ПРЕДУПРЕЖДЕНИЕ КОМПЬЮТЕРНЫХ АТАК ТИПА MAN IN THE MIDDLE, СОВЕРШАЕМЫХ С ИСПОЛЬЗОВАНИЕМ ГЕНЕРАТИВНОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Жарова А. К.[1], Елин В. М.[2], Аветисян Б. Р.[3]

DOI: 10.21681/2311-3456-2024-6-28-41

Целью статьи *является представление научному сообществу разработанной авторской методики выявления/предотвращения компьютерной атаки по типу «злоумышленник посередине» (MITM).*

Метод исследования: *для достижения поставленной цели авторы использовали методы математического моделирования, сравнительного анализа, табличный метод, а также методы экспериментально-теоретического уровня.*

Результат: *проведен сравнительный анализ программных решений, представленных в виде исходного кода на площадках по типу GITHUB, которые обеспечивают реализацию атаки злоумышленник посередине как в локальных, так и глобальных сетях, а также анализ некоторых методик предотвращения атаки по типу MITM, использующих сервисы искусственного интеллекта (ИИ). На основе данного анализа определены различные логические реализации атаки по типу MITM, а также представлены уязвимости информационных систем перед компьютерной атакой MITM. На основании проведенного анализа существующих методов противодействия этим атакам и выявленных слабых сторон этих методов, предложена авторская методика предотвращения атаки по типу MITM, которая включает обучение ИИ на дата-сетах, подключенных к библиотекам разных языков программирования и алгоритмизированных эвристических моделях, реагирующих на изменение логики поведения пользователей, либо активности персонального компьютера, сетевого оборудования.*

Практическая ценность *состоит в разработанной авторской методике выявления/предотвращения компьютерной атаки по типу MITM с использованием «предиктивных» сетевых технологий, которые основаны на применении нейронных сетей, обученных методами машинного обучения.*

Ключевые слова: *дата-сеты, MITM, методики предотвращения атаки, эвристические модели, поведение пользователей, предиктивные сетевые технологии.*

Введение

Развитие инженерной мысли позволило интегрировать достижения в сфере науки и техники в сферу жизнедеятельности человека, формируя тем самым новые общественные отношения. Цифровые технологии могут работать на благо человека, но нередко используются и в противозаконных целях [1]. Чем чаще человек пользуется информационными технологиями, тем больше он оставляет цифровых следов и становится более уязвимым перед злоумышленниками. Интерес злоумышленников представляют и организации, они также становятся уязвимыми перед различными информационными угрозами [2].

В соответствии с отчетом, составленным Kaspersky ICS CERT, за первое полугодие 2023[4] компьютерным атакам подверглись промышленные компании в следующих отраслях экономики.

С каждым годом в мире наблюдается рост интенсивности компьютерных инцидентов. По исследованию, проведенному компанией Positive Technologies, «утечка конфиденциальной информации стала одним из самых распространенных последствий кибератак. Ее доля в I квартале 2024 г. выросла до 72 %, тогда как за тот же период 2023 г. этот показатель составлял 59 %. За первые 2,5 месяца этого года также утекли данные 170 компаний — это 40 % от всего числа инцидентов[5] в 2023 г.»

Причем в руках злоумышленников находятся различные решения, которые позволяют им получить

1 Жарова Анна Константиновна, доктор юридических наук, профессор Финансового университета при Правительстве Российской Федерации, Москва, Россия. E-mail: anna_jarova@mail.ru

2 Елин Владимир Михайлович, кандидат юридических наук, доцент Финансового университет при Правительстве Российской Федерации; доцент кафедры информационной безопасности Московского университета МВД России имени В.Я. Кикотя, Москва, Россия. E-mail: elin_vm @mail.ru

3 Аветисян Борис Рафаелович, главный научный сотрудник, Научно-исследовательский институт образования и науки, Москва, Россия. E-mail: Boris.Avetisyan@gmail.com

4 Первое полугодие 2023 года — краткий обзор основных инцидентов промышленной кибербезопасности // https://ics-cert.kaspersky.ru/publications/reports/2023/10/05/h1-2023-a-brief-overview-of-main-incidents-in-industrial-cybersecurity/ (дата обращения 20.09.2024)

5 Информационная опасность: доля утечек личных данных выросла до 72% // https://iz.ru/1696887/elizaveta-krylova/informatcionnaia-opasnost-dolia-utechek-lichnykh-dannykh-vyrosla-do-72 (дата обращения 20.09.2024)

незаконный доступ к данным. Существуют различные методы проведения компьютерных атак, одним из таких методов является атака типа «злоумышленник посередине» (MITM). На данный момент невозможно привести статистику ущерба происходящего именно от компьютерных атак типа MITM, поскольку каждый случай атаки имеет свои уникальные характеристики и последствия. В научной литературе изучается специфика этой атаки и предлагаются различные методы ее предотвращения. Так, наиболее эффективным методом предотвращения атаки по типу MITM являются «предиктивные» сетевые технологии, которые основаны на применении нейронных сетей, обученных методами машинного обучения. Предиктивные сетевые технологии формируют прогнозы на основе анализа целого спектра данных: от анализа сетевых портов, используемых в рамках исследуемого сегмента сетевой инфраструктуры, производительности систем, до анализа доменной зоны (геолокации серверной части – «цифровой юрисдикции»). Полученные данные выступают базисом для анализа и прогнозировании инцидентов, а также для формирования вероятностной оценки угрозы утечки информации ограниченного доступа по техническим каналам связи и основаниями расчета ущерба от несанкционированного доступа третьих лиц.

Предиктивные сетевые технологии целесообразны при предотвращении атак типа MITM в компьютерных сетях. Одним из компонентов обеспечения безопасности обрабатываемых и передаваемых данных выступают правила, реализуемые в цифровых сертификатах и криптографических протоколах SSL/TLS шифрования. Решение SSL (Secure Sockets Layer) или его более современной версии TLS (Transport Layer Security) используется для защиты передаваемых данных между клиентом и сервером.

Искусственный интеллект (ИИ) используется не только для предотвращения/выявления компьютерных атак, но и в противоположных целях [3]. С применением ИИ [4] злоумышленники ускоряют процесс поиска уязвимостей.

Анализируя существующие методы противодействия компьютерным атакам и, в частности, атаке по типу MITM, авторы статьи поставили задачу разработки авторской методики выявления/предотвращения компьютерной атаки по типу MITM.

В статье проводится сравнительный анализ программных решений, представленных в виде исходного кода на площадках по типу GITHUB, которые обеспечивают реализацию атаки по типу MITM как в локальных, так и глобальных сетях, а также анализ некоторых методик предотвращения атаки по типу MITM, использующих сервисы ИИ. На основе данного анализа определяются различные логические реализации атаки по типу MITM, а далее в целях предотвращения атаки по типу MITM проводится обучение ИИ на дата-сетах, подключенных к библиотекам разных языков программирования и алгоритмизированных эвристических моделях, реагирующих на изменение как логики поведения пользователей, так и активности персонального компьютера и сетевого оборудования.

Понятие компьютерной атаки

Понятие компьютерной атаки раскрывается стандартом ISO/IEC 27000:2014[6] как «попытка уничтожения, раскрытия, изменения, блокирования, кражи, получения несанкционированного доступа к активу или его несанкционированного использования». Ученые относят компьютерную атаку к инструменту кибероперация [5], проводимых против конкретных лиц или организаций.

К классическим компьютерным атакам можно отнести атаки типа «отказ в обслуживании» (DoS) и распределенные атаки «отказ в обслуживании» (DDoS); атаки «злоумышленник посередине» (MITM); фишинг; целевые фишинговые атаки; атаки путем внедрения (SQLI и XSS); глушение, подслушивающие атаки; и атаки вредоносных программ.

Определенной новеллой выступают компьютерные атаки, основанные на технологиях ИИ, результатом которых является некорректная классификация данных, генерация синтетических данных, незаконный доступ к данным и их анализ [5]. Так, в компьютерной атаке по типу MITM субъектом может выступать ИИ, который перехватывает и анализирует передаваемый трафик. В целях противодействия такой атаке исследователи предлагают производить анализ и классификацию сетевого трафика и обнаружение аномалий, на основе которых формируются прогнозы о возможной атаке [6].

