Anomaly Detection Through Blind Flow Analysis Inside a Local Network

Ron McLeod, BCSc, MCSc.
Director - Corporate Development Telecom Applications Research Alliance
Doctoral Student, Faculty of Computer Science, Dalhousie University
Vagishwari Nagaonkar, BCSc
Senior Systems Engineer, Wipro Technologies
Graduate Student, Faculty of Computer Science, Dalhousie University

Abstract

In the August of 2006, 4 months of Netflow [1] records that were collected inside a sma ll private network were subjected to a Blind Flow Analysis. Such an analysis is characterized by having access to the flow records from inside the network but no access to the payload data and no physical access to the hosts generating the traffic. Experiments were conducted to discover if useful behavioural clusters could be constructed with such minimal access and whether individual classes of hosts could be clustered into standard ranges including clusters indicative of compromised hosts. Early results are promising in that hosts may be clustered into User Workstations, Servers, Printers and hosts Compromised by Worms.

Overview

In a network environment where a single managed network is used by many different individuals and/or corporate entities, network monitoring for security purposes can be problematic. Each entity may have specific concerns with respect to privacy or corporate confidentiality. A network analyst employed by the network provider may be specifically forbidden from capturing the payload data in the traffic. Furthermore, the analysts may not be granted access to specific hosts by the host owner and may not even be able to receive information as to the type and nature of the host in question (i.e. is this a server, a workstation or a printer).

In this environment the analyst may be restricted to analyzing only packet header data or flow records. The authors decided to test the ability of the analyst to form useful characterizations in such a restricted environment. While the work is in its early stages it has already proven useful in that a previously undetected worm was found in the network. As well, early evidence of another worm was found in historical traffic records one month prior to the time that the worm's presence was actually detected when it attacked an external host.

Within the boundaries of locality, flow profiles were developed for user workstations, network servers, network printers and hosts compromised by worms.

Clusters by File Size

The first level of separation in the flow data occurred at the directory level by simply comparing the file size of Bag Files generated Silktools [2]. Figure 1 identifies significant groupings in the data for Bag Files of Destination Ports, Destination IP's and Protocols for non-port 80 traffic during the month of February. These anomalies were analyzed by the authors. A hypothesis was developed and the IP number of the host in question along with the hypothesis was sent to the network owner. For the purpose of testing experimental results, the network owners were asked to confirm (on a voluntary basis) the hypothesis. The Anomaly labeled Game Server represented a compromised host on the network that was being used to support worldwide on-line gaming. The anomaly labeled Host 22 was believed to be a VLAN gateway, but this hypothesis has yet to be confirmed.

