Visualization for Monitoring Networks

Sean McKenna & Nivedita Viswanath

December 14, 2013

1 Introduction

Networks consist of ever-changing and always-flowing information, and this information is something that network security analysts wish to protect.[1–3] As networks continue to grow in size, the amount of data that gets sent across these networks increases, along with the overall complexity of the data that one can ascertain from network traffic. To further complicate matters, network traffic is chaotic, making it particularly tricky to pick out threats among the noise.[3]

Thus, network monitoring and intrusion detection is hard for a human to do. Nevertheless, humans play a critical role in network defense, due to our useful skill of pattern finding. This is also a reason why visualization can be helpful for this kind of data. Human analysts are just better at detecting new attacks, more so than an automated system. But humans can only analyze so much data, so intrusion detection systems (IDS) are used to automatically filter the network data in order to try and flag and prioritize the network traffic for analysts.

Since it can prove to be difficult to collect data without violating privacy, we chose to explore a dataset provided by the IEEE VAST 2011 Mini-Challenge #2.[4] An overview of this network is provided in Figure 1.

This dataset provides three days of anonymized data from a corporate network. The data consists of firewall logs, IDS logs, a Nessus scan log, and an aggregated file of syslogs for all hosts. With this data, we visualized the network activity and attacks that occurred during this time frame. Additionally, the dataset has a ground truth solution that is provided for validation of our findings.

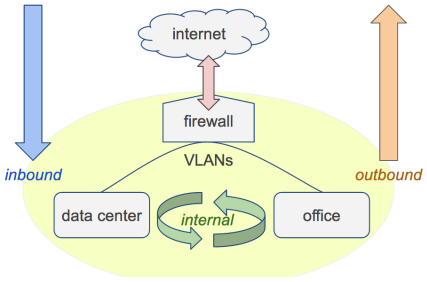

Figure 1: Overview of the network topology of the VAST 2011 challenge dataset.

For our project, we explored several different methods of monitoring, filtering, and visualizing traffic on a network in order to better support human analysts. Our contributions from this project include:

- scripts to pre-process, synchronize, and correlate multiple types of network data
- scripts to flag critical events and events that violate the network terms of service
- visualization of several key events
- validation of findings from the ground truth
- identification of two events not presented in the ground truth solution

1.1 Threat Model

In our threat model, the adversary is capable of carrying out any known attack on the network. The adversary has all permissions and access rights required to conduct these events. However, the adversary is incapable of modifying and/or deleting the logs which record the events. In other words, the adversary does not care if the attack or event is detected.

2 Related Work

Previously, various attempts have been made to visualize netwok traffic patterns, focusing on detection and prevention of malicious activity on the network. The TVi tool[2] combines multiple visualizations of network traces and facilitates customized querying required to make sense of the complex network data. It incorporates machine learning algorithms for anomaly detection. NAVIGATOR (Network Asset VIsualization: Graphs, ATacks, Operational Recommendations)[5] is another tool which uses attack graphs to model the impact of client-side, credential-based and trust-based attacks. The work by W. Lian et. al[6] uses visualization motifs to recognize application protocols in unidentified traffic.

However, little work currently exists for effective visualization of streaming data. One new idea for visualizing this type of data is visual sedimentation, based on the geological concept of sediment.[7] Tokens get represented with tiny graphical elements (circles) and interact physically in some limited area of a graph. These tokens eventually get aggregated into the graph, like in Figure 2 as a bar graph.

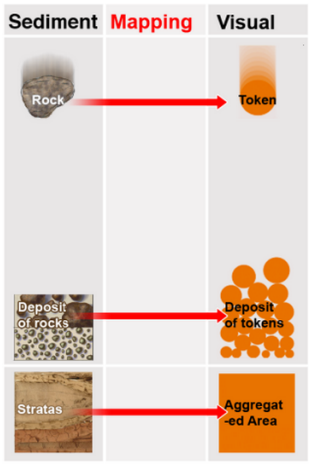

Figure 2: The principle behind visual sedimentation.

