available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/cose

Computers & Security

SVision: A novel visual network-anomaly identification technique

*Iosif-Viorel Onut, Ali A. Ghorbani**

Faculty of Computer Science, University of New Brunswick Fredericton, Canada

ARTICLE INFO

Article history:
Received 16 October 2005
Revised 21 September 2006
Accepted 4 October 2006

Keywords:
Anomaly visualization
Network security
Intrusion detection
Service based data visualization
Network monitoring

ABSTRACT

We propose a novel graphical technique (SVision) for intrusion detection, which pictures the network as a community of hosts independently roaming in a 3D space defined by the set of services that they use. The aim of SVision is to graphically cluster the hosts into normal and abnormal ones, highlighting only the ones that are considered as a threat to the network. Our experimental results conducted on DARPA 1999 and 2000 intrusion detection and evaluation datasets as well as real network data captured between 2003 and 2005 from the University of New Brunswick main link, and also a private network, show the proposed technique as a good candidate for the detection of various network threats such as vertical and horizontal scanning attacks, Denial of Service (DoS) attacks, Distributed DoS (DDoS) attacks, as well as worm propagation attack. Finally, the visualization technique proves to cope with high number of hosts in the network, the experimental results using network data of up to 1,000,000 distinct IPs per time interval.

1. Introduction

Data visualization represents a fundamental part of the current network security practices, providing the network administrators with important information regarding the state of the network as well possible threats that exist. Frost and Sullivan (2004), recently reported that only 11.6% of the available Intrusion Prevention Systems (IPSs) in 2003 were set to prevention mode by the administrators. Consequently, in all the other cases, the network administrator is the one that decides upon the proper response that has to be enforced. In order to do that, he/she has to have a deep understanding of the current state of the network, and this is mostly achieved through different network visualization techniques. Thus, despite all the existing criticisms against the visualization techniques as a detection method, we do not anticipate its possible replacement in the near future.

We propose a network visualization technique that allows the security personnel to easily identify potential anomalies in the network. The network is depicted as a community of hosts that are roaming inside a three dimensional space. Since a network might have hundreds of hosts, the proposed view highlights only the ones that might represent a potential threat to the network, while the normal hosts overlap near the center of the view.

Our experimental results conducted on two of the well known intrusion detection and evaluation datasets (i.e. DARPA, 1999, 2000) empirically proved the technique to be successful against main types of Denial of Service (DoS) attacks, Distributed DoS (DDoS) attacks, as well as vertical and horizontal scanning attacks.

This paper is organized as follows: Section 2 presents some of the important existing visualization techniques. Section 3 describes the proposed visualization technique presenting

* *Corresponding author.*
E-mail addresses: onut.viorel@unb.ca (I.-V. Onut), ghorbani@unb.ca (A.A. Ghorbani).
0167-4048/$ – see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cose.2006.10.001

the main outcomes and drawbacks of the representation. The implementation and deployment of our system is described in Section 4. Next, Section 5 presents the empirical results against the common attacks such as probing, DoS, DDoS, and worm propagation. Finally, the last section summarizes the conclusions and presents possible future improvements.

2. Background review

Humans tend to be very comfortable with data presented to them in a form of charts or graphics. It is this reason that lead researchers in their attempt to graphically model anything related with network security. Network administrators tend to be very comfortable with network data presented in the form of charts, functions, and tables. The network visualization techniques do target most aspects of the network security including topology representation, protocol communication, and congestion control, to name a few.

Spencer (2005) proposed a visualization technique that displays the network topology, assisting the security personnel in detecting possible failure points and checking the availability of the devices within the network. Erbacher (2001) proposed a similar technique that uses a glyph based approach in order to represent not only the topology of the network but also its load. Each node represents a host, server or router, while edges represent connections. The load is depicted by a gray filling inside the nodes or edges. In the same line of work, Estrin et al. (1999) proposed a visualization system that shows network topologies animations, measuring packet loss rates for various links in order to detect potential connectivity problems.