Понятие компьютерной атаки по типу «злоумышленник посередине» (MITM)

Понятие компьютерная атака по типу «злоумышленник посередине» (MITM) является собирательным, описывает ситуацию, когда субъект использует различные методики и технические решения, направленные на получение доступа к трафику и его декодировки. Данная компьютерная атака реализуется с применением алгоритмов перехвата трафика, передаваемого между двумя оконечными устройствами.

Корпорация по управлению доменными именами и IP-адресами (Internet Corporation for Assigned Names and Numbers, ICANN) в атаке «злоумышленник

6 ISO/IEC 27000:2014 Информационные технологии. Методы и средства обеспечения информационной безопасности. Системы менеджмента информационной безопасности. Общий обзор и терминология (Information technology. Security techniques. Information security management systems. Overview and vocabulary)

посередине» качестве посредника определяет как человека, так и устройство, которые имеют возможность перехватывать или модифицировать данные, пересылаемые между двумя абонентами системы связи. ICANN приводит два примера атак MITM в Интернете.

Первый — это «клонирование» или подмена точки доступа (иногда такой тип атаки именуется «злой двойник».

Второй тип атак называется «противник в браузере»[7].

ФСТЭК России связывает проведение MITM-атаки с уязвимостью реализации протокола инкапсуляции Ethernet, которая позволяет объединять заголовки. Эксплуатация данной уязвимости позволяет действовать удаленно и вызывать необходимые технические сбои с последующей реализацией атаки (MITM)»[8].

Исследователи Keeper Security считают, что MITM это тип компьютерной атаки, при «которой злоумышленник перехватывает данные, передаваемые между двумя устройствами, компьютером или мобильным терминалом, на котором запущен веб-браузер и главный сервер»[9].

Классическая MITM-атака проходит в два этапа. Первый – перехват данных, когда преступник интегрируется в среду передачи данных. Далее при помощи спуфинга реализует подмену IP-адресов, ARP[10]-сообщений, сервера доменных имен и т.д. Атаки MITM зачастую используют ARP-кэш, который представляет собой локальный кэш с назначенными IP-адресами и сопоставленными физическими уникальными идентификаторами устройств в сети (MAC-адресами). В результате реализуются задачи получения сведений о структуре исследуемой сети и сопоставления локальных идентификаторов и универсальных идентификаторов сети (MAC-адресов).

Второй этап – дешифрация, т.е. получение доступа к зашифрованным данным. Поскольку существует большое разнообразие методик проведения атаки, существуют и методики противодействия, одной из которых является анализ исходного кода (как устанавливаемых приложений, так и исследование передаваемых данных в рамках «песочницы» на виртуальной машине и без наличия выхода в открытую сеть). Такой анализ может осуществляться вручную специалистом, либо анализироваться при помощи решений, созданных на основе обученного по соответствующему направлению ИИ.

Требования по безопасности информации к средствам защиты информации от воздействий

ФЗ «О техническом регулировании»[11] в части обеспечения информационной безопасности не устанавливает требований об обязательной сертификации средств защиты информации. Требования об обязательной сертификации средств защиты информации (технические, криптографические, программные и другие средства, предназначенные для защиты сведений, составляющих государственную тайну, средства, в которых они реализованы, а также средства контроля эффективности защиты информации) определены Постановлением Правительства РФ «О сертификации средств защиты информации»[12]. Но, как мы можем отметить, эти требования касаются информационных технологий, обрабатывающих государственную тайну.

Условно, с 2018 года ФСТЭК России начала формировать требования к поставщикам средств защиты информации, при этом были определены требования к средствам защиты информации от воздействий, направленных на отказ в обслуживании информационных (автоматизированных) систем[13]. В 2020 г. ФСТЭК России сформулировала требования, устанавливающие уровни доверия к средствам технической защиты информации и средствам обеспечения безопасности информационных технологий[14].

С целью предотвращения перехвата трафика и обеспечения информационной безопасности пользователей сети ФСТЭК России определила требования о сертификации межсетевых экранов. В 2023 г. ФСТЭК России установила соответствия классов защиты многофункциональных межсетевых экранов уровня сети уровням доверия[15]. В 2024 г. в СМИ была опубликована информация, что «Росреестр рассматривает возможность заказать продукты в области межсетевых экранов нового поколения (NGFW). Стоимость проекта оценивается в 1 млрд руб. Другим крупным заказчиком NGFW является ВТБ»[16].

7 Что такое атаки типа «злоумышленник в середине» или, как их еще называют, атаки посредника (Man in the Middle Attack, MIMA)? // https://www.icann.org/ru/blogs/details/what-is-a-man-in-the-middle-attack-2-11-2015-ru (дата обращения 20.09.2024)

8 BDU:2022-05987: Уязвимость реализации протокола инкапсуляции Ethernet, связанная с возможностью объединения заголовков, позволяющая нарушителю вызвать отказ в обслуживании или реализовать атаку «человек посередине» (MITM) // https://bdu.fstec.ru/vul/2022-05987

9 Что такое атаки «злоумышленник в середине»? // https://www.keepersecurity.com/blog/ru/2023/10/16/how-to-detect-man-in-the-middle-attacks/ (дата обращения 20.09.2024)

10 ARP — протокол в компьютерных сетях, предназначенный для определения MAC-адреса другого компьютера по известному IP-адресу.

11 ФЗ «О техническом регулировании» от 27 декабря 2002 г. № 184-ФЗ // СЗ РФ 2002. № 52 (Ч. 1). Ст. 5140.

12 Постановление Правительства РФ «О сертификации средств защиты информации» от 26 июня 1995 г. № 608 // СЗ РФ 1995. № 27. Ст. 2579.

13 Требования по безопасности информации к средствам защиты информации от воздействий, направленных на отказ в обслуживании информационных (автоматизированных) систем (утв. приказом ФСТЭК России от 30.07.2018 N 132) (Документ опубликован не был) // СПС «КонсультантПлюс».

14 Требования по безопасности информации, устанавливающие уровни доверия к средствам технической защиты информации и средствам обеспечения безопасности информационных технологий (утв. приказом ФСТЭК России от 02.06.2020 N 76) (Документ опубликован не был) // СПС «КонсультантПлюс».

15 Требования по безопасности информации к многофункциональным межсетевым экранам уровня сети (утв. приказом ФСТЭК России от 07.03.2023 № 44) (Документ опубликован не был) // СПС «КонсультантПлюс».

16 Росреестр закажет средства защиты информации на миллиард // https://www.cnews.ru/news/top/2024-07-08_rosreestr_zakazhet_sredstva (дата обращения 20.09.2024)

Некоторые методики выявления атаки MITM

Поскольку атака MITM направлена на получение несанкционированного доступа, то для обнаружения этой атаки могут использоваться сигнатурные и эвристические анализаторы, входящие в состав систем обнаружения атак (IDS). Так, исследователи в 2017 г. предположили, что со временем может состояться переход к эвристическим решениям в средствах обнаружения/предотвращения вторжений (IDS/IPS). В настоящее время применяются как сигнатурные, так и эвристические анализаторы, поскольку обе модели обнаружения атак MITM имеют свои сильные и слабые стороны в зависимости от конкретной ситуации.

В основе работы эвристического анализатора заложена схема, в которой в режиме обучения формируются «правильные» шаблоны поведения системы, а в режиме анализа – обнаруживаются отклонения от этих шаблонов. За счет этого эвристический анализатор может обнаружить вредоносную активность, не попавшую ни под какую конкретную сигнатуру.

Исследования в области разработки современных систем обнаружения компьютерных атак показывают, что методы ИИ и машинного обучения могут быть применимы в области обнаружения/предотвращения атак по типу MITM [7]. Одним из главных преимуществ эвристических анализаторов IDS, использующих методы машинного обучения, является их способность выявлять новые виды атак, в отличие от сигнатурных анализаторов [8]. Основным компонентом эвристического анализа вредоносных программ является искусственная нейронная сеть в виде многослойного перцептрона с иммунным обучением [9]. Для решения задачи обучения авторы статьи использовали модель кодирования настраиваемых параметров в виде адаптивного структурированного мультиантитела, что позволило уменьшить количество нейронов в скрытом слое и устранить, таким образом, избыточность нейронной сети.

