МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н.Г.ЧЕРНЫШЕВСКОГО»

Кафедра теории функций и стохастического анализа

АНАЛИЗ СЕТЕВОГО ТРАФИКА КАК ИНСТРУМЕНТ ОПРЕДЕЛЕНИЯ УГРОЗ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

АВТОРЕФЕРАТ МАГИСТЕРСКОЙ РАБОТЫ

студента 2 курса 248 группы
направления 09.04.03 — Прикладная информатика

механико-математического факультета
Хуссейн Соха Бадр

Научный руководитель
доцент, к.э.н. _____ А.Р.Файзлиев

Заведующий кафедрой
д.ф.-м.н., доцент _____ С.П.Сидоров

Саратов 2023

ВВЕДЕНИЕ

С развитием сетей пятого поколения и технологий искусственного интеллекта появились новые угрозы и вызовы для систем беспроводной связи, особенно в области информационной безопасности.

В этом исследовании предоставляется обзор подходов для обнаружения атак методом глубокого обучения. В частности, сначала суммируется основные проблемы сетевой безопасности и обнаружения вторжений и представляем несколько успешных связанных приложений, использующих структуру глубокого обучения.

Основываясь на категоризации методов глубокого обучения, уделяется особое внимание методам обнаружения атак, основанным на различных архитектурах, таких как Автоэнкодер (Autoencoder), состязательные генеративные сети, рекурсивные нейронные сети и сверточные нейронные сети. Затем представляется несколько сравнительных наборов данных с описаниями и сравнивается производительность представленных подходов, чтобы показать текущее состояние методов обнаружения атак с использованием сред глубокого обучения.

Актуальность темы исследования заключается в том, при интенсивном использовании сети Интернет сетевая безопасность становится очень важной, когда целью защиты является предотвращение несанкционированного доступа и модификации. Растущее число подключенных к Интернету систем в финансовом, электронной коммерции и военном секторах делает их мишенями для сетевых атак с высоким риском и большим ущербом.

Предметом исследования являются методы машинного обучения для классификации сетевых атак без предварительного знания их конкретных характеристик.

Научная новизна исследования состоит в использовании нейронных сетей для распознавания сетевых угроз.

Практическая значимость заключается в возможности использования результатов исследования для решения практических задач, связанных с анализом угроз сетевой безопасности.

Цель и задачи исследования является обзор методов машинного обучения для обнаружения сетевых угроз. Для достижения поставленной цели в работе ставятся следующие задачи:

1. Обзор методов машинного обучения (классификация, кластеризация, регрессия, поиск ассоциативных правил) обнаружения атак.
2. Описание метрик для определения эффективности алгоритмов обнаружения атак.
3. Сравнение алгоритмов обнаружения атак по скорости, эффективности и стабильности.

Объектом исследования является оценка качества распознания сетевых атаках.

Методы исследования. Решение поставленных в работе задач осуществляется с использованием методов теории вероятностей, математической статистики, численных методов, методов классификации, а также моделей нейронной сети.

Во введении к диссертационной работе обоснована актуальность темы исследования, цель и задачи исследования, объект и методы исследования.

Первая глава магистрской работы посвящена аналитическому обзору сетевой безопасности и важности сетевой безопасности, архитектура интернета и аспекты уязвимости безопасности, модель сетевой безопасности, распространенные методы интернет-атак, Технология интернет-безопасности, и также безопасность в разные сетях.

Непрерывное развитие и широкое использование Интернета принесло множество преимуществ многим пользователям сети.

Между тем, при интенсивном использовании сети сетевая безопасность становится гораздо более важной. Сетевая безопасность тесно связана с компьютерами, сетями, программами, различными данными и т. Д., где целью

защиты является предотвращение несанкционированного доступа и модификации. Однако растущее число подключенных к Интернету систем в финансовом, электронной коммерции и военном секторах делает их мишенями для сетевых атак с высоким риском и большим ущербом.

