

ИИ: Защита информации



АЛЕКСЕЙ ГОЛЬДМАН

Алексей Гольдман

ИИ: Защита информации

«Издательские решения»

Гольдман А.

ИИ: Защита информации / А. Гольдман — «Издательские решения»,

ISBN 978-5-00-677645-6

Эта книга — практическое руководство по внедрению ИИ в кибербезопасность. Для CIO, CISO, SOC-команд. Обнаружение угроз: ИИ для SIEM, UEBA, EDR/XDR, облака, OT/IoT. Прогноз и автоматизация: Предвидение атак, SOAR, автономная защита. Внедрение: Пошаговая стратегия, выбор решений, ROI, безопасность ИИ. Фокус на реальные кейсы, работающие методики и измеримые результаты. Превратите ИИ из инструмента в стратегический актив. Ваше настольное руководство для эпохи киберугроз.

ISBN 978-5-00-677645-6

© Гольдман А.
© Издательские решения

Содержание

Алексей Гольдман	6
Глава 1: кибервойна и эволюция угроз: почему традиционные методы проигрывают	6
Глава 2: базовые технологии ИИ/мо для иб: не просто «магия»	8
Глава 3: интеллектуальный soc: автоматизация и ускорение обнаружения угроз (detection)	12
Глава 4: прогнозирование и проактивная защита: предвидеть атаку до начала	16
Конец ознакомительного фрагмента.	19

ИИ: Защита информации

Алексей Гольдман

© Алексей Гольдман, 2025

ISBN 978-5-0067-7645-6

Создано в интеллектуальной издательской системе Ridero

Алексей Гольдман

ИИ: Защита информации

Глава 1: кибервойна и эволюция угроз: почему традиционные методы проигрывают

Представьте себе команду пожарных, вооруженных только ведрами воды, перед лицом стремительного лесного пожара, подпитываемого ураганным ветром. Примерно так сегодня чувствуют себя многие команды информационной безопасности (ИБ), пытающиеся защитить свои организации от современных кибер-угроз с помощью устаревших инструментов и подходов. Мир кибербезопасности переживает настоящую революцию угроз, и традиционные методы защиты все чаще оказываются неэффективными. Почему так происходит? Давайте разберемся.

Растущая сложность и скорость: враг становится быстрее и умнее

Главная проблема сегодняшнего дня – невероятная сложность и скорость кибератак. Злоумышленники больше не одиночки-хулиганы. Это хорошо финансируемые, высокоорганизованные группы, часто поддерживаемые государствами. Их атаки, известные как *advanced persistent threats (apt)*, могут длиться месяцами, оставаясь незамеченными. Они тщательно исследуют жертву, используют несколько точек входа и постоянно меняют тактику, чтобы обойти защиту.

Добавьте к этому разрушительные атаки вымогателей (*ransomware*), которые за считанные минуты шифруют критически важные данные и парализуют работу целых предприятий, больниц или городов. Или *zero-day* уязвимости – бреши в программном обеспечении, о которых не знает даже производитель, а значит, нет и заплатки. Атаковать через них можно мгновенно после обнаружения, не оставляя времени на реакцию.

Человеческий фактор: нехватка рук и перегруженность

Пока угрозы множатся и усложняются, нехватка квалифицированных специалистов по ИБ становится все острее. Центры управления безопасностью (*soc*) хронически недоукомплектованы. Аналитики буквально тонут в потоке предупреждений от систем мониторинга. Большинство этих предупреждений оказываются ложными срабатываниями (*false positives*) – безобидными событиями, ошибочно принятыми за угрозу. Поиск реальной атаки в этом «шуме» похож на поиск иголки в стоге сена. Это приводит к выгоранию сотрудников и неизбежным ошибкам, когда настоящая угроза просто упускается из виду из-за перегрузки.

Сложная инфраструктура: везде, но нигде контроля полного

Современная ит-инфраструктура – это сложный гибрид. Традиционные серверы в собственном дата-центре (*on-prem*), публичные и частные облачные сервисы (*aws, azure, gcp*), тысячи устройств интернета вещей (*iot*) от камер до умных кофеварок, и критически важные промышленные системы управления (*ot/ics/scada*) на производстве. Каждый из этих элементов представляет свою уязвимость и требует специфической защиты. Обеспечить единую видимость и контроль безопасности над всей этой разнородной, постоянно меняющейся средой традиционными методами – задача почти невыполнимая.

