**Левин Илья Израилевич** – Южный федеральный университет; e-mail: levin@superevm.ru; г. Таганрог, Россия; тел.: +78634612111; кафедра интеллектуальных и многопроцессорных систем; зав. кафедрой; д.т.н.; профессор.

**Дудников Евгений Александрович** – e-mail: everlast-83@mail.ru; тел.: +79185914907; кафедра интеллектуальных и многопроцессорных систем; аспирант.

**Levin Ilya Izrailevich** – Southern Federal University; e-mail: levin@superevm.ru; Taganrog, Russia; phone: +78634612111; the Department of Intellectual and Multiprocessor Systems; head of the department; dr. of eng. sc.; professor.

**Dudnikov Evgeny Alexandrovich** – e-mail: everlast-83@mail.ru; phone: +79185914907; the Department of Intellectual and Multiprocessor Systems; graduate student.

УДК 004.056 DOI 10.18522/2311-3103-2024-5-58-68

**С.Ю. Мельников, Р.В. Мещеряков, В.А. Пересыпкин**

## НЕКОТОРЫЕ АСПЕКТЫ ПРИМЕНЕНИЯ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ЗАДАЧАХ ЗАЩИТЫ ИНФОРМАЦИИ (ОБЗОР)

*Технологии искусственного интеллекта (ИИ) являются одной из наиболее динамично развивающихся областей обработки информации. Технологии ИИ используются как для обеспечения защиты информации, так и для организации атак на средства ее защиты. Сами системы ИИ могут содержать уязвимости и быть подвержены атакам различного рода. В статье анализируются некоторые аспекты применения технологий ИИ в задачах защиты информации. В рамках задачи биометрической идентификации рассматриваются угрозы подделки биометрических идентификационных признаков с целью получения прав доступа, и способы противодействия таким угрозам. Анализируются преимущества использования ИИ при защите информации в компьютерных системах и сетях по сравнению с традиционными средствами защиты. На примере акустического канала утечки информации от клавиатуры иллюстрируется использование технологий ИИ для обработки данных из технических каналов утечки. Рассматриваются методы повышения информативности таких каналов, использующие временные сверточные сети и модели классификации изображений, а также способы противодействия им. Отдельное внимание уделено вопросам информационной безопасности в набирающих популярность системах сжатия и передачи информации без значительных смысловых потерь (направление Semantic Communications). Рассматриваются ряд вопросов информационной безопасности, возникающих при использовании больших языковых моделей типа ChatGPT, способных массово генерировать уникальный «человекоподобный» контент и использовать его для организации фишинговых и других атак социальной инженерии. Описана атака на системы ИИ с использованием скрытого канала. Уделено внимание необходимости развития технологий доверенного искусственного интеллекта.*

*Информационная безопасность; кибербезопасность; технический канал утечки; искусственный интеллект; доверенный искусственный интеллект.*

**S.Yu. Melnikov, R.V. Meshcheryakov, V.A. Peresypkin**

## SOME ASPECTS OF APPLICATION OF ARTIFICIAL INTELLIGENCE TECHNOLOGIES IN INFORMATION SECURITY (REVIEW)

*Artificial intelligence (AI) technologies are one of the most dynamically developing areas of information processing. AI technologies are used both to ensure the information security and to organize attacks on information security tools. AI systems themselves may contain vulnerabilities and be susceptible to various types of attacks. The article analyzes some aspects of the use of AI technologies in information security tasks. Within the framework of the task of biometric identification, threats of falsification of biometric identification characteristics in order to obtain access rights, and ways to counter such threats are considered. The advantages of using AI in protecting information in computer systems and networks in comparison with traditional means of protection are analyzed. Using the example of an acoustic channel of information leakage from a keyboard, the use of AI technologies for processing data from technical leakage channels is illustrated. Methods for increasing the information content of such channels using*

*temporary convolutional networks and image classification models, as well as ways to counter them, are considered. Special attention is paid to information security issues in increasingly popular systems for compressing and transmitting information without significant semantic losses (Semantic Communications). A number of information security issues that arise when using large language models such as ChatGPT, capable of massively generating unique "human-like" content and using it to organize phishing and other social engineering attacks, are considered. An attack on AI systems using a covert channel is described. Attention is paid to the need to develop trusted artificial intelligence technologies.*

*Information security; cybersecurity; technical leakage channel; artificial intelligence; trusted artificial intelligence.*

**Введение.** Под искусственным интеллектом понимается комплекс технологических решений, позволяющий имитировать когнитивные функции человека и получать при выполнении конкретных задач результаты, сопоставимые с результатами интеллектуальной деятельности человека или превосходящие их [1].

Технологии ИИ являются динамично развивающийся областью обработки информации, включающей в себя обработку естественного языка, компьютерное зрение, распознавание и синтез речи, интеллектуальную поддержку принятия решений и др. К основным задачам, которые могут решаться с использованием методов ИИ, относятся: классификация, кластеризация, распознавание образов, обнаружение аномалий, прогнозирование, обработка естественного языка, инженерия знаний, создание экспертных систем.

Стремительное развитие технологий ИИ, которые используются как для обеспечения защиты информации, так и для организации атак на средства ее защиты, в ближайшее время приведет к значительным изменениям ландшафта индустрии информационной безопасности. Кроме того, сами системы ИИ могут содержать уязвимости и быть подвержены атакам различного рода.

**Угрозы безопасности информации, создаваемые с использованием технологий ИИ, и способы противодействия им**

1. Первые методы ИИ разрабатывались для решения идентификационных задач, связанных с обработкой речи и текста [2]. В настоящее время биометрическая идентификация используется для управления доступом, например, в современных мобильных телефонах и планшетах.