Другие исследователи [9] пришли к выводу, что большинство используемых методов глубокого обучения в области обнаружения вторжений показывают хорошие результаты, независимо от того, используется какой-то один вид нейронной сети (например, recurrent neural network – RNN) или их сочетание (например, convolutional neural network CNN-RNN). Сочетания призваны устранить недостатки конкретных методов или в целом улучшить степень автоматизации всего процесса выявления атак. Использование методов глубокого обучения с учетом всех предварительных и вспомогательных приемов является более эффективным, чем просто использование этих методов перед классическими методами машинного обучения. Нейронные сети, особенно при их комбинировании с другими, не относящимися к глубокому обучению методами, обычно демонстрируют хорошие результаты.

Авторы статьи [10] подчеркивают, что популярностью у исследователей пользуются RNN и CNN и их сочетания, но все чаще в новых исследованиях разработчики обращаются к таким архитектурам технологий как автокодировщики, графовые нейронные сети, трансформеры.

Сетевая система обнаружения вторжений с применением машинного обучения позволяет выявлять широкий спектр веб-атак, производимых на сетевом уровне [11]. Другие исследователи предлагают для решения этой задачи разработанный ими алгоритм выявления атак по типу MITM для статически назначаемых IP-адресов хоста, а также IP-адресов, назначаемых через DHCP. Этот алгоритм, как пишут авторы статьи, они реализовали с использованием асинхронного метода диспетчеризации для снижения затрат на производительность [12].

Однако, несмотря на то что предлагаются различные методики выявления/предотвращения атаки по типу MITM, наиболее эффективным методом предотвращения атаки по типу MITM являются предиктивные сетевые технологии, которые основаны на алгоритмах искусственного интеллекта и машинного обучения.

Обзор некоторых российских решений, анализаторов исходного кода

Идет постоянный поиск наиболее удачных решений, которые обсуждаются как на теоретическом, так и на практическом уровнях. Разнообразие атак по типу MITM эксплуатирует уязвимости информационных технологий, наиболее сложными для обнаружения являются программные закладки, и эффективность этого зависит от уровня их встраивания. Программные закладки, встроенные на этапе производства, практически не поддаются выявлению[17]. Существуют три основных типа анализаторов исходного кода программного обеспечения (ПО) на наличие уязвимостей и закладок:

1. Анализаторы кода веб-приложений, которые помогают предотвратить уязвимости на веб-сайтах.
2. Анализаторы встраиваемого кода, которые позволяют найти проблемы в исходных текстах модулей, расширяющих функциональность корпоративных систем, таких как 1С, CRM и SAP.
3. Анализаторы исходного кода на других языках программирования, не связанных с бизнес- и веб-приложениями.

Наибольшего результата в области анализа исходного кода, позволяет достичь применение двух

17 Кое-что о закладках, или Как АНБ следит за пользователями // https://www.cryptopro.ru/en/blog/2015/11/10/koe-chto-o-zakladkakh-ili-kak-anb-sledit-za-polzovatelyami (дата обращения 20.09.2024)

основных технологий анализа – динамический анализатор (DAST – Dynamic Application Security Testing) и статический анализатор (SAST – Static Application Security Testing), разновидностью которого является бинарный анализ.

Предлагаемое на российском рынке компанией Solar решение Solar appscreener, технологическая основа которого представлена на рисунке 1, позволяет применять технологии динамических и статических анализаторов кода.

Рис. 1. Представленная разработчиками технологическая основа «Solar appScreener»

В основе подхода, реализованного Solar, применяется единая технологическая платформа, обеспечивающая комплексный анализ безопасности приложений[18]. В нее входит ядро платформы, технологические модули, коннекторы, единый интерфейс для удобного управления сканированиями, корреляция результатов разных видов анализа и функция получения подробного отчета. В технологическом решении используется технология Fuzzy Logic Engine для сокращения ложных срабатываний.

Сканер уязвимостей в Yandex выступает еще одним российским решением в области анализа исходного кода. При этом он позволяет хранить и распространять Docker-образы[19], размещаемые в отказоустойчивом хранилище. Для всех данных настроена автоматическая репликация при редактировании, создании и удалении Docker-образа меняется каждая копия. Docker-образы передаются по протоколу HTTPS. Сканер уязвимостей анализирует Docker-образ и сравнивает его содержимое с базами уязвимостей CVE[20].

Следующим решением является система обнаружения вторжений, осуществляющая мониторинг и обработку событий внутри хоста – ViPNet IDS HS от INFOTECS, которое использует сигнатурный и эвристический методы анализа атак на основе правил и сигнатур, разработанных в России. За счет централизованного управления агентами, настройками и группами правил на хостах администраторы по информационной безопасности могут оперативно реагировать на события безопасности в сети[21].

Статическим анализатором исходного кода для поиска ошибок и уязвимостей в программах на языке C, C++ и C# выступает анализатор PVS-Studio разработанный компанией ООО «СиПроВер»[22].

На рынке также представлены анализаторы с открытым исходным кодом, например, SonarQube, как платформа для непрерывной оценки качества кода путем статического анализа и измерения качества программного кода. В возможности платформы входит анализ кода и поиск ошибок согласно правилам стандартов программирования некоторых языков[23].

Обзор вредоносных систем ИИ предназначенных для совершения компьютерных атак различного типа

В настоящее время разработаны и применяются ряд систем ИИ изначально предназначенных для совершения компьютерных атак различного типа, некоторые наиболее часто применяемые представлены в таблице 1. По каждой из систем проводились исследования, направленные на изучение возможностей и особенностей ее применения для осуществления незаконного анализа данных и нецелевого использования информационных систем.

Вредоносные системы используются для совершения компьютерных атак в различных сферах, в частности в различных сферах экономики. Подтверждение этому представлено в табл. 2.

Основным методом осуществления представленных атак является состязательное машинное обучение (Adversarial Machine Learning (AML) как метод, основанный на машинном обучении, суть которого заключается в использовании существующих «слепых зон» между обрабатываемыми в процессе обучения модели совокупности данных. При обучении вредоносного ИИ определяются слабые стороны защищаемой системы и вносятся небольшие изменения в ее массивы данных. В связи с этим, в защищаемой модели не формируются устойчивые связи между целевыми значениями, что в дальнейшем приводит к неправильным классификациям с пересечением границы принятия решения и ошибочного

18 SOLARAPPSCREENER // https://rt-solar.ru/products/solar_appscreener (дата обращения 20.09.2024)

19 Шаблон (исполняемый пакет), из которого создаются Docker-контейнеры. Образ содержит всё необходимое для запуска приложения, помещённого в контейнер: код, среду выполнения, библиотеки, переменные окружения и конфигурационные файлы.

20 Yandex Container Registry // https://yandex.cloud/ru/services/container-registry?utm_source (дата обращения 20.09.2024)

21 О продукте //https://infotecs.ru/products/vipnet-ids-hs-versiya-1/#:~:text (дата обращения 20.09.2024)

22 Как PVS-Studio ищет ошибки: методики и технологии // https://habr.com/ru/companies/pvs-studio/articles/319382/ (дата обращения 20.09.2024)

23 Keep AI generated code clean // https://www.sonarsource.com/products/sonarqube/ (дата обращения 20.09.2024)

Таблица 1.

Инструменты на базе искусственного интеллекта, использующие анализ данных для совершения компьютерных преступлений

Наименование	Область применения
DeepHack	Инструмент на базе искусственного интеллекта для создания шаблонов атак с инъекциями для приложений баз данных[24]
DeepLocker	Инструмент на базе искусственного интеллекта, который эмулирует APT для запуска сложных кибератак[25]
GyoiThon	Инструмент на базе искусственного интеллекта для сбора информации и автоматической эксплуатации[26]
EagleEye	Инструмент на базе искусственного интеллекта для разведки информации в социальных сетях с использованием алгоритмов распознавания лиц[27]
Malware-GAN	Инструмент на базе искусственного интеллекта, используемый для создания вредоносного ПО, которое может обходить механизмы обнаружения безопасности[28]
uriDeep	Инструмент на базе искусственного интеллекта, который генерирует поддельные домены для использования в различных сценариях атак[29]
Deep Exploit	Инструмент на базе искусственного интеллекта, который автоматизирует Metasploit для сбора информации, сканирования и последующей эксплуатации[30]
DeepGenerator	Инструмент на базе искусственного интеллекта для создания шаблонов атак с инъекциями для веб-приложений

Таблица 2.