3 Methodology

3.1 Data Pre-Processing

Initially, the log files were not synchronized and events were recorded based on different criteria for each log file. Synchronization elimated any false positives that are found to be reported in only one of the logs. Generally, most attacks or suspicious activities tend to be recorded in multiple logs. This synchronization was done using timestamps, IP addresses, and port numbers. The combination of these fields was found to be unique for each event.

Further processing was required to detect any malicious activity that may have been logged in these files. In order to do this, we wrote scripts in Python to detect high priority events such as DoS and port scan attacks in the IDS logs. For the firewall logs, we wrote a script to count the total number of packets that were being sent to any destination IP address from different source IP addresses. This acts as a confirmation for the occurance of any DoS attack. Furthermore, any "Emergency" elerts and "Warnings" that appear in the firewall logs were noted. The Nessus scan logs were checked for the current patch status and any reported system vulnerabilities. The system logs were checked for violation of file access policies, failed logon attempts and any suspicious event.

3.2 Visual Sedimentation

For the firewall log dataset, we used visual sedimentation [7] in order to visualize the packets that are going into, out of, and across the corporate network. We were simply interested in the number of packets to find load-intensive attacks with this technique. To handle the limitations of the size of our dataset (tens of millions of packets), we chose to mimic the idea behind IP traceback and probabilistically select only a subset of packets from the firewall to visualize. [8] This turned out to be about one in every ten-thousand packets to get the visualization to display our subset of packets in the web-based tool. We also tested many different randomly selected datasets and visually inspected them to ensure that all the same patterns were preserved during this probabilistic selection.

4 Analysis

4.1 Visual Analytics

We also used the concept of visual sedimentation for visual analytics of the firewall logs. This highlighted two key events: both a denial of service attack and an internal port scan on the network. These two events dominated the amount of packets getting sent across the network when aggregated from the three days.

The first event was the DoS attack, shown in Figure 3, where there is a significant increase in the number of incoming packets from a few different IP addresses hitting the corporation's web server. This coincided with the attack presented in the ground truth solution for this dataset, and this is a critical event that disrupts an important node on the corporation's network: its web server. It is very likely that its website went down during this attack around noon on the first day, and the visualization easily shows why.

The second attack seen is a port scan from one VLAN to the other, as seen in Figure 4. Interestingly, there are two spikes in this port scan, about when the port scan begins and when it ends. It is as if the port scan tried to adapt to the network and slow down to possibly avoid detection, though that theory would not explain the second peak. This attack also appears in the ground truth solution as an event that is critical, where a worm could be scanning the network for other machines to infect, or an outside attacker trying to find other machines that may be vulnerable.

Lastly, we used this visual analytic tool to analyze every single outbound packet on the network, as also shown in Figure 4. What is easily seen is that our second day logs have a significant increase in the number of these outbound packets, or our baseline in the timeline view. This traffic all occurs on that second day from one node: the network's DNS server. This is likely a significant event that is not listed or explained in the ground truth solution for this dataset. We were unclear why this might be, but it definitely merits further exploration.

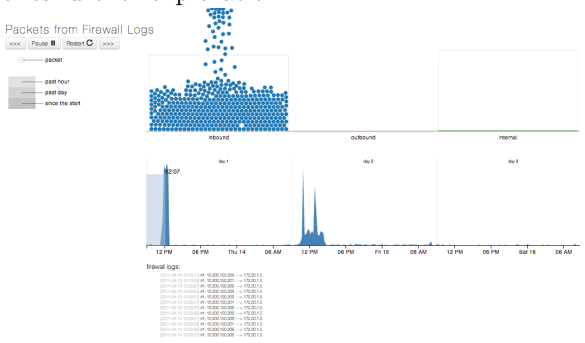

Figure 3: Denial of Service attack seen in a visual sedimentation example.

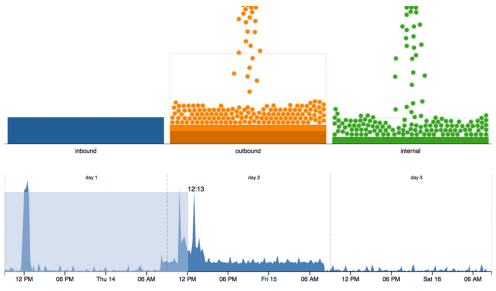

Figure 4: Showing all outbound packets detected by the firewall along with an internal port scan, all occuring on day two.

