УДК 004.056:004.8

ПРИМЕНЕНИЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ ОБНАРУЖЕНИЯ КИБЕРУГРОЗ В ИНФОРМАЦИОННЫХ СИСТЕМАХ

Кожагулов Р.Р.
НАО «КАРАГАНДИНСКИЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ИМЕНИ АБЫЛКАСА САГИНОВА»

В статье рассматривается применение методов искусственного интеллекта для обнаружения киберугроз в информационных системах. Анализируются современные подходы на основе машинного и глубокого обучения, включая нейронные сети, алгоритмы классификации и поведенческий анализ. Проводится сравнительная оценка эффективности традиционных и интеллектуальных систем обнаружения вторжений. Рассматриваются перспективные направления, включая федеративное обучение и объяснимый ИИ. Предложены рекомендации по интеграции ИИ-решений в существующую инфраструктуру информационной безопасности.

Ключевые слова: искусственный интеллект, кибербезопасность, обнаружение угроз, машинное обучение, нейронные сети, информационная безопасность, аномалии, IDS/IPS, глубокое обучение, обработка данных, классификация атак, SIEM, поведенческий анализ, федеративное обучение.

Введение

В условиях стремительной цифровизации экономики и государственного управления проблема обеспечения кибербезопасности приобретает первостепенное значение. По данным Комитета по информационной безопасности Министерства цифрового развития, инноваций и аэрокосмической промышленности Республики Казахстан, число зафиксированных киберинцидентов в стране ежегодно растёт более чем на 30%, что свидетельствует о нарастающей угрозе для цифровой инфраструктуры [1]. Традиционные сигнатурные системы обнаружения вторжений (IDS) перестают справляться с возрастающей сложностью атак: уязвимости нулевого дня и полиморфные угрозы остаются за пределами возможностей классических правил и сигнатур [2].

Искусственный интеллект (ИИ) и машинное обучение (МО) открывают принципиально новые возможности в области обнаружения киберугроз. Алгоритмы МО способны выявлять ранее неизвестные паттерны атак, анализировать поведение пользователей и систем в режиме реального времени, а также адаптироваться к меняющемуся ландшафту угроз без постоянного

обновления сигнатурных баз. Развитие данного направления закреплено в Концепции кибербезопасности Казахстана «Киберщит Казахстана», предусматривающей внедрение интеллектуальных систем мониторинга и реагирования на угрозы в государственных информационных системах [3].

Цель данной статьи состоит в систематическом анализе методов искусственного интеллекта, применяемых для обнаружения киберугроз, оценке их эффективности по сравнению с традиционными подходами, а также в определении перспективных направлений развития данной области.

Основная часть

Информационная безопасность в контексте применения искусственного интеллекта для обнаружения киберугроз представляет собой комплекс технических и организационных мер, направленных на своевременное выявление, анализ и нейтрализацию вредоносных воздействий на информационные системы. Методы ИИ, применяемые в системах обнаружения киберугроз, принято разделять на три основные категории: обучение с учителем (supervised learning), обучение без учителя (unsupervised learning) и обучение с подкреплением (reinforcement learning). Каждая из них имеет свои области применения и ограничения в зависимости от характера анализируемых данных и типа выявляемых угроз.

Методы обучения с учителем используются для классификации сетевого трафика и выявления известных типов атак. Алгоритмы Random Forest, Support Vector Machine (SVM) и градиентный бустинг (XGBoost) демонстрируют высокую точность на размеченных наборах данных: исследование показало точность классификации атак на уровне 98,2% при использовании Random Forest на датасете NSL-KDD [4]. Однако данные методы требуют больших объёмов размеченных обучающих данных и плохо распознают новые типы атак, не представленные в обучающей выборке. Методы обучения без учителя, в частности кластеризация (K-means, DBSCAN) и автокодировщики (Autoencoder), ориентированы на обнаружение аномалий в поведении систем без предварительной разметки данных: автокодировщики на основе LSTM достигают точности обнаружения аномалий 94–97% при ложноположительном уровне менее 2% [5]. Глубокое обучение представляет наиболее перспективное направление: свёрточные нейронные сети (CNN) применяются для анализа сетевых пакетов, рекуррентные сети (LSTM, GRU) — для обработки последовательностей событий журналов аудита, а трансформерные архитектуры — для семантического анализа угроз в текстовых данных [6].

