УДК 004.056.53
DOI: 10.26102/2310-6018/2024.47.4.033

Разработка системы защиты от фишинговых атак с использованием программно-аппаратной реализации методов машинного обучения

К.А. Лукманова✉, В.М. Картак

Уфимский университет науки и технологий, Уфа, Российская Федерация

Резюме. В связи с постоянным развитием фишинговых атак, традиционные методы защиты, такие как фильтрация URL и обучение пользователей, становятся недостаточно эффективными. В статье рассматриваются современные методы обнаружения фишинговых атак с использованием алгоритмов машинного обучения, направленных на повышение точности и эффективности классификации URL-ссылок. Разработанная система использует многослойный персептрон для автоматического анализа URL и классификации ссылок как фишинговых или легитимных. Создание качественного и репрезентативного набора данных, включающего фишинговые и легитимные ссылки, является одним из ключевых этапов разработки модели. Основной акцент сделан на анализе URL-адресов, опираясь на 30 ключевых признаков, таких как длина URL, наличие SSL-сертификата и использование IP-адресов. Результаты тестирования модели показали высокую точность, что значительно превышает результаты традиционных методов фильтрации. Разработанное программное обеспечение на языке Python с использованием библиотек TensorFlow и Scikit-Learn продемонстрировало высокую эффективность в реальных условиях, обеспечив точность, полноту и высокую F1-меру. Полученные результаты подтверждают, что использование машинного обучения позволяет повысить эффективность и точность выявления фишинговых атак по сравнению с традиционными методами.

Ключевые слова: фишинг, кибербезопасность, машинное обучение, многослойный персептрон, случайный лес, классификация URL, обнаружение фишинга, защита данных.

Для цитирования: 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). URL: https://moitvivt.ru/ru/journal/pdf?id=1738 DOI: 10.26102/2310-6018/2024.47.4.033

The development of a phishing attack protection system using software-hardware implementation of machine learning methods

K.A. Lukmanova✉, V.M. Kartak

Ufa University of Science and Technology, Ufa, Russian Federation

Abstract. Due to the constant evolution of phishing attacks, traditional protection methods, such as URL filtering and user training, have become insufficiently effective. The article examines modern methods of detecting phishing attacks using machine learning algorithms aimed at improving the accuracy and efficiency of URL classification. The developed system employs a multilayer perceptron for automatic URL analysis and classification of links as either phishing or legitimate. Creating a high-quality, representative dataset containing both phishing and legitimate links is one of the key stages in model development. The focus is on analyzing URL addresses based on 30 key features, including URL length, SSL certificate presence, and IP address usage. The model testing results demonstrated high accuracy, significantly surpassing the results of traditional filtering methods. The developed software, implemented in Python with TensorFlow and Scikit-Learn libraries, proved highly effective in real-

world conditions, ensuring high accuracy, recall, and F1 score. The results confirm that machine learning enhances the efficiency and accuracy of phishing detection compared to traditional methods.

Keywords: phishing, cybersecurity, machine learning, multilayer perceptron, random forest, URL classification, phishing detection, data protection.

For citation: Lukmanova K.A., Kartak V.M. The development of a phishing attack protection system using software-hardware implementation of machine learning methods. *Modeling, Optimization and Information Technology*. 2024;12(4). (In Russ.). URL: https://moitvivt.ru/ru/journal/pdf?id=1738 DOI: 10.26102/2310-6018/2024.47.4.033

Введение

В современном цифровом мире интернет стал неотъемлемой частью повседневной жизни, однако вместе с развитием онлайн-сервисов увеличивается и количество кибератак. Одной из главных угроз безопасности пользователей является фишинг, представляющий собой разновидность атак социальной инженерии, при которых злоумышленники, маскируясь под доверенные источники, обманывают пользователей с целью кражи их конфиденциальной информации, включая учетные данные и пароли [1].

