УДК 004.056.5

О.С. Терновой, А.С. Шатохин

Раннее обнаружение DDOS-атак статистическими методами при учете сезонности

Рассмотрены вопросы снижения ошибки и раннего обнаружения DDOS-атак статистическими методами; учет сезонности при использовании статистических методов; эффективное выделение периодов сезонности.

Ключевые слова: DDOS-атака, бот-сеть, среднеквадратичное отклонение, статистический анализ.

DDOS-атаки – распределенные атаки, направленные на отказ в обслуживании, продолжают оставаться одной из важнейших угроз в сети. Атаки такого типа могут быстро истощить сетевые ресурсы или мощности сервера, что приведет к невозможности получить доступ к ресурсу и вызовет серию негативных последствий: упущенная прибыль, невозможность воспользоваться услугами и произвести различные транзакции и т.д. [1].

В DDOS-атаке в роли атакующего выступает так называемая бот-сеть, или зомби-сеть. Зомби-сеть может насчитывать от нескольких десятков до тысяч хостов. Обычно это нейтральные компьютеры, которые в силу каких-то причин (отсутствие файрвола, устаревшие базы антивируса и т.д.), были заражены, вредоносными программами. Программы, работая в фоновом режиме, непрерывно посылают запросы на атакуемый сервер, выводя его таким образом из строя [2].

В настоящий момент не существует какого-то универсального средства для противодействия DDOS-атакам. Даже такие крупные компании, как Microsoft, eBay, Amazon, Yahoo, страдают от DDOS-атак и не всегда могут с ними справиться [2].

Для противодействия распределенным атакам, направленным на отказ в обслуживании, требуется выполнение двух основных задач [3].

1. Диагностировать DDOS-атаку на самых ранних стадиях. Чем раньше будет обнаружена DDOS-атака, тем раньше сможет включиться в игру сетевой администратор и тем раньше можно будет начать проводить антиDDOS-мероприятия. Кроме того, при обнаружении DDOS-атаки можно будет, не дожидаясь реагирования администратора, автоматически запустить мероприятия по противодействию: задействовать резервные каналы связи, включить фильтры и т.д.

2. Вторая задача связана с разделением общего потока трафика на вредоносный и обычный. Поняв, какие из клиентских запросов являются результатом DDOS-атаки, можно будет создать соответствующие правила для межсетевого экрана или ACL правила для маршрутизатора или же, в случае масштабной атаки, передать эти данные на вышестоящие маршрутизаторы.

Первая из этих задач является достаточно новой. Несколько лет назад основной являлась именно задача по «сортировке» трафика. Однако злоумышленники постоянно совершенствуют способы проведения атак такого типа. И современные атаки отличаются сложностью и наличием этапа подготовки. Во время подготовительного этапа злоумышленник пытается выявить наиболее уязвимые для атаки места. Например, для web-сервера такими местами могут быть определенные скрипты, которые совершают большое количество запросов к базе данных или чрезмерно используют процессорное время. Для выявления этих мест злоумышленник может совершать серию мини-DDOS-атак на различные скрипты, отслеживая при этом время ответа сервера и время выполнения скрипта. Найдя уязвимое место, злоумышленник сможет парализовать работу сервера, используя бот-сеть меньшего размера. С другой стороны, если диагностировать атаку удастся уже на этом этапе, можно будет задействовать автоматические средства предотвращения атаки, а у системного администратора будет время подготовиться – оптимизировать скрипты, чрезмерно загружающие ресурсы компьютера, создать фильтры и т.д.

Для обнаружения DDOS-атак и создания специальных фильтров для отсечения вредоносного трафика применяются разнообразные методы и подходы.

Среди основных методов можно выделить методы, базирующиеся на статистическом анализе. Это количественный анализ, анализ среднеквадратичных отклонений, кластерный анализ и т.д. Все

эти виды анализа могут оценивать различные параметры сетевой активности и диагностировать начало атаки либо определять вредоносный трафик.

