представляет собой гармоническое среднее точности (доля корректных идентификаторов среди выданных) и полноты (доля найденных идентификаторов среди эталонных). Метрика RAGAS Faithfulness оценивает долю утверждений в ответе, подкрепленных информацией из контекста; низкие значения указывают на присутствие «галлюцинаций».

Прирост EM для всех комбинаций, что подтверждает устойчивость преимущества GraphRAG с межонтологическими связями. Влияние выбора embedding-модели превышает влияние выбора LLM, что указывает на критичность качества retrieval для GraphRAG-систем. Комбинация Gemma3:27B и BGE-M3 обеспечила наилучшие абсолютные показатели.

Практическая значимость определяется возможностью применения разработанной системы для поддержки принятия решений аналитиками центров мониторинга информационной безопасности при разборе TTP-отчетов, формировании рекомендаций по мерам защиты в соответствии с требованиями российского законодательства и атрибуции инцидентов с использованием международных таксономий.

Список использованных источников

1.Котенко И.В., Абраменко Г.Т. Использование больших языковых моделей для поиска угроз кибербезопасности на основе методов глубокого обучения: анализ современных исследований // Правовая информатика. 2024. №. 3. С. 32-42.

2.Lewis P. et al. Retrieval-augmented generation for knowledge-intensive NLP tasks // Advances in neural information processing systems. 2020. Vol. 33. P. 9459-9474.

3.Singh J.P., Agrawal S. Automating threat intelligence analysis with retrieval augmented generation rag for enhanced cybersecurity posture // International Journal of Science and Research (IJSR). 2024. Vol. 13, no. 5. P. 251-255.

4.Rajapaksha S., Rani R., Karafili E. A RAG-based question-answering solution for cyber-attack investigation and attribution // European Symposium on Research in Computer Security. 2024. P. 238-256.

Луцкович Андрей Иванович, ст. препод. каф. ВТиЗИ, andrey.lutskovich@gmail.com.
Вульфин Алексей Михайлович, д.т.н., профессор каф. ВТиЗИ, vulfin.alexey@gmail.com.
Кириллова Анастасия Дмитриевна, к.т.н., доцент каф. ВТиЗИ, kirillova.ad@ugatu.su.

УДК 004.7.056; 004.7:004.8

МУЛЬТИМОДАЛЬНАЯ НЕЙРОСЕТЕВАЯ СИСТЕМА ОБНАРУЖЕНИЯ ФИШИНГОВЫХ ВЕБ-САЙТОВ

А.В. Минко, А.М. Вульфин, А.Д. Кириллова
Уфимский университет науки и технологий, г. Уфа

Ключевые слова: фишинг, мультимодальное обнаружение, позднее слияние, объяснимый искусственный интеллект, кибербезопасность

Традиционные методы обнаружения (черные списки, статический

анализ URL) не способны своевременно выявлять новые фишинговые кампании и приводят к значительному числу ложных срабатываний [1, 2]. Перспективным является использование методов машинного обучения, которые позволяют классифицировать URL в пространстве характерных признаков. Решение задачи обнаружения фишинговых веб-сайтов при этом сводится к задаче бинарной классификации: определить принадлежность URL к легитимному или фишинговому классу. Современные подходы на основе глубокого обучения (CNN, LSTM, трансформеры) и мультимодального анализа [3-5] позволяют привлекать дополнительные источники информации: HTML-код, визуальный образ страницы, метаданные домена, что существенно повышает достоверность классификации. Подобные модели могут быть отнесены к классу мультимодальных, где под модальностью понимается тип входных данных и способ извлечения признаков для последующей обработки в модели машинного обучения.

Цель работы – разработка мультимодальной нейросетевой системы обнаружения фишинговых веб-ресурсов, которая.

Система включает восемь подсистем: 1) предобработка данных и обогащение контекста; 2) создание и управление обучающими данными; 3) создание и управление ML-моделями; 4) управление жизненным циклом моделей; 5) классификация и принятие решений; 6) объяснение решений; 7) интеграция с Threat Intelligence и SIEM; 8) валидация решений аналитиками SOC.