Давление регуляторов: соответствовать или заплатить

Параллельно с техническими сложностями растет и регуляторное давление. Законы и стандарты, такие как *gdpr* в Европе, *nis2* на горизонте, требования в России или *hipaa* в здравоохранении, предъявляют жесткие требования к защите персональных данных, конфиденциальной информации и критической инфраструктуры. За нарушения грозят огромные

штрафы, исчисляемые миллионами долларов или евро, а также колоссальный репутационный ущерб. Доказать регуляторам, что вы сделали все возможное для защиты, используя устаревшие методы, становится все сложнее.

Цифры говорят сами за себя: неприемлемые задержки и ущерб

Чем же оборачиваются все эти проблемы на практике? Ключевые метрики безопасности рисуют тревожную картину:

Среднее время обнаружения бреши (mean time to detect – mttdd): это время, которое проходит от момента начала атаки до момента ее обнаружения. По данным авторитетных исследований (например, от ibm или ponemon institute), это время все еще составляет недели или даже месяцы. Представьте, что вор проник в ваш дом, и вы узнали об этом только через полгода! За это время злоумышленники успевают сделать все, что планировали: украсть данные, установить скрытые «бэкдоры» для постоянного доступа, подготовиться к масштабной атаке.

Среднее время реагирования (mean time to respond – mtrr): это время, которое требуется команде ИБ, чтобы понять, что делать с обнаруженной атакой, и начать принимать меры для ее устранения (изоляция зараженных систем, удаление вредоносного ПО, восстановление данных). Даже после обнаружения реагирование часто занимает дни или недели. Это как обнаружить пожар, но не мочь найти огнетушитель или воду.

Результат? Неприемлемо высокий ущерб. Успешная атака вымогателей может обойтись компании в миллионы долларов прямых затрат (выкуп, простои, восстановление) и многомиллионные косвенные потери из-за ущерба репутации, потери доверия клиентов и судебных исков. Утечка персональных данных ведет к гигантским штрафам по gdpr или другим законам. А атака на промышленную систему может привести к физическим разрушениям, остановке производства и угрозе жизни людей.

Ии как «умножитель силы»: необходимость нового подхода

В условиях этой «идеальной кибер-бури» – сложных и быстрых атак, нехватки специалистов, гибридной инфраструктуры и регуляторного пресса – становится очевидным: традиционные, ручные и сигнатурные методы безопасности больше не справляются. Нужен качественный скачок. Нужен инструмент, который сможет:

Автоматизировать рутинные задачи: фильтровать ложные срабатывания, обрабатывать горы данных.

Обеспечить беспрецедентную скорость анализа: обнаруживать аномалии и угрозы в реальном времени, пока еще не поздно.

Прогнозировать риски: видеть слабые места и возможные векторы атак до того, как ими воспользуются.

Управлять масштабом: эффективно контролировать безопасность огромных и сложных гибридных сред.

Таким инструментом, «умножителем силы» для истощенных команд ИБ, становится искусственный интеллект (ИИ) и его ключевые технологии, прежде всего машинное обучение (МО). ИИ – это не «магия» и не замена человеку. Это мощный помощник, который берет на себя тяжелую аналитическую работу, позволяет аналитикам сосредоточиться на действительно сложных инцидентах и дает шанс успевать за эволюцией угроз. В следующих главах мы подробно разберем, как именно технологии ИИ решают описанные здесь критически важные проблемы современной кибербезопасности.

Глава 2: базовые технологии ИИ/ мо для иб: не просто «магия»

В первой главе мы увидели, почему традиционные методы защиты трещат по швам. Надежда – на искусственный интеллект (ии) и машинное обучение (мо). Но прежде, чем погружаться в конкретные применения в soc, сети или облаке, давайте разберемся с фундаментом. Что скрывается за модным словом «ИИ» в контексте безопасности? Это не волшебная палочка и не черный ящик, из которого вылетают готовые ответы. Это набор мощных, но вполне объяснимых технологий, которые стали «суперсилой» для аналитиков иб.

Машинное обучение (ml) и глубокое обучение (dl): учимся видеть угрозы

Представьте себе нового сотрудника soc. Сначала он не знает, что нормально для вашей сети, а что – нет. Но он очень наблюдателен и учится быстро. Вы показываете ему миллионы событий: «смотри, вот это – нормальный вход пользователя. А вот это – попытка взлома. Вот – легитимный скачок трафика. А вот – начало ddos-атаки». Через какое-то время он начинает замечать закономерности и сам отличать одно от другого. Машинное обучение (ml) – это и есть процесс «обучения» компьютера на примерах данных. В ИБ этими данными чаще всего являются логи событий безопасности.

Как это работает в ИБ?

Обучение на «хорошем» и «плохом»: вы «кормите» алгоритм ml огромным количеством помеченных данных. Например, логами сетевого трафика, где известно, какие сессии были нормальными, а какие – атаками (или логамиedr с чистыми и зараженными файлами).

