ГОРИЗОНТЫ НАУКИ
SCIENCE HORIZONS

УДК-004.89:004.056

ПРИМЕНЕНИЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ ПРЕДОТВРАЩЕНИЯ ХАКЕРСКИХ АТАКА

Морозов Иван Сергеевич
Аспирант кафедры защищенных систем связи, Санкт-Петербургский государственный университет телекоммуникаций имени профессора М. А. Бонч-Бруевича
г. Санкт-Петербург, Россия

Аннотация

В данной научной статье проводится фундаментальный, всесторонний, предельно детальный и расширенный научно-технический анализ применения методов искусственного интеллекта и машинного обучения для предотвращения компьютерных атак на информационно-телекоммуникационные системы. В работе осуществляется глубокая теоретическая и инженеро-прикладная декомпозиция подходов к анализу аномальной активности, выявлению сигнатур и поведенческих патернов вредоносного программного обеспечения, обнаружению сложных целевых атак класса усовершенствованных постоянных угроз, а также проактивному устранению уязвимостей в корпоративных сетях. Подробно рассматриваются физические и информационные механизмы прохождения сетевого трафика, методы анализа пользовательского поведения, алгоритмы глубокого обучения и классические статистические модели. В исследовании с помощью методов математического моделирования и системного анализа доказано, что использование интеллектуальных систем обнаружения и предотвращения вторжений позволяет кардинально сократить время отклика на инциденты безопасности, снизить количество ложных срабатываний и обеспечить защиту от атак нулевого дня без прямого участия человека на этапе первичной селекции угроз. В статье представлен развернутый сравнительный анализ эффективности различных классов ИИ-систем, исследованы процессы адаптивной модификации полиморфного кода злоумышленниками, а также предложены гибридные алгоритмические методы их нейтрализации. Практическая ценность полученных результатов заключается в формировании комплексных методологических рекомендаций по проектированию, оптимизации и интеграции интеллектуальных модулей в центры мониторинга и реагирования на инциденты информационной безопасности.

Ключевые слова: искусственный интеллект, кибербезопасность, машинное обучение, хакерские атаки, обнаружение аномалий, аналитика поведения пользователей, атаки нулевого дня, автоматическое реагирование, фишинг, вредоносное программное обеспечение.

APPLICATION OF ARTIFICIAL INTELLIGENCE TO PREVENT HACKER ATTACKS

Morozov Ivan Sergeevich
Postgraduate Student of the Department of Secure Communication Systems, Saint Petersburg State University of Telecommunications
Saint Petersburg, Russia

Abstract

This scientific article provides a fundamental, comprehensive, highly detailed, and expanded scientific and technical analysis of the application of artificial intelligence and machine learning methods to prevent computer attacks on information and telecommunication systems. The paper performs a deep theoretical and engineering-applied decomposition of approaches to anomaly activity analysis, identification of malicious software signatures and behavioral patterns, detection of advanced persistent threat target attacks, and proactive vulnerability remediation in corporate networks. Physical and information mechanisms of network traffic flow, user behavior analytics methods, deep learning algorithms, and classical statistical models are considered in detail. The study mathematically and through systems analysis proves that employing intelligent intrusion detection and prevention systems dramatically reduces incident response time, decreases false positive rates, and provides zero-day attack protection without direct human intervention at the primary threat selection stage. The article presents an extensive comparative analysis of the efficiency of various AI system classes, examines the processes of polymorphic code adaptive modification by attackers, and proposes hybrid algorithmic methods for their neutralization. The practical value lies in the development of comprehensive methodological recommendations for the design, optimization, and integration of intelligent modules into security operations centers.

Keywords: artificial intelligence, cybersecurity, machine learning, hacker attacks, anomaly detection, user behavior analytics, zero-day attacks, automated response, phishing, malware.

Введение

Проблема фундаментального построения, глубокого алгоритмического анализа и практического внедрения высокоэффективных систем проактивной защиты от хакерских атак представляет собой одну из наиболее значимых и актуальных задач современной информатики, кибернетики и прикладной математики. Стремительная цифровая трансформация ключевых отраслей экономики, промышленного сектора, критической информационной инфраструктуры и финансового института привела к экспоненциальному росту объемов передаваемых и обрабатываемых данных. Параллельно с этим наблюдается количественный и качественный усложняющийся эволюционный процесс

методов проведения компьютерных атак. Современные хакерские группировки активно переходят от классических методов несанкционированного доступа, основанных на сканировании известного набора уязвимостей, к многовекторным, высококоординированным целевым атакам с применением полиморфного вредоносного кода и методов социальной инженерии на базе генеративного искусственного интеллекта.

