Group Behavior Metrics for P2P Botnet Detection

John Felix Charles Joseph and Ali. A. Ghorbani

ABSTRACT

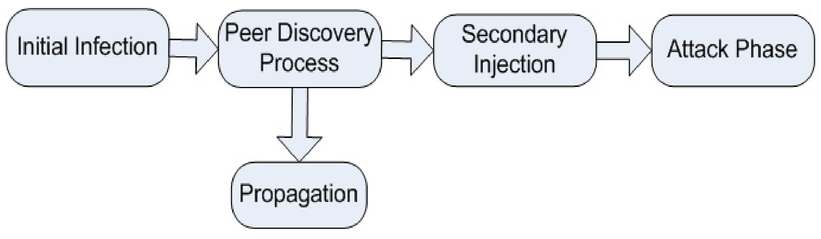
We propose a set of metrics for efficient botnet detection. The proposed metrics captures the unique group behavior that is inherent in bot communications. Our premise for proposing group behavior metrics for botnet detection is that, group behavior observed in botnets are unique and this unique group behavior property is inherent in the botnet architecture. The proposed group behavior metrics uses three standard network traffic characteristics, namely, topological properties, traffic pattern statistics and protocol sequence and usage to derive the proposed metrics. We derive six group behavior metrics and illustrate the efficiency of botnet detection using these metrics. It was observed that, group behavior metrics offers a promising solution for botnet detection.

BACKGROUND

Botnet detection is a non-trivial problem. Experience with respect to centralized IRC/HTTP based botnet detection and mitigation will prove that. Now, the challenge of botnet detection has become harder due to the advent of P2P botnets. Research is yet to provide standard and efficient system for botnet detection. Existing botnets detection methodologies suffer from the tactics used by attackers to thwart detection. Just like software, the bot malware is constantly updated and new revised versions are released, periodically. With P2P technology, this update process is distributed and autonomous, which make botnet resilient to detection and mitigation.

Existing work on group behavior based detection of botnets is very few. Group behavior is often overlooked due to the intuitive belief that presence of groups of bots within the same subnet is highly unlikely. However, a look at the traffic through an ISP gateway will prove otherwise. Due to bot propagation mechanism, it is highly likely that more than few bots exist in the traffic of a subnet.

FIGURE 1. BOT DEVELOPMENT CYCLE



GROUP BEHAVIOR IN BOTNETS

The generic development cycle of a malicious botnet consists of three primary stages, as shown in Figure 1. Group behavior among bots is inherent due to the botnet architecture. After the initial infection, each phase in the development cycle of the malicious botnet adds strong group behavior properties to the bot behavior.

Once a host is being infected with the malicious bot code, the bot tries to propagate itself to other hosts that are connected to the infected host. As the hosts that are infected using the propagation mechanism are infected with the same bot code, the bots' network behavior is completely identical.

In the second stage of botnet development cycle, the malicious bot installed in the infected host tries to connect to other bots (peers). This process in P2P terminology is referred to as peer discovery process. This peer discovery process causes a bot to exhibit strong group behavior with respect to common network connectivity.

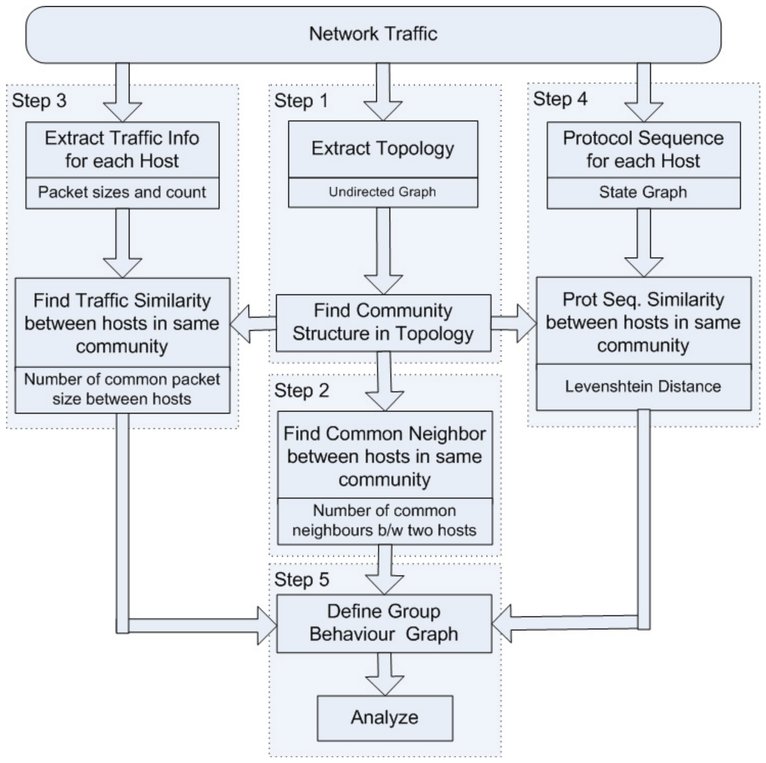
During the attack phase of a botnet, bots exhibit strong group behavior. This is primary due to the fact that attacks are coordinated using a set of bots.

GROUP BEHAVIOR METRICS

The group behavior metrics for hosts in the network are derived using three network traffic characteristics, namely, topology, traffic pattern and protocol usage. The process of deriving the group behavior metrics from the network traffic is illustrated in Figure 2. The process comprises of five stages.