К классическим угрозам, реализуемым с использованием ИИ, можно отнести [3] подделку биометрических идентификационных признаков с целью получения прав доступа путем формирования идентификационных признаков другого лица. К возможным современным угрозам следует отнести формирование ложных видео/речевых/текстовых сообщений, имитирующих конкретного человека; создание ложных фото и видео с участием конкретных лиц; подделку почерка; подделку авторского стиля текстов и др. Способы создания фальшивого контента приведены на рис. 1, результаты их применения – на рис. 2 [4].


*Рис. 1. Способы создания фальшивого контента.*

Угрозы биометрическим системам с использованием подходов на основе ИИ предполагают нарушение не только конфиденциальности и доступности информации, но и ее целостности, т.к. может производиться подмена информации. Помимо подделки внешних идентификационных признаков, которые различимы, возможна подделка более сложных «внутренних» признаков [4], связанных с социальными и иными поведенческими реакциями (рис. 2).

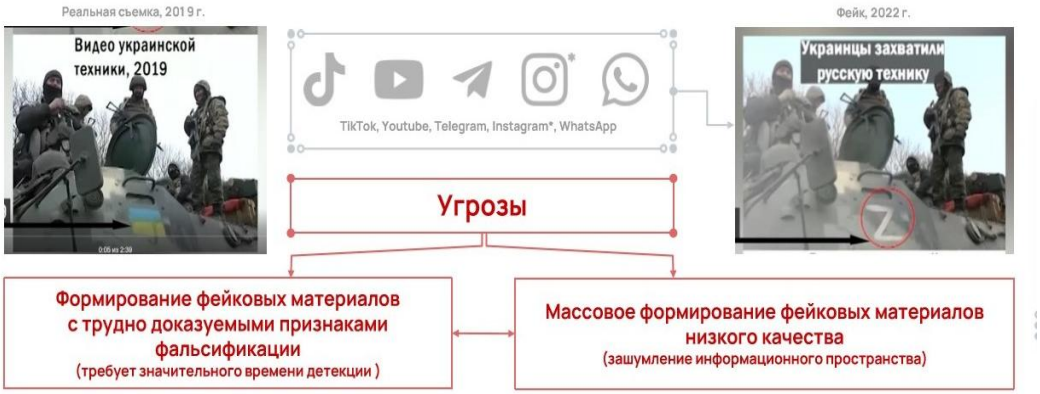


*Рис. 2. Иллюстрация результатов применения технологии deepfake*

В связи с вышесказанным возникают задачи противодействия угрозам безопасности информации с использованием технологий ИИ [5]. К ним, в частности, относятся: выявление источника угроз, который использует технологии ИИ (рис. 3) [4]; классификация угроз в отношении защищаемого объекта и формирование нового пространства признаков угроз; моделирование действий злоумышленника, в том числе с помощью поведенческого анализа [6]; формирование стратегии защиты для выявления аномалий, генерируемых системами с ИИ; разработка новых методов защиты от атак с использованием генеративных моделей ИИ, в том числе больших мультимодальных языковых моделей.


*Рис. 3. Основные способы выявления deepfake*

2. К основным задачам защиты информации в компьютерных системах и сетях с использованием ИИ относятся [5]: обнаружение компьютерных атак и вредоносных программ; обнаружение модификаций данных; предотвращение утечек конфиденциальных данных по тем или иным техническим каналам; повышение надежности и киберустойчивости компьютерных систем и сетей.

Недостатки традиционных систем информационной безопасности во многом связаны с тем, что они используют заблаговременно сформированные подходы к выявлению угроз и варианты реакций на них. Это влечет за собой как неспособность быстро реагировать на новые угрозы, так и появление большого числа ложных срабатываний. Другим важным ограничением являются объемы анализируемых данных. Существующие системы могут генерировать значительное количество показателей, связанных с событиями информационной безопасности, их вычислительно сложно анализировать в реальном времени.

К важнейшим преимуществам технологий ИИ следует отнести [7, 8]:

- ♦ возможность быстрой обработки больших массивов данных для раннего предупреждения о критических событиях безопасности;
- ♦ возможность одновременного анализа данных из нескольких источников, включая сетевой трафик, системные журналы и данные о поведении пользователей, чтобы выявлять подозрительные аномалии;
- ♦ возможность автоматического реагирования на угрозы, включая автоматический запрет доступа к скомпрометированной системе с целью предотвращения дальнейшего ущерба.

Средства интеллектуализации также позволяют провести моделирование различных угроз и поведения нарушителя периметра безопасности систем контроля управления доступом (СКУД), а также обеспечить на рубежах охраны средства контроля с использованием распознавания видеобразов пользователей СКУД.

3. С позиций нападающей стороны перспективным представляется использование технологий ИИ для обработки данных из технических каналов утечки информации [9]. Технологии ИИ могут использоваться как для «проигрывания» различных сценариев кибератак с использованием технических каналов утечки, так и для обработки (распознавания) сигналов, регистрируемых в этих каналах. На рис. 4. представлен фрагмент сигнала, полученного по акустическому каналу от клавиатуры компьютера [10]. Использование технологий ИИ позволяет значимо повысить качество классификации фрагментов сигнала, соответствующих нажатиям на клавиши, обеспечив практически однозначное определение истинных клавиш.

