УДК 004.056.55

ИСПОЛЬЗОВАНИЕ ГЕНЕРАТИВНЫХ МОДЕЛЕЙ ИИ ДЛЯ МАСКИРОВКИ АНОМАЛЬНОЙ СЕТЕВОЙ АКТИВНОСТИ В РАСПРЕДЕЛЁННЫХ СИСТЕМАХ

М. А. Баранчик, Е. Г. Крупенич, С. А. Мигалевич

Белорусский государственный университет информатики и радиоэлектроники, Минск, Беларусь, runolero@gmail.com

Рассматривается новая потенциальная угроза кибербезопасности, связанная с использованием генеративных моделей искусственного интеллекта (GAN, VAE и др.) для маскировки аномальной сетевой активности в распределённых системах. Теоретически обосновывается возможность использования моделей, обученных на нормальном пользовательском поведении, для сокрытия вредоносных действий. Предлагаются возможные меры противодействия.

Ключевые слова: генеративные модели; информационная безопасность; аномальная сетевая активность; распределённые системы.

USING GENERATIVE AI MODELS TO MASK ANOMALOUS NETWORK ACTIVITIES IN DISTRIBUTED SYSTEMS

M. A. Baranchik, E. G. Krupenich, S. A. Migalevich

Belarusian State University of Informatics and Radioelectronics, Minsk, Belarus, runolero@gmail.com

A new potential cybersecurity threat is considered, associated with the use of generative artificial intelligence models (GAN, VAE, etc.) to mask anomalous network activity in distributed systems. The possibility of using models trained on normal user behavior to hide malicious actions is theoretically substantiated. Possible countermeasures are proposed.

Keywords: generative models; information security; anomalous network activity; distributed systems.

1. Введение

Современные распределённые информационные системы характеризуются высокой степенью масштабируемости, неоднородностью компонентов и динамичной структурой взаимодействия между узлами [1]. Компоненты распределённой инфраструктуры обмениваются значительными объёмами данных, часто в реальном времени, что создаёт множество точек

потенциального внедрения злоумышленника. Угрозы, использующие такие уязвимости, проявляются в виде статистически или поведенчески отклоняющегося от нормы трафика.

Одним из основных инструментов мониторинга и защиты сетевого периметра являются системы обнаружения вторжений (Intrusion Detection Systems, IDS), действующие на основе анализа сетевого трафика [2]. IDS выполняют автоматизированный сбор и обработку информации о проходящих через сеть пакетах данных, сопоставляя полученные метаданные с заранее известными сигнатурами атак или оценивая статистические и поведенческие аномалии (Heuristics-Based Anomaly Detection).

Среди систем обнаружения вторжений (IDS/IPS) выделяют два типа.

– Синтаксические (сигнатурно-ориентированные) выявляют атаки путём сравнения сетевого трафика с базой известных шаблонов (сигнатур), где каждая сигнатура описывает последовательность байтов или специфические фрагменты пакетов, характерные для конкретного вида угрозы. Такие системы эффективны против ранее задокументированных атак, но оказываются бесполезными при появлении новых, неизвестных видов вредоносного ПО.

– Поведенческие (анализ аномалий) строят модель «нормального» трафика на основе статистических, временных или семантических характеристик, а затем ищут отклонения от этой базы. Ключевой недостаток данного подхода заключается в высокой вероятности ложных срабатываний при изменении легитимных сценариев работы системы и возможном адаптивном поведении злоумышленника.

С развитием генеративных моделей ИИ появляется новая угроза – возможность маскировки вредоносной активности под легитимное поведение, обучая модели на «норме» и встраивая отклонения в рамки ожидаемых параметров.

2. Аномальная активность как форма вредоносного воздействия

В настоящее время наиболее распространены следующие формы вредоносного воздействия.

1. Сетевое сканирование и разведка. Атаки начинаются с непрерывного отправления запросов к большому количеству портов и служб с целью выявить активные сервисы и определить уязвимости на удалённых узлах. Часто такие сканирования выполняются распределённо (распределённые сканирующие боты), что позволяет снизить уровень подозрительности с точки зрения традиционных IDS.

2. Эксплуатация уязвимостей приложений и служб. После разведки злоумышленник использует конкретные уязвимости (например, RCE,

SQL-инъекции, ошибки в протоколах) для внедрения вредоносного кода или получения несанкционированного доступа. Данная активность проявляется в виде аномальных последовательностей запросов, отклоняющихся от шаблонов легитимного трафика.

