Peer to Peer Botnet Detection Based on Node

Suyu Gu, ISCX UNB

Introduction

Since 1989, botnets have evolved from benign assistant tool to the most severe threats to Internet security. A botnet is an organized network of compromised computers (bots) that share the same C&C infrastructure and controlled by a remote command botmaster.

Through the C&C channel, the botmaster can issue commands to his bots army and carry out a series of malicious activities, including stealing personal or sensitive information, engaging in spamming and conducting a DDoS attack.

Objectives

The objective of this thesis is to develop a detection technique based on node-level traffic behavior analysis which allows us to identify P2P botnet activity in real time by examining the features of these active nodes' behavior in small time windows. It contains three sub-objectives.

- First, exploit a set of features of botnet traffic for the purpose of detection by analyzing the network traffic based on node level.
- Second, according to these extracted features implement the detection framework using Artificial Neural Networks techniques.
- Third, evaluate both the feature set and classification techniques to generate an optimized combination that reaches higher discriminating power.

Methods

Feature Set Selection

An feature is some characteristic of a collection of flows in a given time window T which may be represented as numeric or nominal value. Table 1 lists the set of 11 features we have selected for the purposes of our detection.

Attribute	Description
SrcIp	Flow source IP address
SrcPort	Flow source port address
DstIp	Flow destination IP address
DstPort	Flow destination port address
Protocol	Transport layer protocol
NP	Number of protocols used in time interval
NF	Number of flows in time interval
NPS	Number of packets sent in time interval
RNP	Ratio of number of packets sent to number of packets received
ALPS	Average length of packets sent
RLP	Ratio of average length of packets sent to average length of packets received in time interval

Classifier Selection

We use a feature vector to depict a node in the network, and take this feature vector as the input of our classifiers, which adopt two different ANN algorithms.

- **SMO (Sequential Minimal Optimization)**

SVM (Supported Vector Machine) is a kind of blend of linear modeling and instance-based learning. A SVM selects a small number of critical boundary samples from each class and builds a linear discriminant function that separates them as widely as possible. SMO indicates sequential minimal optimization. SMO(SVM) implements SMO algorithm for training a support vector classifier using polynomial kernels.

- **Back Propagation Neural Network**

BP neural network involves three stages: the feed-forward of the input training pattern, the calculation of the associated error, and the adjustment of the weights. The BP network used herein has an input layer, an intermediate hidden layer, and an output layer. The nodes in this network are all sigmoid.

Results

The experiments for now are supervised style, which means a training set and a testing set are included.

Training Set

- Our training set contains 3000– 4000 Bot feature vectors and 3000 – 4000 Normal feature vectors. And among the Bot vectors, half of them are Strom Bot and half of them are Waledac Bot.

Testing Set

- Our testing set contains 800 -- 900Bot feature vectors and 400 – 500 Normal feature vectors. And among the Bot vectors, half of them are Strom Bot and half of them are Waledac Bot.

For the classification, we divide it into two levels:

2-Class Classification

• Nodes are classified into to two classes: Bot node and Normal node. The detection rate for SMO is shown in Figure 1, and for BPNN is shown in Figure 2.

Figure 1. Detailed Accuracy for SMO

```
=== Detailed Accuracy By Class ===

                 TP Rate   FP Rate   Precision
                   0.944     0          1
                   1         0.056      0.896
Weighted Avg.      0.962     0.018      0.966
```

Figure 2. Detailed Accuracy for BPNN

```
=== Detailed Accuracy By Class ===

                 TP Rate   FP Rate   Precision
                   0.944     0          1
                   1         0.056      0.896
Weighted Avg.      0.962     0.018      0.966
```

3-Class Classification

- Nodes are classified into to three classes: Storm Bot node, Waledac Bot node and Normal node. The detection rate for SMO is shown in Figure 3, and for BPNN is shown in Figure 4.

Figure 3. Detailed Accuracy for SMO

```
=== Detailed Accuracy By Class ===

                 TP Rate   FP Rate   Precision
                   0.726     0.016      0.936
                   0.804     0.112      0.847
                   1         0.112      0.81
Weighted Avg.      0.849     0.089      0.856
```

Results

Figure 2. Detailed Accuracy for BPNN

```
=== Detailed Accuracy By Class ===

                 TP Rate   FP Rate   Precision
                   0.979     0.005      0.985
                   0.909     0.004      0.995
                   1         0.058      0.892
Weighted Avg.      0.955     0.021      0.959
```

All the experiment results above are based on a 5s time window.

Conclusions

We proposed a detection model for detecting bot activity based on the observation of its network flow features for specific time intervals. We use the concept of time intervals to limit the duration we would have to observe any particular flow before we may raise our suspicions about the nature of the traffic.

We showed that using an ANN classifier, we were able to not only successfully detect botnet activity but also different botnet activity with high accuracy by simply observing small portions of a full network flow.

For our future work, we note that the current detection framework is a supervised system, which doesn't equip the ability to detect unknown bots. In order to make up for this defect, we plan to additionally produce an unsupervised system which is capable to recognize unknown bot trace and dynamically expand its detection ability without requiring an offline training process.

ISCX
Information Security Centre *of* Excellence