**Обнаружение и классификация вредоносного ПО с применением искусственного интеллекта**

**Амангелді Ернұр**

[Salomat423@gmail.com](mailto:Salomat423@gmail.com)

*Кафедра систем информационной безопасности, факультет информационных технологий,*

*Евразийский национальный университет им. Л.Н. Гумилева, Астана, Казахстан*

*Научный руководитель – PhD, и.о. доцента Муханбеткалиева А.К.*

**Аннотация**

В последние годы, искусственный интеллект (ИИ) превратился в мощный инструмент для борьбы со стремительно растущим количеством вредоносных программных обеспечений (ПО). Эта статья представляет краткий обзор возможностей на основе ИИ для автоматизации обнаружения вредоносных программ и классификаций за период с 2018 по 2025 год.

Исследование суммирует главные нейронные архитектуры, ключевые датасеты, практические результаты, существующие сложности и большинство направлений исследований. Особое внимание уделено на сравнение основанных на ИИ подходах со стандартными методами защиты, оценке их преимуществ и ограничениях в реальный сценариях.

С каждым годом, вредоносное программное обеспечение становится более сложным для его обнаружения. Стандартные методы обнаружения основанные на сигнатурах и заранее установленных правил показали неэффективность в идентификации новых, непредвиденных атак, особенно в уязвимостях нулевого дня. Методы статистического анализа легко обходятся через упаковки, обфускации или незначительных изменений кода, что позволяет киберпреступникам генерировать бесконечные вариации уникальных образцов вредоносного ПО [1, 2, 4].

Выходит, что искусственный интеллект, в частности машинное и глубокое обучение стали многообещающей альтернативой. Эти методы могут выявлять сложные зависимости функций, учиться с больших датасетов и адаптироваться к возрастающим типам угроз. К 2024 году больше половины коммерческих продуктов кибербезопасности стали интегрировать компоненты основанные на ИИ, включая облачные антивирусы и EDR системы [2, 9, 10] .

**Введение**

За последние годы в сфере изучении опознания и идентификации ранее незримых вариантов вредоносных программ значительно улучшились методы машинного обучения (МО) и глубокого обучения (ГО), сверточные нейронные сети (СНС) и гибридные модели при помощи обучения комплексных поведенческих и структурных паттернов из огромных датасетов [1, 2], изучаются графические модели определения структуры вредоносного кода [3].

Сверточные нейронные сети показали свою эффективность в статистическом анализе и конвертации исполняемых файлов в цветовой режим изображений, которые отображаются в оттенках серого цвета или байтовой последовательности для распознавания образов [6], пока рекуррентная нейронная сеть (РНС) и основанные на долгой краткосрочной памяти (ДКП) сети эффективно работают в анализе таких динамических поведенческих особенностей как системные вызовы [7]. Основанные на трансформных архитектурах исследованы возможности обнаружения запутанных кодов и сложные угрозы при помощи фиксации долгосрочных зависимостей в последовательных данных [10]. Относительно недавние исследования также отметили потенциал возможностей графовых нейронных сетей (ГНС) в моделировании структур кода и обнаружения вирусов [3]. Разработаны тесты для обучения CIC-MalMem [2], EMBER [5], Malimg и Microsoft BIG 2015 [6].

Однако большинство проблем в кибербезопасности остаются в силе: враждебные атаки, запутывание, дисбаланс данных и ход связей моделей значительно снижают уровень надежности и достоверности моделей [1, 2, 4], а также высокие вычислительные требования и ограниченные возможности доступности данных остаются практическим барьером для внедрения систем [1].

Сравнение с традиционными подходами основанных на сигнатурах показывает, что пока МО модели превосходят статистические методы в обнаружениях, они восприимчивы к уклонению от состязательности и требуют постоянной переподготовки [1, 4]. Тем не менее, реальные ИИ приложения все чаще интегрируются в антивирусные обеспечения, системы обнаружения конечных точек и реагирования на них и инструменты сетевой безопасности [9, 10].

Новые тенденции объяснимого искусственного интеллекта (ОИИ), обучение с помощью нескольких шагов, мета-обучение и адаптация моделей в режиме реального времени показывают, что необходимо создать более устойчивые и адаптируемые системы обнаружения вредоносных программ [2, 3, 10]. Исследование нацелено на обобщение недавних достижений и идентификацию будущих направлений на создание интеллектуальных и надежных систем безопасности.

**Подходы и архитектуры ИИ**

Современные системы обнаружения с возрастсанием полагаются на архитектуры ГО способные на обработку как статических так, и динамических возможностей функции ПО. Среди самых частых используемых моделей являются свёрточные нейронные сети (СНС), рекурентные нейронный сети (РНС), трансформеры и гибридные архитектуры [1][2][10].

СНС используются для анализа бинарных файлов, представляющих изображения или байтовые последовательности. Этот метод доказал свою эффективность в таких датасетах как Malimg и Microsofft BIG 2015, где классификация вредоносных ПО достигла 96.6% точности на основе визуализации PE файлов [6]. РНС хорошо справляется с динамическим анализом, к примеру классификации поведении ПО через последовательности в системных вызовах [7].

С 2020 года трансформерные модели захватили большое внимание в связи с их возможности фиксировать долгосрочные зависимости. Они могут анализировать последовательности вызовов API, токены и поведения в PowerShell или JavaScript скриптах. Благодаря механизмам саморегулирования, трансформерные модели имеют высокий уровень точности даже при наличии помех или небольших изменений [10].

