Scalable Visualization of Propagating Internet Phenomena

Alfonso Valdes valdes@sdl.sri.com

Martin Fong mwfong@sdl.sri.com

ABSTRACT

The Internet has recently been impacted by a number of large distributed attacks that achieve exponential growth through self-propagation. Some of these attacks have exploited vulnerabilities for which advisories had been issued and for which patches and detection signatures were available. It is increasingly apparent, however, that such prevention and detection mechanisms are inadequate, and that the attacker's time to exploit is shrinking relative to the defender's ability to learn of a new attack and patch systems or update intrusion detection signatures. We introduce visual, scalable techniques to detect phenomena such as distributed denial-of-service attacks and worms. It is hoped that these new approaches will enable detection of such events at an early stage and enable local response actions even before the publication of advisories about a new vulnerability and the availability of patches.

Categories and Subject Descriptors

C.2.0 [**Computer-Communications Networks**]: General – *security and protection*. C.2.3 [**Computer-Communications Networks**]: Network Operations – *network monitoring*.

General Terms

Management, Measurement, Performance, Experimentation, Security.

Keywords

Intrusion Detection, Internet Worms, Data Mining, Scalable Visualization.

1. INTRODUCTION

Several recent and highly publicized examples of Internet attacks have consisted of self-propagating phenomena ([2, 5] describe just two of the better known examples). Such attacks have the potential to saturate the vulnerable population of hosts on the Internet in a very short amount of time [7, 8], and the probes launched with the objective of self-propagation may significantly degrade local and Internet-wide quality of service. For example, although our site has no hosts vulnerable to the Slammer worm [5], and the firewall rejects the malicious UDP packets, our users experienced significant performance degradation during the period of maximum Slammer attack intensity. August and September of 2003 witnessed still greater frequency and impact of these events, placing the Internet into a "Worm of the Week" repair mode.

These attacks stress enterprise-level intrusion detection systems. We observed more than 100,000 alerts at our site related to the Blaster and Nachi attacks [6] in a twenty-four hour period, a rate of almost 1.5 alerts per second. We seek a global characterization of such attacks not based on IDS signatures that might be visible at an early stage in the attack and at various scales on the Internet, from the enterprise to the backbone level.

Detection and response to these events to date has consisted of publication of a vulnerability to some service followed by administrative response (blocking the service where not needed, applying vendor patch, updating intrusion detection rules). However, time to note the vulnerability and undertake the administrative response is comparatively long, and response is not always effective or universally applied. As such, self-propagating attacks now feature a rise time that has shrunk from days to hours or minutes. As the defender's decision cycle time shrinks, it is more likely that the attack will hit before the advisory is posted or at least before the defender is aware of it. Lacking an advisory, administrators may not be sensitive to unusual activity for a port until the attack is well established. Detection and prevention by means other than an advisory or IDS update may be useful if it can be done early in the rise time of the event.

To this end, we introduce a scalable visualization technique for detection. The basis of our approach is to represent the networks under consideration as images, where coordinates are obtained by a suitable mapping function of the address space or ports. At peering points, the source and destination address space may be considered the same. At gateways between open and private networks, we may wish to represent sources from the open network and destinations in the secure network. Pixels in this image can themselves represent networks. These images are maintained by a rapidly updated fading memory mechanism, so that the image at any given time represents a situational picture of what has happened in the very recent past (fading memory time is a tunable parameter).

The remainder of this paper is organized as follows. We first describe our scalable visualization approach, and provide initial results (including a view of the Kuang 2 attack). This is followed by a discussion of related work, and a summary in which we outline future directions, principally the concept of searching for interesting events in process innovations after appropriate de-trending.

VizSEC/DMSEC'04, October 29, 2004, Washington, DC, USA.

2. Scalable Visualization

We propose techniques to view propagating network phenomena as analogous to image processing, where multi-scale processing techniques and high-performance hardware coprocessors are more highly developed. We arbitrarily map sources to the vertical axis and destinations to horizontal. The techniques presented here as suitable for detection at peering points and are inherently scalable and suitable to distributed computation.

The basis of our approach, building on our fast abstract presented at DSN 04 [9], is to represent the networks under examination as images, where coordinates are obtained by a suitable mapping function of the address space. At peering points, the source and destination address space may be considered the same. At gateways between the Internet and the enterprise, we may wish to represent sources from the network and destinations in the enterprise, perhaps using a different address mapping function for each.