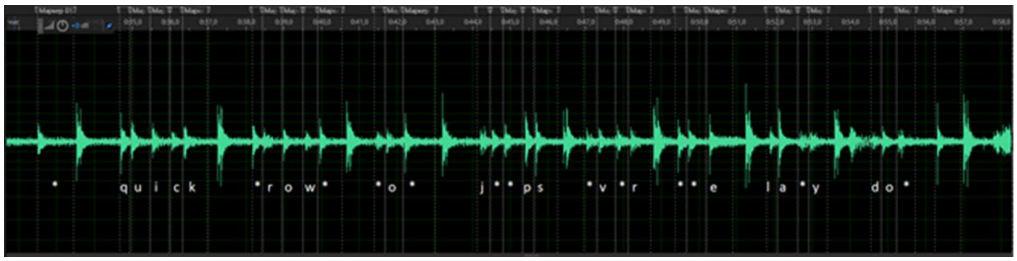


*Рис. 4. Результаты нейросетевого распознавания фрагментов сигнала*

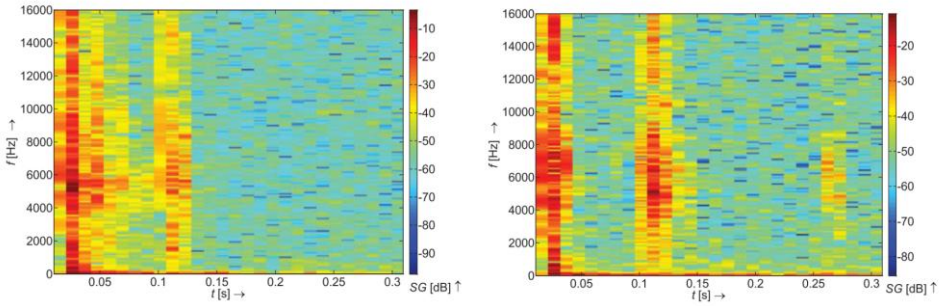


*Рис. 5. Спектрограммы звуков нажатия клавиш «L» и «D» ([11])*

На графиках на рис. 5 представлены спектрограммы звуков нажатий на клавиши «L» и «D». Ось абсцисс соответствует времени (в секундах), ось ординат – частоте (в герцах), а цвета соответствуют амплитудам наблюдаемых частот в данный момент времени (яркие цвета обозначают большие амплитуды). Эту спектрограмму можно рассматривать как изображение, что сводит [11] проблему распознавания звуков отдельных клавиш к проблеме классификации изображений. В [12] использовалась модель глубокого обучения для распознавания нажатий клавиш на ноутбуке. Звуки нажатия клавиш фиксировались с помощью встроенного микрофона рядом расположенного смартфона. В качестве входных данных для модели классификации изображений CoAtNet использовались мел-спектрограммы.

Еще один подход с использованием временных сверточных сетей (Temporal Convolutional Network, TCN) изложен в [13]. Предложенный классификатор достиг точности 95% без использования языковой модели. Для постобработки, то есть коррекции ошибок, применяются те или иные вероятностные модели текста [10, 14], которые могут значительно повысить точность распознавания.

В работе [15] предложен способ добавления специальных «фальшивых» звуков нажатия клавиш для противодействия рассматриваемым атакам.

4. Развитие ИИ привело к появлению принципиально новых технологий сжатия и передачи информации без значительных смысловых потерь (направление Semantic Communications) [16]. В свою очередь, расширяющееся использование таких технологий в системах связи и прежде всего, в сетях интернета вещей (IoT) [17], делает необходимым проведение тщательного анализа возникающих задач в области информационной безопасности. Семантическая коммуникация направлена на адресную и точную передачу смысла между различными интеллектуальными агентами, как людьми, так и машинами. В отличие от обычных систем связи, в которых приоритетна точность передачи исходящих данных, при семантической коммуникации приоритет принадлежит семантической точности. Семантическая коммуникация делает упор на извлечение и передачу информации семантического уровня из сообщения с целью «сохранения смысла». Современные коммуникационные технологии направлены на передачу большего количества данных с меньшим количеством ошибок с минимумом затрат (в рамках шенноновской парадигмы), тогда как семантическая коммуникация стремится передать максимальный объем семантики с наименьшими затратами коммуникационных ресурсов, быстро доставляя семантический контент (уже вне шенонновской парадигмы). В [18], например, для извлечения смысла передаваемого текста использовалась архитектура Transformer. Для извлечения смысла при передаче изображений в [19] использована сверточная нейронная сеть. Однако по мере развития таких систем они сталкиваются со значительными проблемами безопасности, конфиденциальности и доверия при интеграции технологий ИИ в интеллектуальные коммуникационные приложения. Перехват или зашумление сообщений при семантической коммуникации создает более серьезные угрозы безопасности [20], чем в обычных системах передачи информации на уровне бит или символов. Разрабатываются новые подходы к безопасности таких систем, в том числе использующие физический уровень канала (Physical Layer Security, PLS) [21], предлагаются метрики для оценки безопасности семантических коммуникаций [22].

5. Технологии ИИ могут быть использованы не только в технических областях. Использование больших языковых моделей типа GPT позволяет повысить эффективность деятельности аналитических подразделений организаций и предприятий. Спектр применения таких моделей весьма широк и позволяет не только совершенствовать аналитическую деятельность и конкурентную разведку, но также массово генерировать уникальный «человекоподобный» контент для проведения атак социальной инженерии [23, 24]. В последнее время широкое распространение больших языковых моделей приводит к росту социально-культурных рисков для гражданского общества [25].

Растущие возможности систем генеративного ИИ, таких как ChatGPT, привели к увеличению генерации синтетического контента, что имеет последствия для различных секторов общественной жизни, включая средства массовой информации, кибербезопас-

ность, образование, социальные сети, художественное творчество и др. В [26] проведен мониторинг основных мировых новостных веб-сайтов, и показано, что за последние полтора года доля новостных публикаций, созданных генеративным ИИ, удвоилась. Особую тревогу у авторов вызывает рост числа дезинформирующих и фейковых публикаций. Отметим однако, что существует и альтернативная точка зрения ([27]), согласно которой распространенные представления об опасности генеративного ИИ для производства фейков и дезинформации являются преувеличенными.