Поиск закономерностей: алгоритм анализирует эти данные, выявляя сложные, часто неочевидные для человека, закономерности и связи между различными параметрами (например, время суток, источник, порт, размер пакета, частота запросов, последовательность действий).

Построение модели: на основе найденных закономерностей алгоритм строит математическую модель. Это и есть «искусственный интеллект» для конкретной задачи – например, модель обнаружения вторжений или модель классификации вредоносных файлов.

Применение к новым данным: когда на вход модели поступают новые, невиданные ранее логи, она применяет выученные закономерности, чтобы классифицировать событие (нормальное/подозрительное/злонамеренное), обнаружить аномалии (сильное отклонение от выученной «нормы») или даже спрогнозировать вероятность атаки на основе текущих сигналов.

Глубокое обучение (deep learning – dl) – это более сложная и мощная разновидность ml, вдохновленная работой человеческого мозга (нейронные сети). Dl особенно хорош для анализа очень сложных, многомерных данных, таких как:

Сырой сетевой трафик (пакеты): для обнаружения скрытых каналов связи или сложных атак, маскирующихся под легитимный трафик.

Поведение процессов на конечной точке: для выявления файловых вредоносных без сигнатур по их действиям в системе (например, попытки шифрования файлов, манипуляции с реестром, подозрительные сетевые вызовы).

Изображения (для анализа скриншотов вредоносного по или фишинговых страниц).

Пример из жизни: представьте день распродажи «черная пятница» на сайте крупного ритейлера. Трафик взлетает в десятки раз. Традиционная система обнаружения ddos на основе простых пороговых значений завалит soc тысячами ложных срабатываний. Мl-модель, обученная на исторических данных (включая прошлые «черные пятницы» и реальные ddos-атаки), видит разницу. Она анализирует не только объем, но и характер трафика: источники (разбросанные по миру боты vs. Концентрация реальных пользователей из одного региона), типы запросов (однообразные запросы к уязвимому api vs. Разнообразные запросы к каталогам, кор-

зине), временные паттерны. Она понимает, что это легитимный всплеск, а не атака, и не поднимает ложную тревогу. И наоборот, если придет настоящая ddos, модель распознает ее характерные признаки, даже если общий объем трафика не будет рекордным.

Обработка естественного языка (nlp): понимаем, о чем говорят данные

Логи, тикеты, отчеты сканеров, новости об угрозах, посты в даркнете – огромный объем данных в иб представлен в виде текста. Человеку физически невозможно проанализировать все это вручную. Обработка естественного языка (natural language processing – nlp) – это технология ИИ, которая позволяет компьютерам «понимать», извлекать смысл и работать с человеческим языком.

Как это работает в иб?

Анализ логов и тикетов: nlp автоматически классифицирует инциденты безопасности, извлекает ключевые сущности (ip-адреса, домены, имена пользователей, названия уязвимостей) и даже определяет тональность или срочность описания в тикете. Это ускоряет рутину sos и направляет аналитиков к самым важным событиям.

Мониторинг threat intelligence: nlp сканирует тысячи открытых и закрытых (включая даркнет) источников информации об угрозах: форумы, блоги, отчеты, базы данных уязвимостей. Она ищет упоминания вашей компании, ваших продуктов, ваших доменов, используемого вами по, специфических ttp (тактик, техник и процедур) злоумышленников, актуальных для вашего сектора.

Автоматическое обогащение контекстом: найденная в тексте информация (например, новый эксплойт для используемого вами сервера) автоматически связывается с событиями в ваших системах (siem) и добавляется как контекст к алертам, делая их гораздо информативнее.

Анализ фишинговых писем и вредоносных скриптов: nlp помогает определить подозрительные языковые конструкции в письмах или обфусцированный (намеренно запутанный) код в скриптах.

Пример из жизни: специалист по угрозам (threat intelligence) банка получает уведомление от nlp-системы. Система, непрерывно мониторившая русскоязычные даркнет-форумы, обнаружила свежий пост. В нем упоминается внутреннее название банковской системы перевода средств, которое используется только сотрудниками, и выложен фрагмент базы данных с реальными именами и номерами телефонов клиентов банка. Nlp выделила ключевые сущности (название системы, имена клиентов) и определила высокую степень релевантности и угрозы. Это позволило sos немедленно начать расследование (проверить доступы к системе, искать подозрительную активность) и уведомить клиентов о потенциальной утечке, минимизируя ущерб.

Поведенческая аналитика (ueba/nta): знаем «норму», видим отклонения

Это, пожалуй, самое мощное применение ИИ/мо в иб для борьбы с современными, целенаправленными атаками, которые обходят сигнатурные защиты. Идея проста: сначала понять, что является «нормальным» поведением для пользователей, устройств, приложений или сетевого трафика в вашей уникальной среде. А затем найти любое существенное отклонение от этой нормы, которое может сигнализировать об атаке или инциденте, даже если сама активность не соответствует известным шаблонам вредоносного по.