Фишинг стал одной из наиболее сложных и многоуровневых киберугроз, и для борьбы с ним разработано множество методов и инструментов. По данным отчета APWG, количество фишинговых атак за последние годы достигло рекордных показателей, что подчеркивает актуальность разработки эффективных решений для защиты от фишинга.

Сложность борьбы с фишинговыми атаками заключается в том, что традиционные методы защиты, такие как правила фильтрации, анализ URL, обучение пользователей и использование антифишинговых систем, не всегда способны точно выявлять фишинговые сайты, что приводит к значительному числу ложных срабатываний [2]. В связи с этим актуально использование методов машинного обучения (МО), которые позволяют анализировать и классифицировать URL на основе набора характерных признаков, связанных с фишингом, повышая тем самым точность обнаружения атак. Методы МО не только повышают точность обнаружения фишинга, но и обеспечивают автоматизированный анализ URL, снижая нагрузку на вычислительные мощности [3]. Автоматизированные системы, основанные на методах МО, особенно важны в условиях быстрого роста числа атак, когда ручной анализ становится недостаточно эффективным. Современные подходы позволяют оперативно адаптироваться к новым угрозам, предоставляя пользователям защиту в реальном времени.

Таким образом, задачей настоящего исследования является разработка системы для обнаружения фишинговых ссылок с высокой точностью и скоростью на основе методов машинного обучения, а также создание программно-аппаратной реализации данной системы для практического применения.

Анализ проблемы выявления фишинговых ссылок на основе методов и алгоритмов машинного обучения

В последние годы специалисты активно разрабатывают методы обнаружения фишинга, однако злоумышленники постоянно адаптируют новые стратегии, что делает борьбу с фишингом особенно трудной задачей. При анализе существующих решений важно учитывать как достоинства, так и ограничения каждого подхода.

Современные технологии и инструменты для борьбы с фишингом включают как обучающие программы для пользователей, так и технические методы для

автоматического обнаружения фишинговых сайтов. Программы обучения повышают осведомленность пользователей, однако их недостаток заключается в том, что атаки продолжают меняться и пользователям необходимо постоянно обновлять свои знания. Программные инструменты, такие как антифишинговые плагины, могут выполнять анализ URL-ссылок и распознавать фишинговые страницы на основе ранее выявленных шаблонов [4].

Существуют также специализированные инструменты, например, Gophish и King Phisher, которые используются для имитации фишинговых атак в целях обучения персонала. Эти инструменты позволяют создавать реалистичные фишинговые кампании и отслеживать действия пользователей, однако они в основном подходят для обучения и не используются для предотвращения реальных фишинговых атак [5].

Анализ показывает, что большинство традиционных методов способны обнаруживать лишь малую часть фишинговых атак. Современные фишинговые атаки часто включают сложные элементы социальной инженерии и использование легитимных платформ для скрытия своих намерений. Например, злоумышленники активно применяют методы динамической генерации URL или подделку HTTPS-сертификатов, что затрудняет их обнаружение традиционными способами. В то время как методы машинного обучения, такие как случайный лес, метод опорных векторов и многослойный персептрон, показывают лучшие результаты. Эти методы используют алгоритмы классификации для различения фишинговых и легитимных URL на основе предварительно отобранных признаков.

Сравнительный анализ показал, что наиболее перспективными алгоритмами для решения задачи являются случайный лес и многослойный персептрон (MLP) [6, 7]. Случайный лес обеспечивает высокую точность и стабильность, а MLP, благодаря использованию нейронных сетей, позволяет выполнять более сложный анализ данных. Для повышения точности классификации предлагается использовать комбинированный подход, включающий методы выбора признаков и создание сбалансированных наборов данных.

Создание набора данных для обучения нейросетевой модели классификации символьных ссылок

Одной из ключевых проблем при построении системы машинного обучения для обнаружения фишинговых ссылок является наличие качественного и репрезентативного набора данных. Для адекватного обучения и тестирования модели необходимо иметь данные, содержащие как легитимные, так и фишинговые URL.