Актуальной становится проблема обнаружения контента, созданного LLM. Одной из новых и важных задач в области обработки текстов сейчас является определение того, написан ли данный текст человеком или генеративным ИИ. Это задача бинарной классификации, в которой анализируемый текст должен быть отнесен к одному из двух классов: тексты, имеющие естественное или искусственное происхождение. Для человека эта задача оказывается весьма сложной. Так, в экспериментах, проведенных в [28], точность решений, принимаемых экспертами-людьми, составила всего 61%. Однако эта задача хорошо решается алгоритмически, точность современных методов детектирования машинно-сгенерированных текстов достигает 90% и более [29–31].

**Обеспечение безопасности систем искусственного интеллекта.** Отдельным направлением обеспечения информационной безопасности является обеспечение безопасности самих систем ИИ. Новыми вызовами для государства, как отмечается в Национальной стратегии развития ИИ, являются, в том числе, «возникновение в сфере разработки, создания и использования ИИ новых типов угроз информационной безопасности, нехарактерных для других сфер применения информационных технологий». Технологии ИИ обладают той особенностью, что алгоритм решения задачи не фиксирован заранее, а формируется в процессе ее решения и существенным образом зависит от входных данных (обучающей выборки). Это приводит [32] к новым возможным атакам на системы ИИ, таким, как атаки на обучающие данные, искажение разметки, атаки, направленные на установление принадлежности конкретных данных обучающей выборке, атаки, направленные на получение данных из обученной модели, атаки на уровне вычислительных платформ и др. (рис. 6) [4].

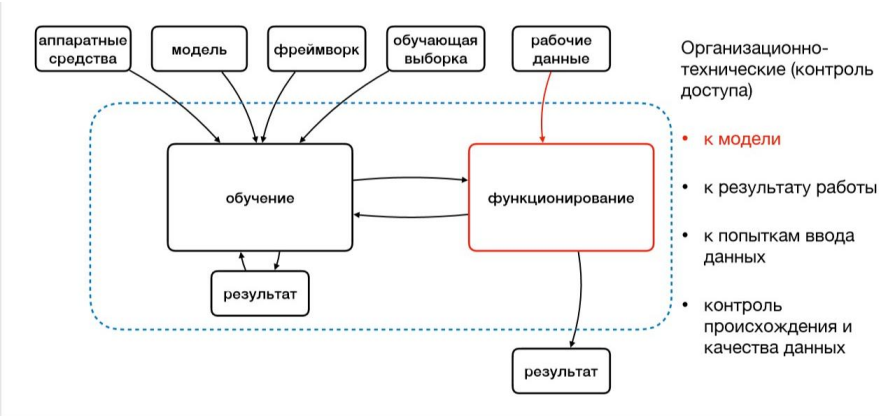


*Рис. 6. Организационно-технические средства защиты систем ИИ*

При построении информационных систем возможны атаки с помощью т.н. «скрытых каналов» [33]. Отметим недавно предложенную атаку [34] с использованием скрытого канала, напоминающего акустический технический канал утечки от клавиатуры. Атака направлена на приложения (помощники) с искусственным интеллектом, которые все интенсивнее проникают в нашу жизнь, их используют для получения совета или помощи в личных и конфиденциальных вопросах. В качестве скрытого канала рассматривается просто длина токенов сообщения. Информационный обмен пользователя с LLM, которая выполняется на удаленном сервере, защищен, и осуществляется в зашифрованном виде.

Однако LLM генерирует и отправляет ответ в виде серии токенов, причем каждый токен передается по мере его создания. Несмотря на шифрование, размер пакетов может раскрыть длины использованных токенов. Определение содержания ответа только на основе последовательности длин токенов, конечно, является непростой задачей и может допускать несколько вариантов решения. Однако, используя специально настроенную для решения такой задачи LLM, злоумышленник в ряде случаев может восстановить передаваемые тексты. Доля успешных атак на ChatGPT-4 от OpenAI и Copilot от Microsoft с помощью такого подхода составила от 17 до 53%.

Ряд систем генеративного искусственного интеллекта (рис. 7) [4] подвержена не только “переобучениям”, но и “галлюцинациям” – выдается по запросу информация, которой нет в обучающей выборке. Указанная уязвимость может быть эксплуатируема злоумышленниками, что еще раз подчеркивает актуальность задачи создания доверенного искусственного интеллекта [35].

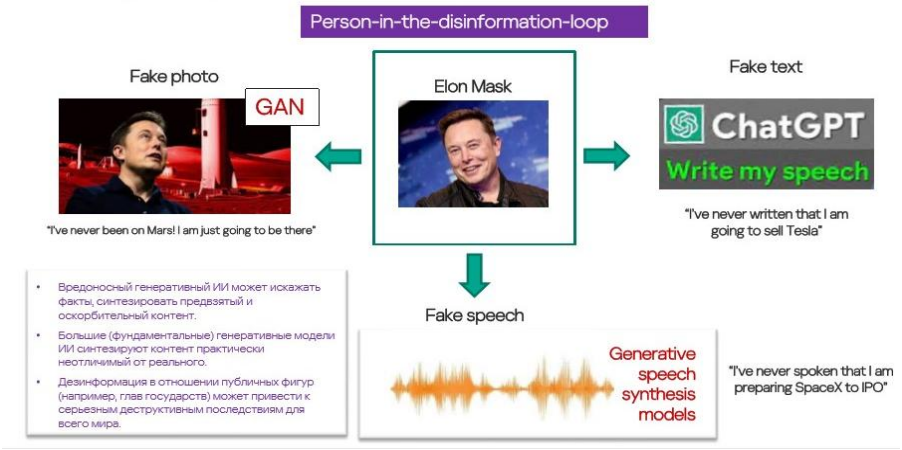


*Рис. 7. Пример системы генеративного ИИ для дискредитации личности*

Для защиты систем ИИ наряду с типовыми средствами защиты информации должны использоваться и специализированные технологии, и средства защиты, к основным из которых можно отнести: повышение надежности обучающих выборок, оценка доверия к принимаемым решениям, интерпретируемость результатов, контроль процессов обучения и верификации, повторяемость, отсутствие галлюцинаций и др. Основным направлением здесь должно стать создание систем доверенного искусственного интеллекта.