Вредоносные алгоритмы, используемые для подмены данных и в целях обхода решений на основе ИИ

Объект воздействия	Способ воздействия
Дорожные знаки	Неправильная классификация дорожного знака алгоритмами ИИ может привести к дорожно-транспортным происшествиям на автономных автомобилях[31]
Данные медицинских изображений	Неправильная классификация медицинских отклонений алгоритмами ИИ может привести к ложной диагностике состояния здоровья [11].
Данные изображений лица	Неправильная классификация изображений лиц может привести к аутентификации [13].
Цифровая рекомендация системы	Внесение ложных данных алгоритмами ИИ может привести к неверным рекомендациям [14].
Данные КТ сканирования	Неправильная классификация подделанных 3D-изображения компьютерной томографии может привести к ложной диагностике [15].
Речевые аудиоданные	Состязательная атака на голосовую активацию персональной помощи может нарушить ее функциональность [16].
Системы обнаружения сетевых вторжений	Генерация вредоносного трафика для обхода защиты систем обнаружения сетевых вторжений на базе искусственного интеллекта [17].

отнесения данных к другому классу. В дальнейшем другой ИИ при анализе защищаемой системы, не покажет наличие вредоносных данных, поскольку произошла подмена классификации данных.

Ряд авторов [16], анализируя этап машинного обучения, выделяют четыре основных направления атак AML:

- атаки на решение классификатора, включая отравляющие (причинные) атаки на этапе обучения и исследовательские (уклоняющиеся) атаки обученной модели на этапах тестирования;
- атаки либо на целостность модели, приводящие к неправильной классификации, либо на пригодность модели при наличии высокой частоты неправильных классификаций;
- целенаправленные атаки, когда состязательные выборки нацелены на достижение определенного целевого значения, или неизбирательные атаки, когда выборки не нацелены на определенное целевое значение;
- атаки на конфиденциальность, целью которой выступает извлечение информации из классификатора.

Иной подход к классификации атак предлагается на основе:

- сложности по критерию последствий, варьирующихся от незначительного снижения достоверности прогнозов модели до неправильной классификации всех невидимых точек данных;

24 Bishopfox/deephack: Poc code from def con 25 presentation. (дата обращения 20.09.2024)

25 Cyberwarefare/deeplocker: Deeplocker – deep learning based malware. (дата обращения 20.09.2024)

26 Gyoisamurai/gyoithon: Gyoithon is a growing penetration test tool using machine learning. (дата обращения 20.09.2024)

27 Thoughtfuldev/eagleeye: Stalk your friends. find their instagram, fb and twitter profiles using image recognition and reverse image search. (дата обращения 20.09.2024)

28 Yanminglai/malware-gan: Realization of paper: «generating adversarial malware examples for black-box attacks based on gan» 2017. (дата обращения 20.09.2024)

29 Mindcrypt/urideep: Unicode encoding attacks with machine learning. (дата обращения 20.09.2024)

30 Machine learning security/deepexploit at master · 13o-bbr- bbq/machine learning security. (дата обращения 20.09.2024)

31 Nicolas Papernot, Patrick McDaniel, Ian Goodfellow, Somesh Jha, Z Berkay Celik, and Ananthram Swami. Practical black-box attacks against machine learning. In Proceedings of the 2017 ACM on Asia conference on computer and communications security, pages 506–519. ACM, 2017.

- полученного противником знания, например по типу «атака белого ящика»[32], в целях получения злоумышленником знаний об обучающей модели (ее архитектуре, сетевом трафике, который она анализирует, и ее функциям, которые используются для поддержки обучения) [17].

AML атаки исследователи предлагают классифицировать как таргетированные, которые направлены на изменение предсказания классификатора к определенному классу. Исследование особенностей осуществления автоматически генерируемых AML атак, позволило прийти ученым к выводу, что система защиты от таких атак может быть разработана на основании анализа алгоритмов машинного обучения при применении состязательных выборок.

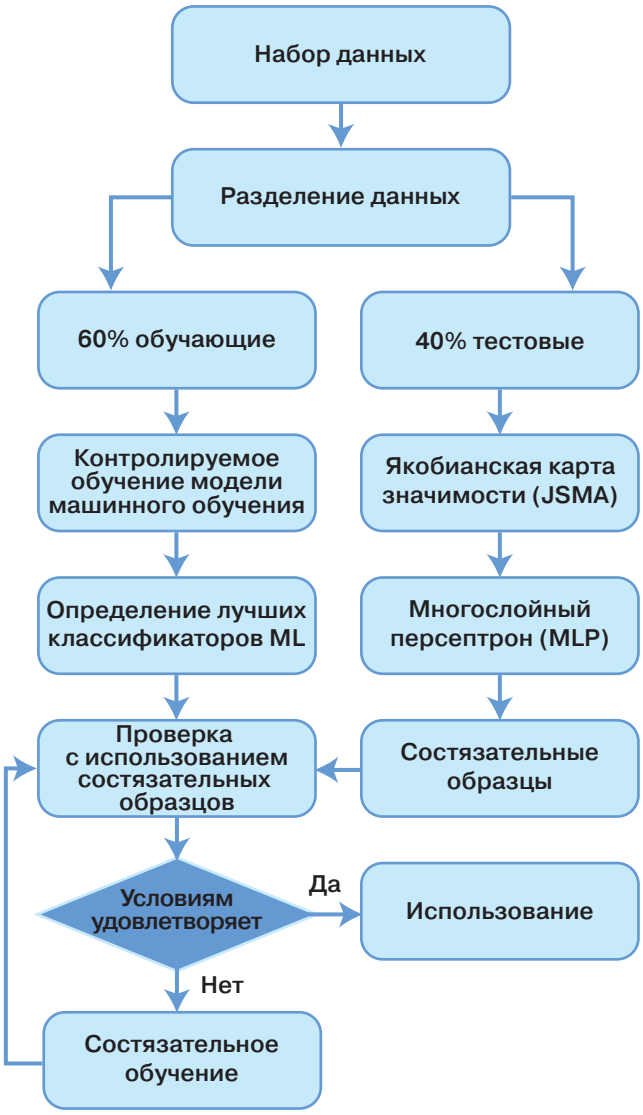

Рис. 2. Применение состязательных выборок для построения системы защиты от атак

32 Противоположностью выступает атака по типу черного ящика, когда у противника нет информации о внутренней работе целевой модели.

Для обнаружения вторжений в систему ученые предлагают разделять данные на обучающий и тестовый наборы, в соотношении – 60 % и 40 % соответственно (рис. 2). Далее производится оценка моделей машинного обучения под наблюдением с установлением наиболее эффективных моделей. Производится генерация состязательных выборок с использованием метода карты значимости на основе метода Якобиана, оценивается производительность системы, обученной на сгенерированных состязательных выборках. Процент состязательных выборок включается в обучающие данные и производится повторное обучение и оценка моделей [17].

Различные методы генерации состоятельных выборок целесообразно классифицировать по сложности, скорости генерации данных, и их производительности.

Наиболее простой (также наиболее трудоемкий и наименее точный) подход заключается в ручном изменении входных точек данных. К популярным методам автоматической генерации возмущенных выборок относят метод быстрого градиентного знака (Fast Sign Gradient Method, FGSM) и метод использования карты значимости на основе метода Якобиана (Jacobian saliency map, JSMA). Оба метода используют алгоритм, согласно которому при добавлении небольших изменений (δ) к исходной выборке (X) результирующая выборка X^* может демонстрировать состязательные характеристики $X^* = X + \delta$. Оба метода также обычно применяются при использовании предварительно обученного многослойного перцептрона (Multilayered perseptron, MLP) в качестве базовой модели для генерации состязательной выборки.

Метод FGSM воздействует на входные данные путем добавления определенного количества возмущения, когда знаки от функции градиента исходной функции потерь умножаются на некоторый ε. Шум возмущения вычисляется градиентом функции стоимости J по отношению к входным данным. Пусть θ представляет параметры модели, x — входные данные для модели, y — метки, связанные с входными данными, ε — значение, которое представляет степень применяемого шума, а $J(\theta,x,y)$ — функция стоимости, используемая для обучения целевой нейронной сети.

$$X' = X + e * sign(\nabla_x J(X, ytrue))$$

В JSMA методе считается прямая производная, на основании чего строится карта градиентов. На карте каждому параметру объекта соответствует критерий и его удельный вес, направленный на изменение конечного результата работы алгоритма. Тем самым, метод позволяет изменить как можно меньше параметров в атакуемом объекте.

Таблица 3.