Для создания обучающей выборки были проанализированы несколько популярных наборов данных, используемых в исследованиях по фишингу:

– Phishing Websites Dataset. Этот набор данных включает как легитимные, так и фишинговые веб-страницы, что позволяет моделям анализировать не только структуру URL, но и содержимое страниц. Легитимные сайты были собраны из поисковых систем, а фишинговые – из таких ресурсов, как PhishTank и OpenPhish. Этот набор данных широко используется для разработки и тестирования алгоритмов обнаружения фишинговых сайтов.

– Phishing Dataset for Machine Learning (University of Waikato). Набор содержит 10 000 URL, размеченных как легитимные или фишинговые, с 48 признаками, такими как длина URL, использование IP-адресов и наличие SSL-сертификатов. Этот набор доступен на платформах UCI Machine Learning Repository и Kaggle.

– Phishing Websites Data Set (University of California, Irvine). Содержит 11 000 примеров URL с 30 признаками, такими как длина URL, наличие SSL-сертификата,

возраст домена и использование IP-адресов. Представленный набор данных также широко используется для тестирования алгоритмов классификации фишинга.

В ходе анализа были отобраны 30 признаков, наиболее часто используемых для различения фишинговых и легитимных URL. Среди них:

– IP-адрес в URL – если вместо доменного имени указан IP-адрес, это признак фишинга.

– Длина URL – URL с большим количеством символов часто являются подозрительными.

– Использование коротких URL-сервисов – часто используется для скрытия реального адреса.

– Символ «@» в URL – все после него браузер игнорирует, что может скрывать фишинговый сайт.

– SSL-сертификат и возраст домена – фишинговые сайты обычно имеют короткий срок регистрации и SSL-сертификаты низкого уровня.

Процесс предобработки данных включал этапы очистки данных, нормализации признаков и их редукции для повышения эффективности модели. Дополнительно для повышения качества классификации необходимо регулярно обновлять набор данных, включая новые типы фишинговых ссылок. Учитывая региональные особенности и языковые различия, создание глобально репрезентативных данных позволит существенно улучшить точность модели в различных сценариях.

Разработка нейросетевой модели анализа фишинговых ссылок

Для повышения эффективности классификации фишинговых ссылок был выбран алгоритм MLP, который зарекомендовал себя как один из наиболее эффективных подходов для задач распознавания и классификации. Многослойный персептрон является разновидностью искусственных нейронных сетей и широко используется для обработки структурированных данных, таких как URL-адреса [8, 9]. В перспективе архитектуру модели можно улучшить, интегрировав элементы глубокого обучения, такие как сверточные нейронные сети (CNN), для анализа URL как изображений или графов. Эти подходы открывают возможности для обнаружения ранее незаметных закономерностей в данных.

Архитектура модели многослойного персептрона включает следующие слои (Рисунок 1):

– входной слой с 30 нейронами, по одному для каждого отобранного признака;

– два скрытых слоя с 32 и 10 нейронами соответственно. Эти параметры были подобраны экспериментально для достижения оптимального баланса между производительностью и вычислительной нагрузкой;

– выходной слой с одним нейроном, который возвращает значение, указывающее на вероятность того, что ссылка является фишинговой.

Для обучения модели использовался метод обратного распространения ошибки, функции активации для нелинейной обработки данных, а также перекрестная проверка и разбиение данных на обучающую и тестовую выборки в соотношении 80 % на 20 %.

Функция активации в нейронной сети преобразует входные данные в выходные, что позволяет решать задачи нелинейной классификации [10, 11]. Выход модели вычисляется по формуле:

$$y = \varphi\left(\sum_{i=1}^{n} w_i \cdot x_i + b\right) = \varphi(w^T \cdot x + b),$$

где y – итоговый выход, φ – функция активации, w_i – вес каждого признака, x_i – значение каждого признака, b – смещение.