Основными параметрами, по которым проводится анализ, могут быть:

- Количество запросов за определенный период.
- Скорость поступления запросов.
- Количество запросов с определенного источника или из определенной сети.
- Количество запросов к определенному пункту назначения (для web-сервера это конкретный скрипт).
- Время между запросами.
- Другие различные параметры сетевой активности.

С помощью среднеквадратичного отклонения можно рассчитать допустимую границу для одного из параметров сетевой активности, например, для количества запросов за какой-то период времени. В случае если граница будет нарушена, это станет свидетельством начала атаки. Так как в разное время нагрузка на сетевой ресурс, так же может быть разной, то для раннего обнаружения атаки необходим постоянный мониторинг и пересчет границ для каждого временного шага. Постоянный мониторинг позволит определить атаку, если она начнется в период небольшой сетевой активности, или, если злоумышленник ищет потенциально уязвимые места на сервере, проводя мини-DDOS-атаки и изучая поведения сервера. В случае если верхняя граница задана строго и злоумышленник проводит мини-атаки в период наименьшей сетевой активности, он может не нарушать заданную границу, и его действия будут не обнаружены. Атака будет обнаружена тогда, когда злоумышленник найдет потенциально уязвимое место, и предпримет на него атаку. Постоянный мониторинг активности и перерасчет допустимых границ позволяет этого избежать. В период меньшей сетевой активности верхняя граница снизится. Однако и этот метод имеет ряд минусов.

Во-первых, злоумышленник может начать атаку постепенно. Показатели активности на каждом шаге будут плавно повышаться, но при этом не будут нарушать границ. Так как при расчете среднеквадратичного отклонения используются последние n интервалов, в том числе и те, которые уже содержат данные атаки, то злоумышленник, постепенно увеличивая интенсивность атаки, будет отодвигать границу.

Во-вторых, выбор размера периода n для расчета среднеквадратичного отклонения, не является однозначным.

Если n будет слишком велико, полученная граница будет слишком высоко, если используется малое значение n, то возможны частые срабатывания. Выбрать же оптимально значение n в этой ситуации будет невозможно, так как любое его значение может захватывать два разных периода. Например, даже малое значение n, рассчитываемое в начале рабочего дня, будет захватывать данные и ночного периода, связанного с низкой активностью, и дневного периода, который характеризуется большей нагрузкой на сетевые ресурсы.

Для того чтобы предотвратить ложное срабатывание, связанное с началом рабочего дня, потребуется использовать такое значение n, в котором будут данные за несколько дней. Либо контролировать сразу несколько временных периодов – при срабатывании на минутных интервалах, рассмотреть часовые или суточные периоды. Но это в свою очередь приведет к уменьшению точности, и атака будет определена с опозданием [4].

В качестве гипотезы можно предположить, что более высокую точность в этих случаях может дать учет различных периодов активности и сравнение сходных между собой периодов.

Пусть x_i количество запросов к серверу за один час. Сервер испытывает стабильную суточную нагрузку. Количество суточных периодов n. Тогда запросы к серверу можно записать в виде матрицы:

$$x_{1\,1}, x_{1\,2}, x_{1\,3} \ldots x_{1\,24},$$
$$x_{2\,1}, x_{2\,2}, x_{2\,3} \ldots x_{2\,24},$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$x_{n\,1}, x_{n\,2}, x_{n\,3} \ldots x_{n\,24}.$$

Каждая строка матрицы включает в себя суточные данные о количестве запросов. Первая строка отражает данные текущих суток, в этой связи она может быть заполнена не до конца. Расчет среднеквадратичного отклонения в этом случае может проводиться двумя способами [5]:

- Обычным способом с учетом определенного числа последних значений, например так:

$$x_{2\,1},\ x_{2\,2}, x_{2\,3} \ldots x_{2\,24}, x_{1\,1},\ x_{1\,2}, x_{1\,3}, x_{1\,4}.$$

Значения берутся из строк матрицы.