Развитие функционала вредоносной GAN по некоторым отраслям экономики

Год	Название	Способ воздействия	Тип данных
2016	TextGAN	Синтетическая генерация текста посредством состязательного обучения	Текстовый
2017	FM-GAN	Генерация синтетического текста с помощью состязательных признаков	Текстовый
2017	MidiNet	Генерация синтетического звука	Аудио
2017	Age-cGAN	Предсказание возраста лица с помощью условных генеративных состязательных сетей	Визуальный
2017	CVAE-GAN	Генерация синтетического изображения лица	Визуальный
2017	SenseGen	Модель глубокого обучения для генерации синтетических данных датчиков	Текстовый
2018	WGAN	Генерация синтетического изображения МРТ мозга	Визуальный
2018	ACGAN	Генерация синтетического медицинского изображения печени	Визуальный
2018	Predestrian Synthesis GAN	Генерация синтетических данных пешеходов	Визуальный
2018	HP-GAN	Генерация синтетических данных для прогнозирования движения человека	Визуальный
2018	VAE-GAN	Генерация синтетического видео из текста	Визуальный
2018	WaveGAN	Состязательный синтез звука	Аудио
2019	DermGAN	Генерация синтетического изображения кожи	Визуальный
2019	CT-GAN	Генерация синтетического медицинского изображения МРТ	Визуальный
2019	X2CT-GAN	Генерация синтетического медицинского изображения рентгена	Визуальный
2020	D-NET [20]	Генерация биометрических данных радужной оболочки глаза	Визуальный
2021–2022	GPT-Chatbot[33]	Генерация предвзятого, неэтичного и опасного материала	Визуальный, текстовый, аудио
2023	ChatGPT от OpenAI[34]	Анализ запросов сотен миллионов людей по всему миру и сопоставление их с данными, снимаемыми с конечных устройств, в том числе с информацией о транзакциях, выполненных с помощью Apple Pay, геолокации, голосовыми командами и тысячами других дата-маркеров	Визуальный, текстовый, аудио
2024	Midjourney[35]	Использование дипфейков (поддельных изображений, видео и аудио) для манипуляций общественным мнением и дискредитации соперников	Визуальный, текстовый, аудио

Якобиан используется для вычисления карты значимости, которая определяет какие особенности входных данных являются наиболее релевантными для модельного решения. Эти характеристики, если их изменить, скорее всего, повлияют на классификацию целевых значений.

Учитывая, что методу JSMA может потребоваться несколько итераций для генерации состязательных выборок, FGSM быстрее в вычислительном отношении, несмотря на то что он изменяет каждую функцию. Кроме того, в отличие от FGSM, JSMA является более сложным подходом, но наиболее точно представляет

33 Администрация президента США выпустила 5 положений о защите людей от ИИ // https://www.tadviser.ru/index.php/Статья:Риски_использования_искусственного_интеллекта#2019:_.D0.A1.D0.B5.D0.BA.D1.81.D0.B8.D0.B7.D0.BC_.D0.B8_.D1.88.D0.BE.D0.B2.D0.B8.D0.BD.D0.B8.D0.B7.D0.BC_.D0.B8.D1.81.D0.BA.D1.83.D1.81.D1.81.D1.82.D0.B2.D0.B5.D0.BD.D0.BD.D0.BE.D0.B3.D0.BE_.D0.B8.D0.BD.D1.82.D0.B5.D0.BB.D0.BB.D0.B5.D0.BA.D1.82.D0.B0._.D0.9F.D0.BE.D1.87.D0.B5.D0.BC.D1.83_.D1.82.D0.B0.D0.BA_.D1.81.D0.BB.D0.BE.D0.B6.D0.BD.D0.BE_.D0.B5.D0.B3.D0.BE_.D0.BF.D0.BE.D0.B1.D0.BE.D1.80.D0.BE.D1.82.D1.8C.3F

34 Ваш карманный манипулятор: чем опасен генеративный ИИ в смартфонах // https://trends.rbc.ru/trends/industry/6698fef79a79472609486cff?from=copy

35 Искусственный интеллект и генеративные инструменты меняют американскую политику // https://www.securitylab.ru/news/538553.php (дата обращения 20.09.2024)

атаки, поскольку он в течении длительного времени пошагово изменяет небольшой процент функций. В связи с этим, точность JSMA в значительной степени зависит от количества входных функций. Чем больше пространство признаков, тем больше итераций требуется для определения наиболее успешного подхода при генерации состязательных выборок, влияющих на производительность модели.

Традиционно для выявления атак используются алгоритмы Naive Bayes, Random Forest, SVM, и J4. Наиболее современными инструментами обнаружения атак по типу AML являются Recurrent Neural Networks.

Вредоносный ИИ может применяться и в такой модели нейронной сети как генеративно-состязательная сеть (generative adversarial network, GAN). В ней обучаются одновременно две сети (одна – генерация изображений, вторая – отраслевая визуализация). Архитектура GAN включает генератор и дискриминатор, каждый представляет собой сети с разным задачами. Генератор изучает распределение данных и генерирует образцы для сети дискриминатора. Дискриминатор определяет происходит ли выборка генератора из исходных данных или из сети генератора, на его вход поступают два типа выборок: из исходных данных и сгенерированные сетью-генератором.

При этом функция GAN заключается в генерации данных (визуальных, текстовых и аудиальных) с помощью приложений, например, рисования видео, синтеза звука, суперразрешения, интеллектуального анализа текста и синтеза обучающих данных для обучения других глубоких сетей. Поскольку эта технология является относительно недорогой, она применяется как в различных отраслях экономики, так и во вредоносных целях. Так, например, функционал GAN может использоваться во вредоносных целях, некоторые из существующих по отраслям экономики представлены в таблице 3.

Во время обучения генератор пытается создать более реалистичные образы, чтобы обмануть дискриминатор, в то время как дискриминатор пытается отличить исходные и синтетические образы. Обучение GAN осуществляется сквозным образом. Предполагается, что сеть будет обучена, когда неточность генератора (неудачная попытка обмануть сеть дискриминатора) будет равна неточности сети дискриминатора (отсутствие дискриминации между реальным и синтетическим образцом). Однако практически очень сложно установить такое равновесие, поскольку функции неточности колеблются вокруг положения равновесия. Обычно через несколько сотен циклов сгенерированные данные проверяются визуально или с помощью соответствующей метрики.

Некоторыми исследователями [19] предлагается использовать облачную инфраструктуру для идентификации атаки по типу MITM, осуществляемой с использованием ИИ. Облачная инфраструктура позволяет предотвратить атаки и создать защищенный административный центр на основании трех показателей: энтропия IP-адреса, местоположение порта и скорость поступления данных. На основании этих показателей ИИ вычисляет вероятность атаки по типу MITM.

Модель использования ИИ в целях выявления аномальных активностей

На эффективность установления субъектов MITM-инцидентов влияет, в том числе, «цифровая юрисдикция», а именно расположение основной серверной части атакующего в рамках контролируемой сетевой инфраструктуры государства или группы государств, в которых возможен сбор и анализ информации, в частности о крипто-транзакциях.

Зачастую лицо, реализующее атаку по типу MITM, использует виртуальной сервер – VPS (Virtual Private Server) на территории неконтролируемой государственными органами. На сервере VPN «развернута» виртуальная частная сеть на основе технологии VPN и маскируется структура сети через стек правил передачи данных по типу NAT, посредством которого реализована процедура преобразования IP-адреса(ов) узла(ов) локальной сети, либо удаленного тунелируемого узла.

При «многоступенчатой структуре», т. е. использовании ряда задействованных в инициализации конкретной атаки терминалов и локальных сетей, следует устанавливать всю цепочку задействованных в событии. активных элементов.

Такой цепочкой может быть: использование на оконечном устройстве – сервере, мобильном терминале (сотовом телефоне), ноутбуке, стационарном персональном компьютере, планшете или айпаде. Иными словами, на любом устройстве с сетевым интерфейсом (сетевой картой) и возможностью использования сетевых протоколов (правил передачи данных), позволяющих реализовывать функцию выхода в глобальную сеть Интернет. А также получение данных о структуре подсетей, классов, а также об организации, которой принадлежит исследуемый идентификатор (IP-адрес), о диапазоне адресов, которым владеет организация. Это возможно сделать, используя специализированное программное обеспечение (сетевые анализаторы), а также алгоритмы, реализованные в функционале терминальных команд по типу «look up».