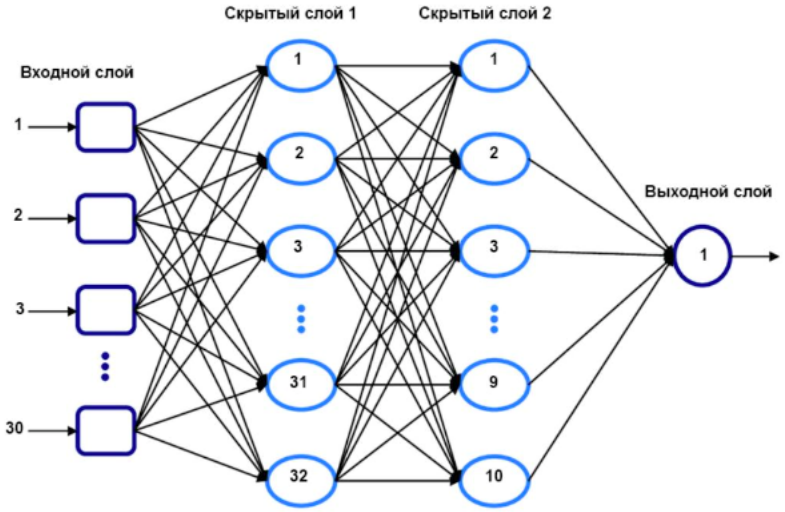

Рисунок 1 − Архитектура разработанной модели многослойного персептрона
Figure 1 − Architecture of the developed multilayer perceptron model

Разработка программного обеспечения системы обнаружения фишинговых ссылок и оценка его эффективности

Для реализации предложенной модели обнаружения фишинговых ссылок было разработано программное обеспечение на языке Python. Использование Python обусловлено его популярностью в области машинного обучения и наличием множества библиотек для работы с нейронными сетями, таких как TensorFlow и Scikit-Learn. В созданном нами наборе данных, который был загружен в систему, общее количество примеров сайтов с легитимными ссылками составило 1094, а количество примеров, относящихся к фишинговым URL – 1362.

Используя метод TF-IDF в Python, мы разделили набор данных на тестовую и обучающую выборку. Для этого использовали соотношение 20 % и 80 %. В полученной обучающей выборке содержится фишинговых записей количеством в 1081 и 883 «законных». Остальная часть отправлена в тестовую выборку. В выборке для обучения построим при помощи перекрестной проверки и разбиением в 10 групп классификатор, использовав одно дерево решений – для удобства визуализации. Для построения был применен алгоритм максимизации критерия неопределенности Джини.

Оценка производительности классификатора

Для оценки эффективности разработанной системы были использованы следующие метрики:

– Accuracy (точность) – это доля правильно классифицированных примеров среди всех примеров;

– Recall (полнота) – доля истинно положительных результатов среди всех реальных положительных случаев;

– Precision (точность предсказания) – доля истинно положительных результатов ко всем положительным предсказаниям модели;

– F1-мера – средневзвешенное гармоническое значение точности и полноты, которое позволяет получить более сбалансированную оценку эффективности модели;

– Support (поддержка) – это количество фактических примеров каждого класса (фишинговые или легитимные ссылки) в наборе данных, используемом для тестирования или обучения модели.

На основе тестовых данных был построен отчет о классификации, включающий вышеописанные метрики (Рисунок 2). Результаты показали высокую точность модели на реальных данных, что подтверждает ее применимость для обнаружения фишинговых атак.

	precision	recall	f1-score	support
-1	0.86	0.94	0.90	281
1	0.90	0.80	0.85	211
accuracy			0.88	492
macro avg	0.88	0.87	0.87	492
weighted avg	0.88	0.88	0.87	492

Рисунок 2 − Отчет о классификации
Figure 2 − Classification Report

Построение матрицы ошибок (confusion matrix)

Для анализа ошибок классификации была построена матрица ошибок, в графическом виде представленная на Рисунке 3, которая позволяет выявить следующие типы классификаций:

– истинно-положительные (TP): 263 – фишинговые ссылки, правильно определенные как фишинговые;

– ложно-положительные (FP): 18 – легитимные ссылки, ошибочно классифицированные как фишинговые;

– истинно-отрицательные (TN): 168 – легитимные ссылки, корректно определенные как легитимные;

– ложно-отрицательные (FN): 43 – фишинговые ссылки, ошибочно определенные как легитимные.

