Обнаружение аномальной активности в сети на основе анализа статистических отклонений в работе системы

Сластухин А. Ю. (ЯрГУ им. П. Г. Демидова, Ярославль)

Научный руководитель: д.ф.-м.н., Савинов Д.А.

Задача

Разработать IDS систему для выявления статистических аномалий в сетевой активности.

Неизменность природы атак

В последние десятилетия, в связи с активным развитием технологий, в частности, персональных компьютеров и сети интернет, у человека появился доступ к огромному океану различной информации, многие проблемы стали решаться в несколько кликов мыши. Это привело к тому, что и злоумышленники получили в своё распоряжение отличные инструменты для совершения зловредных действий.

Природа атак не изменилась с развитием техники, старые подходы, способы обмана и хищения используются при новых инструментах, но старые методы защиты частично перестали работать из-за возросших требований. Мы так же пытаемся обезопасить свои личные устройства или корпоративные сети, но проблема в том, что на любое средство защиты найдётся способ его обойти.

Статический анализ

Среди способов защиты традиционно выделяется статический анализ – анализируется текущее состояние системы на предмет нахождения уже известных сигнатур угроз. В этом отношении такие системы пересекаются с антивирусами, имеют минус – не реагируют на новые уязвимости. Поэтому требуется своевременное получать информацию о новых уязвимостях.

Сигнатурой называется шаблон уже известной угрозы. Помимо сигнатур выделяют состояния: в начальный момент времени считаем, что система в безопасном состоянии, но каждое

действие в системе: несанкционированное действие пользователя, сетевое обращение, установка и активация ПО, особенно драйверов, может привести в скомпрометированное состояние.

Статистический анализ или поиск аномалий

Аномалией является любое действие, которое хоть по каким-то признакам отличается от нормы, что даёт возможность предполагать, что система могла перейти в скомпрометированное состояние.

Системы поиска аномалий способны помочь выявить новую угрозу, в то же время они не дают никакой гарантии, что в принципе сработают, в то время как сигнатурные гарантируют, что конкретные сценарии практически недостижимы - поэтому оба подхода используются вместе.

Портрет сетевой активности

Сущность, учитывающая историю предыдущей активности системы с целью ответить на вопрос: "Является ли новая активность аномальной?".

В основе лежит попытка рассматривать систему с точки зрения индивидуальных особеностей владельца системы, анализ которых позволит ответить на поставленный выше вопрос.

Предполагается, что резкие изменения в жизне человека/активности системы являются поводом обратить на это внимание. При этом незначительные изменения в течение времени допускаются - не считаются аномалией, требующей внимания.

Однако, подобная особенность является потенциальной уязвимостью - если злоумышленник знает, как устроена система анализа аномалий, он может попытаться постепенно менять активность захваченной системы, чтобы изменить портрет её деятельности с целью сокрытия аномальной активности - на это требуется время, индивидуальный подход к конкретной системе, это потенциально отбросит зловреды, которые не контролируются создателем, а даже если создатель сможет изменить профиль активности, то затраченное время и усилия, вероятно, не будут стоить результата.

Новая IDS система

Intrusion Detection System - система обнаружения вторжений. В рамках курсовой работы мне удалось построить прототип такой системы, который позволяет передать на вход Pandas DataFrame с исходными данными, сконфигурировать предустановленные базовые анализаторы, которые умеют анализировать идентификаторы/числа/строки/массивы.

Внутренние альгоритмы обучат систему на переданных данных, подберут оптимальный вес для каждого анализатора. Может оказаться, что пользователь пытается использовать неподходящий анализатор - вес анализатора укажет на то, что на входных данных он работает плохо.

С использованием полученной системы проанализировал DataSet TelecomX[1], который был представлен в 2024 году на хакатоне в рабочем кейсе от компании Arenadata.

Результаты оказались неплохими: высокая скорость работы, низкие затраты памяти, результаты интуитивны и обосновываются статистикой. Однако алгоритм пока не идеален, есть почва для усовершенствования автоматического подбора весов, внедрения новых передовых подходов в построении базовых анализаторов.

Заключение

Считаю развитие систем выявления аномалий в активности пользовательских систем перспективным шагом к безопасному интернет пространству, обеспеченному защищёнными одиночными станциями, и к изучению человека, его поведения и мышления.

В следующем году усовершенствую систему и применю её в анализе активности пользователя на его машине, активность планирую получать в помощью перехвата трафика в WireShark.

Полная версия доклада доступна в открытом GitHub репозитории[2].

Литература

[1] https://habr.com/ru/companies/arenadata/articles/856366/
[2] https://github.com/SkibaSAY/IPS_University/tree/master/IPSLib/Examples/