УДК 159.9

Методология оценки эффективности систем защиты от фишинга на основе когнитивного моделирования

Волокитина Татьяна Сергеевна

Аспирант кафедры информационной безопасности,
Юго-Западный государственный университет,
305040, Российская Федерация, Курск, ул. 50 лет Октября, 94;
e-mail: tativolokitina@gmail.com

Таныгин Максим Олегович

Доктор технических наук,
доцент кафедры информационной безопасности,
Юго-Западный государственный университет,
305040, Российская Федерация, Курск, ул. 50 лет Октября, 94;
e-mail: tativolokitina@gmail.com

Аннотация

В работе рассматривается методология оценки эффективности систем защиты от фишинга, построенная на принципах когнитивного моделирования процессов принятия решений в условиях неопределенности. Экспериментальная часть исследования выполнена на базе социальной сети Одноклассники с анализом выборки, насчитывающей 800 тысяч записей. Авторами предлагается комплексный подход к анализу защитных механизмов, охватывающий оценку фильтров на этапе публикации, анализ эффективности веб-браузеров и изучение поведенческих паттернов пользователей. В ходе работы была разработана математическая модель, позволяющая прогнозировать эффективность защитных систем с учетом временных задержек обновления черных списков. Полученные результаты демонстрируют, что действующие системы защиты обеспечивают блокировку лишь 40% угроз на этапе публикации, в то время как 60% фишинговых атак успешно проходят через существующие фильтры. Использование методов когнитивного моделирования дало возможность выявить ключевые уязвимости системы и сформулировать рекомендации по повышению эффективности защиты до 80%.

Для цитирования в научных исследованиях

Волокитина Т.С., Таныгин М.О. Методология оценки эффективности систем защиты от фишинга на основе когнитивного моделирования // Психология. Историко-критические обзоры и современные исследования. 2025. Т. 14. № 7А. С. 105-112.

Ключевые слова

Фишинг, когнитивное моделирование, информационная безопасность, социальные сети, защитные системы, эвристический анализ, машинное обучение, киберугрозы.

Введение

В современную эпоху цифровизации фишинг остается одной из наиболее острых проблем информационной безопасности. Статистические данные за 2023 год фиксируют рост зарегистрированных фишинговых атак на 35%, причем социальные сети превратились в основную площадку для распространения вредоносного контента [Антонов, Петрова, 2024, с. 15]. Вызывает серьезную озабоченность тот факт, что традиционные методы защиты показывают неудовлетворительную эффективность в условиях непрерывной эволюции угроз.

Вопрос оценки эффективности систем защиты от фишинга становится критически важным применительно к российским социальным сетям, где пользователи регулярно сталкиваются с локализованными атаками на государственные сервисы, банковские системы и популярные интернет-ресурсы. Действующие подходы к анализу защитных механизмов нередко сводятся к статистическому анализу уже свершившихся инцидентов, при этом игнорируется динамика развития угроз и когнитивные особенности восприятия информации пользователями.

Целью настоящего исследования является разработка комплексной методологии оценки эффективности систем защиты от фишинга на основе когнитивного моделирования, которая даст возможность не только оценить текущее состояние защитных механизмов, но и спрогнозировать их эффективность в условиях изменяющейся киберугрозной среды.

Литературный обзор

Обзор современных исследований в сфере защиты от фишинга демонстрирует, что большинство научных работ концентрируется на технических аспектах обнаружения угроз, при этом когнитивные факторы, влияющие на эффективность защитных систем, остаются недостаточно изученными.

В работе Петрова А.В. и соавторов [Петров, Сидоров, Козлов, 2023, с. 45] представлен метод машинного обучения для классификации фишинговых сайтов, который показал точность 87% на тестовой выборке. Тем не менее, предложенный подход не принимает во внимание специфику социальных сетей, где контекст размещения ссылок играет ключевую роль в принятии решений пользователями.

Исследование Сидорова К.М. [Сидоров, 2023, с. 81] было посвящено анализу эффективности современных антифишинговых решений в российском сегменте интернета и выявило существенные пробелы в покрытии локальных угроз. Автор подчеркивает, что международные черные списки охватывают лишь 40% российских фишинговых ресурсов, что формирует серьезную уязвимость для отечественных пользователей.