**Заключение.** Обеспечение высокого уровня информационной безопасности сегодня требует привлечения широкого спектра технологий ИИ. Полноценно владеющий этими технологиями будет превосходить противника вне зависимости от выполняемой функции - атакующей или нападающий. Развитие технологий ИИ для обработки данных различной природы делает необходимым формирование новых требований к моделям угроз и к средствам защиты информации. Основным направлением защиты собственно систем ИИ должно стать развитие технологий доверенного искусственного интеллекта.

*Работа выполнена при финансовой поддержке гранта РНФ № 24-11-00340.*

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Национальная стратегия развития искусственного интеллекта на период до 2030 года, утверждена Указом Президента Российской Федерации от 10 октября 2019 г. № 490.
2. *McCorduck P., Cfe C.* Machines who think: A personal inquiry into the history and prospects of artificial intelligence. – AK Peters/CRC Press, 2004.
3. *Mughal A.A.* Artificial Intelligence in Information Security: Exploring the Advantages, Challenges, and Future Directions // Journal of Artificial Intelligence and Machine Learning in Management. – 2018. – Vol. 2, No. 1. – P. 22-34.

4. Материалы первого форума «Цифровая экономика. Технологии доверенного искусственного интеллекта». Москва, МГУ, кластер «Ломоносов», 23 мая 2023 г. – Режим доступа: https://ib-bank.ru/trust-ai/materials23 (дата обращения: 10.10.2024).
5. *Мещеряков Р.В., Мельников С.Ю., Пересыпкин В.А., Хорев А.А*. Перспективные направления применения технологий искусственного интеллекта при защите информации // Вопросы кибербезопасности. – 2024. – № 4 (62). – С. 2-12. – DOI: 10.21681/2311-3456-2024-4-02-12. – EDN GJWQWP.
6. *Shelke P., Hämäläinen T.* Analysing Multidimensional Strategies for Cyber Threat Detection in Security Monitoring / In M. Lehto, & M. Karjalainen (Eds.) // Proceedings of the 23rd European Conference on Cyber Warfare and Security. – 2024. – 23. – P. 780-787. Academic Conferences International Ltd. Proceedings of the European Conference on Cyber Warfare and Security. – Режим доступа: https://doi.org/10.34190/eccws.23.1.2123 (дата обращения: 10.10.2024).
7. *Аветисян А.И.* Кибербезопасность в контексте искусственного интеллекта // Вестник Российской академии наук. – 2022. – Т. 92, № 12. – С. 1119-1123. – DOI: 10.31857/S0869587322120039. – EDN RYZRRU.
8. *Camacho N.G.* The Role of AI in Cybersecurity: Addressing Threats in the Digital Age // Journal of Artificial Intelligence General science (JAIGS). – 2024. – Vol. 3, No. 1. – P. 143-154. – ISSN: 3006-4023.
9. *Panoff M. et al.* A review and comparison of AI-enhanced side channel analysis // ACM Journal on Emerging Technologies in Computing Systems (JETC). – 2022. – Vol. 18, No. 3. – P. 1-20.
10. *Zhuang L., Zhou F., Tygar J.D.* Keyboard acoustic emanations revisited // ACM Transactions on Information and System Security (TISSEC). – 2009. – Vol. 13, No. 1. – P. 1-26.
11. *Taheritajar A., Harris Z. M., Rahaeimehr R.* A Survey on Acoustic Side Channel Attacks on Keyboards // arXiv preprint arXiv:2309.11012. – 2023.
12. *Harrison J., Toreini E., Mehrnezhad M.* A practical deep learning-based acoustic side channel attack on keyboards //2023 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW). – IEEE, 2023. – P. 270-280.
13. *Spata M. O. et al.* A New Deep Learning Pipeline for Acoustic Attack on Keyboards // IntelliSys 2024. – Cham: Springer Nature Switzerland, 2024. – P. 402-414.
14. *Мельников С.Ю., Пересыпкин В.А*. Об эволюции классических вероятностных моделей языка в естественно-языковых приложениях // Вестник современных цифровых технологий. – 2023. – № 16. – С. 4-14. – EDN YDIGDT.
15. *Rodrigues D. et al.* A Prototype for Generating Random Key Sounds to Prevent Keyboard Acoustic Side-Channel Attacks // 2024 IEEE 22nd Mediterranean Electrotechnical Conference (MELECON). – IEEE, 2024. – P. 1287-1292.
16. *Yang W. et al.* Semantic communications for future internet: Fundamentals, applications, and challenges // IEEE Communications Surveys & Tutorials. – 2022. – Vol. 25, No. 1. – P. 213-250.
17. *Wang Y.* Semantic Communication Networks Empowered Artificial Intelligence of Things // 2024 IEEE Annual Congress on Artificial Intelligence of Things (AIoT). – IEEE, 2024. – P. 189-193.
18. *Xie H. et al.* Deep learning enabled semantic communication systems // IEEE Transactions on Signal Processing. – 2021. – Vol. 69. – P. 2663-2675.
19. *Bourtsoulatze E., Kurka D. B., Gündüz D.* Deep joint source-channel coding for wireless image transmission // IEEE Transactions on Cognitive Communications and Networking. – 2019. – Vol. 5, No. 3. – P. 567-579.
20. *Luo X. et al.* Encrypted semantic communication using adversarial training for privacy preserving // IEEE Communications Letters. – 2023. – Vol. 27, No. 6. – P. 1486-1490.
21. *Nguyen V.L. et al.* Security and privacy for 6G: A survey on prospective technologies and challenges // IEEE Communications Surveys & Tutorials. – 2021. – Vol. 23, No. 4. – P. 2384-2428.
22. *Li Y. et al.* Secure Semantic Communications: From Perspective of Physical Layer Security // IEEE Communications Letters. – 2024. – DOI: 10.1109/LCOMM.2024.3452715.
23. *Hazell J.* Large language models can be used to effectively scale spear phishing campaigns // arXiv preprint arXiv:2305.06972. – 2023.
24. *Greco F. et al.* David versus Goliath: Can Machine Learning Detect LLM-Generated Text? A Case Study in the Detection of Phishing Emails // ITASEC 2024: The Italian Conference on CyberSecurity, Italy. CEUR-WS Vol. 3731. – 2024.
25. *Былевский П.Г.* Социально-культурные риски мультимодальных больших генеративных моделей «искусственного интеллекта» (GenAI) // Культура и искусство. – 2024. – № 6. – С. 213-224. – DOI: 10.7256/2454-0625.2024.6.70926. – EDN: DWMERQ.
26. *Hanley H.W.A., Durumeric Z.* Machine-made media: Monitoring the mobilization of machine-generated articles on misinformation and mainstream news websites // Proceedings of the International AAAI Conference on Web and Social Media. – 2024. – Vol. 18. – P. 542-556.