Как это работает?

Формирование «базовой линии» (baseline): система (используя ml) наблюдает за поведением в течение длительного периода (недели, месяцы). Она учится, как обычно ведет себя каждый пользователь (когда и с каких устройств логинится, к каким системам обращается, какие объемы данных скачивает/загружает), каждое устройство (нормальная загрузка сри, сетевая активность), приложение или сетевой сегмент (типичные порты, протоколы, объемы трафика).

Обнаружение аномалий: алгоритмы ml (часто статистические методы и dl) в реальном времени сравнивают текущую активность с выученной «нормой». Они ищут любые значительные отклонения:

Пользователь заходит в систему в 3 часа ночи из необычной страны и начинает массово скачивать базу данных клиентов.

Сервер, который обычно «молчит», внезапно инициирует тысячи исходящих соединений на неизвестные ip-адреса.

Принтер в бухгалтерии передает неожиданно большой объем данных наружу.

Сетевое устройство начинает общаться по порту, который никогда раньше не использовал.

Оценка риска: система не просто кричит «аномалия!». Она оценивает контекст (кто пользователь? Каковы его права? Какая система?) И тяжесть отклонения. На основе этого присваивает событию уровень риска (низкий, средний, высокий, критический). Это помогает сосприоритезировать расследование.

Два ключевых подхода:

Ueba (user and entity behavior analytics): фокус на поведении пользователей (человеков) и сущностей (учетных записей, хостов, приложений). Основная цель: обнаружение скомпрометированных учетных записей (credential theft), инсайдерских угроз (злонамеренных или небрежных), аномалий в действиях привилегированных пользователей (админов).

Nta (network traffic analysis) с ИИ: фокус на анализе сетевого трафика (метаданных и/или содержимого пакетов). Основная цель: обнаружение скрытых каналов управления ботнетом (с&с), перемещения данных (data exfiltration), сканирования сети злоумышленником, необычных коммуникаций между системами, которые обычно не общаются.

Пример из жизни: в крупной технологической компании ueba-система выдала высокоприоритетный алерт на действия системного администратора сергея. Модель заметила критическое отклонение: сергей, который обычно работал днем и обращался к серверам разработки, ночью в 2:45 подключился к базе данных с персональными данными клиентов (к которой он имел доступ по должности) и начал выгружать огромные объемы данных (в сотни раз больше его обычной активности) не на корпоративный ресурс, а на внешний почтовый сервер через веб-интерфейс. Это сочетание «необычное время», «необычная система», «аномальный объем данных» и «необычное место назначения» заставило систему присвоить инциденту критический уровень риска. Расследование сосподтвердило попытку кражи данных инсайдером. Без ueba эта активность, возможно, затерялась бы в общем ночном трафике или была бы обнаружена слишком поздно.

Итог: ИИ/мо – инструмент, а не волшебство

Ключевой вывод этой главы: ИИ и мо в иб – это не абстрактная «магия», а конкретные, мощные технологии анализа данных. Они не заменяют экспертизу иб-специалиста, а становятся его «цифровым напарником», обрабатывающим гигабайты информации со скоростью света и выделяющим самое важное и подозрительное. Их сила – в способности:

Учиться на ваших собственных данных, понимая уникальную «норму» вашей среды.

Обнаруживать сложные, неизвестные ранее угрозы по их аномальному поведению.

Автоматизировать рутинный анализ (логов, текстовых отчетов), освобождая людей для сложных задач.

Обеспечивать скорость и масштаб, недоступные человеку.

Но важно помнить: «мусор на входе – мусор на выходе» (garbage in, garbage out). Эффективность ИИ-системы в иб напрямую зависит от качества и полноты данных, на которых она обучается и которые анализирует (логи, конфигурации, информация об угрозах). Без хороших данных даже самая продвинутая модель будет бесполезна или, что хуже, опасна ложными сра-

батованиями. В следующих главах мы увидим, как эти фундаментальные технологии воплощаются в реальные решения для защиты soc, сети, конечных точек и облака.

Глава 3: интеллектуальный soc: автоматизация и ускорение обнаружения угроз (detection)

Представьте центр управления безопасностью (soc) в разгар серьезного инцидента. Экраны мигают сотнями, если не тысячами, предупреждений. Аналитики в панике пытаются понять, что реально, а что ложная тревога, где начинается атака и насколько она опасна. Традиционный siem (security information and event management) – основа soc – превращается в источник информационного шума. Именно здесь искусственный интеллект становится не просто полезным, а жизненно необходимым, превращая siem из «сигнальной сирены» в «интеллектуального аналитика». Давайте разберемся, как ИИ революционизирует ключевую функцию soc – обнаружение угроз (detection).

