Аннотация

Корпоративные коммуникационные сети часто подвергаются атакам с использованием сложных и ранее неизвестных вредоносных программ или инсайдерских угроз, что делает передовые защитные механизмы, такие как системы обнаружения вторжений на основе аномалий, необходимыми для обнаружения, оповещения и реагирования на инциденты безопасности. Такие стратегии (как на основе сигнатур, так и для обнаружения аномалий) основаны на признаках, извлеченных из сетевого трафика, для чего необходимы безопасные и расширяемые стратегии сбора, использующие современные многоядерные архитектуры. Существующие решения написаны на низкоуровневых языках системного программирования, требующих ручного управления памятью, и страдают от многих уязвимостей, которые позволяют злоумышленнику удалённо отключить или скомпрометировать сетевой монитор. Другие не были разработаны с учётом исслещдовательских целей и страдают с точки зрения гибкости и доступности данных. Чтобы решить эти проблемы и облегчить будущие эксперименты с методами обнаружения на основе аномалий, будет представлена ​​исследовательская структура для сбора признаков трафика, реализованная на языке, безопасном для памяти. Она обеспечивает доступ к сетевому трафику в виде типобезопасных структурированных данных для определенных протоколов или пользовательских абстракций, путем генерации записей аудита в платформо-независимом формате. Чтобы уменьшить объем памяти, вывод сжат по умолчанию. Подход полностью реализован на языке программирования Go, имеет параллельный дизайн, легко расширяем и может использоваться для захвата данных в реальном времени с сетевого интерфейса или из дамп-файлов PCAP и PCAPNG. Кроме того, платформа предлагает функциональные возможности для создания маркированных наборов данных, ориентированных на применение в машинном обучении с учителем. Для демонстрации разработанного инструментария проводится серия экспериментов по классификации вредоносного поведения в наборе данных CIC-IDS-2017 с использованием Tensorflow и Deep Neural Network.

Введение

Современная жизнь, начиная с изобретения протокола TCP (IP) в 1974 году, сильно зависит от сложных компьютерных сетей и приложений. Эта зависимость варьируется от повседневного использования компьютеров в личных целях до распределенных деловых сетей, медицинских и промышленных приложений и беспилотных автомобилей. В связи с увеличением объема ценной информации, хранящейся в таких системах, они становятся все более и более популярными целями для злоумышленников. Успешные атаки на компьютерные системы часто оказывают прямой экономический ущерб. Злоумышленники варьируются от новичков-энтузиастов до организованных преступников и государственных организаций. В связи с растущей доступностью и изощренностью интернет-технологий и увеличением числа сетевых атак, сетевая безопасность стала важной областью исследований. Сетевая безопасность требует глубоких знаний обо всех протоколах, используемых в сети, поэтому программное обеспечение для захвата данных должно содержать множество сложных синтаксических анализаторов, чтобы правильно интерпретировать собранные данные. Идти в ногу с растущим объемом обрабатываемых данных может быть непросто и требовать эффективных реализаций и алгоритмов. Обработка данных в реальном времени необходима для быстрого обнаружения текущих или прошлых атак и предотвращения или уменьшения экономического ущерба. Новые уязвимости появляются каждый день и используются в составе так называемых «атак нулевого дня». Стратегии обнаружения на основе сигнатур страдают от невозможности обнаружения ранее неизвестных угроз. Тем не менее, методы машинного обучения редко встречаются в общедоступных инструментах, несмотря на то, что последние 20 лет они были популярной областью исследований. Большое количество ложных срабатываний и отсутствие надлежащих наборов данных для обучения и оценки являются хорошо известными проблемами обнаружения вторжений на основе аномалий. Шаблоны трафика сильно меняются, как и программный стэк в сетевых средах. Чтобы учитывать это, для оценки стратегий обнаружения вторжений на основе аномалий должны использоваться современные наборы данных. Обнаружение вторжений в сеть очень помогает в выявлении сетевых нарушений, отслеживании их и последующем принятии мер для выявления и устранения любого нанесенного ущерба. Функции распознавания атак и мониторинга событий систем обнаружения вторжений также оказывают сдерживающее воздействие на злоумышленников, увеличивая их риск быть обнаруженными и привлеченными к ответственности. Наличие системы обнаружения вторжений может убедить злоумышленника искать другую цель, в которую легче проникнуть. Получение блокировки от сетевого монитора или аналитика из-за привлекает нежелательное внимание к злоумышленнику и замедляет его работу. Однако для того, чтобы извлечь выгоду из обнаружения вторжений, необходим надежный и обширный источник данных для точных прогнозов.

1.1 Благодарности  
Я благодарю моего руководителя Дипл.-Информ. Стефана Метцгера и профессора д-ра Хельмута Райзера из Leibniz Rechenzentrum Munich (LRZ) за то, что они предложили мне написать диссертацию в сотрудничестве с институтом, направляя меня в моих исследованиях, критических рецензиях на мои идеи и многих конструктивных дискуссиях. Верн Паксон и Аашиш Шарма из Национальной лаборатории им. Лоуренса Беркли за интересные беседы на Bro Workshop Europe 2018 в Карлсруэ, которые вдохновили меня на многие мои дизайнерские решения для этого исследовательского проекта. «Лабораторию Касперского» за ежегодные студенческие конкурсы, которые побудили меня провести исследование по теме обнаружения сетевых вторжений и опубликовать мои инструменты, чтобы помочь другим исследователям в этой области. И, наконец, моей семье, которая полностью поддерживала и поощряла меня на пути к написанию диссертации и способствовала моему выживанию, часто принося еду в мою компьютерную пещеру.  
1.2 Краткое содержание тезиса

В настоящем введении объясняется мотивация мониторинга сетевой безопасности и стратегий обнаружения на основе аномалий, а также описывается влияние повреждённых данных в памяти на безопасность современной сетевой инфраструктуры. Объяснена общая терминология, а также определение проблемы и описание задачи для определения рамок данного тезиса. После краткого обсуждения связанных работ анализируются требования к механизмам сбора сетевых признаков, после чего проводится оценка существующих решений. После этого будет подробно изложена концепция и реализация исследовательского проекта, а также идеи для будущих улучшений и несколько вариантов использования помимо обнаружения вторжения в сеть. В окончательной оценке показано несколько практических приемов разработанного инструмента, который включает серию экспериментов по классификации злонамеренного поведения с помощью нейронной сети глубокого обучения. После заключения дается взгляд на возможные пути развития проекта.