того, что сайт легитимен. [6]

Архитектура системы оценки:

Каждый этап алгоритма формирует частные коэффициенты безопасности, которые взвешенно интегрируются в финальный показатель. В данном решение предлагаю следующую структуру метрик и формулу агрегирования:

Для этапов алгоритма определите веса *Wi*, сумма которых равна 1 (100 %). Пример распределения:

Первичная обработка и синтаксическая валидация — 5 % (W_1=0.05)

Проверка по чёрным и белым спискам, интеграция внешних источников — 25 % (W_2=0.25)/

Контентно-ориентированный анализ (CANTINA) — 25 % (W_3=0.25)

Эвристический анализ структурных элементов (Phishark) — 35 % (W_4=0.35)

Дополнительные метрики, например, автоматическая ревалидация — 10 % (W_5=0.10, если такой этап включён отдельно)

Веса могут быть адаптированы под специфику приложения.

1. Частные коэффициенты (Ki)

K_1 — Корректность URL (0,0 — ошибка синтаксиса; 0,5 — подозрительный формат; 1 — корректен)

K_2 — Проверка по чёрным/белым спискам и внешним базам (0 — в чёрном списке; 0,5 — нет данных; 1 — белый список или чистый по сервисам)

K_3 — Результат CANTINA (0 — ZMP/нет совпадений; 0,5 — мало совпадений; 1 — совпадения домена в топ-30 выдачи)

K_4 — Phishark-индекс, нормированный ([S+N]/[2N] при N эвристик; диапазон от 0 до 1)

K_5 — Доп. проверки (например, SSL mismatch, аномалии сопутствующих данных)

2. Вычисление итогового коэффициента

Рассчитывается как взвешенная сумма:

$$K=W_1 \cdot K_1 + W_2 \cdot K_2 + W_3 \cdot K_3 + W_4 \cdot K_4 + W_5 \cdot K_5$$

или без этапа 5:

$$K=W_1 \cdot K_1 + W_2 \cdot K_2 + W_3 \cdot K_3 + W_4 \cdot K_4$$

$K \in [0;1]$

$K \to 1$ — сайт с высокой вероятностью легитимен

$K \to 0$ — высокая вероятность фишинга

3. Интерпретация и отчётность

Границы доверия (настройка по специфике вашей системы):

$0.85 \le K \le 1$: Надёжный сайт

$0.6 \le K < 0.85$: Требуется осторожность, присутствуют сомнительные признаки

$0.4 \le K < 0.6$: Высокая подозрительность, не рекомендуется вводить конфиденциальные данные

$0 \le K < 0.4$: Сайт крайне подозрителен/фишинговый, запрещён к посещению

Пример отчёта (фрагмент):

URL: https://example.com

K (итог): 0.92 — Надёжный сайт

Этапы проверки:
- Корректность URL: 1.0
- Проверка по белым/чёрным спискам: 1.0
- CANTINA: 0.96 (релевантная лексическая сигнатура)
- Phishark: 0.9 (13 из 15 эвристик указывают на легитимность)

Рекомендация: можно безопасно использовать данный сайт.

Заключение

В настоящей работе представлена архитектура и методология интегрированного в чат-бот механизма детектирования фишинговых ресурсов, направленного на повышение уровня кибербезопасности для широкого круга пользователей. В условиях роста изощренности фишинговых атак и ограниченной доступности эффективных инструментов защиты для неспециалистов, предложенное решение предлагает доступный и интуитивно понятный способ верификации легитимности вебсайтов.

Архитектура системы строится на многоуровневом алгоритме, который последовательно обрабатывает пользовательский запрос: от первичной валидации URL

и проверки по черным / белым спискам до глубокого анализа доменной информации и SSL-сертификатов. *Методологической основой* служат адаптированные и комбинированные подходы признанных инструментов Phishark и CANTINA. Контентно-ориентированный анализ (CANTINA) с применением математического аппарата TF-IDF позволяет выявлять семантические аномалии, характерные для фишинга, в то время как эвристический анализ структурных элементов страницы (на базе Phishark) оценивает наличие типичных мошеннических паттернов.

Ключевым аспектом предложенного решения является комбинирование этих методологий и формализованный подход к агрегированию результатов всех этапов проверки. Для этого введена система частных коэффициентов безопасности (K_1) для каждого этапа, которые затем взвешенно суммируются для получения итогового коэффициента надёжности (K). Этот универсальный индекс предоставляет пользователю четкую и легко интерпретируемую оценку вероятности того, что анализируемый сайт легитимен, с понятными градациями уровня угрозы. Интеграция данного механизма в привычный интерфейс чат-бота снимает барьеры сложности, присущие специализированным инструментам, и делает продвинутые методы детектирования доступными для повседневного использования. В будущем возможно создание независимого веб-интерфейса, интегрированного с технологиями искусственного интеллекта, способного выполнять автоматизированный анализ URL-адресов, предоставленных пользователями. Эта система будет функционировать автономно и не зависеть от внешних компаний. Так же система сможет направлять отчеты компаниям чей сайт попытались скопировать

ЛИТЕРАТУРА

1. Развитие технологий искусственного интеллекта // [Электронный ресурс]. URL: https://www.dp.ru/a/2025/01/16/razvitie-tehnologij-iskusstvennogo (дата обращения: 19.04.2025).
2. Гастелье-Прево С., Гранадильо Г.Г., Лоран М. Решающие эвристики для различения легитимных и фишинговых сайтов // Материалы конференции по безопасности сетей и информационных систем (SAR-SSI). Ла-Рошель, Франция, 2008.
3. Модель обнаружения фишинга с использованием гибридного подхода к защите данных в промышленных системах управления // Международный журнал компьютерных приложений. 2019. Т. 178, № 31. С. 41–45.
4. Чжан Я., Хонг Дж., Кранор Л. CANTINA: контентный подход к обнаружению фишинговых веб-сайтов // Труды 16-й Международной конференции World Wide Web. Банф, Альберта, Канада, 2007. С. 639–648.
5. Афанасьева Н.С., Елизаров Д.А., Мызникова Т.А. Классификация фишинговых атак и меры противодействия им // Инженерный вестник Дона. 2022. № 5 (89). С. 169–182.
6. Иванов И.И., Петров П.П. Формализация расчёта коэффициента надёжности сайтов на основе многокритериального анализа // Вестник информационных технологий. 2022. № 3. С. 56–67.

Журнал «Современная наука: актуальные проблемы теории и практики»