Measures of interest include traffic volume, either total or for specific protocols of interest. If ubiquitous instrumentation may be assumed (possibly the case in secure networks), we may consider the volume of rejected packets, the rate of alerts of various types, and so forth. The technique is composable in that a distributed component can implement the approach for its domains of interest and propagate its results upward ("all OK" or "I've quarantined this address space").

We generated such images for source IP address/destination IP address and source IP address/destination port. The intensity of the pixels corresponds to the count of connections rejected by our firewall for the source and target (IP address or port). To obtain images 256 pixels on a side, source IP address, destination IP address, and destination ports were hashed into an unsigned byte using the following algorithm:

```
result = 0;
while (input /* != 0 */) {
  result = circLeftShift (result, 1) ^
           lastByte (input);
  input  = rightShift (input, 8);
}
```

where `circLeftShift ()` performs a 1 bit left-circular shift on a byte, `lastByte ()` retrieves the low-order byte from an integer, and `rightShift ()` performs a non-sign extended right-shift of 8 bits. Shifting by a different number of bits enables generation of different image sizes.

The properties of this hashing algorithm include byte-order sensitivity and the use of all input bits. Another relevant property for the port numbers is that values less than 256 were maintained without modification, enabling the easy identification of various well-known services (e.g., SMTP, FTP, HTTP).

To enable grayscale display, the counts were normalized to a maximum value of 255. However, because we discovered that the source IP address/destination IP address/count triplet display had severe banding (corresponding to IP address sweeps), we removed the minimum value across the row for each source IP address bucket. Additionally, because the residual maximum value was orders of magnitude higher than the lowest values, the grayscale value was derived from the logarithm of the counts.

Finally, a histogram of the sum counts for a given row or column was generated and displayed with the grayscale output.

In such a representation, horizontal scans appear as a horizontal feature in the source address/target address map, while propagating phenomena appear as vertical features. Figure 1 shows a snapshot of the first type of image, with several apparent horizontal scans. The vertical feature does not correspond to an attack, but rather a configuration artifact due to an advertised MMX service not available through the firewall. The two histogram spikes, one about one-fourth of the way down the vertical axis and the other three-fourths of the way down, correspond to separate vertical scans of a few addresses (the blackest dots in the image).

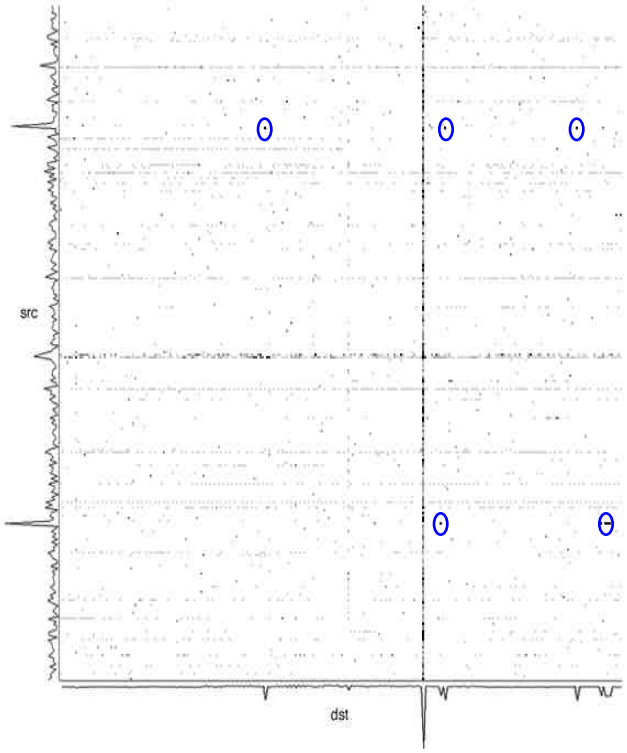

Figure 1: Source IP/Destination IP Map

Figure 2 shows the source IP/destination port map. The Kuang 2 attack results in the most prominent vertical feature, although less pronounced features corresponding to destination port 80 and 135 are apparent as well.