1. Siem на стероидах: ИИ как супер-коррелятор и фильтр ложных срабатываний

Классический siem собирает события (логи) со всего зоопарка устройств: сетевых экранов, серверов, рабочих станций, приложений, систем контроля доступа. Его задача – коррелировать эти события, находя связи между ними, которые могут указывать на атаку. Проблема? Правила корреляции, написанные людьми, часто слишком просты («если 5 неудачных логинов за минуту – тревога») или слишком сложны в поддержке. Они генерируют лавину ложных срабатываний (false positives – fp): легитимная активность ошибочно признается угрозой. Аналитики тратят до 70% времени на их разбор!

Как ИИ трансформирует siem?

Умная корреляция событий: вместо жестких правил ИИ (ml) анализирует исторические данные и контекст. Он учится понимать, какие комбинации событий из разных источников (сеть + конечная точка + приложение + облако) действительно характерны для атак в вашей среде, а какие – просто фоновая активность.

Резкое снижение ложных срабатываний: ИИ действует как интеллектуальный фильтр. Он оценивает вероятность того, что событие или группа событий – это реальная угроза, учитывая десятки факторов (источник, цель, время, предыдущая активность, данные об угрозах). События с низкой оценкой риска автоматически отсеиваются или понижаются в приоритете.

Обогащение контекстом: ИИ автоматически «приклеивает» к событиям важный контекст из внешних источников: является ли ip-адрес известным злоумышленником? Активна ли эксплуатируемая уязвимость в указанном по? Упоминается ли ваш домен в даркнете? Это превращает сырой алерт в информативное сообщение, сразу дающее аналитику ключ к пониманию ситуации.

Выявление сложных, многоэтапных атак (apt): злоумышленники разбивают атаку на множество мелких, на первый взгляд безобидных, шагов. ИИ видит эти шаги в совокупности и во времени, даже если они происходят с разных ip или на разных системах, и поднимает тревогу на всю цепочку, а не на отдельные звенья.

Пример из жизни (финансовый сектор): в крупном банке традиционный siem ежедневно генерировал около 10 000 алертов. После внедрения ИИ-модуля для корреляции и фильтрации, количество событий, требующих немедленного внимания аналитика, упало до 300—500 в день. Как это произошло? Однажды система получила событие: «неудачная попытка входа в доменную учетную запись менеджера среднего звена с необычного ip (геолокация: страна, где у банка нет офиса)». Раньше это был бы один из тысяч fp. Но ИИ-коррелятор связал это с другими, казалось бы, несвязанными событиями за последние 24 часа: сканирование портов на этом же ip в адрес сервера, где работает приложение для управления кредитами; несколько успешных входов в низкоуровневые тестовые учетки с того же ip; и недавнее обновление threat intelligence о фишинговой кампании, нацеленной именно на финансовый сектор из этой страны. ИИ присвоил совокупности событий критический уровень риска и автоматизи-

чески собрал весь контекст в один понятный инцидент. Soc быстро среагировал, заблокировал ip и сбросил пароли затронутых учеток, предотвратив потенциальную кражу данных.

2. Поведенческая аналитика (ueba): ловим инсайдеров и скомпрометированные учетки

Традиционные средства защиты плохо видят угрозы, которые не используют известное вредоносное ПО или эксплойты, а маскируются под легитимную активность. Ueba (user and entity behavior analytics) с ИИ – это ваш «детектор лжи» внутри сети.

Как это работает в soc?

Построение «цифрового днк»: система непрерывно обучается, создавая индивидуальные поведенческие профили (базовые линии) для каждого пользователя, учетной записи, сервера, принтера. Она знает, когда, откуда, к чему и как обычно обращается Сергей-бухгалтер или сервер баз данных.

Обнаружение отклонений в реальном времени: любое существенное отклонение от «нормы» немедленно анализируется ИИ. Примеры:

Кража учетных данных (credential theft): пользователь внезапно логинится ночью из незнакомой страны и начинает массово скачивать файлы, к которым у него есть доступ, но которыми он никогда не интересовался. Или его учетка используется одновременно из двух географически удаленных мест.

Инсайдерская угроза (malicious insider): привилегированный пользователь (админ, менеджер) начинает обращаться к системам или данным, не связанным с его текущими задачами, или выгружает необычно большие объемы конфиденциальной информации наружу.

