УДК 004.62, 004.9

ИССЛЕДОВАНИЕ МЕТОДОВ СТАТИСТИЧЕСКОГО АНАЛИЗА ДЛЯ ПОИСКА АНОМАЛИЙ В СЕТЕВОМ ТРАФИКЕ

Платонова А. В., Белоусов А. А.

Самарский национальный исследовательский университет
имени академика С. П. Королёва, г. Самара

Обнаружение аномалий относится к задаче поиска закономерностей в данных, которые не соответствуют ожидаемым характеристикам. В различных областях применения эти несоответствия часто называют аномалиями, выбросами, исключениями, аберрациями, особенностями или шумом.

Сетевой трафик представляет собой огромный объем данных, передаваемых через компьютерную сеть за определенный период времени. Аномалиями являются необычные изменения в трафике сети. Такими изменениями являются резкие изменения объема трафика, замены IP - адреса источника, адреса назначения, номеров портов и прочее [1].

Причинами аномалий могут служить:

– деятельность злоумышленников;

– некомпетентность и ошибки пользователей;

– неисправность аппаратуры;

– повреждение каналов связи;

– дефекты программного обеспечения.

Система должна быть в состоянии обнаружить вредоносную активность и охарактеризовать ее таким образом, чтобы можно было предпринять соответствующие действия по ее предотвращению. Такая необходимость обусловлена увеличивающимся темпом роста информационного и экономического ущерба, причиненного умышленно или неумышленно атаками, ошибками, дефектами и прочим [2]. Анализ сетевого трафика должен проходить точно и быстро в режиме реального времени, чтобы обеспечить мгновенную реакцию на какие-либо несоответствия в сети. Правильная организация работы с трафиком помогает в поддержке качества обслуживания данной сети.

Исследовались метод k-means и технология гибридной корреляции событий.

Метод k-means является одним из самых простых алгоритмов неконтролируемого обучения, которые решают проблему кластеризации. Метод классифицирует и группирует объекты в K групп. Группировка осуществляется путем минимизации суммы квадратов расстояния между данными и соответствующим центром кластера.

Технология гибридной корреляции событий состоит в автоматическом извлечении знаний из совокупности данных путем обнаружения высоковероятных закономерностей и последующим обнаружением нарушений этих закономерностей. Высоковероятные закономерности определяют нормальное течение событий, а нарушение закономерностей – аномальное: атаку, жульническое или вредоносное течение событий. Важной особенностью этого подхода является возможность обнаружения новых, не встречавшихся ранее атак. Для обнаружения новых атак не требуется наличия обучающего материала, т. е. наличия уже идентифицированного набора атак определенного рода, как это требуется для существующих методов интеллектуального анализа данных [3].

В качестве входных данных при проведении экспериментов использовались искусственно сгенерированные с помощью утилиты для генерации трафика Cat KARAT сетевые трафики и набор данных KDD Cup 1999 Data.

Следует отметить, что результаты работы двух рассматриваемых методов для искусственно сгенерированных трафиков отличаются сильнее, чем для реальных.

Каждый из пяти трафиков содержал информацию о 10000 пакетов. В таблице 1 приведены результаты экспериментов.

Таблица 1. Результаты экспериментов

Номер эксперимента	Всего пакетов	Всего аномальных пакетов	Метод	Количество найденных аномальных пакетов	Количество найденных аномальных пакетов, %
1	10000	496	k-means	342	69
			ГКС	496	100
2		1834	k-means	1830	99,8
			ГКС	1834	100
3		5026	k-means	4987	99,2
			ГКС	5034	100,2
4		6321	k-means	6212	98,3
			ГКС	6261	99,1
5		9940	k-means	9945	100,05
			ГКС	9895	99,6

В заключение стоит сказать, что оба метода прекрасно справились с поставленной задачей, несмотря на небольшое количество псевдоаномалий, возникших в результате работы метода k-means.

Метод k-means показал плохой результат в эксперименте с небольшим количеством аномальных пакетов, не найдя 31% аномалий. Технология гибридной корреляции событий показала хорошие результаты во всех экспериментах, найдя около 100% аномалий.

По результатам экспериментов выявлено, что за счет использования технологии Spark Streaming оба метода показали высокую скорость работы.

Библиографический список

1. Оладько В. С. Причины и источники сетевых аномалий [Текст]/ В. С. Оладько, С. Ю. Микова, М. А. Нестеренко, Е. А. Садовник. — М.: Молодой ученый, 2015. — 160 с.

2. Отчёт об утечках конфиденциальной информации в 2015 году [Электронный ресурс]. — URL: http://www.zecurion.ru/press/analytics/ (дата обращения: 5.05.2016).

3. Витяев Е.Е. Обнаружение закономерностей и распознавание аномальных событий в потоке данных сетевого трафика [Текст] / Е. Е. Витяев, Б. Я. Ковалерчук, А. М. Федотов, В. Б. Барахнин, Д. С. Дурдин, С. Д. Белов, А. В. Демин. // Вестник Новосибирского государственного университета. — Новосибирск, 2008. — Т.6, вып.2. — С.57-68.