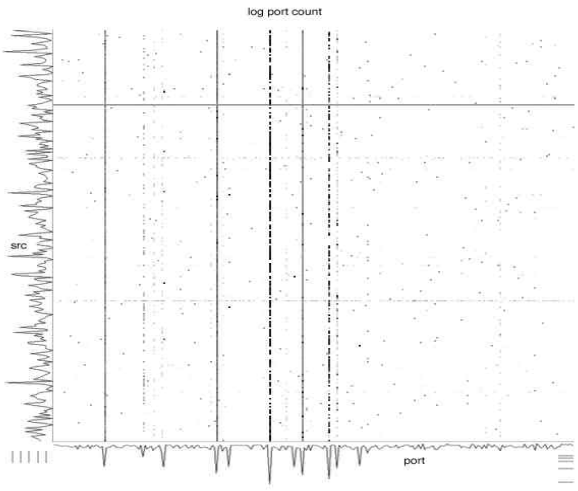

Figure 2: Source IP/Destination Port Map

These images presently represent data over an entire day, but we hope to explore a rapidly updated fading memory mechanism, so that the image at any given time represents a situational picture of what has happened in the very recent past (fading memory time would be a tunable parameter). We will also evaluate frequent snapshots that difference the most recent view from the past, expecting to remove steady state features and highlight novel events that are more likely to be interesting.

The images can be preprocessed by standard techniques such as removal of moving averages, whitening according to an estimated power spectrum, and matched filtering. In particular, worm propagation looks like a line at the source address, while distributed attacks look like a line at the destination address (Figure 2). The degree to which a matched filter rings on these features is statistical, and can be tuned to achieve a constant false alarm rate (CFAR detector). Anomaly detection can examine the image for novel patterns, which may result in the addition of matched filters.

3. Related Work

Staniford and his collaborators [7, 8] have done important foundational work in the theoretical dynamics of propagating phenomena such as worms. Their results raise serious concerns for the Internet community in demonstrating that worm episodes can saturate the vulnerable population in a very short interval of time, an interval of time that is far shorter than the advisory/detect/patch response cycle of today.

The work of Yegneswaran, et al. [10] is an important attempt to address the issue of global characterization of attacks of the nature explored here. The statistics they tabulate are mostly descriptive in nature and produced on batch archives of `syslog` data. Like these authors, we seek a statistical characterization of these attacks that is likely to hold globally, but computable in near real time and summarized as one or a few derived metrics. Moreover, our metric, which is based on the entropy of certain fields in the packet stream, contains no confidential information and can thus be shared among the defenders without reservation.

Burnett [1] proposes displaying resource usage counters with the Multi Router Traffic Grapher (MRTG), noting that many attacks manifest as changes in graphic traces of these counters, which the user identifies visually. The distribution of certain count-based features such as the return code from a Web server may be suitable to the methods presented here.

May [4] detects attacks in large-scale networks by exploiting emergent properties that change statistically between normal and attack states. He employs visualization techniques for large networks, although they are different from what we consider here.

4. Summary and Future Directions

It is apparent that propagating phenomena such as Slammer manifest as changes in process entropy of observable quantities in Internet traffic. Specifically, we observe that Slammer manifests as a sudden increase in source IP entropy and a dramatic decrease in destination port entropy. We introduced an efficient iterative algorithm that calculates these quantities with a minimum of calls to the expensive log function, maintains a manageable state space, and produces a metric that can be exchanged between autonomous domains with no loss of confidentiality. This may enable early detection of propagating phenomena. Moreover, the methodology identifies the states implicated in the process change (in this case, source address and destination port), permitting automated response in the form of throttling requests.

Maximum efficacy was obtained by what amounted to ad-hoc noise removal in the form of discarding UDP requests for port 137. We hope to next explore a more rigorous approach to process mean removal. We observe that the iterative calculation for a new observation does not require that we add an integral count to the calculation (indeed, after aging counts are generally not integral). Instead of updating for the new observation with a count of unity, we update after removing the mean of the category count (which will in this case be its historical probability). The mean removal correction will be capped at a value strictly less than unity to allow the algorithm to work even when a category overwhelms the totals (as in the latter stages of the Slammer trace above). The revised algorithm is given in the appendix.

To this point, we have defined an efficient procedure to compute a potentially useful metric. To use this as the basis for a detection algorithm, we still require tuning to determine adaptation parameters and thresholds at which an entropy change should trigger an alert. We will have to characterize what sort of non-malicious anomalies manifest as entropy

changes, in order to understand false positives from the method. We note that such anomalies may be interesting to administrators even if the underlying cause is not malicious.

We intend to explore how other attacks manifest with respect to entropy. The discovery of scans in the spikes of the entropy traces, for example, was unexpected.