- С учетом сезонности. Расчет проводится по столбцам:

$$x_{n\,1},\ \ldots\ x_{2\,1}, x_{1\,1}.$$

Если мы находимся в i-м периоде, можно рассчитать границу для $(i + 1)$-го периода, используя значение $(i + 1)$-го столбца. Если сетевой ресурс испытывает нагрузку, связанную с недельными или суточными циклами, то необходимо исключить строки, которые соответствуют праздничным и выходным дням. Или даже использовать только каждую седьмую строку, т.е. сравнивать, например, только период с 11:00 до 12:00, для каждого понедельника.

Проверка гипотезы

Апробация данной гипотезы проведена на реальных данных, полученных из лог-файлов различных web-сайтов, которые содержат в себе нормальные данные и данные, соответствующие DDOS-атакам.

Список рассматриваемых сайтов:

http://ankt.ru

http://www.corrupcia.net/

http://litm.ru

Лог-файл представляет собой стандартный файл access_log web-сервера Apache.

Предварительно данные из лог-файлов были обработаны вручную и проанализированы. В результате были выделены сезонные периоды, а также точно обозначено время начала атак.

Диагностирование DDOS-атаки проводилось различными методами:

- Анализ сходных сезонных периодов.
- Анализ последних n периодов, при различных значениях n.
- Анализ последних n периодов, различной размерности (минуты, часы и т.д.), для разных значений n.

В связи с тем, что лог-файлы имеют свой специфичный формат, их анализ стандартными методами является затруднительным. Для проведения анализа был создан скрипт, извлекающий из лог-файла необходимые данные и экспортирующий их в базу данных. Скрипт был реализован с помощью языка PHP (Hypertext Preprocessor). Предпочтение данному языку программирования отдано в связи с наличием богатого инструментария по работе со строками и регулярными выражениями [6]. В качестве системы управления базами данных выбрана свободно распространяемая СУБД MySQL версии 5.5.23. Использование СУБД позволило ускорить процесс обработки и анализа данных и сделать его более гибким.

Для проведения собственно самого анализа также была разработана отдельная программа. Язык реализации программы PHP. Его использование позволило выдержать созданный программный комплекс в одном ключе и дало возможность реализовать web-интерфейс. Web-интерфейс может быть доступен для системного администратора с любого компьютера, что позволяет в удаленном режиме диагностировать атаку и выявлять во входящем трафике различные аномалии. В дальнейшем планируется доработать программу для работы в полностью автоматическом режиме. В этом случае данные из лог-файла будут экспортироваться в базу данных в режиме реального времени. Анализ должен происходить после каждого нового добавления данных. В случае диагностирования начала атаки программа будет рассылать необходимые уведомления и автоматически задействовать мероприятия по противодействию атаке.

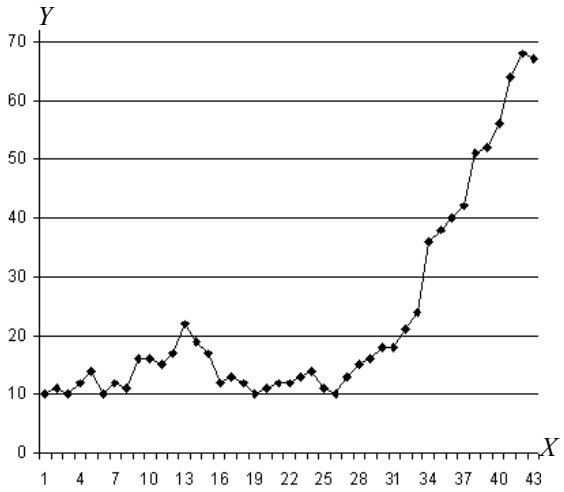

Рис. 1. Количество запросов в период начала DDOS-атаки, 10-минутный интервал

Метод анализа с учетом сезонности показал более высокую точность обнаружения DDOS-атаки и более короткое время, которое прошло с момента начала атаки до её диагностирования.