Зарубежные исследования также свидетельствуют о необходимости комплексного подхода к оценке защитных систем. Johnson M. и Taylor R. [Johnson, Taylor, 2023, с. 235] создали модель временных задержек в обновлении репутационных баз данных, продемонстрировав, что средняя задержка обнаружения новых угроз достигает 12-14 часов, на протяжении которых пользователи остаются уязвимыми.

Особый интерес представляет работа Anderson P. [Anderson, 2024, с. 157], где впервые применены принципы когнитивного моделирования для анализа поведения пользователей при взаимодействии с подозрительными ссылками. Автор установил, что эффективность предупреждающих сообщений браузеров составляет лишь 30% в условиях низкой цифровой

грамотности пользователей.

Chen L. и соавторы [Chen, Wang, Zhang, 2023, с. 2847] предложили гибридную модель оценки киберугроз, сочетающую статистический анализ с элементами искусственного интеллекта. Их подход позволил повысить точность прогнозирования новых атак до 75%, однако модель требует значительных вычислительных ресурсов и не адаптирована для работы в реальном времени.

Российские исследователи Козлов Д.А. и Морозова Е.П. [Козлов, Морозова, 2024, с. 80] разработали методику оценки эффективности корпоративных систем защиты от фишинга, основанную на анализе инцидентов информационной безопасности. Их работа показала важность учета человеческого фактора, который определяет до 80% успешности фишинговых атак.

Несмотря на значительные достижения в области технической защиты от фишинга, остается нерешенной проблема комплексной оценки эффективности защитных систем с учетом когнитивных особенностей восприятия угроз пользователями. Существующие методологии фокусируются на отдельных компонентах защиты, не предоставляя целостной картины функционирования системы безопасности в условиях реальной эксплуатации.

Материалы и методы

Практическое исследование выполнялось на платформе социальной сети Одноклассники в течение периода с 1 по 10 февраля 2025 года. Выбор именно этой платформы обусловлен ее широкой популярностью среди российских пользователей и интенсивной активностью размещения внешних ссылок в публикациях и комментариях.

Для сбора экспериментальных данных нами было создано специализированное приложение, интегрированное с API платформы и состоящее из следующих функциональных компонентов:

– Модуль мониторинга контента - обеспечивал непрерывное отслеживание новых публикаций и комментариев, содержащих внешние ссылки. Система обрабатывала в среднем 80 тысяч записей ежедневно, используя фильтры для исключения дублированного контента.
– Классификатор угроз - применял комбинацию эвристических правил и модели машинного обучения для выявления потенциально вредоносных URL. Алгоритм проводил анализ структуры ссылок, доменных имен, HTTP-заголовков и метаданных страниц.
– Система верификации - выполняла проверку выявленных угроз по международным черным спискам (Google Safe Browsing, OpenPhish, PhishTank) и российским реестрам (Роскомнадзор).
– Модуль поведенческого анализа - осуществлял мониторинг реакции пользователей на размещенные ссылки, фиксируя количество кликов, время экспозиции и паттерны взаимодействия.
– Общий объем собранных данных составил 800 тысяч записей, включающих 500 тысяч публикаций и 300 тысяч комментариев. Из этого массива было извлечено 50 тысяч уникальных URL, из которых 5 тысяч были классифицированы как вредоносные.

Когнитивное моделирование осуществлялось с использованием байесовских сетей, позволяющих учесть неопределенность в процессах принятия решений. Модель включала следующие переменные:

– Технические характеристики угрозы (тип атаки, сложность маскировки, географическая привязка);
– Временные факторы (время публикации, задержки обновления черных списков);
– Пользовательские факторы (уровень цифровой грамотности, опыт работы с социальными сетями);
– Системные параметры (эффективность фильтров, скорость реакции модерации);

Для оценки эффективности защитных систем применялась многокритериальная модель, включающая следующие метрики:

– Коэффициент блокировки на этапе публикации - отношение заблокированных угроз к общему количеству вредоносных ссылок;
– Время реакции системы - средняя задержка между публикацией и обнаружением угрозы;
– Эффективность браузерной защиты - процент угроз, остановленных на этапе перехода пользователя;
– Коэффициент ложных срабатываний - доля легитимных ссылок, ошибочно заблокированных системой;

Статистическая обработка данных проводилась с использованием методов дисперсионного анализа, корреляционного анализа и регрессионного моделирования. Для проверки статистической значимости результатов применялся критерий Стьюдента с уровнем значимости $\alpha = 0,05$.