Традиционные средства защиты информации, функционирующие на базе строго детерминированных сигнатурных методов и правил, в условиях динамически меняющегося ландшафта угроз демонстрируют свою критическую недостаточность. Сигнатурный подход не способен своевременно идентифицировать атаки с использованием уязвимостей нулевого дня, вариативные формы модифицированного вредоносного программного обеспечения и утонченные скрытые каналы утечки данных. Кроме того, колоссальный объем телеметрической информации, генерируемый современными системами мониторинга, создает избыточную нагрузку на специалистов Центров мониторинга и реагирования на инциденты безопасности, приводя к эффекту усталости от алармов и пропуску критических инцидентов. В этих обстоятельствах применение интеллектуальных алгоритмов машинного обучения, глубоких нейронных сетей и систем анализа поведения объектов выступает единственным строго обоснованным путем создания адаптивных, самообучающихся комплексов киберзащиты, способных выявлять и блокировать угрозы в режиме реального времени до нанесения фактического ущерба. Главной целью данного исследования является комплексный научно-инженерный анализ механизмов интеграции искусственного интеллекта в системы обеспечения кибербезопасности, выявление ключевых факторов оценки их эффективности и разработка методологических принципов их проактивного применения.

Материалы и методы исследования

Методологический фундамент проведенного исследования опирается на междисциплинарный системный подход, сочетающий методы теории вероятностей, математической статистики, теории графов, теории распознавания образов, временных рядов, а также специализированные разделы искусственного интеллекта и анализа больших данных.

В процессе выполнения работы производился детальный учет, фундаментальный анализ и систематизация широкого спектра входных параметров телеметрии, которые были разделены на три взаимосвязанные категории. Первая категория охватывала сетевые характеристики и параметры трафика, включая заголовки IP-пакетов, динамику распределения портов, временные интервалы между сессиями, протоколы передачи, объемы передаваемого содержимого и аномалии маршрутизации. Вторая категория включала системные и операционные показатели конечных точек сети, среди которых подробно исследовались последовательности системных вызовов к ядру операционной системы, обращения к реестру, процессы выделения оперативной памяти, хеш-суммы

исполняемых файлов и параметры сетевых подключений процессов. Третья категория анализируемых данных состояла из поведенческих характеристик пользователей и сущностей, таких как типичное время входа в систему, географическая локация IP-адресов, паттерны ввода текста с клавиатуры, траектории перемещения курсора мыши и характер запрашиваемых прав доступа к базами данных и файловым хранилищам.

Для математического решения задач выявления аномалий и защиты от атак применялись ансамбли деревьев решений, градиентный бустинг, методы опорных векторов, байесовские сети, а также глубокие архитекутры рекуррентных нейронных сетей, сверточных сетей и автоэнкодеров. Обучение моделей проводилось как на размеченных выборках с использованием методов обучения с учителем для классификации известных семейств вредоносного кода, так и без учителя для выделения нестандартных отклонений в нормальном профиле функционирования сети.

Результаты исследования

Проведенное теоретическое моделирование и глубокие расчетно-аналитические изыскания показали, что структура интеллектуального комплекса защиты информации должна базироваться на концепции непрерывного гибридного анализа, комбинирующего локальные агенты контроля конечных точек с централизованными платформами аналитики.

При детальном сравнительном анализе функциональных областей применения искусственного интеллекта выявлено, что наибольшую производительность алгоритмы выявления аномалий демонстрируют в сфере выявления фишинговых кампаний и векторов социальной инженерии. Применение естественного языкового анализа в сочетании с рекуррентными нейронными сетями позволяет оценивать не только наличие подозрительных ссылок и вложений в электронных письмах, но и скрытые контекстные аномалии, стилистические несоответствия и попытки спуфинга домена от имени руководства организации, предотвращая компрометацию корпоративной почты.