В конечном счете, у вас должны быть эффективные стратегии для обнаружения и смягчения атак и поддержания сетевой безопасности. Кроме того, различные типы приступов обычно требуют разного лечения. Выявление различных типов сетевых атак становится серьезной проблемой сетевой безопасности, особенно для атак, которые никогда раньше не наблюдались.

В последние годы исследователи использовали различные методы машинного обучения для классификации сетевых атак без предварительного знания их конкретных характеристик. Однако традиционные методы машинного обучения не могут предоставить дескрипторы характерных признаков, описывающие проблему обнаружения атак, из-за ограничений сложности модели .

В последнее время машинное обучение совершило большой прорыв, симулируя человеческий мозг с помощью структуры нейронных сетей, которые называются методами глубокого обучения из-за их общей многослойной архитектуры для решения сложных задач. Среди этих успешных приложений Google AlphaGo является одной из наиболее заметных попыток играть в го, используя мощь типичной среды глубокого обучения, а именно сверточных нейронных сетей.

Поскольку глубокое обучение является сложным в своих исходных предметно-ориентированных инфраструктурах и приложениях, эта статья была написана, чтобы объяснить его тем, кто хочет изучать передовые методы глубокого обучения в области сетевой безопасности.

По сути, существует много предыдущей работы, посвященной обнаружению атак с использованием методов глубокого обучения. Среди них было проведено несколько поисков литературы, чтобы получить идеи по применению глубокого обучения для обнаружения атак, что является основой нашей работе. Например, Берман и др. предоставить набор ресурсов для чтения, описывающих предысторию и историю развития методов глубокого обучения и их соответствующих приложений для обнаружения атак. Он отличается от всестороннего взгляда на эту конкретную область, данного Apruzzese et al. И

фокусируется на объяснении методов обнаружения атак, связанных с обнаружением вторжений, сканированием вредоносных программ и обнаружением спама. В работе Wickramasinghe et al. В основном исследуют методы глубокого обучения для безопасности IoT, которые дают четкое представление о различных типах кибератак и связанных с ними методах обнаружения.

Впоследствии Алиса и соавт. Рассмотрел и проанализировал состояние исследований системы обнаружения вторжений на основе глубокого обучения в четырех крупных базах данных.

В то же время они предлагают систематический поиск в литературе соответствующих работы по ключевым словам «Глубокое обучение», «Вторжение» и «Атака», которые предоставляют исследователям широкий спектр ресурсов.

Ферраг понял, что набор данных очень важен для обнаружения вторжений, описал 35 известных наборов сетевых данных и разделил их на семь категорий.

У них есть семь моделей витрин для каждой категории, в которых они оценивают и сравнивают производительность на основе точности и частоты ложных срабатываний на основе реальных наборов данных о трафике, а именно CSE-CIC-IDS2018 и Bot-IoT. На самом деле, все вышеперечисленные обзоры имеют свою собственную направленность, например, приложения безопасности, тип атак, наборы данных или базы данных.

В отличие от предыдущих методов, будем основывать работу на моделях глубокого обучения, уделяя особое внимание методам обнаружения атак, основанных на различных типах архитектур глубокого обучения. Кроме того, будет предложено честное сравнение и собственный анализ производительности репрезентативных подходов на основе наборов справочных данных.

Данная магистрская работа предоставляет более полный справочник для читателей, заинтересованных в том, как различные архитектуры глубокого обучения влияют на ландшафт обнаружения атак . Чтобы дать представление о том, как эффективно обнаруживать атаки глубокого обучения, необходимо ввести

некоторые базовые знания. Поэтому сначала будет представлено краткое введение в концепции обнаружения атак. Далее будет представлено краткая презентацию успешных приложений кибербезопасности.

Во второй главе работы рассмотрены виды сетевых атак и способы борьбы с ними.