Результаты

Анализ собранных экспериментальных данных показал существенные недостатки в функционировании действующих систем защиты от фишинга. Из 5 тысяч выявленных вредоносных URL лишь 2 тысячи (40%) оказались заблокированы на этапе публикации, что свидетельствует о пропуске 60% потенциальных угроз через первичные фильтры безопасности.

Детальное изучение полученных результатов продемонстрировало стабильность данного показателя вне зависимости от типа контента: эффективность блокировки составила 40% как для публикаций (1,6 тысячи из 4 тысяч угроз), так и для комментариев (400 из 1 тысячи угроз). Подобная закономерность свидетельствует об использовании единого алгоритма фильтрации, который не учитывает специфику размещения контента.

Структурный анализ пропущенных угроз (3 тысячи URL) дал возможность выделить основные категории уязвимостей:

Новые угрозы представили наибольшую долю пропусков - 1,8 тысячи случаев (60% от общего числа пропущенных). Данные ссылки отсутствовали в черных списках на момент публикации, что свидетельствует о реактивном характере действующих систем защиты. Характерным примером может служить фишинговая ссылка, размещенная 4 февраля в 10:00, которая попала в OpenPhish только к 12:00, а в Google Safe Browsing - к 23:00, за это время собрав 50 кликов пользователей.

Локальные атаки на российские сервисы составили 450 случаев (15% пропусков), включая 300 URL с фишингом портала Госуслуг и 150 - банковских сайтов. Низкая эффективность обнаружения связана с недостаточным покрытием российских угроз международными черными списками Google Safe Browsing (40% охвата локальных угроз).

Цепочки редиректов использовались в 240 случаях (8% пропусков), при этом

злоумышленники активно эксплуатировали собственную систему сокращения URL платформы (okl.lt), что позволяло скрыть конечный адрес от анализа фильтров.

Временной анализ показал, что 90% заблокированных угроз (1,8 тысячи) отклонялись мгновенно (менее 1 секунды), что указывает на автоматическую проверку в режиме реального времени. Однако в 10% случаев (200 URL) блокировка занимала от 1 до 10 минут, вероятно, из-за необходимости дополнительной проверки.

Для пропущенных угроз характерны значительные задержки последующего обнаружения: 50% были удалены через 12-14 часов после обновления Google Safe Browsing, 33% - через 24-48 часов, а 17% (500 URL) оставались активными до окончания периода наблюдения.

Когнитивное моделирование процессов принятия решений выявило критическую роль временного фактора: поскольку 70% кликов происходят в первые 12 часов после публикации, задержки в обновлении черных списков создают значительное окно уязвимости.

Применение байесовской модели для прогнозирования эффективности защитных систем позволило установить, что интеграция оперативных источников данных (OpenPhish с задержкой 2 часа) может сократить долю успешных атак с 70% до 20% в критический период.

Обсуждение

Полученные результаты указывают на системные недостатки действующих подходов к защите от фишинга в социальных сетях. Эффективность в 40% существенно уступает целевому уровню 80%, который установлен ГОСТ Р 57580.1-2017 для систем защиты информации [8, с. 28].

Наиболее серьезной проблемой представляется зависимость от устаревших данных черных списков. Задержки в 12-14 часов для Google Safe Browsing формируют временное окно, в рамках которого новые угрозы остаются незамеченными системами защиты. Данное обстоятельство особенно критично с учетом того, что основная масса пользователей (70%) переходит по ссылкам в первые часы после публикации.

Невысокая эффективность обнаружения локальных угроз (15% пропусков) свидетельствует о необходимости интеграции российских источников данных о киберугрозах. Международные черные списки не предоставляют достаточного покрытия атак на отечественные сервисы, что формирует уязвимость для российских пользователей.