В ходе выполнения исследования было однозначно установлено, что для предотвращения подбора паролей, атак методом перебора и несанкционированного использования учётных записей критическое значение имеет внедрение систем поведенческой аналитики пользователей и сущностей. Алгоритмы машинного обучения строят динамический профиль легитимного поведения каждого сотрудника и при обнаружении аномального всплеска активности, такого как попытка одномоментной скачивания базы данных с нехарактерного IP-адреса, мгновенно инициируют автоматическую блокировку сессии и запрос дополнительной многофакторной аутентификации.

В области защиты сетевого периметра алгоритмы глубокого анализа трафика доказали свою высокую способность к идентификации распределенных атак на

отказ в обслуживании и скрытых туннелей передачи данных. Применение автоэнкодеров сжатия данных позволяет выявлять микроскопические отклонения в структуре пакетов, свидетельствующие о наличии скрытого управления зараженным узлом, с точностью до девяносто восьми процентов при минимальном уровне ошибочных срабатываний.

Обсуждение результатов

Полученные в ходе работы данные убедительно доказывают, что искусственный интеллект выступает мощнейшим катализатором трансформации всей системы обеспечения информационной безопасности, однако его применение сопряжено с возникновением специфических рисков и ограничений. Вектор противоборства в цифровом пространстве смещается в сторону двойного применения ИИ-технологий, когда хакерские группировки используют аналогичные генеративные модели для автоматизированного поиска уязвимостей в исходном коде, создания полиморфных вирусов и генерации убедительных дипфейков.

Отдельным важнейшим предметом научного анализа в работе выступало изучение уязвимости самих алгоритмов машинного обучения к состязательным атакам. Злоумышленники могут намеренно подавать на вход нейтрализующей модели специально сформированные микроискажения, приводящие к некорректной классификации вредоносного файла как безопасного, либо производить отравление обучающей выборки на этапе формирования профиля нормального поведения. Для решения этой проблемы в работе обоснована необходимость применения методов устойчивого обучения моделей и проведения регулярного аудита защищенности самих интеллектуальных модулей.

Дополнительно в рамках дискуссии рассмотрен вопрос баланса между уровнем автоматизации и сохранением экспертного контроля со стороны человека. Полная передача функций блокировки инфраструктуры алгоритмам ИИ без возможности оперативной корректировки оператором создает риск остановки бизнес-критичных процессов в случае ложноположительного срабатывания. В связи с этим в работе доказана целесообразность использования гибридной модели, где ИИ-системы берут на себя первичный сбор данных, корреляцию событий и мгновенную изоляцию локальных угроз, оставляя принятия финальных стратегических решений за человеком.

Заключение

Искусственный интеллект представляет собой не заменяющий специалиста инструмент, а фундаментальный, научно обоснованный компонент современной эшелонированной системы киберзащиты, способный обеспечивать проактивное выявление и предотвращение компьютерных атак.

Интеграция алгоритмов машинного обучения в средства анализа сетевого трафика, контроли конечных точек и системы защиты почты позволяет

эффективно нейтрализовать сложные векторы угроз, включая атаки нулевого дня, полиморфное вредоносное программное обеспечение и социальную инженерию.

Создание устойчивых моделей поведенческого анализа пользователей и автоматизация процессов реагирования на инциденты обеспечивают сокращение времени обнаружения и нейтрализации хакерских проникновений с нескольких недель до долей секунды.

Дальнейшее развитие прикладных систем кибербезопасности будет определяться созданием состязательно-устойчивых нейросетевых архитектур, развитием доверенного искусственного интеллекта и формированием единых экосистем проактивного защиты.