Сетевая безопасность является серьезной проблемой для частных лиц, коммерческих и некоммерческих организаций, а также государственных организаций. С цифровым взрывом, который мы наблюдаем в настоящее время, обеспечение сетевой безопасности является настоятельной необходимостью для обеспечения общественного признания тысяч услуг, которые в основном зависят от основы цифровой жизни, поэтому Интернет оказывается насущной необходимостью, а не роскошью.

Несмотря на то, что было внедрено множество методов защиты, некоторые уязвимости по-прежнему используются хакерами, в результате чего администраторы сетевой безопасности находятся в постоянной гонке с сетевыми злоумышленниками.

Методы, основанные на использовании интеллектуальных методов, а именно машинного обучения (ML) и глубокого обучения (DL), были опробованы в различных областях, включая системы здравоохранения, финансовый анализ, образование, высшее образование, энергетику и т. Д.

Действительно, это мотивировало менеджерам по сетевой безопасности для дальнейшего изучения возможностей этих методов для обеспечения требуемого уровня сетевой безопасности.

В результате в последние годы было предложено несколько интеллектуальных методов обеспечения безопасности. Хотя эти методы показали отличные результаты, проблема не решена полностью. Это позволяет нам критически оценивать предлагаемые в настоящее время решения для определения потенциальных направлений исследований, которые могут привести к созданию более безопасных сетевых сред.

Значительный рост использования Интернета и быстрое развитие сетевых технологий связаны с повышенным риском сетевых атак. Сетевые атаки

— это любая форма несанкционированного доступа к сети, включая любые попытки повредить и нарушить работу сети, часто с серьезными последствиями. Обнаружение сетевых атак является активной областью исследований в сообществе кибербезопасности. В литературе доступны различные описания систем обнаружения сетевых атак с использованием различных интеллектуальных методов, включая модели машинного обучения (ML) и глубокого обучения (DL).

Однако, несмотря на то что такие методы оказались полезными в определенных областях, ни один из них не оказался полезным для смягчения всех типов сетевых атак. На самом деле, некоторым интеллектуальным решениям не хватает основных функций, которые делают их надежными системами, способными противостоять различным типам сетевых атак. Это было основной мотивацией этого исследования, в котором оцениваются современные направления исследований, основанные на интеллекте, чтобы заполнить пробел, который все еще существует в этой области. Основными компонентами любой интеллектуальной системы являются обучающие наборы данных, алгоритмы и показатели оценки основные тесты, используемые для оценки интеллектуальных систем, включенных в это исследование.

Недавно был разработан новый протокол под названием DNS over HTTP (DoH) для повышения конфиденциальности пользователей. Этот протокол можно использовать вместо традиционного DNS для преобразования доменных имен с преимуществом шифрования. Однако инструменты безопасности полагаются на информацию, читаемую DNS, для обнаружения таких атак, как вредоносное ПО и ботнеты. Поэтому Сингх и Рой намеревались использовать алгоритмы машинного обучения для обнаружения вредоносного трафика DoH. Используются пять алгоритмов ML: GB, NB, RF, KNN и LR. Команда провела эксперимент с использованием недавно разработанного и опубликованного набора справочных данных Министерства здравоохранения CIRA-CIC-DoHBrw-2020. Он содержал один безвредный файл с 19 807 экземплярами и один вредоносный файл с 249 836 экземплярами. Для извлечения важных функций из файлов PCAP использовался свободно доступный инструмент DoHMeter, разработанный на Python. Для создания модели данные были разделены на 70-

30% в соотношении извлечения/тестирования. Экспериментальные результаты показали, что RF и GB достигли максимальной точности 100% .

В третьей главе проводится сравнение алгоритмов обнаружения атак по скорости, эффективности и стабильности. Также обсуждается несколько способов повышения эффективности обнаружения атак с использованием фреймворков глубокого обучения.

Оценка безопасности динамических сетей затруднена, поскольку сеть и элементы безопасности меняются со временем. Кроме того, трудно оценить, как будут меняться существующие показатели безопасности по мере изменения сети с течением времени, поскольку они использовались только для оценки статических сетей.

