Обнаружения сетевых атак с использованием методов статистического анализа

И. А. Высоцкая, e-mail: i.a.trishina@gmail.com

ВУНЦ ВВС «Военно-воздушная академия имени профессора Н.Е. Жуковского и Ю.А. Гагарина» (г. Воронеж)

Аннотация. *В данной работе рассматриваются статистические методы обнаружения сетевых атак. При моделировании систем защиты от несанкционированных вторжений в компьютерные сети возникают задачи, связанные с обнаружением не известных ранее видов атак. Преимуществами статистических методов является их адаптация к изменению поведения пользователя, а также способность к обнаружению модификаций различных атак.*

Ключевые слова: *компьютерные сети, обнаружение компьютерных атак, статистический анализ.*

Введение

Развитее компьютерных сетей и информационных технологий вызывает непрерывное развитие программного обеспечения, связанного с безопасностью сетевых ресурсов, которое требуют новых подходов.

Аномальные сетевые поведения не всегда можно обнаружить с использованием сигнатурных методов анализа. В частности, такие ситуации могут возникать в случае атак неизвестных типов. Применение статистических методов обнаружения аномального поведения позволяет отслеживать состояние системы и выявлять новые, неизвестные ранее виды атак.

Для описания процесса сетевой атаки существует множество моделей, которые в большинстве основаны представления атаки как последовательного состояния автомата [1]. В целях применения описанного ниже алгоритма можно использовать готовые обучающие средства моделирующие сетевые атаки.

1. Статистические методы обнаружения сетевых атак

Вначале необходимо определить объект анализа, и критерии, по которым в дальнейшем определяется потенциальная угроза сетевой безопасности, построить базовый профиль системы, без аномалий, с использованием методов математической статистически [2]. Профилем является набор различных данных о типичном поведении объекта

исследования, т. е. некоторая выборка $Y = (Y_1, ..., Y_n)$, где Y_i – событие в момент времени t_i, $i = 1, ..., n$, n – установленный период анализа. К таким данным можно отнести (при захвате трафика TCP/IP), поля заголовков IP, TCP, UDP и содержимое полей данных.

Для формирования профиля предлагается для использования программа Wireshark (или любого другого программного продукта, осуществляющего запись и анализ трафика) [3]. Комплекс Wireshark содержит инструмент IO Graphs, для снятия статистики сообщений, распределённых с некоторой частотой.

Далее следует производить сопоставление текущего состояние системы с некими определёнными заранее признаками, характеризующими стационарное функционирование системы [3-7]. При проведении анализа будем опираться на две гипотезы: H_0 – отклонений в трафике нет (трафик подчиняется нормальному закону распределения), H_1 – отклонения в трафике есть. Более подробная формулировка гипотез требует знания распределения вероятностей до предполагаемой атаки.

Предполагаем, что причинами аномалий является существенное изменения трафика. Наблюдения группируются в выборку $X = (X_1, ..., X_n)$, где X_i – событие в момент времени t_i, $i = 1, ..., n$. Для оценки существенности расхождения между частотами выборки $Y = (Y_1, ..., Y_n)$ и выборки $X = (X_1, ..., X_n)$ предлагается использовать критерий χ (формула 1).

$$Q = \sum_{i=1}^{n} \frac{(X_i - Y_i)^2}{Y_i}, \qquad (1)$$

где n – число категорий, X_i – посчитанное число наблюдений в категории i, Y_i – ожидаемое число наблюдений в категории i. По таблицам [2], можно определить предельное верхнее значение при заданном уровне значимости и числе степеней свободы. Если фактическое значение (1) меньше табличного, то расхождение между частотами считают случайными, а гипотезу H_0 принимают.

В результате применения критерия могут возникнуть ошибки первого или второго рода. Ошибку первого рода называют ложным срабатыванием, а второго рода соответственно ложноотрицательным срабатыванием. Таким образом, мы может получить ситуацию, когда авторизированные пользователю классифицируются как нарушители, и наоборот действия нарушители классифицируются как действия

легальных пользователей. Поэтому при моделировании систем защиты всегда стоит задача баланса между сохранения безопасности компьютерной сети и данных и обеспечением стабильного доступа и работы легальных пользователей.

При обнаружении существенных отклонений выдается сообщение о начале сетевой атаки. На качество результатов выявления атак существенно влияет выбор показателей трафика, которые является устойчивым к неисправностям вызванными законными изменениями. Иначе, есть большая вероятность получить множество ложных тревог. Например, администратор сети может применять отладочные утилиты, для диагностики сетевого окружения. Действия подобного рода не должны относиться к злоумышленным, однако системы обнаружения атак может распознать эту деятельность как аномальную сетевую активность.

Критерий χ – квадрат не является единственным, при анализе аномального поведения. Например, можно использовать одновыборочный критерий нормальности Колмагорова-Смирнова. Он основан на максимуме разности между кумулятвным распредлением выборки $X = (X_1, ..., X_n)$ и выборки $Y = (Y_1, ..., Y_n)$, где n – число категорий. Критерий оценки Вилкоксона-Манна-Уити можно использовать в качестве сравнения двух наблюдений на однородность.

2. Оценка эффективности

Для оценки эффективности можно использовать кривую ошибок. Это графическая иллюстрация зависимости между относительным числом правильных срабатываний и числом ложных срабатываний, при изменениях параметров используемого метода. Для построения кривой ошибок используются следующие формулы

$$TPR = TP \Big/ (TP + FN), \tag{2}$$

$$FPR = FP \Big/ (FP + TN), \tag{3}$$

где TP – количество выявленных аномалий, FN – количество пропущенных аномалий, FP – количество ложных срабатываний системы, TN – количество правильных срабатываний системы.

Заключение

При моделировании систем защиты от несанкционированных вторжений в компьютерные сети возникают задачи, связанные с обнаружением не известных ранее видов атак. Использование методов

математической статистики является оптимальным, поскольку для проведения анализа достаточно только статистических данных о потоке пакетов, и нет необходимости в изучении самой атаки и ее методов. Главным преимуществом применения статистического анализа является выявление атак, впервые реализуемых злоумышленниками.

Литература

1. Браницкий А. А. Анализ и классификация методов обнаружения сетевых атак / А. А. Браницкий, И. В. Котенко // Тр. СПИИРАН. – 2016. – № 45. – С. 207-244.
2. Гмурман, В. Е. Теория вероятностей и математическая статистика: учебник для прикладного бакалавриата / В. Е. Гмурман – 12-е изд. – М: Издательство Юрайт, 2014. – 479 с.
3. Шелухин, О. И. Обнаружение вторжений в компьютерные сети (сетевые аномалии). Учебное пособие для вузов / О. И. Шелухин, Д. Ж. Сакалема, А. С. Филинова. – М: Горячая линия-Телеком, 2013. – 220 с.
4. Высоцкая, И. А. Применение средств компьютерного моделирования для проектирования с использованием обобщённого мультипликативного критерия оптимальности / И. А. Высоцкая, А. С. Михалев // Информатика: проблемы, методы, технологии. Материалы XX Международной научно-методической конф. (Воронеж, 13–14 февраля 2020 г.) – Воронеж, 2020. – С. 378-382.
5. Таненбаум, Э., Уэзеролл, Д. Компьютерные сети. 5-е изд. – СПб.: Питер, 2012. – 960 с.
6. Семенов, Н. А. Применение статистических методов обнаружения DoS атак в локальной сети / Н. А. Семенов, А. Ю. Телков // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии. – 2012. – № 1. – С. 82-87.