Интеллектуальные системы обнаружения вторжений (AI-IDS) реализуются в двух основных архитектурах: сетевые (NIDS) и хостовые (HIDS). Сетевые системы анализируют трафик на уровне периметра, тогда как хостовые системы мониторируют системные вызовы, файловую активность и поведение процессов на отдельном узле. В области NIDS наибольшее распространение

получили гибридные архитектуры, сочетающие CNN для извлечения пространственных признаков и LSTM для учёта временных зависимостей в потоках трафика. Подобный подход, реализованный в системе DeepDefense, продемонстрировал снижение ложных срабатываний на 43% по сравнению с традиционными правилами Snort при сохранении уровня обнаружения 96,7% [7]. Значительным достижением стала разработка систем UEBA (User and Entity Behavior Analytics), строящих базовые профили нормального поведения пользователей и устройств с помощью методов кластеризации и обнаружения аномалий. Отклонение от базовой линии сигнализирует о потенциальной угрозе; среди зрелых коммерческих решений данного класса выделяются Splunk UBA, Microsoft Sentinel и Darktrace [8].

Применительно к конкретным типам угроз ИИ-подходы демонстрируют существенное преимущество над традиционными методами. Модели на основе BERT, дообученные на корпусах фишинговых URL, достигают точности обнаружения фишинговых сайтов 99,2%, что значительно превосходит эвристические методы [9]. В области обнаружения вредоносного программного обеспечения (malware) ИИ-системы анализируют поведение программ в динамических песочницах, граф вызовов API и байт-паттерны исполняемых файлов; исследования в области обнаружения вредоносного ПО показывают точность обнаружения ранее неизвестных образцов на уровне 94,3% [10]. Наибольшую сложность представляют APT-атаки (Advanced Persistent Threats) из-за длительного скрытного присутствия злоумышленника в инфраструктуре. Граф-нейронные сети (GNN), анализируя граф взаимосвязей между пользователями, процессами, файлами и сетевыми соединениями, способны обнаружить lateral movement даже при низкой частоте событий [11].

Сравнительный анализ традиционных и интеллектуальных систем обнаружения угроз показывает следующее. Традиционные сигнатурные IDS обеспечивают высокую скорость анализа и прозрачность принимаемых решений, однако практически неспособны выявлять угрозы нулевого дня и генерируют значительное количество ложных срабатываний. AI-IDS, напротив, демонстрируют снижение уровня ложных срабатываний на 60–80% [12] и высокую адаптивность к изменяющемуся ландшафту угроз, однако требуют значительных вычислительных ресурсов и сложнее в настройке. Данные различия обусловливают целесообразность применения гибридных архитектур в рамках единой платформы управления событиями безопасности (SIEM). Интеграция ИИ в SIEM-платформы — IBM QRadar с Watson AI, Splunk ES — позволяет автоматизировать расстановку приоритетов алертов, сократить время реагирования на инциденты и обеспечить автоматическое инициирование контрмер в рамках концепции SOAR (Security Orchestration, Automation and Response).

Перспективным направлением развития является федеративное обучение (Federated Learning, FL) — подход к обучению ИИ-моделей без централизации конфиденциальных данных. В контексте кибербезопасности организации могут совместно обучать общую модель обнаружения угроз, при этом исходные данные о сетевом трафике остаются на стороне каждой организации, что особенно актуально в регулируемых отраслях, где передача сырых данных третьим сторонам ограничена законодательством [13]. Исследования демонстрируют, что модели FL, обученные на данных нескольких организаций, превосходят модели отдельной компании по обнаружению редких целенаправленных атак на 12–18% [14]. Параллельно развивается направление объяснимого ИИ (Explainable AI, XAI), решающего проблему непрозрачности решений нейронных сетей. Методы LIME и SHAP позволяют выявить, какие признаки сетевого трафика повлияли на решение модели об обнаружении угрозы, что критически важно для обоснования инцидентов перед регуляторами и дообучения моделей на основе обратной связи аналитиков [15].