Основной процедурой в механизме анализа выступает технология проверки сетевых пакетов –

DPI (Deep Packet Inspection) с помощью которой представляющий интерес «аномальный трафик» записывается в «логи» (журналы событий) в соответствии с руководящей документацией и внутренними инструкциями.

В случае шифрования OpenVPN его отличительный маркер отслеживается посредством DPI. Функция VPN, позволяющая скрыть зашифрованный трафик OpenVPN, имитируя его в обычный интернет-трафик («обфускация»), реализуется путем удаления данных, связанных с VPN из пакета OpenVPN, и назначения транслируемому шифрованному трафику порта 443, изначально предназначенного для передачи трафика по протоколу HTTPS. После добавления экспорта MTU через API (Application Programming Interface) и обновления сигнатур появляется возможность установить пользователей VPN-протоколов, proxy[36], а также выявить смену User-Agent[37], что позволяет установить уникальные идентификаторы оконечных устройств третьих лиц.

Изучив разные модели выявления неправомерной деятельности в Сети, авторы предлагают систему выявления инцидентов атак по типу MITM, а также несанкционированного доступа к ключевым элементам инфраструктуры распределенных компьютерных сетей (далее по тексту: «Система»). Основной целью создания системы являлась отработка на практике концепции использования ИИ при решении задачи выявления инцидентов компрометации в распределенных компьютерных сетях. Система предназначена для выявления инцидентов в процессе анализа трафика данных с установлением ключевых элементов инфраструктуры компьютерных сетей. Она может быть использована при проверке компьютерных сетей и их элементов на предмет установления инцидентов компрометации – несанкционированного доступа к компонентам критической инфраструктуры.

Особенностью системы является адаптивное восприятие графического интерфейса с логическим представлением структуры локальной сети и визуальным представлением потоков данных анализируемой сети. Ней ронная сеть позволяет на основе атрибутов, полученных из трафика, оценить риск несанкционированного доступа.

Система предназначена для использования в информационно-аналитической деятельности и позволяет автоматизировать труд человека посредством применения технологий нейронных сетей для выявления несанкционированного использования вычислительных мощностей.

Критериями эффективности в этом случае являются:

- автоматизация процесса выявления признаков несанкционированного использования вычислительных мощностей компьютерных сетей;
- оперативность принятия решений на достаточно сформированном перечне признаков, предоставленных со стороны «высоко» (точность идентификация событий, а также поддержка при принятии решений не ниже 99 % по поставленным задачам) обученной нейронной сети;
- повышение эффективности противодействия несанкционированному использованию вычислительных мощностей компьютерных сетей.

Описание системы

1. Сценарии использования

Сценарий использования системы предполагает выполнение последовательности действий Оператора с применением смежных систем. Диаграмма сценариев использования представлена на рисунке 3.

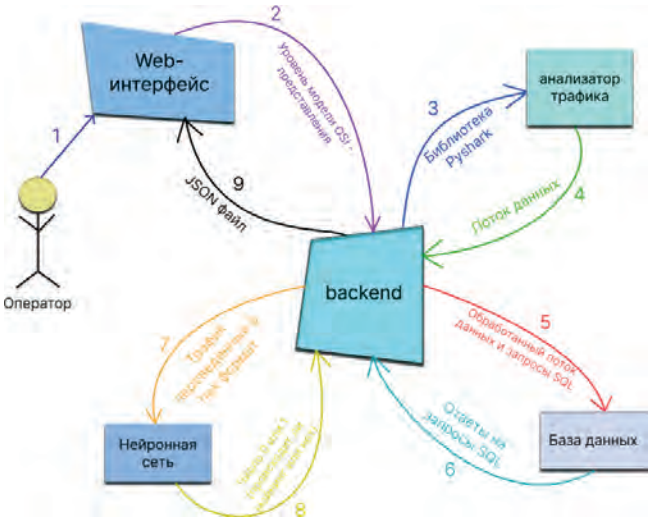

Рис. 3. Диаграмма сценариев использования

Оператор, с целью получения информации о трафике в компьютерных сетях, обращается к приложению, после чего действия осуществляются по следующему алгоритму:

1. Обращение оператора к Web-интерфейсу;
2. Запрос с Web-интерфейса к backend-у для получения информации;
3. До того, как backend отправит информацию Web-интерфейсу, ему необходимо выполнить несколько этапов:
   1) провести анализ трафика «перехватываемых» пакетов с данными;
   2) обработать полученные пакеты на сервере;
   3) сформировать и взаимодействовать с базой данных (PostgreSQL) путем пополнения обработанного на backend трафика;

36 Сетевой «посредник» между узлами
37 Идентификатор браузера

4) отправить полученные с backend обработанные на нейронную сеть, которая в свою очередь классифицирует полученный трафик;

По завершении цикла с backend-а сведения передаются на Web-интерфейс.

2. Архитектура системы

2.1. Описание Web-интерфейса

Система представлена в виде Web-приложения, реализованного на языках HTML, CSS и JavaScript. Используются библиотеки jQuery и anychart-graph. Кроссплатформенность системы обеспечена возможностью запуска на ОС Windows, macOS и семейства Unix-подобных операционных системах на базе ядра Linux.

Инициализируя запуск программного обеспечения посредством браузера, пользователю представляется граф анализируемой локальной сети, на котором отображены: количество устройств, их тип и роль, где основной особенностью разрабатываемого решения выступает возможность пользователя выявить признаки несанкционированного использования вычислительных мощностей конечных устройств в локальной сети.

Взаимодействие с серверной частью осуществляется в виде обмена данными. С сервера поступает файл в формате JSON, в котором содержится информация о состоянии сети. В свою очередь пользователь отправляет на сервер команды для взаимодействия с программой.

2.2. Описание сервера

Для запуска сервера используется фреймворк Django на языке Python. После производится загрузка на сервер данных с frontend и анализ сведений, поступающих из протокола stratum. Указанный протокол, используется большинством пулов для связи между заинтересованным лицом (атакующим субъектом) и сервером пула. Он состоит из набора инструкций, которые сервер может отправить «атакующему субъекту», и другого набора запросов, которые заинтересованное лицо может отправить на сервер. Рассматриваемый протокол реализован поверх TCP. Следует также отметить отсутствие функционирующих портов с вышеуказанным протоколом, и отсутствие возможности установления связей между слоями.

Рассматриваемый протокол - stratum использует формат JSON для всех своих методов и, как правило, использует вызовы:

– *subscribe,*
– *authorize,*
– *extranonce.subscribe,*
– *и submit.*

От сервера к клиенту идут:

– *set_difficulty*
– *и notify.*

Алгоритм взаимодействия пользователя и серверной части можно представить следующим образом:

- authorize: аутентификация субъекта на сервере.
- subscribe: запрос на сбор значимых данных.
- extranonce.subscribe: инициализации атаки по типу MITM.

Вызовы по форме взаимодействия Сервер — Клиент:

1) notify: сервер отправляет всю информацию, необходимую для запуска текущего блока, включая пользовательский идентификатор, используемый в качестве идентификатора при отправке хэша предыдущего блока.
2) set_difficulty: устанавливает сложность идентификации события.

При помощи библиотеки Pyshark осуществляется перехват пакетов в локальной сети, в которых используется протокол TCP, из которых берутся значения полезной нагрузки. IP-адреса устройств с такими пакетами отправляются на Web-интерфейс.

Также все IP-адреса сравниваются с базой данных PostgreSQL, в которой находятся адреса доменов пулов. Если найдено совпадение, можно сделать вывод, что на данном устройстве есть вероятность несанкционированного подключения и утечки информации по техническому каналу связи.

2.3. Сетевой анализатор как модуль комплекса, используемого в качестве поддержки для принятия решений

Модулем, используемым в качестве поддержки в принятии решений в рамках механизма анализа сетевого трафика, является Wireshark – это программное обеспечение имеющее графический пользовательский интерфейс и широкий спектр инструментария по сортировке и фильтрации. Он тем самым предоставляет возможность для оператора просматривать проходящий по сети трафик в режиме реального времени.

Модуль распространяется под свободной лицензией GNU GPL и использует для формирования графического интерфейса кроссплатформенную библиотеку GTK+. Существуют версии для большинства UNIX-подобных систем, в том числе GNU/Linux, Solaris, FreeBSD, NetBSD, OpenBSD, macOS, а также для Windows.

