УДК 004.056

Н.С. НАУМОВА, В.А. РЫЧКОВ

Национальный исследовательский ядерный университет «МИФИ», Москва

МЕТОДЫ ПОИСКА ИНСАЙДЕРА В КОРПОРАТИВНЫХ ИНФОРМАЦИОННЫХ СИСТЕМАХ

Целью работы является анализ и сравнение алгоритмов и моделей машинного обучения, применимых для выявления инсайдерских угроз. Результатом работы является постановка задачи для дальнейшего исследования.

Масштаб и актуальность угрозы инсайдера в крупных корпоративных ИС

Инсайдерские инциденты представляют собой одну из наиболее труднообнаруживаемых и финансово затратных категорий киберрисков. Ключевая особенность внутреннего нарушителя – злоупотребление легитимными правами доступа, что позволяет обходить традиционные периметровые системы защиты.

Согласно глобальным исследованиям, средняя стоимость инцидентов, связанных с инсайдерами, продолжает расти, достигая миллионов долларов на организацию в год. Наиболее частым сценарием является небрежность сотрудников, однако наиболее дорогостоящими - инциденты с компрометацией учетных записей через фишинг.

В российских исследованиях наблюдается дефицит публичных данных об инцидентах, однако аналитика показывает, что наиболее уязвимыми каналами утечки являются мессенджеры, электронная почта и съемные носители.

Постановка задачи

Задача поиска инсайдера сводится к задаче обнаружения аномалий в больших массивах телеметрии корпоративных систем.

Существуют две основные стратегии выявления аномалий:

- Обнаружение выбросов: Идентификация единичных объектов, резко отклоняющихся от общего набора данных. Эффективно для выявления разовых подозрительных событий.
- Обнаружение новизны: Распознавание ранее не встречавшихся паттернов поведения, не соответствующих модели «нормы», построенной на основе обучающих данных. Эффективно для выявления новых, неизвестных ранее сценариев угроз.

Обзор и сравнительная характеристика методов машинного обучения

Алгоритмы для обнаружения выбросов

Изолирующий лес. Быстрый и масштабируемый алгоритм, эффективно выявляющий редкие, многофакторные отклонения в активности пользователя.

Локальный уровень выброса. Оценивает аномальность объекта относительно его локального окружения, что полезно для анализа поведения в рамках отдельных ролей или отделов.

Методы для обнаружения новизны

Одноклассовый метод опорных векторов. Строит границу, отделяющую нормальное поведение от аномального. Чувствителен к настройкам и качеству обучающих данных.

Автокодировщик: Нейросетевая модель, выявляющая аномалии по высокой ошибке реконструкции данных. Способна находить сложные нелинейные зависимости.

Сеть долгой краткосрочной памяти. Анализирует временные последовательности, что позволяет выявлять не одиночные события, а подозрительные цепочки действий, развивающиеся во времени.

Трансформер. Современная архитектура на основе механизма внимания, способная улавливать сложные контекстные и долговременные зависимости в поведенческих последовательностях. Показывает высокую эффективность, но требует значительных вычислительных ресурсов.

Заключение

Выбор алгоритма зависит от конкретной задачи: для поиска разовых отклонений эффективны классические алгоритмы, а для выявления сложных поведенческих аномалий – модели глубокого обучения.

Наиболее перспективным направлением является разработка гибридных систем, которые комбинируют преимущества разных подходов, что позволяет повысить точность и снизить количество ложных срабатываний. Использование трансформеров открывает возможности анализа поведенческих паттернов с учётом контекста и корреляции между событиями.

В дальнейшем представляется целесообразным провести экспериментальную проверку эффективности указанных моделей на данных типа CERT r4.2, а также адаптацию трансформерных архитектур к специфике корпоративных сетей.