На графике отражается количество запросов к серверу за секунду, соответствующие периоду начала DDOS-атаки. Ось X – временной интервал. Одно деление соответствует 10 мин. Ось Y – количеству запросов к серверу за секунду.

На основании IP-адресов, принадлежащих компьютерам бот-сети, которые были выявлены при анализе лог-файлов, удалось точно установить момент начала атаки. На графике он соответствует 24-му периоду. Учет сезонности помог выявить DDOS-атаку уже в 31-м периоде. Другие методы показали худшие результаты. При слишком больших значениях *n* атаку удалось диагностировать только на 42-м периоде, при малых происходило ложно срабатывание в 14-м периоде.

В среднем по всем тестам время обнаружения DDOS-атаки методами с учетом сезонности, сократилось в 4 раза, так же сократилось число ложных срабатываний.

Достаточно большой сложностью, возникающей при использовании данного метода, является правильный выбор сходных между собой периодов. Для апробации были выбраны данные с таких серверов, периоды работы которых однозначно определялись и не вызывали сомнения. Однако определить различные периоды в работе, например, крупного магистрального маршрутизатора, достаточно сложно. Его активность может не подчиняться суточным или недельным периодам, но так же иметь свои периоды, которые могут представлять собой сложные периоды, получаемые в результате сложения активностей различных групп пользователей, например пользователей из разных часовых поясов. Кроме того, уже существующие сезонные периоды могут изменяться, к ним могут добавляться новые периоды, поэтому при постоянном мониторинге трафика необходимо будет проводить его кластеризацию и выявлять новые сезонные периоды в работе.

Литература

1. DDOS-атаки [Электронный ресурс]. – Режим доступа: http://localname.ru/soft/ataki-tipa-otkaz-v-obsluzhivanii-dos-i-raspredelennyiy-otkaz-v-obsluzhivanii-ddos.html, свободный (дата обращения: 24.04.2012).

2. Предотвращение атак с распределенным отказом в обслуживании (DDoS) Официальный сайт компании Cisco [Электронный ресурс]. – Режим доступа: http://www.cisco.com/web/RU/products/ps5887/products_white_paper0900aecd8011e927_.html, свободный (дата обращения: 24.04.2012).

3. Методы защиты от DDOS нападений [Электронный ресурс]. – Режим доступа: http://www.securitylab.ru/analytics/216251.php, свободный (дата обращения: 24.04.2012).

4. Терновой О.С. Раннее обнаружение DDOS-атак методами статистического анализа / Перспективы развития информационных технологий. – Новосибирск: Сибпринт, 2012. – С. 201–212.

5. Боровков А.А. Математическая статистика. Оценка параметров проверки гипотез. – М.: Наука. 1984. – 280 с.

6. Бенкен Е.С. PHP, MySQL, XML. Программирование для Интернета. – СПб.: БХВ-Петербург, 2011. – С. 336.

Терновой Олег Степанович
Аспирант ФТФ, зав. отделом корпоративной сети и компьютерных классов Алтайского государственного университета (АлтГУ)
Тел.: 8 (385-2) 36-35-71
Эл. почта: ternovoy@mc.asu.ru

Шатохин Александр Семенович
Канд. техн. наук, доцент, проректор по информатизации АлтГУ
Тел.: 8 (385-2) 36-86-40
Эл. почта: sas@asu.ru

Ternovoy O.S., Shatohin A.S.
Early detection of DDOS attacks by statistical methods, taking into seasonal variation

Reduce errors, and early detection of DDOS attacks by statistical methods. Accounting for seasonality with by using statistical methods. Efficient detect of periods of seasonality.
Keywords: DDOS attack, bot network, standard deviation, statistical analysis.