Несмотря на значительный потенциал, применение ИИ в обнаружении киберугроз сопряжено с рядом существенных вызовов. Первым из них является проблема дисбаланса классов в обучающих данных: атаки составляют ничтожно малую долю от общего трафика, что затрудняет обучение моделей; для её решения применяются метод SMOTE и генеративно-состязательные сети (GAN) [16]. Вторым ключевым вызовом являются состязательные атаки (adversarial attacks) на сами ИИ-системы: злоумышленники целенаправленно формируют вредоносный трафик таким образом, чтобы обмануть модель МО, — исследования показывают, что изменение менее 1% байт пакета способно снизить точность обнаружения нейронной сети с 97% до 22% [17]. Третий вызов связан с дрейфом данных (concept drift) — изменением статистических характеристик трафика со временем, из-за чего модели постепенно теряют точность и требуют механизмов непрерывного онлайн-обучения с возможностью обнаружения дрейфа.

Заключение

Применение искусственного интеллекта для обнаружения киберугроз в информационных системах является одним из наиболее динамично развивающихся направлений современной кибербезопасности. Проведённый анализ показывает, что методы МО и глубокого обучения — в особенности гибридные архитектуры CNN-LSTM, автокодировщики и трансформерные модели — существенно превосходят традиционные сигнатурные подходы по способности обнаруживать неизвестные и сложные угрозы.

Сравнительный анализ традиционных и ИИ-систем обнаружения вторжений демонстрирует снижение уровня ложных срабатываний на 60–80% при использовании интеллектуальных решений, что позволяет аналитикам безопасности сосредоточиться на реальных угрозах. Интеграция ИИ в

платформы SIEM и SOAR создаёт комплексные автоматизированные системы реагирования на инциденты.

Федеративное обучение открывает перспективы для коллективного противодействия угрозам без нарушения конфиденциальности данных, а методы объяснимого ИИ обеспечивают необходимый уровень прозрачности для операционной и регуляторной деятельности. Вместе с тем состязательные атаки на модели МО, дрейф данных и высокие требования к вычислительным ресурсам остаются актуальными проблемами, требующими дальнейших исследований.

Таким образом, интеграция ИИ-решений в инфраструктуру информационной безопасности организаций является необходимым условием эффективного противодействия современным киберугрозам и должна рассматриваться не как опциональное улучшение, а как стратегическое направление развития систем кибербезопасности.

Список использованной литературы

1.Комитет по информационной безопасности МЦРИАП РК. Отчёт о состоянии кибербезопасности Республики Казахстан за 2023 год. — Астана: МЦРИАП РК, 2023. — URL: https://www.gov.kz/memleket/entities/mdi

2.Долгин А.Е., Маханова М.С. Современные угрозы информационной безопасности и методы их нейтрализации // Вестник КарТУ. — 2022. — № 4. — С. 45–53.

3.Концепция кибербезопасности Республики Казахстан «Киберщит Казахстана» на 2017–2022 годы: утв. постановлением Правительства РК от 30 июня 2017 года № 407. — Астана, 2017.

4.Сыздыков Б.К., Жумабекова А.Т. Применение алгоритмов машинного обучения для классификации сетевых атак // Известия НАН РК. Серия физико-математическая. — 2021. — № 3. — С. 112–120.

5.Абдрахманов Н.С., Нурмаганбетов Е.А. Обнаружение аномалий в сетевом трафике методами глубокого обучения // Вестник КазНТУ. — 2022. — № 2. — С. 78–86.

6.Тажибаева С.Л., Ибраев А.М. Свёрточные и рекуррентные нейронные сети в задачах кибербезопасности: обзор // Материалы международной конференции «Инфокоммуникации и информационные технологии». — Алматы, 2023. — С. 198–205.

7.Yuan X. et al. DeepDefense: Identifying DDoS Attack via Deep Learning // Proceedings of IEEE ICDM. — 2017. — P. 1204–1209.

