

Network Profiling And Data Visualization

Stephen C. Fortier & Lee A. Shombert

AverStar, Inc.

1593 Spring Hill Road, Suite 700

Vienna, Virginia 22182

(703) 827-2606

[sf, las]@averstar.com

31 March 2000

Abstract

Purpose

This paper describes of our research and development effort for network profiling and data visualization. Our premise is that we should be able to characterize network traffic across a firewall in a so-called model of good behavior. We then apply this model to firewall reports to identify deviations from good behavior; these deviations are candidate attacks.

Methods

In the current task, we are examining the Averstar firewall logs to build a model of expected behavior. For example, two days worth of data from the Averstar firewall was examined. This data is a log of all transactions (but not packet content) through the firewall. This data totals about 1.1million records. An initial model was defined and the logs were filtered to eliminate records that conformed to the model. The initial model was quite simple:

1. All traffic originating in an Averstar intranet and destined for an Internet site was removed. Note that a more robust model might look at this traffic for signs of rogue programs communicating back to "home base," but this traffic would appear after an attack had been successfully mounted.
2. All traffic between Averstar intranets was filtered. Again, this traffic could be analyzed to detect compromised systems within the Averstar intranet.
3. All outside traffic destined for boundary servers was eliminated. A boundary server is a server whose function is to provide a service to outside entities. The current filter eliminates from analysis connections to Averstar web servers (port 80) and Averstar SMTP servers (port 25).

By their nature, boundary servers should receive lots of connections and it would be hard to identify attacks buried in the heap. An attack against such servers requires access to the packet contents. For instance, we would care about the web pages actually being requested by an outside client.

Results

Only IP connections that have been accepted were examined. The analysis clearly must be extended to other kinds of firewall records (reject, deny, authorize, etc). The analysis should also be extended to ICMP connections, as these may represent probes. After this first level of filtering, we are left with about 5,500 records. These are then displayed in various dimensions with the SGI Mineset data visualization and mining tool. Mineset provides the capability to examine large amounts of data and quickly identify patterns or relationships in the data. One way to use this "system" is to filter firewall logs periodically (e.g., overnight) and generate the Mineset views offline. The network administrator can review the Mineset reports, and, we hope, quickly identify anomalous traffic.

New or breakthrough aspect of work

The filtering model will evolve. Eventually, this will be modified to a real-time analysis of network traffic, with alarms that are triggered when an attack is in progress. It is certainly desirable for the model to look at data besides firewall logs, too.

Conclusions

We believe that the results of our research will lead to a practical way organizations can profile their network traffic and visualize known good behavior while highlighting anomalous behavior. This methodology would be applied to stand-alone locations as well as multiple node models.