We intend to examine other traffic with this technique. Candidate streams include internal-to-external packet headers and perhaps packet content (e. g., to explore the hypothesis that large sequences of null operations in malicious packets result in lower entropy).

It remains to be shown, that this approach is feasible and gives a useful metric at the level of peering points, ISPs, or major electronic commerce sites. For these, it is normal for a very large number of clients to be active at any one time, but even there it is likely that a large number of client sessions will look different from the accesses triggered by malicious propagating phenomena.

Given the requirement for high-speed detection of worms and other rapidly propagating phenomena in large-scale networks, our preliminary approach for scalable visualization seems promising. The visual representations are both compact and computationally efficient. They reveal both vertical and horizontal scans, as well as exceptional activity in particular hosts and ports. However, because the dynamic range in raw hit counts of our exemplar data set varied by as much as 7 orders of magnitude, we "squashed" the counts when converting them into grayscale values. We anticipate that different types of scaling may help accentuate emerging phenomena, while others may help de-emphasize long-term background noise (e.g., remnants of detected but generally remediated worms).

4.1 Large Data Sets

We have acquired data from DShield [3] and hope to explore the ability of these and related techniques to elucidate useful features and scale to large data sets. Our exploratory set consists of approximately nineteen million records, which represents a day of events voluntarily contributed from multiple autonomous domains. The techniques we have presented may prove useful for the timely generation of recommended block lists at organizations such as DShield.

We have also assembled a large data set of connection attempts to unused address space in our domain. These connections are by definition suspicious if not malicious.

4.2 Temporal Imaging

Because of the dynamic nature of emergent attacks, we will investigate generating and displaying time-dependent images, which can be used, for example, to detect long time constant probes, calculate significant rate changes, and perform A/B comparisons between different time slices or between local and global venues.

4.3 Trend Removal

Because of the dynamic nature of emergent attacks, we will investigate generating and displaying time-dependent images, which can be used, for example, to detect long time constant probes, calculate significant rate changes, and perform A/B comparisons between different time slices or between local and global venues.

4.4 Frequency Space methods

We have begun examining images in which source IP is represented vertically and a combination of destination IP and port is represented horizontally. In this representation, a destination IP is a horizontal line segment in which ports appear at a deterministic offset from the origin of the line segment. This construction induces a periodicity in the port index across host indices. Subtle patterns of activity to the same port or sets of ports across a range of addresses may be apparent via frequency space methods such as Fast Fourier Transforms (FFTs).

5. ACKNOWLEDGMENTS

This work was supported ARDA under Air Force Research Laboratory Contract No. F30602-03-C-0234. The views expressed are those of the authors and do not necessarily reflect the views of the supporting agencies.

6. REFERENCES

[1] Burnett, M. "MRTG for Intrusion Detection With IIS6", http://www.securityfocus.com/1721, August 2003.

[2] CE01 CERT, "Code Red II: Another Worm Exploiting Buffer Overflow In IIS Indexing Service DLL", Incident Note IN-2001-09, Aug. 6, 2001. http://www.cert.org/incident_notes/IN-2001-09.html

[3] DShield Distributed Intrusion Detection System, http://www.dshield.org.

[4] May, J., Peterson, J., and Bauman, J. "Attack Detection in Large Networks", Proceedings of the Second DARPA Information Security Conference and Exposition (DISCEX II), Anaheim, CA, June 2001.

[5] Moore, D., Paxson, V., Savage, S., Shannon, Colleen, Staniford, S., and Weaver, N. "The Spread of the Sapphire/Slammer Worm", http://www.cs.berkeley.edu/~nweaver/sapphire, 2003.

[6] Microsoft Knowledge Base Article – 826234, "Virus Alert About the Nachi Worm", http://support.microsoft.com/default.aspx?kbid=826234, August 2003.

[7] Staniford, S, Grim, G., Jonkman, R. "Flash Worms: Thirty Seconds to Infect the Internet", http://www.silicondefense.com/flash/

[8] Staniford, S., Paxson, V., and Weaver, N. "How to Own the Internet in Your Spare Time", Proceedings of the 11th USENIX Security Symposium, 2002.

[9] Valdes, A. and Fong, M. "Scalable, Signature-Free Characterizations of Propagating Internet Phenomena", Fast abstract presented at Dependable Systems and Networks (DSN04), Florence, Italy, July 2004.

[10] Yegneswaran, V., Barford, P., and Ullrich, J. "Internet Intrusions: Global Characteristics and Prevalence", SIGMETRICS03, ACM, 2003.