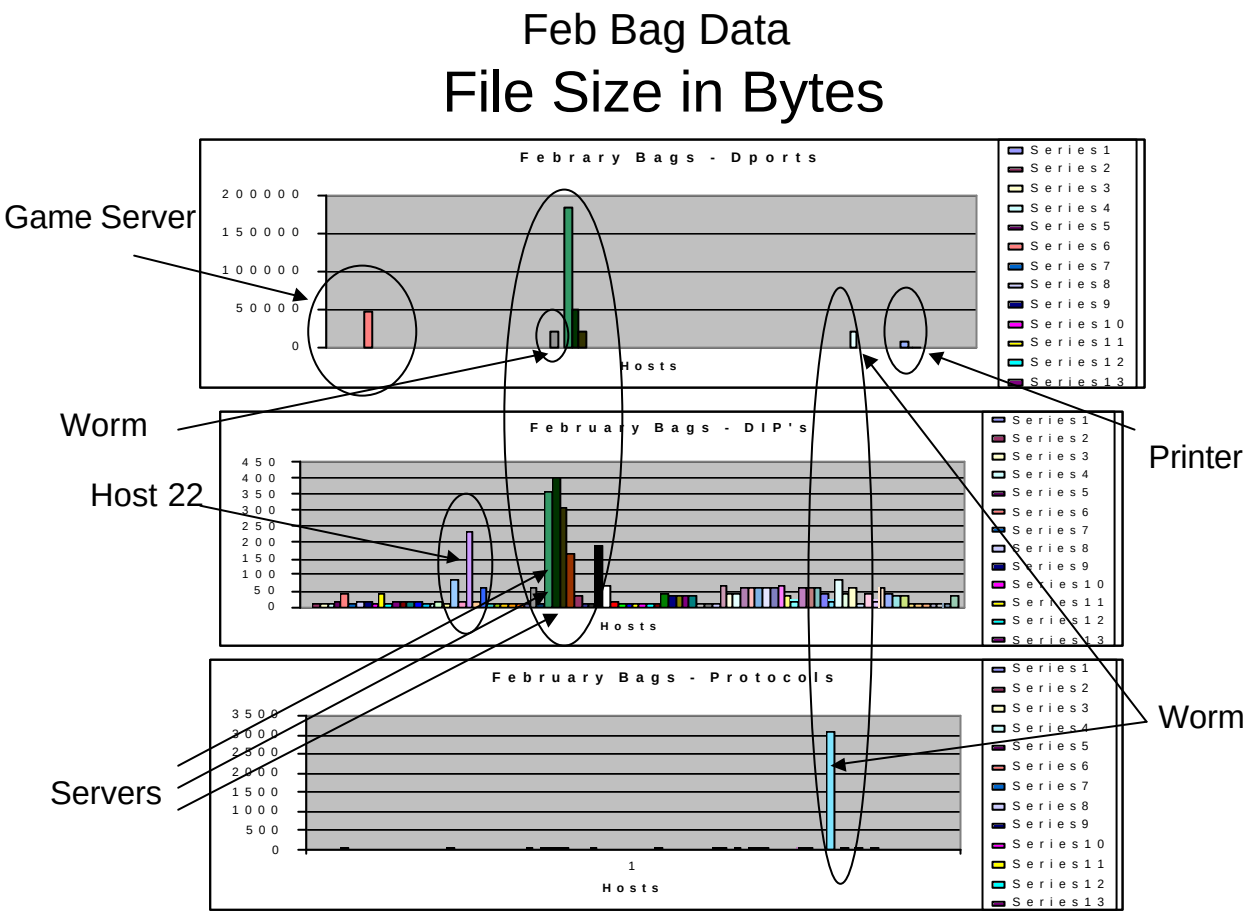

FIGURE 1 – Bag File Sizes

Localized Characterizations by Cluster

Below are some suggested descriptive Rules to be used in identifying for the hosts observed. Actual values will be different for each local Network, but we currently expect that the relative relationship between the values may be somewhat consistent.

Possible Workstation Rule for Classification

IF Bytes Transferred in one month < 20 meg

AND (Internal DIPs < 5 AND External DIP's < 10)

AND (Greatest Byte Volume is to Internal DIP's)
(Randomly Distributed above 1000 bytes)

AND Protocols: 1 ~ 1 % +/- A Threshold
6 ~ 29 % +/- A Threshold
17 ~ 70 % +/- A Threshold

AND Dports: 25% < Port # 1024
50 % < Port # 1024 - 5000
25 % > Port # 5000

THEN HOST is a user Workstation

FIGURE 2 – A User Workstation Rule

Possible Worm Rule for Classification

IF Bytes Transferred in one month > 20 meg

AND (Most Internal DIPs < 5 AND External DIP's > 10)

AND (Greatest Individual Byte Volume is to Internal DIP's
AND Randomly Distributed above 1000 bytes)
(Greatest total Byte Volume to External IP's
AND uniformly Distributed below 1000 bytes)

AND Protocols: 1 ~ 1 % +/- Threshold
6 ~ 29 % +/- Threshold
17 ~ 70 % +/- Threshold
[and optionally small % to all Protocols]

AND Dports: 10 % < Port # 1024
90 % < Port # 1024 - 64K

THEN HOST is a user workstation compromised by a worm

FIGURE 3 – A Worm Rule

On-Going Work

As of the date of submission, this work is on-going and final results are not yet available. However, a parallel effort by one of the authors shows significant early evidence that, using a subset of same data with the DIP removed, a neural classifier can be trained to distinguish between individual user workstations.

This knowledge has lead to an optimistic view by the authors that this classification attempt described herein will prove useful.

References

[1] CISCO IOS Netflow, http://www.cisco.com/en/US/products/ps6601/products_ios_protocol_group_home.html, as accessed on August14, 2006.

[2] SilkTools, http://silktools.sourceforge.net/, as accessed August 14, 2006.