3. DDoS-атаки. Распространённая форма вывода сервиса из строя путём создания огромного объёма трафика с большого числа скомпрометированных устройств. Такое поведение характеризуется резким ростом частоты и объёмов пакетов, и традиционные IDS/IPS хорошо детектируют подобные всплески, однако при «мягких» DDoS-сценариях (slowloris, HTTP-флуд с малой скоростью) аномалия становится менее заметной.

4. Эксфильтрация данных. Скрытая передача конфиденциальной информации из внутренней сети на внешние ресурсы. Злоумышленники часто используют зашифрованные каналы (HTTPS, DNS-туннели), пытаясь замаскировать объёмы передаваемых данных и их частоту под обычный пользовательский трафик.

5. Brute-force атаки (перебор аутентификационных данных). Повторные попытки входа в систему с разными учётными данными генерируют аномально большое количество запросов к механизму аутентификации. Хотя отдельные точки остаются незаметными, совокупность таких попыток создаёт характерный паттерн, легко обнаруживаемый поведенческими IDS [3].

3. Механизм маскировки аномальной активности с использованием генеративных моделей

Современные генеративные модели искусственного интеллекта, включая генеративные состязательные сети (GAN), вариационные автокодировщики (VAE) и диффузионные модели, обладают способностью воссоздавать сложные распределения данных, имитируя поведение, неотличимое от нормального [4]. В контексте сетевого трафика это означает возможность синтезировать потоки, статистически соответствующие легитимной активности, при этом скрывая действия, обладающие вредоносной направленностью. Такой подход создаёт угрозу подмены поведения на уровне сетевого взаимодействия, при которой зловредные операции маскируются под фоновую активность, признанную системой мониторинга допустимой.

Злоумышленник, обладая доступом к параметрам нормального сетевого поведения (частоте обращений к определённым портам, типам протоколов, средним и медианным задержкам в ответах, объёмам передаваемых данных и т.п.), может обучить генеративную модель, которая будет имитировать эти характеристики [5]. Затем она используется для внедрения вредоносного поведения в поток, сохраняющий признаки

легитимности. Таким образом, активность, которая по своей сущности нарушает политику безопасности, остаётся невидимой для средств обнаружения, работающих на основе отклонения от нормы.

С практической точки зрения маскировка может реализовываться в форме постобработки вредоносного трафика (перекрытие статистических характеристик) или путём генерации целого массива пакетов, среди которых зловредные команды распределяются во времени и потоке данных так, чтобы не нарушать допустимые пороги, заданные системой обнаружения. Например, эксфильтрация данных может осуществляться малыми порциями, встроенными в имитируемую службу мониторинга или обновления, при этом сигнатура атаки отсутствует, а поведение в целом соответствует допустимому шаблону.

4. Потенциальная уязвимость информационной инфраструктуры

Сигнатурные системы обнаружения полагаются на известные шаблоны атак, включающие специфические последовательности пакетов, характерные байтовые структуры и определённые порты или адреса. При этом любая модификация структуры трафика, не затрагивающая ключевые сигнатуры, остаётся незамеченной. В случае использования генеративной модели злоумышленник имеет возможность формировать уникальные варианты атаки, не подпадающие под существующие шаблоны. Аналогично, детекторы аномалий оценивают отклонения от предварительно зафиксированных норм. Однако генеративный трафик, обладающий той же статистикой, что и реальный, оказывается вне зоны их чувствительности.

Поведенческий анализ основан на предположении, что отклонение от устоявшихся шаблонов действий – признак потенциальной угрозы. При использовании генеративных моделей возникает феномен «синтетической правдоподобности», при котором вредоносное поведение копирует паттерны легитимных процессов [6]. Это подрывает саму основу эвристического анализа, поскольку создаётся впечатление непрерывности и согласованности действий, в действительности нарушающих политику безопасности. В результате формируется ложное чувство защищённости, при котором поведенческие средства мониторинга не только не фиксируют угрозу, но и способствуют её незаметному распространению.

5. Контрмеры и направления развития систем обнаружения вторжений

Одним из перспективных направлений в борьбе с маскирующим вредоносным трафиком является выявление признаков синтетичности сетевых потоков. Несмотря на высокую степень достоверности, достигаемую

современными генеративными моделями, сформированные ими данные могут содержать скрытые следы искусственного происхождения. В частности, при использовании GAN или VAE в сгенерированном трафике возможно присутствие микроскопических регулярностей, отличающихся от естественных флуктуаций живого сетевого взаимодействия. К ним могут относиться, например, неестественно низкая энтропия временных интервалов между пакетами, одинаковые значения редко изменяющихся полей, а также слабо выраженные корреляции между полями различных уровней сетевого стека [7].