4.2 Script based Log File Analysis

With the help of the scripts that were written, the DoS attack and the port scan attacks that we found were easily detected as these were listed as high prioriy events in the IDS logs. The next step in the analysis of log files was to check for any vulnerable systems that were discovered in the Nessus scan logs. We found five unpatched systems (192.168.2.171-5) to be security holes that are highly susceptible to attacks. On further examination, we see that these are the very same systems that initiate the port scan attack, which we detected on day two. Clearly, these unpatched systems pose a significant risk to the organization's security.

A second important event that we noticed on the second day was a large number of failed logon attempts. The reason we report this event here is due to to the sheer number of these attempts coupled with the fact that this occurance was seen only on this day, during this period, and no where else. This event was also not clearly listed in the ground truth solution either.

A third noteworthy event was observed in the firewall logs. We detected the presence of an IP (192.168.2.251) that does not belong to any legitimate address pool, as described by the network description. It is unclear what the history of this address is, however, any activity that it engages in is highly suspicious. This is also presented as a risk in the ground truth solution.

Another event worth mentioning that was detected in the firewall logs was a remote-desktop connection. The company policy prohibits remote desktop connection from outside the network. This event is flagged as an "Emergency" alert in the firewall logs and appeared in our scans. It may or may not be actual malicious activity, but this event was also flagged in the ground truth solution of the contest data.

5 Conclusion

Using the VAST 2011 contest data, we were able to successfully identify several key attacks and events that were also listed in the ground truth solution of this dataset. We were able to easily find and identify these using a combination of both scripting methods and visual analytics to solidify what events were occurring on the network over time. We also were able to identify unidentified potential attacks or events on the network that were not present in the network (increase in outbound packets and the failed logon attempts, both on day two). These attacks warrant futher exploration to better understand and explain their threat and impact.

Additionally, our method for probabilistic filtering of packets for visualization was simple, at best. It worked, but there are better ways, such as a statistical analysis of packets using similarity, principal component analysis, and a variety of machine learning algorithms. This would provide better feature selection and support certain kinds of attacks to be visually analyzed. Additionally, we only pursued packets sent across our firewall. Looking at all the network traffic would have enabled seeing any kind of attack on the network, rather than just ones across VLAN's.

References

[1] Y. Livnat, J. Agutter, S. Moon, and S. Foresti, "Visual correlation for situational awareness," in *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*, pp. 95–102, IEEE, 2005.

[2] A. Boschetti, L. Salgarelli, C. Muelder, and K.-L. Ma, "Tvi: a visual querying system for network monitoring and anomaly detection," in *Proceedings of the 8th International Symposium on Visualization for Cyber Security*, p. 1, ACM, 2011.

[3] R. F. Erbacher, D. A. Frincke, P. C. Wong, S. Moody, and G. Fink, "A multi-phase network situational awareness cognitive task analysis," *Information Visualization*, vol. 9, no. 3, pp. 204–219, 2010.

[4] SEMVAST, "Computer networking operations at all freight corporation." `http://hcil.cs.umd.edu/localphp/hcil/vast/archive/task.php?ts_id=152`, 2011. [Online; accessed 12-Dec-2013].

[5] R. L. S. W. S. B. Matthew Chu, Kyle Ingols, "Visualizing attack graphs, reachability, and trust relationships with navigator," in *Proceedings of the 7th International Symposium on Visualization for Cyber Security*, p. 1, ACM, 2010.

[6] F. M. Wilson Lian and J. McHugh, "Traffic classification using visual motifs: an empirical evaluation," in *Proceedings of the 7h International Symposium on Visualization for Cyber Security*, p. 1, ACM, 2010.

[7] S. Huron, R. Vuillemot, and J.-D. Fekete, "Visual sedimentation," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 19, no. 12, pp. 2446–2455, 2013.

[8] S. Savage, D. Wetherall, A. Karlin, and T. Anderson, "Network support for ip traceback," *Networking, IEEE/ACM Transactions on*, vol. 9, no. 3, pp. 226–237, 2001.