На вход системы поступает URL анализируемого веб-ресурса, выполняется загрузка HTML заглавной страницы и создание скриншота. Подсистема предобработки извлекает более 89 признаков URL (лексические, синтаксические, семантические, фонетические, статистические), собирает метаданные (WHOIS/RDAP, SSL/TLS, репутационные базы VirusTotal, Google Safe Browsing) и формирует мультимодальный контекст для четырех ветвей анализа.

Модели M1-M4 обучаются независимо и выдают вероятности принадлежности к классу «фишинг». На втором уровне обучаемый мета-классификатор M5 реализует позднее слияние по схеме стекинга [6].

В заключительном эксперименте полный мультимодальный контур (M1-M4 + мета-классификатор M5) достигает F1 = 0,989 на собственном наборе D1 и F1 = 0,953 на публичном наборе данных MTLP.

Научная новизна заключается в одновременном сочетании мультимодального позднего слияния четырех гетерогенных модальностей с обучаемым мета-классификатором и локальной LLM для генерации

объяснений, а также промышленной интеграции с SOC/TI/SIEM.

Состав ML-моделей представлен в таблице 1.

Таблица 1 – Состав ML-моделей системы

№	Модальность	Модель	Описание
1	URL + Метаданные	M1: CatBoost	Подбор гиперпараметров через Optuna, k-cross-validation (k = 5). ROC-AUC = 0,924. Компиляция в C через LLVM (+26% производительности).
2	Символьный URL	M2: CNN1D	Три параллельных сверточных слоя (ядра 3, 4, 5), 64 фильтра, Dropout 0,5. F1 = 0,914 на MTLP.
3	HTML-код	M3: CodeBERT	Fine-tuning предобученного CodeBERT, токенизация до 512 токенов. F1 = 0,950 на MTLP.
4	Скриншот	M4: EfficientNet-B7	Fine-tuning на ImageNet, BCEWithLogitsLoss, AdamW, OneCycleLR. F1 = 0,932 на MTLP.
5	Позднее слияние	M5: Мета-классификатор	Стекинг (stacked generalization): линейный классификатор второго уровня, оптимизация F1 для класса “фишинг”.

Направления дальнейших исследований включают: переход к гибридному слиянию с кросс-модальным вниманием для пар модальностей «HTML + скриншот»; систематическое исследование устойчивости к целенаправленным adversarial-атакам; расширение RAG-подсистемы; применение federated learning для совместной защиты с сохранением приватности данных.

Список использованных источников

1.Dutta A.K. Detecting phishing websites using machine learning technique // PloS one. 2021. Vol. 16(10). e0258361.

2.Basit A., Zafar M., Liu X. et al. A comprehensive sur-vey of AI-Enabled phishing attacks detection techniques // Telecommunication Systems. 2021. Vol. 76(1). P. 139-154.

3.Кротов Е.Ю. Применение методов глубокого обучения для обнаружения фишинговых веб-сайтов // Актуальные исследования. 2025. № 21 (256).

4.Alsaedi M. et al. Cyber threat intelligence-based malicious URL detection model using ensemble learning // Sensors. 2022. Vol. 22(9). 3373.

5.Wangchuk T., Gonsalves T. Multimodal Phishing Detection on Social Networking Sites: A Systematic Review // IEEE Access. 2025.

6.Younis E.M.G. et al. Evaluating ensemble learning methods for multi-modal emotion recognition // Sensors. 2022. Vol. 22(15). 5611.

Минко Александр Васильевич, аспирант каф. ВТиЗИ, minko.av@ugatu.su.
Вульфин Алексей Михайлович, д.т.н., профессор каф. ВТиЗИ, vulfin.alexey@gmail.com.
Кириллова Анастасия Дмитриевна, к.т.н., доцент каф. ВТиЗИ, kirillova.ad@ugatu.su.