ПОВЕДЕНЧЕСКИЙ АНАЛИЗ ТРАФИКА ДЛЯ ДЕТЕКТИРОВАНИЯ БОТНЕТ - АКТИВНОСТИ

С.А.БУРНЫШЕВА, В.И.МОИСЕЕВ

Пермский государственный национальный исследовательский университет, 614990, Пермь, Букирева, 15

Ботнет – это сеть компьютеров, зараженных вредоносной программой. Она позволяет удаленно управлять зараженными машинами без ведома пользователя откуда угодно: из другого города, страны или даже с другого континента, а организация Интернета позволяет делать это анонимно. Хозяин зараженной машины, как правило, даже не подозревает о том, что его машина используется злоумышленниками. [1]

В настоящее время для обнаружения ботнетов применяются разнообразные свободно распространяемые и коммерческие средства. Многие из них выполняют анализ данных о трафике. Другие - используют методы анализа поведения, анализ журнала DNS-серверов и создание систем-приманок. [2]

В данной работе проведен поведенческий анализ трафика в сети ПГУ для детектирования ботнет – активности, основываясь на изучении характеристик потока, таких как время начала и окончания потока, средний размер пакета в потоке, отклонение от среднего размера пакета и номер порта назначения. Этот метод позволяет отслеживать тот трафик подозрительной активности, который не видит автоматизированная система, так как она считает его вполне нормальным и не блокирует.

Мы основываемся на методе, который был рассмотрен в статье [3] исследователями Кембриджского университета. Они получили хорошие результаты в ходе поиска искусственно созданного ботнета. В работе [3] отмечена сильная зависимость эффективности от выбранных характеристик потока.

На начальном этапе было выдвинуто предположение: потоки, которые будут сильно коррелировать, у которых будут общие закономерности, вероятно, и будут являться деятельностью ботнета (части ботнета).

Ход действий:

1. Была создана программа на языке C++, которая захватывает пакеты сети ПГУ и записывает информацию о каждом пакете в базу данных MySQL. Причем захватываются только те пакеты, которые соответствуют протоколу tcp (transmission control protocol) и port 3389 (порт протокола удаленного рабочего стола rdp). Порт 3389 был выбран в качестве начальной фильтрации пакета, так как поток трафика очень большой, а

протокол rdp является весьма популярным среди жертв ботнета. В целом, было захвачено 5 млн. пакетов за 20 минут.

2. На следующем этапе необходимо было все пакеты объединить в потоки, где совпадают ip – адреса и номера портов (из 5 млн. пакетов образовалось 744 потока). Также найти характеристики потока.

3. Далее рассматривался вектор потока всех характеристик как точка в n – мерном пространстве и использовалась мера «расстояние» для определения корреляции между потоками. Однако значения характеристик различны, поэтому их необходимо было нормализовать, прежде чем они могли быть использованы. Для этого использовался следующий способ нормализации: norm_diff=$\frac{|a1-a2|}{a1+a2}$, где a1 (a2) – характеристики первого (второго) потоков.

Расстояние между двумя потоками вычислялось с использованием формулы извлечения квадратного корня из суммы квадратов: Distance=$\sqrt{\sum_{i=1}^{n}(norm\_dif\,f_i)^2}$, где n – число характеристик потока. Всего получилось 276396 пар потоков.

