1. УДК 004.056.53
2. Методы обнаружения аномалий в телекоммуникационных сетях
3. Автор: Киселев Д.А.; место учебы - Университет ИТМО, г. Санкт-Петербург
4. Научный руководитель: Грудинин В.А., Университет ИТМО, г. Санкт Петербург
5. С ростом количества телекоммуникационных сетей растет и количество атак, совершаемых в отношении них. Существуют способы обнаружения вредоносного поведения, как аномалий.

Цель работы: разбор существующих методов анализа и поиска аномалий - атак, выявить наиболее эффективную из них.

Базовые положения исследования: атака на информационную систему — это совокупность преднамеренных действий злоумышленника, направленных на нарушение одного из трех свойств информации — доступности, целостности или конфиденциальности. Атаки отличаются по методам и целям воздействия. Например, атака «отказа в обслуживании» (DoS) направлена на ограничение ресурсов сервера, которые необходимы для правильной его работы во время обработки данных; черви и вирусы используют узлы сети, компрометируя, получают привилегированный доступ к хосту, используя преимущества известных уязвимостей.

Основанный на поиске аномалий подход к предотвращению атак решает проблему, возникающую из-за большого количества трафика данных в сетях, благодаря тому что он позволяет задавать шаблоны “правильного” поведения в сети. Обнаружение аномалий широко используется для определения мошеннического поведения в банковских сетях, в телефонных сетях, а также в сетях военного назначения.

Промежуточные результаты:

Виды аномалий:

1. Точечная аномалия. Отдельный экземпляр данных, который был признан аномальным по отношению к другим.
2. Контекстная аномалия. Экземпляр данных, который был признан аномальным в определенном контексте. Контекст - структура в наборе данных и имеет 2 атрибута: контекстный и поведенческий.
3. Коллективная аномалия. Набор данных, признанный аномальным по отношению к другим данным, но каждый экземпляр данных, выполняясь в наборе, аномальным не будет.

Методы обнаружения аномалий:

1. Статистические тесты. Используют для точечных аномалий. Для отдельных выбросов находятся экстремальные значения при помощи стандартизованной оценки (Z-оценка, Z-value, standart score) или ккоэффициента эксцесса (Kurtosis measure).
2. Модельные тесты. Модель описывает данные, точки, отклонившиеся от модели и есть аномалии. Расчет происходит при помощи неполного сингулярного разложения. В результате ищется максимально похожая матрица.
3. Итерационные методы. В ходе итераций выбирается группа объектов, с наибольшим отклонением.
4. Метрические методы. Существуют некоторые метрики, например расстояние до k-го соседа, которые помогают определить отклонение.
5. Методы подмены задачи. В данных методах используется индукция, и задача кластеризуется на более мелкие, за счет чего и находятся аномалии.
6. Методы машинного обучения:
    - Метод опорных векторов для одного класса (OneClassSVM)
    - Изолирующий лес (IsolationForest)

    - Эллипсоидальная аппроксимация данных (EllipticEnvelope)
7. Ансамбли алгоритмов. Обнаружение аномалии заключается в средней оценке нескольких алгоритмов обнаружения.

Данные методы можно классифицировать также и по способу принятия решения.

Сигнатурные методы — это методы, основанные на шаблонах сетевых атак. Чем больше шаблонов, которые основаны на заголовках и содержимом пакетов, тем точнее данный метод. Преимущества:

- Высокая производительность;
- Эффективное определение атак;
- Отсутствие ложных срабатываний;
- Надежность;
- Точные параметры сигнатуры(шаблона).

Недостатки:

- Обновление базы сигнатур;
- Невозможно определить атаку, если она не описана в системе.

К сигнатурным методам относятся методы контекстного поиска и анализа состояния сигнатур.

Поведенческие методы — это методы, основанные на моделях информационных атак. Принцип работы заключается в поиске расхождений между эталонным и текущим режимами работы системы. Преимущества:

- Определение атаки без сигнатуры;
- Создание информации, на основе которой можно сформировать сигнатуры;
- Высокая чувствительность.

Недостатки:

- Ложные сигналы;
- Временные затраты на обучение.

Основной результат: в результате выполненной работы, были изучены основные типы методов обнаружения сетевых аномалий, было произведено сравнение на основе схемы принятия решения. Как приоритетный был выбран метод машинного обучения, так как он исключает человеческий фактор и способен быстрее адаптироваться на растущем количестве данных.

Автор: Киселев Д.А.

Научный руководитель: Грудинин В.А.

Декан факультета: Хоружников С.Э.