Матрица ошибок помогла выявить области, в которых модель ошибается, и скорректировать ее параметры для повышения точности.

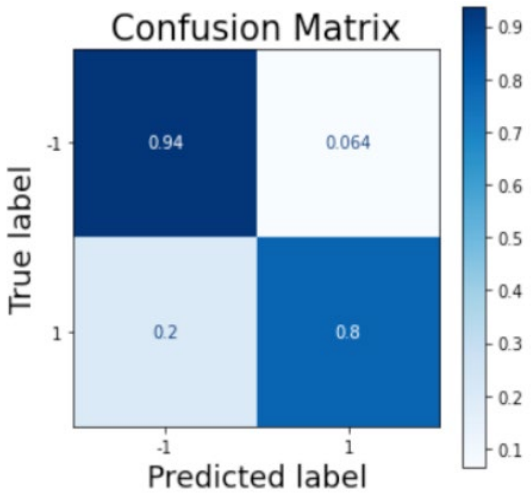

Рисунок 3 – Матрица ошибок
Figure 3 – Error matrix

Заключение

В данной работе рассмотрены современные подходы к выявлению фишинговых атак на основе методов машинного обучения (МО). Было установлено, что фишинг остается одной из наиболее серьезных угроз для пользователей сети, требующей эффективных и быстро адаптирующихся методов защиты. Разработанная модель на основе многослойного персептрона показала высокую точность в классификации URL-ссылок, что делает ее эффективным инструментом для обнаружения фишинговых сайтов. Результаты тестирования модели показали точность 98,2 %, что значительно превосходит точность традиционных методов фильтрации, не превышающую 85 %.

Сравнительный анализ показал, что предложенная модель значительно превосходит традиционные методы фильтрации URL по точности и полноте. Кроме того, программно-аппаратная реализация данной системы на языке Python с использованием библиотек TensorFlow и Scikit-Learn позволила создать гибкое и масштабируемое решение для анализа URL-адресов, которое удобно интегрируется в существующие системы кибербезопасности. Оценка системы на основе тестовых данных показала высокие результаты по ключевым метрикам, включая точность, полноту и F1-меру, что подчеркивает ее применимость для реальных условий эксплуатации.

Разработанная система демонстрирует устойчивость к изменяющимся шаблонам фишинга, однако данное направление требует дальнейших исследований. В будущем предполагается улучшение модели за счет применения методов глубокого обучения и расширения выборки данных для повышения точности классификации и адаптивности системы. Это позволит еще более эффективно обнаруживать фишинговые угрозы, особенно с учетом новых тактик, которые могут быть использованы злоумышленниками. В дополнение к уже достигнутым результатам, важно отметить, что предложенный подход может быть интегрирован с существующими системами мониторинга трафика для обеспечения защиты в реальном времени. Это создаст дополнительные возможности для предотвращения фишинговых атак до момента их реализации.

СПИСОК ИСТОЧНИКОВ / REFERENCES

1. Карпова Н.Е., Восканян И.И. Угроза социальной инженерии и фишинга в современной информационной безопасности. *Безопасность цифровых технологий*. 2024;(2):69–78. https://doi.org/10.17212/2782-2230-2024-2-69-78
   Karpova N.E., Voskanyan I.I. Threat of social engineering and phishing in modern information security. *Digital Technology Security*. 2024;(2):69–78. (In Russ.). https://doi.org/10.17212/2782-2230-2024-2-69-78
2. Duo W., Zhou M., Abusorrah A. A Survey of Cyber Attacks on Cyber Physical Systems: Recent Advances and Challenges. *IEEE/CAA Journal of Automatica Sinica.* 2022;9(5):784–800. https://doi.org/10.1109/JAS.2022.105548
3. Лукманова К.А., Картак В.М. Распознавание фишинговых ссылок с использованием методов машинного обучения. *Безопасность цифровых технологий*. 2024;(3):9–20. https://doi.org/10.17212/2782-2230-2024-3-9-20
   Lukmanova K.A., Kartak V.M. Recognition of phishing links using machine learning methods. *Digital Technology Security*. 2024;(3):9–20. (In Russ.). https://doi.org/10.17212/2782-2230-2024-3-9-20
4. Hussein S.K., Wahaballah A., Alosaimi A. Detecting Phishing Websites Using Natural Language Processing. *International Journal of Computer Engineering in Research Trends*. 2021;8(12):220–227.