4. Результаты корреляции

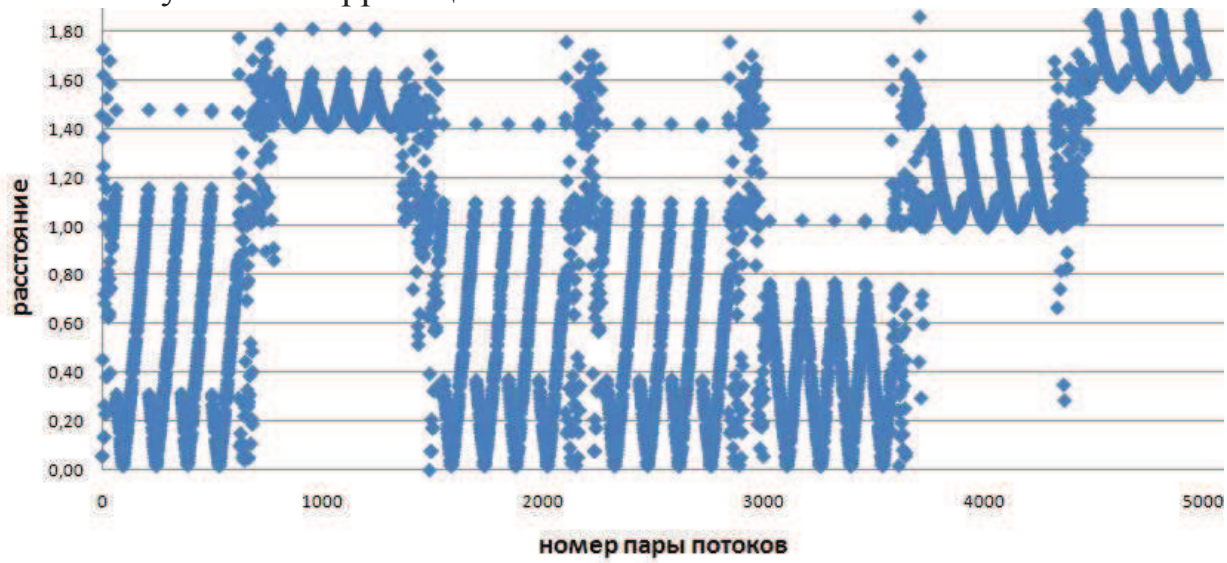

Рис. 1. *Зависимость расстояния от номера пары потоков*

Из рис. 1 видно, что есть всплески на расстоянии, близком к 0. Это означает, что в одно и то же время существуют взаимосвязанные характеристики в различных потоках. Существует 839 пар потоков с нулевым расстоянием, и именно на этот набор направлен наш интерес. Нулевое расстояние интерпретируется как существование одновременно двух потоков с идентичными характеристиками.

5. Топологический анализ.

В топологическом анализе рассматривались только те пары, которые тесно взаимосвязаны между собой (расстояние равно 0). На этом шаге были построены графы связности пар потоков (рисунок 2), где каждое ребро соединяет две высоко корреляционные пары потоков. Был получен 141 граф.

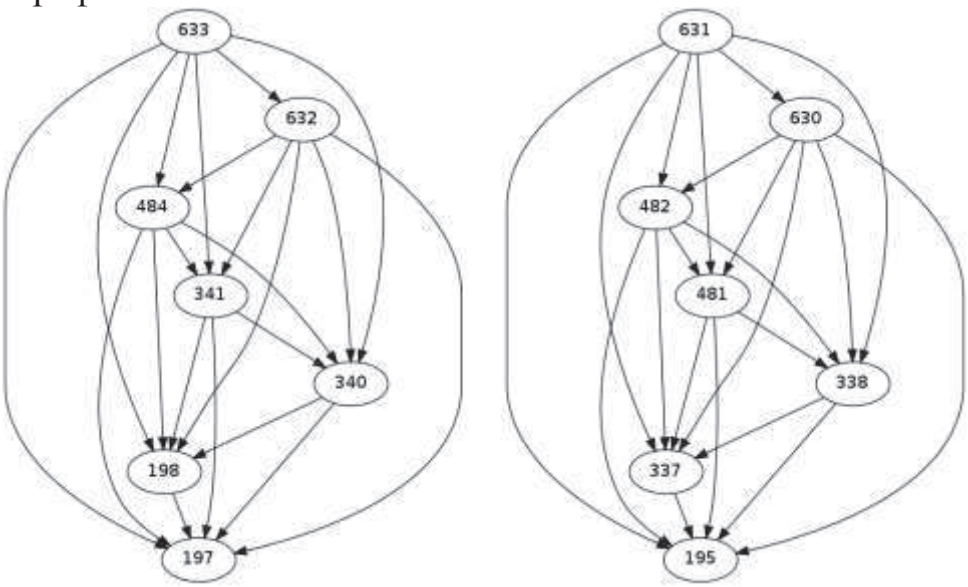

Рис. 2. *Графы связности пар потоков.*

6. Анализ графов.