Небрежность или ошибка (negligent insider): сотрудник случайно загружает конфиденциальный файл в публичное облачное хранилище. Ueba видит аномальную активность этого пользователя (новый тип действия) и необычное место назначения данных.

Оценка риска и приоритезация: ИИ не просто кричит «аномалия!». Он анализирует контекст: критичность данных, уровень привилегий пользователя, серьезность отклонения. На выходе – инцидент с четким уровнем риска (high/critical), требующий немедленного внимания soc.

Пример из жизни (технологическая компания): ueba-система выдала алерт высокого риска на действия главного системного инженера Алексея. Модель обнаружила несколько тревожных отклонений от его обычного поведения за последние 2 дня: 1) входы в систему в 2:30 ночи (обычно он работал с 9 до 18); 2) множественные подключения к серверам, хранящим исходный код нового продукта (к которым у него был доступ, но к которым он редко обращался в последние месяцы); 3) создание большого архива данных (2 гб) с этих серверов; 4) попытка отправить этот архив на личный почтовый ящик через веб-интерфейс (аномальный объем исходящих данных + необычный адрес назначения). ИИ объединил все эти аномалии в один критический инцидент «потенциальная утечка интеллектуальной собственности». Soc немедленно связался с Алексеем и службой hr. Выяснилось, что сотрудник собирался уволиться к конкурентам и попытался украсть исходный код. Инцидент был остановлен до нанесения ущерба.

3. Аналитика сетевого трафика (nta) с ИИ: видим невидимое

Сеть – это кровеносная система ит. Злоумышленники стремятся либо проникнуть через нее, либо использовать ее для скрытого управления и выкраденных данных. Nta (network traffic analysis) с ИИ – это рентгеновское зрение для вашей сети, выявляющее угрозы, которые обходят традиционные брандмауэры и ids/ips.

Как ИИ усиливает nta в soc?

Обнаружение скрытых каналов связи (с&с): ботнеты и продвинутые угрозы (apt) используют зашифрованный или маскирующийся под легитимный (например, https, dns) трафик для связи с сервером управления. ИИ анализирует метаданные трафика (время, объем, пакеты, интервалы, географию) и аномальные паттерны (например, тысячи dns-запросов к несуществу-

ющим поддоменам из одной машины), чтобы выявить такую скрытую активность, даже не расшифровывая содержимое.

Выявление перемещения данных (data exfiltration): кража данных редко выглядит как гигабайтный поток в одну секунду. Злоумышленники «просачивают» данные медленно, маскируясь под нормальный трафик. ИИ замечает отклонения в объеме исходящего трафика для конкретного хоста или пользователя, необычные протоколы или порты для передачи данных (например, icmp, нестандартные tcp-порты), или связь с подозрительными внешними ip/доменами (из threat intelligence).

Обнаружение сканирования и разведки: перед атакой злоумышленник исследует сеть. ИИ видит аномально высокое количество попыток подключения к разным портам/хостам с одного источника или паттерны сканирования уязвимостей в трафике.

Выявление аномальных внутренних коммуникаций: сервер бухгалтерии внезапно начинает активно общаться с инженерным сервером? Контроллер на производстве связывается с сервером в dmz? ИИ видит эти отклонения от нормальных паттернов общения между сегментами сети и хостами, что может указывать на горизонтальное перемещение злоумышленника внутри сети.

Пример из жизни (энергетика): nta-система с ИИ на электростанции обнаружила аномалию в трафике с одного из промышленных контроллеров (плк), управляющего турбиной. Контроллер начал генерировать необычно регулярные, небольшие пакеты данных на внешний ip-адрес, расположенный в стране, не связанной с поставщиками оборудования. Паттерн напоминал «heartbeat» – сигналы «жизни», характерные для ботнетов. При этом сам контроллер функционировал нормально, не вызывая подозрений у операторов. ИИ присвоил событию высокий риск. Soc-аналитики совместно с инженерами от изолировали контроллер для глубокого анализа. Обнаружилось, что он был скомпрометирован через уязвимость в старом ПО и использовался как точка сбора данных и потенциальная «бомба замедленного действия» для будущей диверсии. Угроза была нейтрализована.

Практический совет: как «накормить» ИИ хорошими данными?

Эффективность ИИ в soc (будь то siem, ueba или nta) напрямую зависит от качества данных, на которых он учится и которые анализирует. «мусор на входе – мусор на выходе».

«чистые» данные для обучения: старайтесь обучать модели на периодах, когда не было крупных известных инцидентов. Фильтруйте из обучающих данных заведомо «шумные» или нерелевантные источники. Чем точнее «норма» – тем точнее обнаружение аномалий.

Полнота источников: чем больше релевантных логов вы собираетесь (сеть, хосты, приложения, облако, активные директории, базы данных), тем полнее картина для ИИ. Не экономьте на сборе данных с критичных систем.