Помимо этого, синтетический трафик может демонстрировать недостаточную реакцию при изменении контекста – например, к текущей нагрузке, неожиданным отказам узлов, изменению политик маршрутизации. Таким образом, сочетание анализа скрытых структур данных и мониторинга контекстной адаптивности взаимодействия позволяет выявлять подозрительные потоки, даже если они формально соответствуют статистике нормального поведения.

Как показала практика использования генеративных моделей, наиболее эффективно синтетичность удаётся определять при наличии обученного «дискриминатора» – нейросетевой модели, специализированной на различении реальных и сгенерированных образцов [8]. Интеграция таких моделей в архитектуру системы обнаружения вторжений позволяет не просто анализировать признаки аномальности, а проводить полноценную верификацию достоверности источников трафика.

Подобные дискриминаторы могут быть обучены на специально размеченных выборках, содержащих как настоящие, так и синтетические потоки, сформированные с использованием известных моделей (например, GAN, VAE, diffusion-based). При достаточном разнообразии обучающих данных дискриминатор способен выявлять даже адаптированные к обходу IDS модели, определяя искусственность их поведения не по явным признакам, а по совокупности вторичных характеристик.

6. Заключение

Использование генеративных моделей искусственного интеллекта (GAN, VAE и др.) представляет собой принципиально новый вектор развития угроз кибербезопасности в распределённых системах. За счёт способности имитировать легитимный сетевой трафик с высокой степенью достоверности, такие модели позволяют злоумышленникам эффективно скрывать вредоносную активность, не выходя за рамки поведенческой или статистической нормы, принятой в системах мониторинга. Это подрывает надёжность как сигнатурных, так и поведенческих систем обнаружения вторжений, в корне меняя принципы определения аномального поведения.

Предложенные в работе методы противодействия – в частности, разработка IDS нового поколения с интеграцией дискриминаторов, обученных на различение синтетического и реального трафика – открывают перспективное направление повышения устойчивости информационной инфраструктуры к новым типам угроз. Анализ латентных признаков, поведенческой пластичности и структурных особенностей трафика, дополненный механизмами верификации достоверности источников, позволяет значительно повысить чувствительность к атакам, маскирующимся под норму.

Современные вызовы кибербезопасности требуют переосмысления традиционных подходов к обнаружению угроз и перехода к более гибким, адаптивным системам защиты, способным учитывать появление синтетических форм цифровой активности. В этом контексте критически важно не только развитие технологий обнаружения, но и формирование междисциплинарной исследовательской базы, объединяющей знания в области ИИ, сетевой безопасности и анализа поведения.

Библиографические ссылки

1. Распределенные информационные системы: особенности применения и построения. URL: https://apni.ru/article/6996-raspredelennie-informatsionnie-sistemi-osoben (дата обращения: 05.06.2025).

2. Как обнаруживать вредоносный сетевой трафик. URL: https://labex.io/ru/tutorials/wireshark-how-to-detect-malicious-network-traffic-419254 (дата обращения: 05.06.2025).

3. Advanced Persistent Threat (APT) Таргетированные или целевые кибератаки “Развитая устойчивая угроза”. URL: https://www.tadviser.ru/index.php/Статья:APT_-_Таргетированные_или_целевые_атаки (дата обращения: 05.06.2025).

4. Что такое Generative AI. URL: https://www.sap.com/central-asia-caucasus/products/artificial-intelligence/what-is-generative-ai.html (дата обращения: 05.06.2025).

5. Сетевые аномалии и их обнаружение: как современные системы анализируют трафик. URL: https://www.mobilis.ru/news/setevye-anomalii-i-ikh-obnaruzhenie-kak-sovremennye-sistemy-analiziruyut-trafik-opisanie-tekhnologiy/ (дата обращения: 05.06.2025).

6. *Чио К., Фримэн Д.* Машинное обучение и безопасность: защита систем с помощью данных и алгоритмов (пер. с англ. А. В. Снастина). М. : ДМК Пресс, 2020.

7. Корреляция между временными рядами: что может быть проще? URL: https://habr.com/ru/articles/542638/ (дата обращения: 05.06.2025).

8. Принципы работы генеративных моделей для создания текста и изображений. URL: https://habr.com/ru/companies/fix_price/articles/836206/ (дата обращения: 05.06.2025).

9. Цифровая безопасность как стратегический приоритет: новые вызовы и решения. URL: https://roscongress.org/materials/tsifrovaya-bezopasnost-kak-strategicheskiy-prioritet-novye-vyzovy-i-resheniya/ (дата обращения 05.06.2025).