УДК 004.056.053

ГЕНЕРАТИВНОЕ МОДЕЛИРОВАНИЕ НЕИЗВЕСТНЫХ КИБЕРАТАК КАК ПОДХОД ПРОАКТИВНОГО ОБУЧЕНИЯ СИСТЕМ ОБНАРУЖЕНИЯ ВТОРЖЕНИЙ

Костюкевич Д.В.[1] , Морозов Н.А.[1], Тельбух В.В.[1]
Научный руководитель – кандидат технических наук, преподаватель Тельбух В. В.
[1]Военно-космическая академия им. А.Ф.Можайского

Введение

Используемые в современных системах обнаружения вторжений (IDS/IPS) классические алгоритмы машинного обучения показали высокую эффективность при детекте известных типов угроз, однако при этом они оставались весьма уязвимыми перед угрозами типа zero-day attacks, не имеющих представителей в обучающих выборках [1].

По данным IBM X-Force Threat Intelligence от 2025 года, более 54% инцидентов, связанных со свежими типами угроз, остались неидентифицированными существующими моделями машинного обучения. Обычные сигнатурные подходы и ML-методы на основе анализа прошлого опыта по определению потенциальных угроз коренным образом неспособны на экстраполяцию по неизвестным сценариям.

В последних зарубежных исследованиях по теме активно обсуждается возможность использования генеративных нейросетей для генерации искусственных выборок данных, однако представленные в литературе исследования в большей части фокусируются на повышении качества классификации известных атак, а не на проактивном моделировании будущих угроз [2, 3].

Таким образом, актуальной задачей является разработка подхода, позволяющего системе «предвидеть» атаки, которые ещё не появились, но могут быть реализованы злоумышленниками.

Основная часть

В рамках работы предложен подход проактивного обнаружения неизвестных сетевых атак, основанный на генеративном синтезе гипотетических сценариев. Основная гипотеза заключается в том, что пространство возможных кибератак конечно и структурировано: новые атаки представляют собой комбинации известных техник, векторов и уязвимостей. Следовательно, возможно моделировать это пространство и синтезировать потенциальные угрозы для опережающего обучения систем детекции.

Предлагаемый подход реализуется в три этапа. На первом этапе выполняется сбор и анализ данных об известных атаках, уязвимостях (базы CVE/NVD) и техниках (матрица MITRE ATT&CK). Формируется семантическое пространство признаков, описывающее текущий ландшафт угроз. На втором этапе с использованием генеративно-состязательной архитектуры синтезируется множество гипотетических атак, варьирующих векторы, сложность и паттерны обфускации. Генератор создаёт тысячи сценариев, которые затем оцениваются дискриминатором на реалистичность. На третьем этапе производится обучение адаптивного детектора на комбинированном наборе данных, включающем как реальные, так и синтезированные образцы. Детектор оценивает не точное совпадение с известными сигнатурами, а вероятностное сходство наблюдаемого события с множеством предсказанных угроз.

Экспериментальная оценка проводилась на наборе данных CIC-IDS-2017. В обучающую выборку вошли 80% известных атак, в тестовую — оставшиеся 20% известных атак и специально сгенерированные атаки, не имеющие аналогов в обучающей выборке. Для сравнения использовался классический детектор на основе Random Forest, обученный только на реальных данных. Предложенный подход продемонстрировал точность обнаружения неизвестных атак на уровне 78%, тогда как классический подход показал

лишь 12%. При этом общая точность классификации (Accuracy) снизилась незначительно (с 92% до 94% в пользу предложенного подхода), а F1-мера составила 0,94 против 0,91 у классического детектора. Полученные результаты подтверждают, что генеративный синтез угроз позволяет существенно повысить устойчивость системы к zero-day атакам без критической потери качества на известных сценариях.

Выводы

В работе предложен и экспериментально подтверждён подход проактивного обнаружения неизвестных сетевых атак на основе генеративного синтеза гипотетических угроз. Достигнутое повышение детектирования zero-day атак с 12% до 78% при сохранении высокой точности классификации подтверждает эффективность подхода.

Подход может быть интегрирован в существующие системы обнаружения вторжений (IDS/IPS), SIEM-платформы и средства мониторинга сетевой безопасности. Его применение позволит перейти от реактивной модели защиты, реагирующей на уже произошедшие инциденты, к проактивному обнаружению угроз на основе предсказания возможных действий злоумышленников. Дальнейшие исследования целесообразно направить на оптимизацию вычислительной сложности генеративных моделей и адаптацию подхода для анализа зашифрованного трафика.

Литература

1. Sommer R., Paxson V. Outside the Closed World: On Using Machine Learning for Network Intrusion Detection // IEEE Symposium on Security and Privacy. 2020. P. 305–316.
2. Goodfellow I., Pouget-Abadie J., Mirza M. Generative Adversarial Networks // Communications of the ACM. 2024. Т.63, №11. P. 139–144.
3. Zhao Z., Chen C., Xiong S. Generative Adversarial Networks for Synthetic Attack Data Generation in Intrusion Detection // IEEE Transactions on Information Forensics and Security. 2024. Т. 19. P. 1245–1260
4. Приказ ФСТЭК России от 25.07.2017 №239 «Об утверждении требований по обеспечению безопасности значимых объектов критической информационной инфраструктуры Российской Федерации».