Обогащение threat intelligence: интегрируйте актуальные данные об угрозах (ioc – индикаторы компрометации, ttp – тактики, техники и процедуры злоумышленников) из надежных источников (коммерческие и открытые платформы). Это позволяет ИИ сразу пометить известные вредоносные ip, домены, хэши файлов и искать признаки конкретных атак в вашем трафике.

Нормализация данных: убедитесь, что события из разных источников приводятся к единому формату. ИИ должен легко понимать, что «failed login» из cisco asa и «logon failure» из windows security log – это, по сути, одно и то же.

ИИ – это не замена soc, а его мощнейший усилитель. Он берет на себя тяжелейшую работу: фильтрацию информационного шума, корреляцию событий из десятков источников в понятные инциденты, выявление самых коварных угроз по их аномальному поведению (скомпрометированные учетки, инсайдеры, скрытый c&c, утечки данных). Результат для soc – радикальное снижение mttд (время обнаружения), резкое падение ложных срабатываний, освобождение времени аналитиков для расследования реальных, сложных угроз и повышение

уровня защищенности организации в целом. Интеллектуальный soc, вооруженный ИИ, перестает быть «реагирующим пожарным депо» и становится «центром прогнозирования и превентивной защиты». В следующей главе мы посмотрим, как ИИ позволяет не только обнаруживать, но и предвидеть атаки.

Глава 4: прогнозирование и проактивная защита: предвидеть атаку до начала

Обнаружить атаку быстро – это хорошо. Остановить ее до того, как она нанесет ущерб – еще лучше. Но что, если бы вы могли предсказать, куда нанесет удар злоумышленник завтра или на следующей неделе, и успеть закрыть брешь сегодня? Это не фантастика. Это реальность, которую приносит искусственный интеллект в арсенал современной безопасности. В этой главе мы покидаем область реактивного обнаружения (detection) и вступаем на территорию прогнозирования (prediction) и проактивной защиты. Узнаем, как ИИ превращает вашу ИБ-команду из пожарных в метеорологов, предсказывающих кибер-бурю и готовящих укрепления.

1. Анализ ttp: предугадываем следующий ход противника

Злоумышленники, особенно продвинутые (apt), действуют не хаотично. Они следуют определенным тактикам, техникам и процедурам (tactics, techniques, and procedures – ttp). Это их «почерк» или «плей-бук». Например, сначала фишинг для получения доступа, затем поиск привилегированных учеток, движение по сети, сбор данных, их вывоз. ИИ, обученный на огромных массивах данных об атаках (открытые отчеты, даркнет, данные сенсоров тысяч организаций, базы вроде mitre att&ck), умеет распознавать эти ttp в действии.

Как ИИ предсказывает следующий шаг?

Распознавание начальных стадий: ИИ выявляет в вашей среде активность, характерную для известных ttp. Например, обнаружена попытка фишинга на ключевых сотрудников

Контекст и цель атакующего: модель анализирует, на кого нацелена атака (финансовый отдел? R&d?), какие системы уже затронуты, какие данные представляют интерес в этом сегменте.

Прогноз следующих действий: на основе выученных шаблонов тысяч атак и понимания текущего контекста, ИИ предсказывает наиболее вероятные следующие шаги злоумышленника. Например: «после успешного фишинга на сотрудника отдела продаж, с вероятностью 85% атакующий попытается получить доступ к crm-системе и выгрузить базу клиентов в течение следующих 48 часов».

Рекомендации по упреждающим действиям: система не просто предупреждает. Она предлагает конкретные меры: «усилить мониторинг активности учетных записей в crm; проверить журналы доступа к базе клиентов за последние 72 часа; применить временные правила брандмауэра для ограничения исходящего трафика с рабочих станций отдела продаж; сбросить пароли потенциально скомпрометированным учеткам».

Преимущество: вместо того чтобы ждать, пока злоумышленник доберется до ценных данных, sos может упреждающе заблокировать его путь, усилить защиту на прогнозируемом направлении атаки и начать активный поиск (threat hunting) по конкретным сигнатурам предполагаемых действий.

2. Оценка уязвимости инфраструктуры: где ждет брешь?

Традиционный vulnerability management часто сводится к сканированию и получению длинного списка уязвимостей с приоритетами, основанными в основном на их критичности (cvss score). Но реальная опасность уязвимости зависит от множества факторов, которые ИИ учитывает комплексно:

Контекст системы: насколько критична система, где найдена уязвимость? (доменный контроллер vs. Тестовый сервер). Какие данные она обрабатывает? (pii, финансы, интеллектуальная собственность). Кто имеет к ней доступ?