The most common visualization technique remains the two-dimensional graphs where one dimension represents the time coordinate (e.g. usually x axis), while the second axis represents a particular feature of the network. Moreover, by the use of colors, multiple graphs can be mixed in a single view (Plonka, 2000; Estrin et al., 1999; Oetiker and Rand, 2005; Q1Labs, 2005). Such visualization tools have been widely used by network administrators to monitor the network links and identify abnormal external behavior such as DDoS, DoS, Scanning, and Worms, as well as improper internal activity such as P2P file sharing.

As more powerful computation capability becomes available, visual representation of network has evolved from 2D to 3D in order to incorporate more complex information. In the 3D graphical technique proposed by Fisk et al. (2003), internal and external IP addresses are mapped into a 3D space, the connection between an internal host and an external host is presented by a line with particular color and length, representing information such as service type, duration of that connection, and source/destination IPs. This visualization technique proved to be very efficient in the case of scanning attacks. The CICHLID Data Visualization Software (NLANR, 2005) is an example of 3D visualization tool for network data processing. It consists of various views with colored bars representing different network characteristics such as IP Address Utilization, Packet Length Distributions for major IP protocols, and bytes vs. time for major TCP/UDP port numbers, to name a few.

3. The SVision visualization technique

The proposed visualization technique (SVision) pictures the network as a community of hosts that independently roam in a 3D space defined by the set of services that they use. The aim of SVision is to graphically cluster them into normal and abnormal ones. Spheres are used as graphical elements for representing the hosts' position, while gray and black colors[1] are used to distinguish between the internal and external hosts, respectively. Furthermore, the intensity of the host's color changes from dark to light with respect to the time (i.e. once a host is detected it has a dark color; as the time passes by its color changes to a lighter nuance).

The view require the use of only six fields from the packet datagram (i.e. source and destination IP, source and destination Port, packet length, and IP protocol type) making the approach feasible for working under real traffic loads such as tens of megabits per second.

The hypothesis of this view is that a system administrator is able to identify a set of most important/critical services that have to be closely monitored for a given network. Moreover, for an organization like a bank or factory, the number of critical services is normally less than 10, usually being among the most common ones such as HTTP, FTP, DNS, VoIP, to name a few. Let Ψ represent this particular set of services.

Let *sparsely-active* (*constantly-active*) represent a host who is seldomly (constantly) using any number of the Ψ services in a predefined time window interval τ. Our proposed graphical model uses a two-dimensional plane (i.e. *Service Usage Plane*) to discriminate between sparsely-active and constantly-active hosts with respect to the selected set Ψ of services. Furthermore, let us define the *service point* as the graphical point where a certain service will be displayed in the view (see Fig. 1). All of the *service points* are placed on a circle centered in the origin θ of the view. Moreover, the points are positioned equally distant among themselves. The number of services in Ψ will define the shape of the *Service Usage Plane* where hosts will move (e.g. triangle, pentagon, and hexagon for 3, 5, and 6 services, respectively).

The idea behind our host clustering technique is that the more a host is using a service in a predefined *time window interval* τ, the closer it will be from that *service point*. Consider the case of a host H_j who is mostly using the kth service from the Ψ set (i.e. S_k). Consequently, its position in the view will be attracted by the *service point* of S_k. Thus, if the host is continuously using that service it will eventually end up in the same spot where the S_k *service point* is. Similarly, if the host is using n services, it will be attracted by all of them in the same time.

Let us define the *attraction force* as the force that a particular service S_k attracts a host H_j.

$$\vec{F}_{k,j} = A_k \cdot L_{k,j} \begin{bmatrix} \cos(\alpha_{k,j}) \\ \sin(\alpha_{k,j}) \end{bmatrix}, \quad (1)$$

where $L_{k,j}$ is the load of the host H_j with respect to the service S_k, and A_k is a predefined *anomaly factor* for service S_k. The

[1] Due to printing limitations, we use gray and black colors instead of the original colors that the system works with (i.e. blue and red, respectively).