27. *Simon F.M., Altay S., Mercier H.* Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown // Harvard Kennedy School Misinformation Review. – 2023. – Vol. 4, No. 5.
28. *Liu Y. et al.* ArguGPT: evaluating, understanding and identifying argumentative essays generated by GPT models // arXiv preprint arXiv:2304.07666. – 2023.
29. *Wu J. et al.* A survey on llm-gernerated text detection: Necessity, methods, and future directions // arXiv preprint arXiv:2310.14724. – 2023.
30. *Ghosal S.S. et al.* A Survey on the Possibilities & Impossibilities of AI-generated Text Detection // Transactions on Machine Learning Research. – No. 1. – 2024.
31. *Sadasivan V.S. et al.* Can AI-generated text be reliably detected? // arXiv preprint arXiv:2303.11156. – 2023.
32. *Маршалко Г.Б., Романенков Р.А., Труфанова Ю.А*. Анализ безопасности проекта национального стандарта «Нейросетевые алгоритмы в защищенном исполнении. Автоматическое обучение нейросетевых моделей на малых выборках в задачах классификации» // Тр. Института системного программирования РАН. – 2023. – Т. 35, № 6. – С. 179-188. – DOI: 10.15514/ISPRAS-2023-35(6)-11. – EDN HNDIYD.
33. *Грушо А.А.* Скрытые каналы и безопасность информации в компьютерных системах // Дискретная математика. – 1998. – Т. 10, № 1. – С. 3-9.
34. *Weiss R. et al.* What Was Your Prompt? A Remote Keylogging Attack on AI Assistants // arXiv preprint arXiv:2403.09751. – 2024.
35. *Турдаков Д.Ю., Аветисян А.И., Архипенко К.В. [и др.]*. Доверенный Искусственный интеллект: вызовы и перспективные решения // Доклады Российской академии наук. Математика, информатика, процессы управления. – 2022. – Т. 508, № 1. – С. 13-18. – DOI: 10.31857/S2686954322070207. – EDN CVIVCS.

REFERENCES

1. Natsional'naya strategiya razvitiya iskusstvennogo intellekta na period do 2030 goda, utverzhdena Ukazom Prezidenta Rossiyskoy Federatsii ot 10 oktyabrya 2019 g. № 490 [National Strategy for the Development of Artificial Intelligence through 2030, approved by the Decree of the President of the Russian Federation, October 10, 2019, No. 490].
2. *McCorduck P., Cfe C.* Machines who think: A personal inquiry into the history and prospects of artificial intelligence. AK Peters/CRC Press, 2004.
3. *Mughal A.A.* Artificial Intelligence in Information Security: Exploring the Advantages, Challenges, and Future Directions, *Journal of Artificial Intelligence and Machine Learning in Management,* 2018, Vol. 2, No. 1, pp. 22-34.
4. Materialy pervogo foruma «Tsifrovaya ekonomika. Tekhnologii doverennogo iskusstvennogo intellekta». Moskva, MGU, klaster «Lomonosov», 23 maya 2023 g. [Proceedings of the first forum "Digital Economy. Technologies of Trusted Artificial Intelligence". Moscow, Moscow State University, Lomonosov Cluster, May 23, 2023]. Available at: https://ib-bank.ru/trust-ai/materials23 (accessed on 10 October 2024).
5. *Meshcheryakov R.V., Mel'nikov S.Yu., Peresypkin V.A., Khorev A.A.* Perspektivnye napravleniya primeneniya tekhnologiy iskusstvennogo intellekta pri zashchite informatsii [Promising areas of application of artificial intelligence technologies in information security], *Voprosy kiberbezopasnosti* [Cybersecurity Issues], 2024, No. 4 (62), pp. 2-12. DOI: 10.21681/2311-3456-2024-4-02-12. EDN GJWQWP.
6. *Shelke P., Hämäläinen T.* Analysing Multidimensional Strategies for Cyber Threat Detection in Security Monitoring, In M. Lehto, & M. Karjalainen (Eds.), *Proceedings of the 23rd European Conference on Cyber Warfare and Security,* 2024, 23, pp. 780-787. Academic Conferences International Ltd. Proceedings of the European Conference on Cyber Warfare and Security. Available at: https://doi.org/10.34190/eccws.23.1.2123 (accessed 10 October 2024).
7. *Avetisyan A.I.* Kiberbezopasnost' v kontekste iskusstvennogo intellekta [Cybersecurity in the Context of Artificial Intelligence],*Vestnik Rossiyskoy akademii nauk* [Bulletin of the Russian Academy of Sciences], 2022, Vo.. 92, No. 12, pp. 1119-1123. DOI: 10.31857/S0869587322120039. EDN RYZRRU.
8. *Camacho N.G.* The Role of AI in Cybersecurity: Addressing Threats in the Digital Age, *Journal of Artificial Intelligence General science (JAIGS*), 2024, Vol. 3, No. 1, pp. 143-154. ISSN 3006-4023.
9. *Panoff M. et al.* A review and comparison of AI-enhanced side channel analysis, *ACM Journal on Emerging Technologies in Computing Systems (JETC),* 2022, Vol. 18, No. 3, pp. 1-20.
10. *Zhuang L., Zhou F., Tygar J.D.* Keyboard acoustic emanations revisited, *ACM Transactions on Information and System Security (TISSEC),* 2009, Vol. 13, No. 1, pp. 1-26.