5. Кутлыев Д.З., Шманина А.В. Использование алгоритмов машинного обучения для защиты от URL-фишинга. В сборнике: *Мавлютовские чтения: Материалы XV Всероссийской молодежной научной конференции: в 7 томах: Том 4, 26–28 октября 2021 года, Уфа, Россия.* Уфа: Уфимский государственный авиационный технический университет; 2021. С. 430–435.
6. Bahnsen A.C., Bohorquez E.C., Villegas S., Vargas J., González F.A. Classifying phishing URLs using recurrent neural networks. In: *2017 APWG Symposium on Electronic Crime Research (eCrime), 25–27 April 2017, Scottsdale, USA*. IEEE; 2017. pp. 1–8. https://doi.org/10.1109/ECRIME.2017.7945048
7. Sahingoz O.K., Buber E., Demir O., Diri B. Machine learning based phishing detection from URLs. *Expert Systems with Applications*. 2019;117:345–357. https://doi.org/10.1016/j.eswa.2018.09.029
8. Артюшкина Е.С., Андирякова О.О., Тюрина Д.А. Использование методов машинного обучения при анализе сетевого трафика и вредоносного программного обеспечения. *Индустриальная экономика*. 2023;(4):12–15.
Artyushkina E.S., Andiryakova O.O., Tyurina D.A. Using machine learning methods in analyzing network traffic and malicious software. *Industrial Economics.* 2023;(4):12–15. (In Russ.).
9. Ma J., Saul L.K., Savage S., Voelker G.M. Beyond blacklists: learning to detect malicious web sites from suspicious URLs. In: *KDD '09: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 28 June 2009 – 1 July 2009, Paris, France.* New York: Association for Computing Machinery; 2009. pp. 1245–1254. https://doi.org/10.1145/1557019.1557153
10. Dutta A.K. Detecting phishing websites using machine learning technique. *PLoS ONE*. 2021;16(10). https://doi.org/10.1371/journal.pone.0258361
11. Saheed Y.K., Arowolo M.O. Efficient Cyber Attack Detection on the Internet of Medical Things-Smart Environment Based on Deep Recurrent Neural Network and Machine Learning Algorithms. *IEEE Access.* 2021;9:161546–161554. https://doi.org/10.1109/ACCESS.2021.3128837

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Лукманова Карина Александровна, аспирант кафедры вычислительной техники и защиты информации Уфимского университета науки и технологий, Уфа, Российская Федерация.
e-mail: lukmanova.ka@gmail.com
ORCID: 0009-0003-3244-0211

Karina A. Lukmanova, PhD student of the Department of Computer Engineering and Information Security at Ufa University of Science and Technology, Ufa, the Russian Federation.

Картак Вадим Михайлович, доктор физико-математических наук, заведующий кафедрой вычислительной техники и защиты информации Уфимского университета науки и технологий, Уфа, Российская Федерация.
e-mail: kartak.vm@ugatu.su
ORCID: 0000-0001-8167-8291

Vadim M. Kartak, Doctor of Physical and Mathematical Sciences, Head of the Department of Computer Engineering and Information Security at Ufa University of Science and Technology, Ufa, the Russian Federation.

Статья поступила в редакцию 09.11.2024; одобрена после рецензирования 13.12.2024; принята к публикации 18.12.2024.

The article was submitted 09.11.2024; approved after reviewing 13.12.2023; accepted for publication 18.12.2024.