Мы ввели время в десять существующих показателей кибербезопасности и изучили влияние этих изменений на существующие показатели безопасности. На основе результатов администратор безопасности может определить метрику безопасности, которая эффективно отражает состояние безопасности сети при изменении конфигурации сети.

Анализ потока сетевого трафика выполняется с помощью обнаружения неправильного использования, обнаружения аномалий и анализа протокола с отслеживанием состояния.

Обнаружение злоупотреблений использует предопределенные сигнатуры и фильтры для обнаружения атак. Он полагается на человеческий ввод для постоянного обновления базы данных сигнатур. Этот метод точен для обнаружения известных атак, но совершенно неэффективен против неизвестных атак. Обнаружение аномалий использует эвристику для поиска неизвестных вредоносных действий. В большинстве случаев нахождение аномалия дает высокий уровень ложноположительных результатов .

Наиболее распространенными проблемами в существующих решениях, основанных на моделях машинного обучения, являются:

- Во-первых, модели дают высокий уровень ложных срабатываний при более широком диапазоне атак.

- Во-вторых, модели не создаются общении, поскольку существующие исследования в основном использовали только один набор данных для отчета о производительности модели машинного обучения.
- В-третьих, модели, изученные до сих пор, совершенно не учитывают сегодняшний огромный сетевой трафик; и, наконец, необходимы решения, чтобы выдержать сегодняшние быстрорастущие размеры, скорость и динамику высокоскоростных сетей.

Эти проблемы являются основным стимулом для этой работы с исследовательским акцентом на оценку эффективности различных классических классификаторов машинного обучения и глубоких нейронных сетей (DNN), применяемых к NIDS и HIDS. Коммерческие NIDS в основном используют статистику или пороговые значения, рассчитанные для наборов функций, таких как длина пакета, время между поступлениями, размер потока и другие параметры сетевого трафика, для их моделирования. Эффективно в Лечение определенного временного окна. Они страдают от высокой частоты ложноположительных и ложноотрицательных срабатываний.

Высокий уровень ложноотрицательных предупреждений указывает на то, что NIDS может не обнаруживать атаки чаще, а высокий уровень ложноположительных предупреждений означает, что NIDS может обязательно предупреждать об отсутствии атак.

Следовательно, эти коммерческие решения неэффективны для современных атак. Самообучающаяся система является одним из эффективных методов борьбы с современными атаками.

При этом используются контролируемые, полуконтролируемые и неконтролируемые механизмы машинного обучения для изучения моделей различных обычных и злонамеренных действий с большим набором обычных сетевых событий и событий атаки, а также событий на уровне хоста.

Хотя в литературе можно найти различные решения на основе машинного обучения, их применимость к коммерческим системам находится на ранних стадиях. Существующие решения на основе машинного обучения выдают высокий уровень ложных срабатываний с высокими вычислительными затратами. Это связано с тем, что классификаторы машинного обучения изучают характеристики простых функций TCP/IP локально.

В заключении сформулированы основные результаты магистерской работ, отмечены ее научная значимость и практическая ценность, определены перспективы дальнейшей работы.

В магистрской работе была проведен обзор подходов для обнаружения атак методом глубокого обучения. Основываясь на категоризации методов глубокого обучения, уделяется особое внимание методам обнаружения атак, основанным на различных архитектурах, таких как состязательные генеративные сети, рекурсивные нейронные сети и сверточные нейронные сети. Кроме того, были рассмотрены классические методы классификации, такие как случаный лес, дерево решений, байесовский подход, для решения задачи распознавания сетевых атак. Были представлены несколько сравнительных наборов данных с описаниями и сравнивается производительность представленных подходов, чтобы показать текущее состояние методов обнаружения атак с использованием сред глубокого обучения. В результате было показано, что методы случаный лес и дерево решений не уступают в качестве прогнозов нейронным сетям и требуют намного меньше временных затрат.