11. *Taheritajar A., Harris Z. M., Rahaeimehr R.* A Survey on Acoustic Side Channel Attacks on Keyboards, *arXiv preprint arXiv:2309.11012,* 2023.
12. *Harrison J., Toreini E., Mehrnezhad M.* A practical deep learning-based acoustic side channel attack on keyboards, *2023 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE, 2023, pp. 270-280.
13. *Spata M. O. et al.* A New Deep Learning Pipeline for Acoustic Attack on Keyboards, *IntelliSys 2024.* Cham: Springer Nature Switzerland, 2024, pp. 402-414.
14. *Mel'nikov S.Yu., Peresypkin V.A.* Ob evolyutsii klassicheskikh veroyatnostnykh modeley yazyka v estestvenno-yazykovykh prilozheniyakh [On the evolution of classical probabilistic language models in natural language applications], *Vestnik sovremennykh tsifrovykh tekhnologiy* [Bulletin of modern digital technologies], 2023, No. 16, pp. 4-14. EDN YDIGDT.
15. *Rodrigues D. et al.* A Prototype for Generating Random Key Sounds to Prevent Keyboard Acoustic Side-Channel Attacks, *2024 IEEE 22nd Mediterranean Electrotechnical Conference (MELECON)*. IEEE, 2024, pp. 1287-1292.
16. *Yang W. et al.* Semantic communications for future internet: Fundamentals, applications, and challenges, *IEEE Communications Surveys & Tutorials,* 2022, Vol. 25, No. 1, pp. 213-250.
17. *Wang Y.* Semantic Communication Networks Empowered Artificial Intelligence of Things, *2024 IEEE Annual Congress on Artificial Intelligence of Things (AIoT)*. IEEE, 2024, pp. 189-193.
18. *Xie H. et al.* Deep learning enabled semantic communication systems, *IEEE Transactions on Signal Processing,* 2021, Vol. 69, pp. 2663-2675.
19. *Bourtsoulatze E., Kurka D. B., Gündüz D.* Deep joint source-channel coding for wireless image transmission, *IEEE Transactions on Cognitive Communications and Networking,* 2019, Vol. 5, No. 3, pp. 567-579.
20. *Luo X. et al.* Encrypted semantic communication using adversarial training for privacy preserving, *IEEE Communications Letters,* 2023, Vol. 27, No. 6, pp. 1486-1490.
21. *Nguyen V.L. et al.* Security and privacy for 6G: A survey on prospective technologies and challenges, *IEEE Communications Surveys & Tutorials,* 2021, Vol. 23, No. 4, pp. 2384-2428.
22. *Li Y. et al.* Secure Semantic Communications: From Perspective of Physical Layer Security, *IEEE Communications Letters,* 2024. DOI: 10.1109/LCOMM.2024.3452715.
23. *Hazell J.* Large language models can be used to effectively scale spear phishing campaigns, *arXiv preprint arXiv:2305.06972,* 2023.
24. *Greco F. et al.* David versus Goliath: Can Machine Learning Detect LLM-Generated Text? A Case Study in the Detection of Phishing Emails, *ITASEC 2024: The Italian Conference on CyberSecurity, Italy*. CEUR-WS Vol. 3731, 2024.
25. *Bylevskiy P.G.* Sotsial'no-kul'turnye riski mul'timodal'nykh bol'shikh generativnykh modeley «iskusstvennogo intellekta» (GenAI) [Socio-cultural risks of multimodal large generative models of "artificial intelligence" (GenAI)], *Kul'tura i iskusstvo* [Culture and Art], 2024, No. 6, pp. 213-224. DOI: 10.7256/2454-0625.2024.6.70926. EDN: DWMERQ.
26. *Hanley H.W.A., Durumeric Z.* Machine-made media: Monitoring the mobilization of machine-generated articles on misinformation and mainstream news websites, *Proceedings of the International AAAI Conference on Web and Social Media,* 2024, Vol. 18, pp. 542-556.
27. *Simon F.M., Altay S., Mercier H.* Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown, *Harvard Kennedy School Misinformation Review,* 2023, Vol. 4, No. 5.
28. *Liu Y. et al.* ArguGPT: evaluating, understanding and identifying argumentative essays generated by GPT models, *arXiv preprint arXiv:2304.07666,* 2023.
29. *Wu J. et al.* A survey on llm-gernerated text detection: Necessity, methods, and future directions, *arXiv preprint arXiv:2310.14724,* 2023.
30. *Ghosal S.S. et al.* A Survey on the Possibilities & Impossibilities of AI-generated Text Detection, *Transactions on Machine Learning Research,* No. 1*,* 2024.
31. *Sadasivan V.S. et al.* Can AI-generated text be reliably detected?, *arXiv preprint arXiv:2303.11156,* 2023.
32. *Marshalko G.B., Romanenkov R.A., Trufanova Yu.A.* Analiz bezopasnosti proekta natsional'nogo standarta «Neyrosetevye algoritmy v zashchishchennom ispolnenii. Avtomaticheskoe obuchenie neyrosetevykh modeley na malykh vyborkakh v zadachakh klassifikatsii» [Security analysis of the draft national standard "Neural network algorithms in secure implementation. Automatic training of neural network models on small samples in classification problems"], *Tr. Instituta sistemnogo programmirovaniya RAN* [Proceedings of the Institute for System Programming of the Russian Academy of Sciences], 2023, Vol. 35, No. 6, pp. 179-188. DOI: 10.15514/ISPRAS-2023-35(6)-11. EDN HNDIYD.
33. *Grusho A.A.* Skrytye kanaly i bezopasnost' informatsii v komp'yuternykh sistemakh [Covert channels and information security in computer systems], *Diskretnaya matematika* [Discrete Mathematics], 1998, Vol. 10, No. 1, pp. 3-9.