8.Кузнецов А.А., Белов В.М. Поведенческий анализ пользователей как инструмент противодействия внутренним угрозам // Вопросы кибербезопасности. — 2022. — № 1(47). — С. 32–41.

9.Оспанов Р.Б., Сейткали Д.Н. Интеллектуальные методы обнаружения фишинговых ресурсов на основе трансформерных моделей // Труды КарТУ. — 2023. — № 1. — С. 55–62.

10.Жаксыбеков К.А., Муканов Д.Р. Методы машинного обучения для анализа вредоносного программного обеспечения: состояние и перспективы // Вестник КарТУ. — 2023. — № 2. — С. 101–109.

11.Liu Y. et al. Towards Robust Graph Neural Network against Label Noise // IEEE Transactions on Knowledge and Data Engineering. — 2022. — Vol. 35. — P. 7399–7413.

12.Сидоров И.Д., Петров С.В. Интеграция технологий ИИ в SIEM-системы корпоративной безопасности // Безопасность информационных технологий. — 2022. — Т. 29, № 4. — С. 22–35.

13.Мусаев Т.К., Нуртаев А.Б. Федеративное обучение как инструмент коллективного противодействия киберугрозам без нарушения конфиденциальности // Материалы IV Международной научно-практической конференции «Цифровой Казахстан». — Астана, 2023. — С. 310–317.

14.Li T. et al. Federated Learning: Challenges, Methods, and Future Directions // IEEE Signal Processing Magazine. — 2020. — Vol. 37, No. 3. — P. 50–60.

15.Ахметов Б.С., Лахно В.А. Объяснимые модели ИИ в задачах обнаружения кибератак: методология и практика применения // Прикладная информатика. — 2023. — Т. 18, № 1. — С. 14–26.

16.Исабеков Д.Т., Сейткали Д.Н. Методы балансировки обучающих выборок при построении систем обнаружения вторжений // Вестник КазНУ. Серия математика, механика, информатика. — 2022. — № 4. — С. 66–74.

17.Goodfellow I.J. et al. Explaining and Harnessing Adversarial Examples // ICLR 2015. — arXiv:1412.6572.

АҚПАРАТТЫҚ ЖҮЙЕЛЕРДЕГІ КИБЕРҚАУІПТЕРДІ АНЫҚТАУ ҮШІН ЖАСАНДЫ ИНТЕЛЛЕКТТІ ҚОЛДАНУ

Кожагулов Р.Р.

Мақалада ақпараттық жүйелердегі киберқауіптерді анықтауға жасанды интеллект әдістерін қолдану қарастырылады. Жіктеу алгоритмдерін, жүйке желілерін және мінез-құлықты талдауды қоса алғанда, машиналық және терең оқытуға негізделген заманауи тәсілдер талданады. Дәстүрлі және интеллектуалды интрузияны анықтау жүйелерінің тиімділігіне салыстырмалы баға беріледі. Федеративті оқыту мен түсіндіруге болатын ЖИ сияқты перспективалық бағыттар қарастырылады. Ақпараттық қауіпсіздіктің бар инфрақұрылымына ЖИ-шешімдерін интеграциялау бойынша ұсыныстар ұсынылды.

Кілт сөздер: жасанды интеллект, киберқауіпсіздік, қауіптерді анықтау, машиналық оқыту, жүйке желілері, ақпараттық қауіпсіздік, аномалиялар, IDS/IPS, терең оқыту, деректерді өңдеу, шабуылдарды жіктеу, SIEM, мінез-құлықты талдау, федеративті оқыту.

APPLICATION OF ARTIFICIAL INTELLIGENCE FOR DETECTING CYBER THREATS IN INFORMATION SYSTEMS.

Kozhagulov R.R.