Данный модуль выступает в качестве поддержки принятия решений в вопросах обеспечения безопасности компонентов критической информационной инфраструктуры, идентификации аномальных процессов, возникающих при инициализации посторонних устройств в контролируемом сегменте обслуживаемой сети, с точностью идентификации инцидентов компрометации (о совокупности признаков) ~90 %.

2.4. Описание базы данных

Для хранения информации была использована база данных PostgreSQL. Всего будут использованы две базы данных:

- первая будет хранить в себе домены pool и их IP-адреса,
- вторая база данных будет содержать трафик, получаемый сервером, в котором будут известны IP источника и IP назначения, порт источник и порт назначения, а также количество передаваемой информации и время.

Изначально «используется» трафик, без признаков реализации атаки MITM. Далее он сравнивается с подозрительным трафиком. Эти данные запрашивает backend с помощью языка запросов SQL.

2.5. Описание нейронной сети

Мы используем нейросеть с реализованной сверточной моделью, состоящей из разных видов слоев: сверточные слои, субдискретизирующие слои и слои «обычной» нейронной сети – перцептрона. Первые два типа слоев, чередуясь между собой, формируют входной вектор признаков для многослойного перцептрона. Сверточные сети являются наиболее эффективными решениями при анализе конвертированного трафика.

Разработанная Система по выявлению инцидентов по типу MITM выступает в качестве MVP (Minimum Viable Product), ее функции позволяют отследить: преобразование нагрузки сетевого трафика из HEX в изображение; проанализировать полученное изображение с помощью сверточной модели нейронных сетей; получить ответ серверной части приложения коэффициентом идентичности исходного трафика с аномальным (исследуются также порты).

Разработанное решение (в виде MVP) включает преобразование необработанного сетевого трафика, собранного с помощью инструмента – сетевого анализатора трафика. Точность выявления инцидентов по совокупности признаков составляет около ~90 %.

Литература

1. *Жарова, А. К. Обеспечение права на доступ к Интернету и забвение в цифровом пространстве Российской Федерации / А. К. Жарова, В. М. Елин // Мониторинг правоприменения. – 2021. – № 2(39). – С. 48–53. – DOI 10.21681/2226-0692-2021-2-48-53. – EDN NEDFXI.*
2. *Жарова, А. К. Парадигма цифрового профилирования деятельности человека: риски, угрозы, преступления / А. К. Жарова, В. М. Елин, А. В. Минбалеев. – Москва: Общество с ограниченной ответственностью «Русайнс», 2022. – 240 с. – ISBN 978-5-466-00766-4. – EDN DNKVPR.*
3. *Zharova, A. The Bayes model for the protection of human interests / A. Zharova, V. Elin, M. Levashov // International Journal of Electrical and Computer Engineering. – 2023. – Vol. 13, No. 6. – P. 6419–6425. – DOI 10.11591/ijece.v13i6.pp6419-6425. – EDN CFNXXA.*
4. *Карцхия, А. А. Правовые горизонты технологий искусственного интеллекта: национальный и международный аспект / А. А. Карцхия, Г. И. Макаренко // Вопросы кибербезопасности. – 2024. – № 1(59). – С. 2-14. – DOI 10.21681/2311-3456-2024-1-2-14. – EDN JTGKFM.*
5. *Добрышин, М. М. Особенности применения информационно-технического оружия при ведении современных гибридных войн / М. М. Добрышин // I-methods. – 2020. – Т. 12, № 1. – С. 1–11. – EDN PPGYRU.*
6. *Yamin M. M., Ullah M., Ullah H., Katt B.Weaponized AI for Cyber Attacks // https://ntnuopen.ntnu.no/ntnu-xmlui/bitstream/handle/11250/3021130/Weaponized_AI_for_Cyber_Attacks__2_.pdf?sequence=1 (Дата обращения 20.09.2024)*
7. *Сычев, Д. И. Методы машинного и глубокого обучения для систем обнаружения вторжений: обзор и анализ / Д. И. Сычев // Международный журнал информационных технологий и энергоэффективности. – 2023. – Т. 8, № 4(30). – С. 9–17. – EDN CFCXQS.*
8. *Talukder, M. A., Islam, M. M., Uddin, M. A. et al. Machine learning-based network intrusion detection for big and imbalanced data using oversampling, stacking feature embedding and feature extraction. J Big Data 11, 33 (2024). https://doi.org/10.1186/s40537-024-00886-w*
9. *Шиловский, Г. В. Возможность реализации правдоподобных алгоритмов глубокого обучения на небольших нейронных сетях со скрытыми слоями / Г. В. Шиловский, В. М. Юлкова // Вестник компьютерных и информационных технологий. – 2020. – Т. 17, № 12(198). – С. 14–19. – DOI 10.14489/vkit.2020.12.pp.014-019. – EDN KJJLTW.*
10. *Getman A. I., Goryunov M. N., Matskevich A. G., Rybolovlev D. A., Nikolskaya A. G. Deep Learning Applications for Intrusion Detection in Network Traffic. Trudy ISP RAN/Proc. ISP RAS, vol. 35, issue 4, 2023 pp. 65–92 (in Russian). DOI: 10.15514/ISPRAS-2023-35(4)-3.*
11. *Avishek Joey Bose and Parham Aarabi. Adversarial attacks on face detectors using neural net based constrained optimization. In 2018 IEEE 20th Interna-tional Workshop on Multimedia Signal Processing (MMSP), pages 1–6. IEEE, 2018. 30*
12. *Способы осуществления специальных программных воздействий на радиоэлектронные объекты. Атаки Man-In-The-Middle / И. Г. Головенкин, Ю. Ю. Громов, Ю. А. Губсков, О. Г. Иванова // Промышленные АСУ и контроллеры. – 2018. – № 9. – С. 11–18. – EDN MAAYRV.*
13. *Samuel G Finlayson, Hyung Won Chung, Isaac S Kohane, and Andrew L Beam. Adversarial attacks against medical deep learning systems. arXiv preprint arXiv:1804.05296, 2018.*
14. *Christakopoulou K. and Banerjee A. Adversarial attacks on an oblivious recommender. In Proceedings of the 13th ACM Conference on Recommender Systems, pages 322–330, 2019.*
15. *Yisroel Mirsky, Tom Mahler, Ilan Shelef, and Yuval Elovici. Ct-gan: Malicious tampering of 3d medical imagery using deep learning. arXiv preprint arXiv:1901.03597, 2019.*
16. *Juncheng B Li, Shuhui Qu, Xinjian Li, J Zico Kolter, and Florian Metze. Adversarial music: Real world audio adversary against wake-word detection system. arXiv preprint arXiv:1911.00126, 2019.*

17. Aritran Piplai, Sai Sree Laya Chukkapalli, and Anupam Joshi. Nattack! adversarial attacks to bypass a gan based classifier trained to detect network intrusion // arXiv preprint arXiv:2002.08527, 2020.
18. Eirini Anthi, Lowri Williams, Matilda Rhode, Pete Burnap, Adam Wedgbury. Adversarial attacks on machine learning cybersecurity defences in Industrial Control Systems //Journal of Information Security and Applications 58 (2021) 102717
19. Anthi E., Williams L., Rhode M., Burnap P., Wedgbury A. Adversarial attacks on machine learning cybersecurity defences in Industrial Control Systems //Journal of Information Security and Applications 58 (2021) 102717
20. Fadi Boutros, Naser Damer, Kiran Raja, Raghavendra Ramachandra, Florian Kirchbuchner, and Arjan Kuijper. Iris and periocular biometrics for head mounted displays: Segmentation, recognition, and synthetic data generation. Image and Vision Computing, 104:104007, 2020.
21. Chowdary P., Challa Y., Jitendra M.Identification of MITM Attack by Utilizing Artificial Intelligence Mechanism in Cloud Environments // International conference on computer vision and machine learning IOP Conf. Series: Journal of Physics: Conf. Series 1228 (2019) 012044 IOP Publishing. doi:10.1088/1742-6596/1228/1/012044

PREVENTION OF COMPUTER ATTACKS SUCH AS MAN IN THE MIDDLE, COMMITTED USING GENERATIVE ARTIFICIAL INTELLIGENCE

Zharova A. K.[38], Elin V. M.[39], Avetisyan B. R.[40]