34. *Weiss R. et al.* What Was Your Prompt? A Remote Keylogging Attack on AI Assistants, *arXiv preprint arXiv:2403.09751,* 2024.
35. *Turdakov D.Yu., Avetisyan A.I., Arkhipenko K.V. [i dr.].* Doverennyy Iskusstvennyy intellekt: vyzovy i perspektivnye resheniya [Trusted Artificial Intelligence: Challenges and Promising Solutions], *Doklady Rossiyskoy akademii nauk. Matematika, informatika, protsessy upravleniya* [Reports of the Russian Academy of Sciences. Mathematics, informatics, control processes], 2022, Vol. 508, No. 1, pp. 13-18. DOI: 10.31857/S2686954322070207. EDN CVIVCS.

Статью рекомендовал к опубликованию д.т.н. Г.Е. Веселов.

**Мельников Сергей Юрьевич** – Российский университет дружбы народов имени Патриса Лумумбы; e-mail: melnikov@linfotech.ru; г. Москва, Россия; кафедра теории вероятностей и кибербезопасности института компьютерных наук и телекоммуникаций факультета физико-математических и естественных наук; д.ф.-м.н.; доцент.

**Мещеряков Роман Валерьевич** – ИПУ РАН; e-mail: mrv@ieee.org; г. Москва, Россия; г.н.с.; д.т.н.; профессор.

**Пересыпкин Владимир Анатольевич** – Академия криптографии РФ; e-mail: info@cryptoacademy.gov.ru; г. Москва, Россия; действительный член; д.т.н.

**Melnikov Sergey Yur'evich** – Patrice Lumumba Peoples' Friendship University of Russia; e-mail: melnikov@linfotech.ru; Moscow, Russia; the Department of Probability Theory and Cybersecurity, Institute of Computer Science and Telecommunications, Faculty of Physics, Mathematics and Natural Sciences; dr. of phys. and math. sc.; associate professor.

**Meshcheryakov Roman Valer'evich** – IPU RAS; e-mail: mrv@ieee.org; Moscow, Russia; chief researcher; dr. of eng. sc.; professor.

**Peresypkin Vladimir Anatol'evich** – Academy of Cryptography of the Russian Federation; e-mail: info@cryptoacademy.gov.ru; Moscow, Russia; Full Member; dr. of eng. sc.

УДК 004.382.2 DOI 10.18522/2311-3103-2024-5-68-78

**Е.А. Титенко, Э.И. Ватутин, М.А. Титенко, Э.В. Мельник, А.П. Локтионов**

# АППАРАТНО-ОРИЕНТИРОВАННЫЙ МЕТОД УСКОРЕННОГО ПОИСКА ВХОЖДЕНИЙ ОБРАЗЦА НА ОСНОВЕ СТРУКТУРНО-ПРОЦЕДУРНЫХ ВЫЧИСЛЕНИЙ

*Операция поиска вхождений образца в тексте является общезначимой в современных вычислительных средствах при решении проблемно-поисковых задач. Наибольший интерес представляют аппаратно-программные решения, имеющие однородную структуру и регулярные связи между вычислительными блоками. Целью работы является сокращение временных затрат на поиск вхождений на основе применения параллельного поиска в ассоциативной памяти и метода распараллеливания по итерациям. Предлагаемый метод использует ассоциативную память для параллельного поиска вхождений и динамическую реконфигурации структуры исходной строки из одномерного вида в матричную форму. Вовлечение в реконфигурацию всех элементов влечет избыточные затраты внутренней блочной памяти на последовательный просмотр частичных вхождений по одному множеству стартовых позиций, кратных длине образца (второй символьный операнд. Вместо этого предложен метод совмещения во времени поиска частичных вхождений по двум наборам подстрок, кратных длине образца, с одновременным пропорциональным уменьшением элементов разрядного среза ассоциативной памяти по каждому набору, что позволяет на текущем шаге поиска обрабатывать несколько символов образца. Количественные оценки времени поиска определяются количеством операций сравнения и записи подстрок в общем цикле работы, а также пропорциями времени данных операций. Показано, что для образцов более 10 элементов временной выигрыш составляет примерно в 1,8-2 раза. Данный эффект получен за счет исключения шагов последовательного сдвига с переходами между граничными элементами строк. Разработанный метод обеспечивает конвейерную обработку потока строковых операндов с совмещением просмотра на текущем шаге поиска неединичного множества символов обрабатываемой строки. Сокращение времени поиска обеспечивается введением конвейера, количество ступеней*