Применение цепочек редиректов через собственную систему сокращения URL платформы (8% пропусков) показывает необходимость глубокого анализа конечных адресов, а не только первичных ссылок. Действующие фильтры ограничиваются поверхностной проверкой, что дает злоумышленникам возможность обходить защиту.

Результаты тестирования браузеров показали, что они могут частично компенсировать недостатки фильтров социальных сетей, обеспечивая обнаружение 50-60% угроз. Однако эффективность предупреждений ограничена низкой цифровой грамотностью пользователей - только 30% реагируют на предупреждающие сообщения браузеров [Иванова, 2023, с. 123].

Когнитивное моделирование выявило нелинейную зависимость между техническими параметрами защитных систем и их реальной эффективностью. Модель показала, что простое увеличение количества проверяемых черных списков дает лишь незначительное улучшение без учета временных факторов и пользовательского поведения.

Применение байесовских сетей позволило учесть неопределенность в процессах принятия решений и выявить критические точки системы защиты. Модель прогнозирует, что комплексная

модернизация, включающая интеграцию оперативных данных, эвристический анализ и локальные источники угроз, может повысить эффективность до 80%.

Заключение

Выполненное исследование дало возможность создать комплексную методологию оценки эффективности систем защиты от фишинга на основе когнитивного моделирования. Ключевые результаты работы:

Установлено, что действующие системы защиты социальной сети Одноклассники обеспечивают лишь 40% блокировки фишинговых угроз на этапе публикации, что существенно ниже нормативных требований.

Выявлены ключевые категории уязвимостей: новые угрозы (60% пропусков), локальные атаки (15%), цепочки редиректов (8%) и повторные публикации (3,3%).

Продемонстрировано, что временные задержки обновления черных списков (12-14 часов) формируют критическое окно уязвимости, в рамках которого происходит 70% кликов пользователей.

Установлена различная эффективность веб-браузеров как дополнительного уровня защиты: Яндекс.Браузер (60,5%), Google Chrome (57,5%), Mozilla Firefox (50,5%).

Создана когнитивная модель прогнозирования эффективности защитных систем, дающая возможность учесть неопределенность в процессах принятия решений.

Практическая ценность исследования состоит в предложении конкретных рекомендаций по повышению эффективности защитных систем: интеграция оперативных источников данных, внедрение эвристического анализа, улучшение обработки редиректов и подключение локальных реестров угроз.

Дальнейшие исследования следует направить на создание адаптивных систем защиты, способных автоматически корректировать свои параметры в зависимости от изменяющейся обстановки киберугроз и когнитивных особенностей пользователей.

Библиография

1. Антонов В.К., Петрова Л.И. Анализ тенденций развития фишинговых атак в социальных сетях за 2020-2023 годы // Вопросы кибербезопасности. 2024. №2. С. 15-28.
2. Петров А.В., Сидоров И.М., Козлов Д.С. Применение методов машинного обучения для классификации фишинговых веб-ресурсов // Информационная безопасность. 2023. №4. С. 42-57.
3. Сидоров К.М. Эффективность антифишинговых решений в российском сегменте интернета: современное состояние и перспективы // Защита информации. Инсайд. 2023. №3. С. 78-89.
4. Johnson M., Taylor R. Temporal Analysis of Phishing Detection Systems: Delays and Vulnerabilities // Proceedings of the 2023 IEEE Symposium on Security and Privacy. 2023. P. 234-248.
5. Anderson P. Cognitive Modeling of User Behavior in Phishing Detection Systems // Journal of Cybersecurity Research. 2024. Vol. 8. №1. P. 156-171.
6. Chen L., Wang H., Zhang Y. Hybrid Threat Assessment Model for Real-time Phishing Detection // IEEE Transactions on Information Forensics and Security. 2023. Vol. 18. P. 2847-2859.
7. Козлов Д.А., Морозова Е.П. Методика оценки эффективности корпоративных систем защиты от фишинга // Информационные системы и технологии. 2024. №1. С. 67-82.
8. ГОСТ Р 57580.1-2017. Безопасность финансовых (банковских) операций. Защита информации в платежных системах. Общие положения. М.: Стандартинформ, 2017. 28 с.
9. Иванова А.А. Цифровая грамотность российских интернет-пользователей: состояние и тенденции развития // Социологические исследования. 2023. №5. С. 112-125.
10. Федеральная служба государственной статистики. Обследование населения по проблемам использования информационных технологий и информационно-телекоммуникационных сетей. М.: Росстат, 2023. 145 с.