For each of three network traffic characteristics, we use features that capture group behavior in network behavior. Common connectivity among hosts is derived from the topological properties of the network and is used for capturing the group connectivity. Similarity in packet sizes and frequency is used to measure the group behavior in traffic patterns. Similarity between protocol sequences exhibited by hosts in their network traffic is used to measure the group behavior in protocol usage. Thus, the process uses the three primary characteristics of network traffic to derive the group behavior metrics.

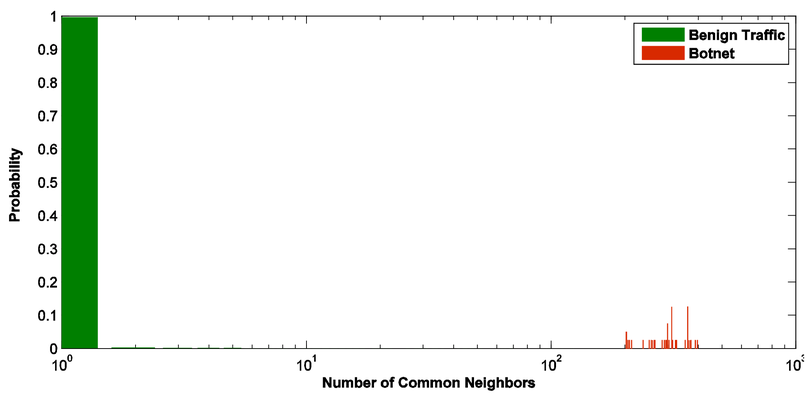
FIGURE 2. GROUP BEHAVIOR METRICS



TOPOLOGICAL PROPERTIES

In the first step, the topology of the network that is represented in the network is extracted, as shown in Figure 2.

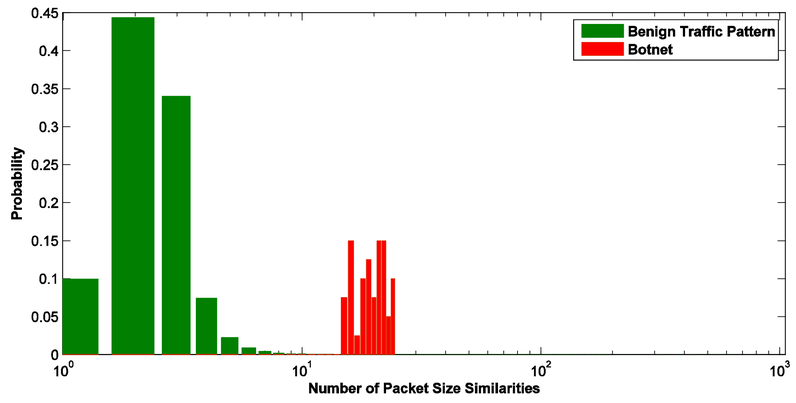
At stage two, the group behavior within the topological properties of the network is evaluated. The common connectivity of nodes is derived by computing the number of common neighbors between two hosts within the topology. The common neighbor for every node pairs in the topology is computed. The property of bot to have high common connectivity is captured using this metric.



TRAFFIC PATTERN STATISTICS

For representing the traffic pattern, we primarily use packet size feature of the network traffic. At stage three, we extract the traffic information for each host in the network. For each host, we record different packet sizes that are observed within the host's network traffic. Additionally, the frequency of packet sizes within the host's communication is also extracted.

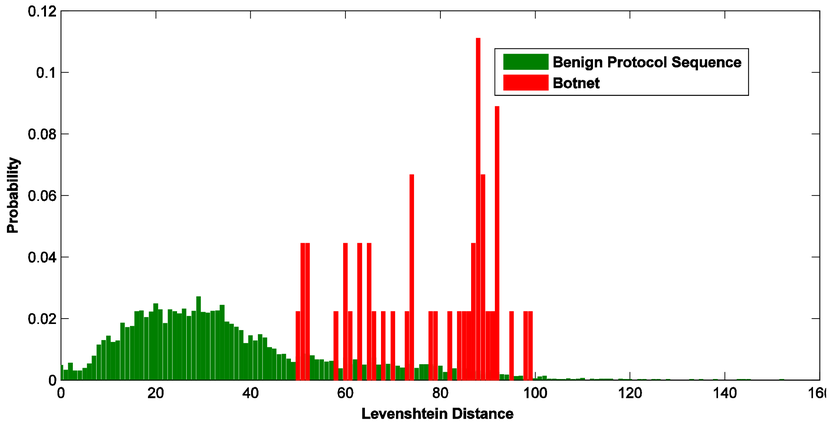
After the traffic pattern is extracted, the common traffic pattern is evaluated for every two hosts within every group in the topology. Similarity in traffic pattern using the packet size representation is computed.



PROTOCOL SEQUENCE SIGNATURE

At stage four, the protocol usage and sequence is extracted from the network traffic. First, we identify the protocol of each packet in the network traffic by using wireshark's protocol dissectors.

After protocol analysis, for each host, the sequence of protocol communication is captured in a state graph. With the protocol sequence state graphs defined for all hosts in the network, we then compute the similarity in protocol sequence between every two hosts in a network topology community. The similarity between state graphs of two hosts is used to compute the common behavior in protocol sequence and usage. For measuring the similarity between two state graphs, we use two different similarity measures, namely, Levenshtein distance and Jaccard similarity.



CONCLUSIONS

The property of bots to exhibit similar communication patterns is exploited to derive the proposed metrics. Three network properties, namely, topological characteristics, traffic statistics and protocol usage sequence is used to derive the group behavior for each host in the network. It is observed that, group behavior of bots is distinctly captured by these metrics.

2011 Research Expo, Fredericton, NB, April 15, 2011