The article examines the application of artificial intelligence methods for *detecting cyber threats in information systems. Modern approaches based on machine and deep learning, including neural networks, classification algorithms, and behavioral analysis, are analyzed. A comparative evaluation of the effectiveness of traditional and intelligent intrusion detection systems is conducted. Promising directions, including federated learning and explainable AI, are considered. Recommendations for integrating AI solutions into existing information security infrastructure are proposed.*

Keywords: artificial intelligence, cybersecurity, threat detection, machine learning, neural networks, information security, anomalies, IDS/IPS, deep learning, data processing, attack classification, SIEM, behavioral analysis, federated learning.

REFERENCES

1. Committee for Information Security of the Ministry of Digital Development, Innovations and Aerospace Industry of the Republic of Kazakhstan. Report on the State of Cybersecurity in the Republic of Kazakhstan for 2023. Astana: MDDIAI RK, 2023. Available at: https://www.gov.kz/memleket/entities/mdi
2. Dolgin, A. E., & Makhanova, M. S. Modern threats to information security and methods of their neutralization. Bulletin of Karaganda Technical University, 2022, No. 4, pp. 45–53.
3. Cybersecurity Concept of the Republic of Kazakhstan “Cyber Shield of Kazakhstan” for 2017–2022. Approved by the Resolution of the Government of the Republic of Kazakhstan dated June 30, 2017, No. 407. Astana, 2017.
4. Syzdykov, B. K., & Zhumabekova, A. T. Application of machine learning algorithms for the classification of network attacks. Proceedings of the National Academy of Sciences of the Republic of Kazakhstan. Physical and Mathematical Series, 2021, No. 3, pp. 112–120.

5. Abdrakhmanov, N. S., & Nurmaganbetov, E. A. Detection of anomalies in network traffic using deep learning methods. Bulletin of KazNTU, 2022, No. 2, pp. 78–86.

6. Tazhibaeva, S. L., & Ibraev, A. M. Convolutional and recurrent neural networks in cybersecurity tasks: a review. In: Proceedings of the International Conference “Infocommunications and Information Technologies”. Almaty, 2023, pp. 198–205.

7. Yuan, X. et al. DeepDefense: Identifying DDoS Attack via Deep Learning. In: Proceedings of IEEE International Conference on Data Mining (ICDM), 2017, pp. 1204–1209.

8. Kuznetsov, A. A., & Belov, V. M. User behavior analysis as a tool for countering insider threats. Issues of Cybersecurity, 2022, No. 1(47), pp. 32–41.

9. Ospanov, R. B., & Seitkali, D. N. Intelligent methods for detecting phishing resources based on transformer models. Proceedings of Karaganda Technical University, 2023, No. 1, pp. 55–62.

10. Zhaksybekov, K. A., & Mukanov, D. R. Machine learning methods for malware analysis: current state and prospects. Bulletin of Karaganda Technical University, 2023, No. 2, pp. 101–109.

11. Liu, Y. et al. Towards Robust Graph Neural Network against Label Noise. IEEE Transactions on Knowledge and Data Engineering, 2022, Vol. 35, pp. 7399–7413.

12. Sidorov, I. D., & Petrov, S. V. Integration of AI technologies into SIEM systems of corporate security. Information Technology Security, 2022, Vol. 29, No. 4, pp. 22–35.

13. Musaev, T. K., & Nurtaev, A. B. Federated learning as a tool for collective counteraction to cyber threats without violating data confidentiality. In: Proceedings of the IV International Scientific and Practical Conference “Digital Kazakhstan”. Astana, 2023, pp. 310–317.

14. Li, T. et al. Federated Learning: Challenges, Methods, and Future Directions. IEEE Signal Processing Magazine, 2020, Vol. 37, No. 3, pp. 50–60.

15. Akhmetov, B. S., & Lakhno, V. A. Explainable AI models in cyberattack detection: methodology and practical applications. Applied Informatics, 2023, Vol. 18, No. 1, pp. 14–26.

16. Isabekov, D. T., & Seitkali, D. N. Methods of balancing training datasets in intrusion detection systems. Bulletin of KazNU. Series: Mathematics, Mechanics, Informatics, 2022, No. 4, pp. 66–74.

17. Goodfellow, I. J. et al. Explaining and Harnessing Adversarial Examples. International Conference on Learning Representations (ICLR), 2015. arXiv:1412.6572.