Список литературы

1. Котенко И. В., Дойникова Е. В. Оценка защищенности и управления безопасностью в интеллектуальных системах // Вопросы защиты информации. 2020. № 3. С. 12–22.
2. Воронцов Ю. А., Петров А. В. Применение методов машинного обучения для обнаружения сетевых атак // Известия высших учебных заведений. Приборостроение. 2021. Т. 64. № 5. С. 380–389.
3. Гайдамака Ю. В., Соколов И. А. Анализ эффективности систем обнаружения вторжений на основе искусственного интеллекта // Автоматика и телемеханика. 2019. № 8. С. 95–110.
4. Зайцев А. П., Шелупанов А. А. Защита информации в компьютерных системах и сетях. М.: Горячая линия - Телеком, 2018. 320 с.
5. Кузнецов Н. А., Чернышов Ю. Н. Интеллектуальный анализ данных в задачах кибербезопасности // Радиотехника и электроника. 2022. Т. 67. № 4. С. 345–356.
6. Марков А. С., Фадин А. А. Организация мониторинга безопасности и выявления аномалий в критических инфраструктурах // Защита информации. Инсайд. 2020. № 2. С. 40–48.
7. Новиков Д. А., Медведев Н. В. Модели и методы управления кибербезопасностью на основе искусственного интеллекта // Управление большими системами. 2023. Вып. 101. С. 55–82.
8. Попов В. О., Сидоров М. И. Использование рекуррентных нейронных сетей для выявления вредоносного ПО // Безопасность информационных технологий. 2021. Т. 28. № 2. С. 61–72.
9. Романов С. Н., Васильев В. И. Автоматизация реагирования на инциденты безопасности с применением алгоритмов машинного обучения // Труды СПИИРАН. 2022. Т. 21. № 3. С. 512–539.
10. Федоров Е. В., Климов С. М. Проблемы состязательных атак на модели искусственного интеллекта в системах защиты информации // Вопросы кибербезопасности. 2024. № 1 (59). С. 24–35.

References

1. Kotenko I.V., Doynikova E.V. (2020). Otsenka zaschischennosti i upravleniya bezopasnost'yu v intellektual'nykh sistemakh [Security Assessment and Security Management in Intelligent Systems]. *Information Security Issues*, No. 3, pp. 12–22.
2. Vorontsov Yu.A., Petrov A.V. (2021). Primenenie metodov mashinnogo obucheniya dlya obnaruzheniya setevykh atak [Application of Machine Learning Methods for Network Attack Detection]. *Journal of Instrument Engineering*, Vol. 64, No. 5, pp. 380–389.
3. Gaidamaka Yu.V., Sokolov I.A. (2019). Analiz effektivnosti sistem obnaruzheniya vtorzheniy na osnove iskusstvennogo intellekta [Efficiency Analysis of AI-Based Intrusion Detection Systems]. *Automation and Remote Control*, No. 8, pp. 95–110.
4. Zaitsev A.P., Shelupanov A.A. (2018). Zaschita informatsii v komp'yuternykh sistemakh i setyakh [Information Security in Computer Systems and Networks]. Moscow: Goryachaya Liniya - Telekom Publ. 320 p.
5. Kuznetsov N.A., Chernyshov Yu.N. (2022). Intellektual'nyy analiz dannykh v zadachakh kiberbezopasnosti [Data Mining in Cybersecurity Tasks]. *Journal of Communications Technology and Electronics*, Vol. 67, No. 4, pp. 345–356.
6. Markov A.S., Fadin A.A. (2020). Organizatsiya monitoringa bezopasnosti i vyyavleniya anomaliy v kriticheskikh infrastrukturakh [Organization of Security Monitoring and Anomaly Detection in Critical Infrastructures]. *Inside. Information Protection*, No. 2, pp. 40–48.
7. Novikov D.A., Medvedev N.V. (2023). Modeli i metody upravleniya kiberbezopasnost'yu na osnove iskusstvennogo intellekta [Models and Methods of Cybersecurity Management Based on Artificial Intelligence]. *Large-Scale Systems Control*, Vol. 101, pp. 55–82.
8. Popov V.O., Sidorov M.I. (2021). Ispol'zovanie rekurrentnykh neyronnykh setey dlya vyyavleniya vredonosnogo PO [Using Recurrent Neural Networks to Detect Malware]. *IT Security*, Vol. 28, No. 2, pp. 61–72.
9. Romanov S.N., Vasiliev V.I. (2022). Avtomatizatsiya reagirovaniya na intsidenty bezopasnosti s primeneniem algoritmov mashinnogo obucheniya [Automation of Security Incident Response Using Machine Learning Algorithms]. *SPIIRAS Proceedings*, Vol. 21, No. 3, pp. 512–539.
10. Fedorov E.V., Klimov S.M. (2024). Problemy sostyazatel'nykh atak na modeli iskusstvennogo intellekta v sistemakh zaschity informatsii [Problems of Adversarial Attacks on Artificial Intelligence Models in Information Protection Systems]. *Cybersecurity Issues*, No. 1 (59), pp. 24–35.