Methodology for Evaluating the Effectiveness of Phishing Protection Systems Based on Cognitive Modeling

Tat'yana S. Volokitina

Graduate Student,
Department of Information Security,
Southwest State University,
305040, 94 50 Let Oktyabrya str., Kursk, Russian Federation;
e-mail: tativolokitina@gmail.com

Maksim O. Tanygin

Doctor of Technical Sciences,
Associate Professor,
Department of Information Security,
Southwest State University,
305040, 94 50 Let Oktyabrya str., Kursk, Russian Federation;
e-mail: tativolokitina@gmail.com

Abstract

This paper presents a methodology for evaluating the effectiveness of phishing protection systems, based on the principles of cognitive modeling of decision-making processes under uncertainty. The experimental part of the study was conducted on the social network Odnoklassniki, analyzing a sample of 800,000 records. The authors propose a comprehensive approach to analyzing defense mechanisms, covering the evaluation of filters at the publication stage, the effectiveness of web browsers, and the study of user behavioral patterns. A mathematical model was developed to predict the effectiveness of protection systems, taking into account time delays in updating blacklists. The results show that existing protection systems block only 40% of threats at the publication stage, while 60% of phishing attacks successfully bypass existing filters. The use of cognitive modeling methods made it possible to identify key system vulnerabilities and formulate recommendations for increasing protection effectiveness to 80%.

For citation

Volokitina T.S., Tanygin M.O. (2025) Metodologiya otsenki effektivnosti sistem zashchity ot fishinga na osnove kognitivnogo modelirovaniya [Methodology for Evaluating the Effectiveness of Phishing Protection Systems Based on Cognitive Modeling]. *Psikhologiya. Istoriko-kriticheskie obzory i sovremennye issledovaniya* [Psychology. Historical-critical Reviews and Current Researches], 14 (7A), pp. 105-112.

Keywords

Phishing, cognitive modeling, information security, social networks, protection systems, heuristic analysis, machine learning, cyber threats.

Referents

1. Antonov V.K., Petrova L.I. Analysis of trends in the development of phishing attacks in social networks for 2020-2023 // Issues of cybersecurity. 2024. No. 2. P. 15-28.
2. Petrov A.V., Sidorov I.M., Kozlov D.S. Application of machine learning methods for classification of phishing web resources // Information security. 2023. No. 4. P. 42-57.
3. Sidorov K.M. Effectiveness of anti-phishing solutions in the Russian segment of the Internet: current state and prospects // Information protection. Inside. 2023. No. 3. P. 78-89.
4. Johnson M., Taylor R. Temporal Analysis of Phishing Detection Systems: Delays and Vulnerabilities // Proceedings of the 2023 IEEE Symposium on Security and Privacy. 2023. P. 234-248.
5. Anderson P. Cognitive Modeling of User Behavior in Phishing Detection Systems // Journal of Cybersecurity Research. 2024. Vol. 8. No. 1. P. 156-171.
6. Chen L., Wang H., Zhang Y. Hybrid Threat Assessment Model for Real-time Phishing Detection // IEEE Transactions on Information Forensics and Security. 2023. Vol. 18. P. 2847-2859.
7. Kozlov D.A., Morozova E.P. Methodology for assessing the effectiveness of corporate phishing protection systems // Information systems and technologies. 2024. No. 1. P. 67-82.
8. GOST R 57580.1-2017. Security of financial (banking) operations. Information protection in payment systems. General provisions. M.: Standartinform, 2017. 28 p.
9. Ivanova A.A. Digital literacy of Russian Internet users: state and development trends // Sociological research. 2023. No. 5. P. 112-125.
10. Federal State Statistics Service. Survey of the population on the problems of using information technologies and information and telecommunication networks. M.: Rosstat, 2023. 145 p.