Конфигурация и окружение: есть ли работающие меры защиты (waf, ips), которые уже могут смягчать угрозу? Находится ли система в изолированном сегменте или открыта в интернет? Используется ли уязвимое по активно?

Внешняя угроза: есть ли доказательства, что уязвимость активно эксплуатируется в дикой природе (exploited in the wild)? Упоминается ли она в даркнете? Нацелены ли на нее известные хакерские группы, работающие против вашей отрасли? Есть ли публичный эксплойт (poc)?

Сложность эксплуатации: насколько легко злоумышленнику может быть использовать эту брешь против вашей конкретной конфигурации?

Как ИИ строит точную картину риска?

Агрегация данных: ИИ собирает данные из сканеров уязвимостей, систем управления конфигурациями (cmdb), данных об актуальных угрозах (threat intelligence), логов брандмауэров и ips, информации о сетевой топологии.

Контекстный анализ: модели ml анализируют все перечисленные выше факторы для каждой найденной уязвимости в вашем уникальном окружении.

Прогноз вероятности и воздействия: ИИ рассчитывает не абстрактный cvss, а реальную вероятность успешной эксплуатации уязвимости в вашей среде и прогнозируемый ущерб от такого инцидента (финансовый, репутационный, операционный).

Динамический приоритет: на основе этого прогноза формируется рейтинг риска уязвимостей, который постоянно обновляется по мере поступления новой информации (например, о начале массовой эксплуатации). Это позволяет soc и командам исправлений сосредоточиться на действительно опасных «дырах» здесь и сейчас, а не на всех подряд.

Результат: резкое сокращение времени на исправление критичных с точки зрения реального бизнес-риска уязвимостей и более эффективное использование ресурсов патч-менеджмента.

3. Прогнозирование риска атаки на организацию/отрасль: взгляд в кибербудущее

Ии может подняться еще выше, оценивая риск атаки не только на конкретную систему, но и на организацию в целом или даже целую отрасль. Это макроуровневое прогнозирование.

Какие факторы анализирует ИИ?

Геополитическая обстановка: эскалация конфликтов, санкции, кибервойны между государствами – все это увеличивает риск целевых атак на компании из определенных стран или секторов.

Активность хакерских групп: появление новых групп, смена их фокуса (например, с госструктур на здравоохранение или энергетику), объявление о кампаниях против определенной отрасли в даркнете.

Отраслевые уязвимости и тренды: распространение критической уязвимости в по-широко используемом в вашей отрасли (например, pos-системы в ритейле, scada в энергетике, moveit в любом секторе). Увеличение числа успешных атак на аналогичные компании.

События высокого профиля: крупные конференции, слияния и поглощения, финансовые отчеты – могут сделать компанию более привлекательной мишенью.

Внутренние сигналы (опционально): аномальная разведывательная активность (сканирование) против вашей инфраструктуры, фишинговые волны, нацеленные на сотрудников.

Как используется прогноз?

Повышение готовности: предупреждение soc, ит и бизнес-руководства о повышенном уровне угрозы. Усиление мониторинга, проверка резервных копий, обновление планов реагирования на инциденты (igr).

Упреждающие контрмеры: применение временных дополнительных правил безопасности (более строгий контроль доступа, фильтрация трафика из «горячих» регионов), ускорение критических обновлений по, проведение дополнительного обучения сотрудников по осведомленности.

Стратегическое планирование: информирование о долгосрочных трендах киберрисков для бюджетирования и развития иб-архитектуры.

Пример из жизни: ИИ-система прогнозирования рисков в крупной розничной сети проанализировала несколько ключевых факторов за неделю до «черной пятницы»: 1) резкий всплеск обсуждений в закрытых хакерских чатах, нацеленных на ритейл, с упоминанием конкретной модели кассовых терминалов; 2) появление в базе данных уязвимостей (но еще без патча!) Критичной тсе-уязвимости в по этих терминалов; 3) исторические данные о волнах атак на ритейлеров перед крупными распродажами. Система присвоила красный уровень риска и сгенерировала предупреждение: «высокая вероятность (75%) целенаправленных атак через уязвимость в по кассовых терминалов model x в период с dd.mm по dd.mm с целью установки ransomware или кражи данных карт».

Конец ознакомительного фрагмента.

Текст предоставлен ООО «Литрес».

Прочитайте эту книгу целиком, [купив полную легальную версию](#) на Литрес.

Безопасно оплатить книгу можно банковской картой Visa, MasterCard, Maestro, со счета мобильного телефона, с платежного терминала, в салоне МТС или Связной, через PayPal, WebMoney, Яндекс.Деньги, QIWI Кошелек, бонусными картами или другим удобным Вам способом.