После топологического анализа приступили к изучению структуры каждого графа, и выяснили, что почти в каждом графе ip – адрес источни-

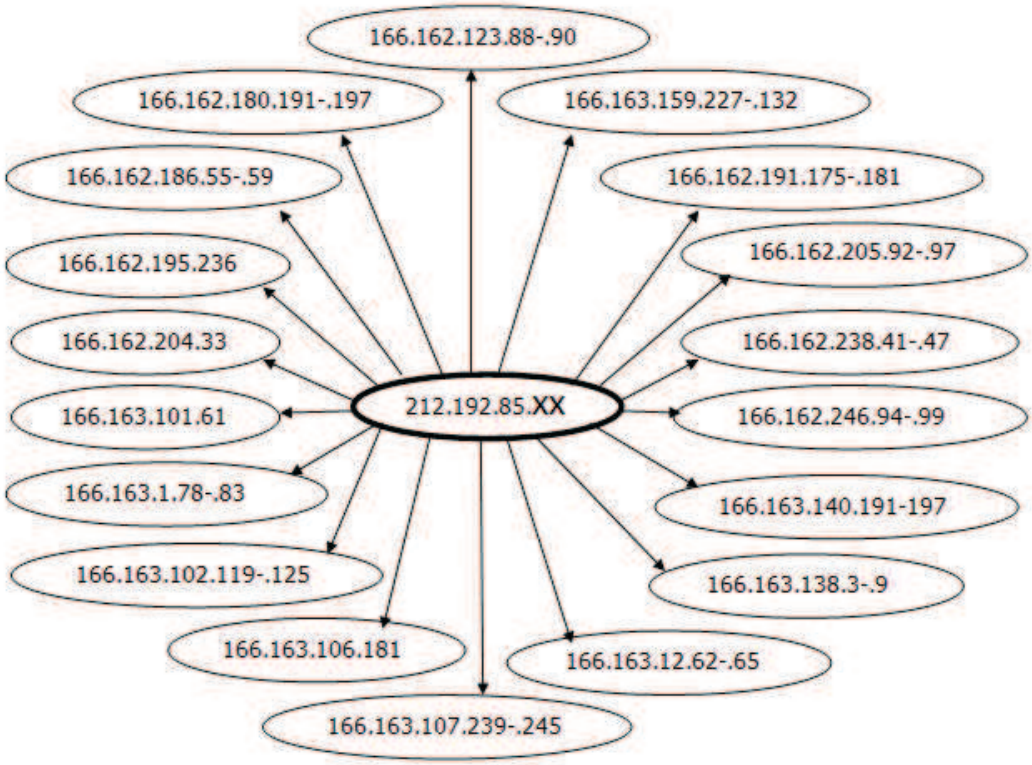

Рис. 3.

ка одинаков. Из этого можно сделать вывод, что источником является злоумышленник, который пытается подключиться к различным удаленным рабочим столам. Он является частью ботнет сети. Поэтому все графы можно собрать в один большой граф, где в центре будет злоумышленник, которого окружают его жертвы (на рисунке 3 отображена только часть такого графа).

В данной работе изучался поведенческий анализ для детектирования ботнет – активности. В ходе исследования был обнаружен бот, который пытался подключиться к удаленным рабочим столам большой группы пользователей. Данный результат показывает, что метод является эффективным, он позволяет отслеживать подозрительный трафик, который не видит автоматизированная система; но в то же время очень трудоемким. Данный анализ имеет перспективы развития. Можно весь алгоритм автоматизировать и запустить в режиме реального времени. Это значительно упростит обработку данных.

СПИСОК ЛИТЕРАТУРЫ

1. http://www.securelist.com/ru/analysis/204007610/Botnety
2. «Ботнеты: новый характер угроз» компания Cisco.
3. «Botnet Detection Based on Network Behavior» W. Timothy Strayer, David Lapsely, Robert Walsh, and Carl Livadas.