The purpose of the article *is to present to the scientific community the developed author's methodology for detecting/preventing a computer attack of the MITM type.*

The research method. *To achieve this goal, the authors used methods of mathematical modeling, comparative analysis, tabular method, as well as methods of experimental and theoretical level.*

Result. *The article conducted a comparative analysis of software solutions presented in the form of source code on sites like GITHUB, which provide the implementation of an attack in the middle in both local and global networks, as well as an analysis of some MITM-type attack prevention techniques using artificial intelligence (AI) services. Based on this analysis, various logical implementations of the MITM-type attack are identified, as well as vulnerabilities of information systems to a MITM computer attack are presented. Based on the analysis of existing methods of countering these attacks and the identified weaknesses of these methods, the authors propose an author's method of preventing MITM-type attacks, which includes training AI on data sets, connected libraries of different programming languages and algorithmized heuristic models that respond to changes in the logic of user behavior, or the activity of a personal computer, network equipment.*

The scientific novelty *of the article consists in the developed author's methodology for detecting/preventing a computer attack of the MITM type using "predictive" network technologies based on the use of neural networks trained by machine learning methods.*

Keywords: *Data sets, MITM, attack prevention techniques, heuristic models, user behavior, predictive network technologies.*

References

1. Zharova, A. K. Obespechenie prava na dostup k Internetu i zabvenie v cifrovom prostranstve Rossijskoj Federacii / A. K. Zharova, V. M. Elin // Monitoring pravoprimeneniya. – 2021. – № 2(39). – S. 48–53. – DOI 10.21681/2226-0692-2021-2-48-53. – EDN NEDFXI.
2. Zharova, A. K. Paradigma cifrovogo profilirovaniya deyatel'nosti cheloveka: riski, ugrozy, prestupleniya / A. K. Zharova, V. M. Elin, A. V. Minbaleev. – Moskva: Obshchestvo s ogranichennoj otvetstvennost'yu \«Rusajns»\, 2022. – 240 s. – ISBN 978-5-466-00766-4. – EDN DNKVPR.
3. Zharova, A. The Bayes model for the protection of human interests / A. Zharova, V. Elin, M. Levashov // International Journal of Electrical and Computer Engineering. – 2023. – Vol. 13, No. 6. – P. 6419-6425. – DOI 10.11591/ijece.v13i6.pp6419-6425. – EDN CFNXXA.
4. Karckhiya, A. A. Pravovye gorizonty tekhnologij iskusstvennogo intellekta: nacional'nyj i mezhdunarodnyj aspekt / A. A. Karckhiya, G. I. Makarenko // Voprosy kiberbezopasnosti. – 2024. – № 1(59). – S. 2–14. – DOI 10.21681/2311-3456-2024-1-2-14. – EDN JTGKFM.
5. Dobryshin, M. M. Osobennosti primeneniya informacionno-tekhnicheskogo oruzhiya pri vedenii sovremennyh gibridnyh vojn / M. M. Dobryshin // I-methods. – 2020. – T. 12, № 1. – S. 1–11. – EDN PPGYRU.
6. Yamin M. M., Ullah M., Ullah H., Katt B.Weaponized AI for Cyber Attacks // https://ntnuopen.ntnu.no/ntnu-xmlui/bitstream/handle/11250/3021130/Weaponized_AI_for_Cyber_Attacks__2_.pdf?sequence=1 (Data obrashcheniya 20.09.2024)
7. Sychev, D. I. Metody mashinnogo i glubokogo obucheniya dlya sistem obnaruzheniya vtorzhenij: obzor i analiz / D. I. Sychev // Mezhdunarodnyj zhurnal informacionnyh tekhnologij i energoeffektivnosti. – 2023. – T. 8, № 4(30). – S. 9–17. – EDN CFCXQS.

38 Anna K. Zharova, Dr.Sc. of Law, Professor of Financial University under the Government of the Russian Federation, Moscow. E-mail: anna_jarova@mail.ru
39 Vladimir M. Elin, Ph.D. in Law, Associate Professor at the Financial University under the Government of the Russian Federation; Associate Professor of the Department of Information Security at the Moscow University of the Ministry of Internal Affairs of Russia named after V. Ya. Kikot, Moscow. E-mail: vm_elin@mail.ru
40 Boris R. Avetisyan, Scientific Research Institute of Education and Science, Chief Researcher, Moscow. E-mail: Boris.Avetisyan@gmail.com

8. *Talukder, M. A., Islam, M. M., Uddin, M .A. et al. Machine learning-based network intrusion detection for big and imbalanced data using oversampling, stacking feature embedding and feature extraction. J Big Data 11, 33 (2024). https://doi.org/10.1186/s40537-024-00886-w*
9. *Shilovskij, G. V. Vozmozhnost' realizacii pravdopodobnyh algoritmov glubokogo obucheniya na nebol'shih nejronnyh setyah so skrytymi sloyami / G. V. Shilovskij, V. M. Yulkova // Vestnik komp'yuternyh i informacionnyh tekhnologij. – 2020. – T. 17, № 12(198). – S. 14–19. – DOI 10.14489/vkit.2020.12.pp.014-019. – EDN KJJLTW.*
10. *Getman A .I., Goryunov M. N., Matskevich A. G., Rybolovlev D. A., Nikolskaya A. G. Deep Learning Applications for Intrusion Detection in Network Traffic. Trudy ISP RAN/Proc. ISP RAS, vol. 35, issue 4, 2023 pp. 65–92 (in Russian). DOI: 10.15514/ISPRAS-2023-35(4)-3.*
11. *Avishek Joey Bose and Parham Aarabi. Adversarial attacks on face detectors using neural net based constrained optimization. In 2018 IEEE 20th Interna-tional Workshop on Multimedia Signal Processing (MMSP), pages 1–6. IEEE, 2018. 30*
12. *Sposoby osushchestvleniya special'nyh programmnyh vozdejstvij na radioelektronnye ob"ekty. Ataki Man-In-The-Middle / I. G. Golovenkin, Yu. Yu. Gromov, Yu. A. Gubskov, O. G. Ivanova // Promyshlennye ASU i kontrollery. – 2018. – № 9. – S. 11–18. – EDN MAAYRV.*
13. *Samuel G Finlayson, Hyung Won Chung, Isaac S Kohane, and Andrew L Beam. Adversarial attacks against medical deep learning systems. arXiv preprint arXiv:1804.05296, 2018.*
14. *Christakopoulou K. and Banerjee A. Adversarial attacks on an oblivious recommender. In Proceedings of the 13th ACM Conference on Recommender Systems, pages 322–330, 2019.*
15. *Yisroel Mirsky, Tom Mahler, Ilan Shelef, and Yuval Elovici. Ct-gan: Malicious tampering of 3d medical imagery using deep learning. arXiv preprint arXiv:1901.03597, 2019.*
16. *Juncheng B Li, Shuhui Qu, Xinjian Li, J Zico Kolter, and Florian Metze. Adversarial music: Real world audio adversary against wake-word detection system. arXiv preprint arXiv:1911.00126, 2019.*
17. *Aritran Piplai, Sai Sree Laya Chukkapalli, and Anupam Joshi. Nattack! adversarial attacks to bypass a gan based classifier trained to detect network intrusion. arXiv preprint arXiv:2002.08527, 2020.*
18. *Eirini Anthi, Lowri Williams, Matilda Rhode, Pete Burnap, Adam Wedgbury. Adversarial attacks on machine learning cybersecurity defences in Industrial Control Systems //Journal of Information Security and Applications 58 (2021) 102717*
19. *Anthi E., Williams L., Rhode M., Burnap P., Wedgbury A. Adversarial attacks on machine learning cybersecurity defences in Industrial Control Systems //Journal of Information Security and Applications 58 (2021) 102717*
20. *Fadi Boutros, Naser Damer, Kiran Raja, Raghavendra Ramachandra, Florian Kirchbuchner, and Arjan Kuijper. Iris and periocular biometrics for head mounted displays: Segmentation, recognition, and synthetic data generation. Image and Vision Computing, 104:104007, 2020.*
21. *Chowdary P., Challa Y., Jitendra M.Identification of MITM Attack by Utilizing Artificial Intelligence Mechanism in Cloud Environments // International conference on computer vision and machine learning IOP Conf. Series: Journal of Physics: Conf. Series 1228 (2019) 012044 IOP Publishing doi:10.1088/1742-6596/1228/1/012044*