Гибридные модели комбинируют СНС, РНС и трансформерные модели также получили широкое распространение. Эти модели захватывают локальные и глобальные зависимости кода [2].  Кроме того, возрастает интерес к графовым нейронным сетям (ГНС), которые могут анализировать контроль графов, функции вызова графов и других структур [3].

Для обучения этих моделей, были использованы разные датасеты: EMBER [5] , CIC-MalMEm [2] , Malimg [6], а также VirusShare и Drebin (вредоносные программы для Android) [4] являются одними из наиболее распространенных. Однако многие из этих наборов данных устарели, несбалансированы или не отражают возникающие угрозы [2, 6].

**Проблемы и перспективы**

Несмотря на явный прогресс в использовании основанных на ИИ программ для обнаружения вредоносного ПО, некоторые сложности остаются не решенными. Одной из ключевых проблем является устойчивость модели к методам обхода, такие как упаковка, обфускация, полиморфизм, а также к враждебным атакам, в те моменты когда образцы вредоносного ПО меняются, модифицируются для попыток обхода классификаторов [2, 4] .

Вторая проблема это отсутствие возможности интерпретации. ГНС модели, особенно трансформерные, часто не дают объяснения почему какой-либо образец помечен как вредоносный [2, 10]. Такие моменты ограничивают уровень доверия к результатам анализа и усложняют внедрения в критические системы.

Также проблемой является недостаток данных для обучения или их недостоверность. Многие часто используемые датасеты устарели или не справляются с отражением новых типов вредоносных ПО, что сильно влияет на саму модель [2, 6].

Однако, открываются потенциально успешные тенденции. Исследования в методах объяснимого искусственного интеллекта (ОИИ) становятся все популярнее. Методы обучения с несколькими кадрами позволяют моделям хорошо работать даже на небольших или ранее неизвестных наборах данных [2, 10]. На даннный момент, изучаются графические модели определения структуры вредоносного кода [3]. Современные исследования работают на повышение устойчивости к обходу и добавлению обновлений в модель в реальном времени без полных переделок.

**Заключение**

Таким образом, использование ИИ значительно улучшило обнаружение вредоносных ПО и возможностей классификации. Современные архитектуры такие как СНС, РНС и трансформеры предоставляют высокую точность и являются способными к идентификации новых непредвиденных угроз. Однако присутствуют значительные угрозы: адверсариальные атаки, обфускация, нехватка интерпретируемости, ограниченность данных. ИИ модели требуют продолжительной адаптации к новым угрозам и оптимизации к реальным угрозам.

Будущие направления включают в себя развитие разработки ОИИ, обучение на малых выборках, адаптации к работе в реальном времени и графовые подходы. Эти инновации открывают путь от исследовательских прототипов к надежным, масштабным, а главное рабочим решениям в области кибербезопасности. ИИ не заменяет стандартные методы обнаружения, а дополняет их, создавая основу для будущих гибридных платформ кибербезопасности.

**Список использованной литературы:**

[1] Bensaoud, A., Kalita, J., & Bensaoud, M. (2024). A survey of malware detection using deep learning. <https://doi.org/10.48550/arXiv.2407.19153>  
[2] Tayyab, U.-e.-H., Khan, F. B., Durad, M. H., Khan, A., & Lee, Y. S. (2022). A survey of the recent trends in deep learning-based malware detection. Journal of Cybersecurity and Privacy, 2(4), 800–829. <https://doi.org/10.3390/jcp2040041>  
[3] Bilot, T., El Madhoun, N., Al Agha, K., & Zouaoui, A. (2023). A survey on malware detection with graph representation learning. ACM Computing Surveys, 56(11), 1–36. <https://doi.org/10.1145/3664649>  
[4] Ucci, D., Aniello, L., & Baldoni, R. (2019). Survey of machine learning techniques for malware analysis. Computers & Security, 81, 123–147. <https://doi.org/10.1016/j.cose.2018.11.001>  
[5] Anderson, H. S., & Roth, P. (2018). EMBER: An open dataset for training static PE malware machine learning models. arXiv:1804.04637. <https://doi.org/10.48550/arXiv.1804.04637>  
[6] Ni, S., Qian, Q., & Zhang, R. (2018). Malware identification using visualization images and deep learning. Computers & Security, 77, 871–885. <https://doi.org/10.1016/j.cose.2018.03.048>  
[7] Kolosnjaji, B., Zarras, A., Martinez, G., & Eckert, C. (2018). Deep learning for classification of malware system call sequences. In Proc. 26th European Signal Processing Conference (EUSIPCO) (pp. 533–537). <https://doi.org/10.23919/EUSIPCO.2018.8553102>  
[8] Onwuzurike, L., Mariconti, E., et al. (2019). MaMaDroid: Detecting Android malware by building Markov chains of API calls. IEEE Transactions on Information Forensics and Security, 14(5), 1309–1322. <https://doi.org/10.1109/TIFS.2018.2870899>  
[9] Kim, J., Lee, J., Ko, S., & Lee, S. (2024). A machine learning-enhanced endpoint detection and response framework for fast and proactive defense against advanced cyber attacks. Soft Computing, 28(13), 7807–7821. <https://doi.org/10.1007/s00500-024-09727-7>  
[10] Alshomrani, M., Albeshri, A., Alturki, B., et al. (2023). Survey of transformer-based malicious software detection systems. Electronics, 13(23), 4677. https://doi.org/10.3390/electronics13234677