К.Л. БЕРЛИЗЕВА, Ю.М. МОНАХОВ

ПОДГОТОВКА ЭКСПЕРИМЕНТОВ ПО ПРОВЕРКЕ РАБОТОСПОСОБНОСТИ МЕТОДА ОБНАРУЖЕНИЯ БОТНЕТОВ НА ОСНОВЕ КЛАСТЕРНОГО АНАЛИЗА И МЕТОДА ОБНАРУЖЕНИЯ БОТНЕТОВ НА ОСНОВЕ СКРЫТОЙ ГИБРИДНОЙ МОДЕЛИ МАРКОВА

Владимирский Государственный Университет им. А.Г. и Н.Г. Столетовых

Доклад посвящён подготовке экспериментов по проверке работоспособности двух методов обнаружения ботнетов: на основе кластерного анализа и на основе скрытой гибридной марковской модели.

Эти способы существенно различаются, но для проведения экспериментов и в первом и во втором случае требуется анализ сетевого трафика. Для эксперимента по первому методу понадобятся такие параметры как ip-адрес, время получения пакета. Для экспериментов по второму методу понадобятся такие параметры как ip-адрес, время получения пакета (для расчёта частоты прохождения пакета), размер пакета (для получения среднего размера пакета).

Для анализа трафика удобно использовать программу Wireshark. Программа предоставляет опцию экспорта файла в формат .txt. В дальнейшем рассматривается возможность написания программного модуля, переносящего значения из текстового файла в базу данных. Расчёт результатов экспериментов удобнее проводить, обращаясь к базе данных.

Краткое описание методов

Метод обнаружения ботнетов при помощи кластерного анализа

Данный метод описан в статье Implementation and Evaluation of Bot Detection Scheme based on Data Transmission Intervals, 2010 (Seiichiro Mizoguchi, Yuji Kugisaki, Yoshaki Hori, Kouichi Sakurai).

В качестве входных данных использовались временные интервалы, между двумя пакетами, от одного и того же пользователя по протоколу IRC. Исследования проводились для обычных пользователей и для заведомых клиентов ботнета. В результате исследования было выявлено, что кривая распределения временных интервалов для клиентов ботнета имеет ступенчатый характер. Это объясняется тем, что боты периодически отправляют сообщения серверу. Для выявления групп сходных временных интервалов авторы статьи предлагают кластерный анализ, который осуществляется путём применения восходящего иерархического алгоритма кластеризации.

Алгоритм кластеризации

1) Каждый элемент данных объявляется кластером. Формируется набор кластеров C=C1, C2,….Cп.
2) Выбирается пара наиболее сходных кластеров Ci и Cj.
3) Ci и Cj – объединяются, создаётся новый кластер Cnew.
4) Cnew добавляется в C, Ci и Cj – удаляются
5) Шаги 2 – 4 повторяются до тех пор, пока число кластеров не станет равным k.

Мера сходства кластеров определяется при помощи центроидного алгоритма. Центроидный алгоритм - нахождение расстояние между центрами тяжести кластеров. В данном случае центром тяжести кластера будет среднее арифметическое его элементов.

Вывод о принадлежности хоста к ботнету осуществляется на основе отношения размерности отдельных кластеров к первоначальному числу кластеров в наборе.

Метод обнаружения ботнетов при помощи скрытой гибридной марковской модели

Авторы данного метода обнаружения исходят из того факта, что хост может находиться в одном из трёх состояний – либо легитимный IRC-клиент,

либо бот, находящийся в состоянии ожидания команды, либо бот в состоянии атаки. Каждому из этих состояний соответствуют специфические значения средней длины пакета и частоты пересылки пакетов.

Это предположение даёт нам количество состояний системы N=3, состояния системы, описываются вектором $S=\{S_1,..,\ S_n\}$. Изменение состояний отслеживаемого сетевого трафика описывается вектором $X=x_1,...,x_T$, где $x_t \in S$. Каждый поток трафика выделяется при помощи некоторого числа временных сетевых характеристик $f \in F_1,...,F_L$, где число наблюдаемых характеристик потока трафика. $V^k = \{v_1^k \ldots v_M^k\}$ – вектор наблюдений. Характеристика F_k – состоит из элементов вектора наблюдений. M – общее число наблюдений характеристики F_k. Гибридная скрытая марковская модель для создания профиля потока включает в себя матрицу переходов между состояниями P, матрицу вероятности наблюдений и матрицу распределения начальной вероятности π. Таким образом, параметры модели обозначаются $\lambda = \{P,Q,\pi\}$.

Матрица переходов P заполняется по формуле:

$$P_{ij} = P(x_{t+1} = s_i \mid x_t = s_j) = \frac{N(s_i \to s_j)}{N(s_i)}, \qquad (1)$$

где $N(s_i->s_j)$ – число переходов из состояния s_i в состояние s_j,

$N(s_i)$ – число состояний s_i.

Q – матрица распределения вероятности появления символов в j-ом состоянии вычисляется по формуле:

$$q_n(V_m) = \prod_{k=1}^{L} P(v_{m_k}^k \mid s_n), \qquad (2)$$

где Vm – элемент вектора наблюдений состоит из $< v_{m1}^1 \ldots v_{mL}^L >$, которые являются значениями рассматриваемых характеристик.

(3)

$$P(v_{m_k}^k \mid s_n) = \frac{1 + N(v_{m_k}^k, s_n)}{M + \sum_{l=1}^{L} N(v_{m_l}^l, s_n)}$$

,

где $N(v_{mk}^k, s_n)$ – число появлений наблюдаемого значения v_{mk}^k в состоянии s_n.

Эксперимент с последующей обработкой результатов методом кластерного анализа

Планируется 6-часовой эксперимент с общим количеством машин, равном 60 хостам. На 20 из них устанавливаются легитимные irc-клиенты. За этими машинами должны быть реальные общающиеся пользователи. На остальные машины устанавливается программа, имитирующая функционирование клиента ботнета. Так же в эксперименте участвует:1 машина – контроллер ботнета (даёт команду на проведение имитации атаки при помощи легитимного irc-клиента), 1 машина – целевой компьютер – жертва, 1 irc-сервер.

Имитацию атаки на целевой компьютер планируется проводить ежечасно. Для этого контроллер ботнета посылает сообщение на канале в следующем формате:

flood [ip-address жертвы][количество ping пакетов]

Эксперимент с последующей обработкой результатов при помощи гибридной скрытой марковской модели

Эксперимент состоит из двух частей обучение (training) и тестирование (testing).

Первая часть нужна для сбора характеристик, а вторая часть для выявления ботнета на основе профиля трафика, полученного с помощью собранных ранее характеристик.

Training

Эта часть эксперимента также подразделяется на две части: сбор трафика ботнета и сбор трафика легитимных пользователей.

Для сбора трафика легитимных пользователей потребуется 15-20 клиентских машин, 1 irc-сервер. Эксперимент длится 3 часа.

Для сбора трафика ботнета требуется сеть, с установленной на 20 хостах программой, имитирующей функциональность клиента ботнета, одна машина – контроллер ботнета, 1 irc-сервер,1 целевой компьютер. Имитация атаки на целевой компьютер производится ежечасно. Эксперимент длится 3 часа.

Testing

Здесь необходим анализ трафика сети, где существуют как ботнеты, так и реальные клиенты. Планируется совместить эту часть эксперимента с экспериментом по методу кластерного анализа, так как условия для эксперимента требуются идентичные.

Сетевая топология не имеет большого значения для проведения экспериментов. В частности, можно провести эксперименты в рамках одной подсети. Но в зависимости от эксперимента менять ip-адреса, irc-сервера, ботов, легитимных машин.

Программа, имитирующая действия клиента ботнета, существует в двух версиях: программа запускает на одной машине один клиент, либо сотню клиентов. Программа работает на платформе Windows. Экспериментально установлено, что для запуска варианта с одним клиентом достаточно машины Windows XP с оперативной памятью 64Мб, для запуска 100 клиентов требуется 98Мб оперативной памяти.

Планируется проведение экспериментов поочерёдно сначала для версии, имитирующей функциональность одного клиента ботнета, потом для версии, имитирующей функциональность 100 клиентов ботнета.

ЛИТЕРАТУРА

1. *Seiichiro Mizoguchi, Yuji Kugisaki, Yoshaki Hori, Kouichi Sakurai* “Implementation and Evaluation of Bot Detection Scheme based on Data Transmission Intervals”, 2010
2. *Ching-Hao Mao, Yu-Jie Chen, Si-Yu Huang, Hahn-Ming Lee* “IRC-Botnet Network Behavior Detection” in Command and Conrol Phase Based on Sequential Temporal Analys, 2010

Сведения об авторах

Берлизева Кира Львовна

Владимирский Государственный Университет им. А.Г. и Н.Г. Столетовых

Студент

Secretik@bk.ru

Монахов Юрий Михайлович

Владимирский Государственный Университет им. А.Г. и Н.Г. Столетовых, кафедра Информатики и Защиты Информации

К.т.н.

Доцент кафедры ИЗИ

ymm@izi.vlsu.ru