Visualization for Security: End of the road or a new beginning?

Josh Hyman
Univ. of California
Los Angeles, CA 90095
josh@cs.ucla.edu

Mohit Lad
Univ. of California
Los Angeles, CA 90095
mohit@cs.ucla.edu

Lixia Zhang
Univ. of California
Los Angeles, CA 90095
lixia@cs.ucla.edu

Abstract

In this position paper, we take an objective look at the usefulness of visualization for the purpose of security and study the current state of the art with a view to identify innovative directions for the future. Based on our study, we find that there is a lot of similarity among different efforts in visualization. We argue that data abstraction and correlating different data sets could open up new frontiers and foster innovation in security visualization.

1 The need for visualization

The traditional role of visualization has been to present data so as to increase human cognition and enable users to identify events of interest. Over the past two decades, various visualization techniques have been proposed to help visual identification of anomalies and security threats. Security breaches can be at various levels: within a host machine, inside a network, or between different networks, and at each level, new techniques are being discovered every day to compromise systems. These security threats do not follow any particular pattern. Further, there is no exhaustive means of enumerating all possible attacks that could occur against a given network. Identifying attacks, and separating the anomalous from the normal requires human intelligence, and this human intelligence is very difficult to capture.

In order to leverage the human ability to reason about the network and identify traffic patterns which may represent an attack, we need visualization. The data sets that administrators and operators have to sift through are increasing in size by the day. In addition to summarizing data, visualization offers new ways of exploring the data and understanding relations within the data. In addition to being able to identify new attacks, humans may also be able to reason about the cause behind a given event. For example, a network prefix being announced to originate from more than on autonomous system (AS) can either be malicious, or the result of a mis-configuration. Only a human, through his knowledge of the usual operations of the other autonomous systems, would be able to decide whether this seeming anomalous event is innocuous or not. Again, clear visualization would greatly help the administrator sift through such instances quickly.

The ability of visualization to summarize large data sets, and to involve the user, make it a necessary component for the development of solutions in monitoring network security. In order to understand the state of the art, we conducted an exhaustive study of research efforts in security visualization. While our sentiment is that visualization is effective, we found a lot of similarity between various efforts. In the remainder of the paper we investigate why this similarity occurs, and propose potential directions to promote innovative research in visualization for security.

2 Adaptation or Innovation?

Most prior work in security visualization uses one or more of the three dominant techniques: parallel axis plots, scatter plots, and graph representation. Though the data being plotted using these strategies may differ between applications, almost all applications choose one or more of these three strategies to display the data they collect. Much of the recent literature provides some incremental improvements on the three basic plotting techniques, or shows how a new data set that is inherently multi-dimensional in nature, can be visualized using one of the above techniques. For example, applications visualizing traffic patterns almost always deal with a fixed number of dimensions, i.e. source IP, source port, desti-

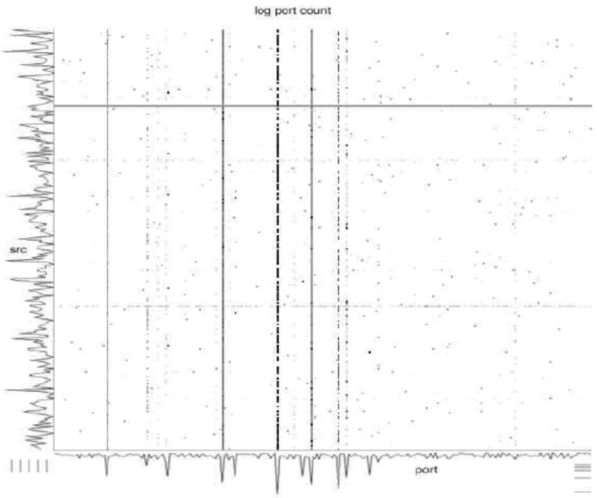


Figure 1: Scatter plot from [1]. This plot has destination port on the x-axis and source IP on the y-axis. Thus, in this plot vertical lines represent large amount of inbound traffic on a particular port, perhaps representative of a worm outbreak (a Kuang 2 outbreak in this case). Histograms along the axes represent quantity of traffic.

nation IP, and destination port. As an example, a straight line along the destination IP dimension would indicate a possible host scan. Such host scans and port scans are thus easy to identify from visual plots. We examine why most applications seem to use very similar techniques and present two potential directions for innovations in visualization for security.

The similarity in flavor of visualization stems from the fact that most of these techniques visualize raw data streams (e.g. netflow data). As a result, visualization of the raw data inherently boils down to visualization of the multiple dimensions in the data. Figure 1 is a hybrid scatter plot of raw network data, plotting not only source IP and destination port, but also a histogram of traffic quantity for each. In this plot a vertical lines represent traffic from many hosts converging on a few specific ports on a single host.

One way to foster innovation in security visualization is to visualize some abstraction of the data rather than the raw data itself. In order to be able to abstract the data, one would need to understand entities and their relations in the data. This type of understanding can be gained by using the current visualization techniques. Thus visualizing the abstracted data set might let us to ask more relevant questions about the relations within the data, possibly resulting in new ways of visualizion. As an example, raw Internet routing data consists of routing update messages indicating the observer, the destination, and the route change from observer to destination. In its raw form, this data is visualized by [4] helping the user interact with the data. On the other hand, abstracted visualization of this data set is carried out by [3]. Specifically, this application produces a derived abstract data set consisting of the number of routes traversing a given link and tracks how the number of routes per link changes. Figure 2 presents an example of raw versus abstract data visualization.

Another possible direction for innovative visualization techniques is to work with composite data sets. There have been a few attempts, specifically [2], to combine two separate data sets together to produce a higher level, and possibly more interesting, data set to visualize. By combining more than one data set, visualization techniques can take advantage of the correlation between the data to add more interesting dimensions. Further, having these extra dimensions allows users to ask more complex and subtle questions about the data. Such questions will possibly lead to new visualization techniques to better display their answers.

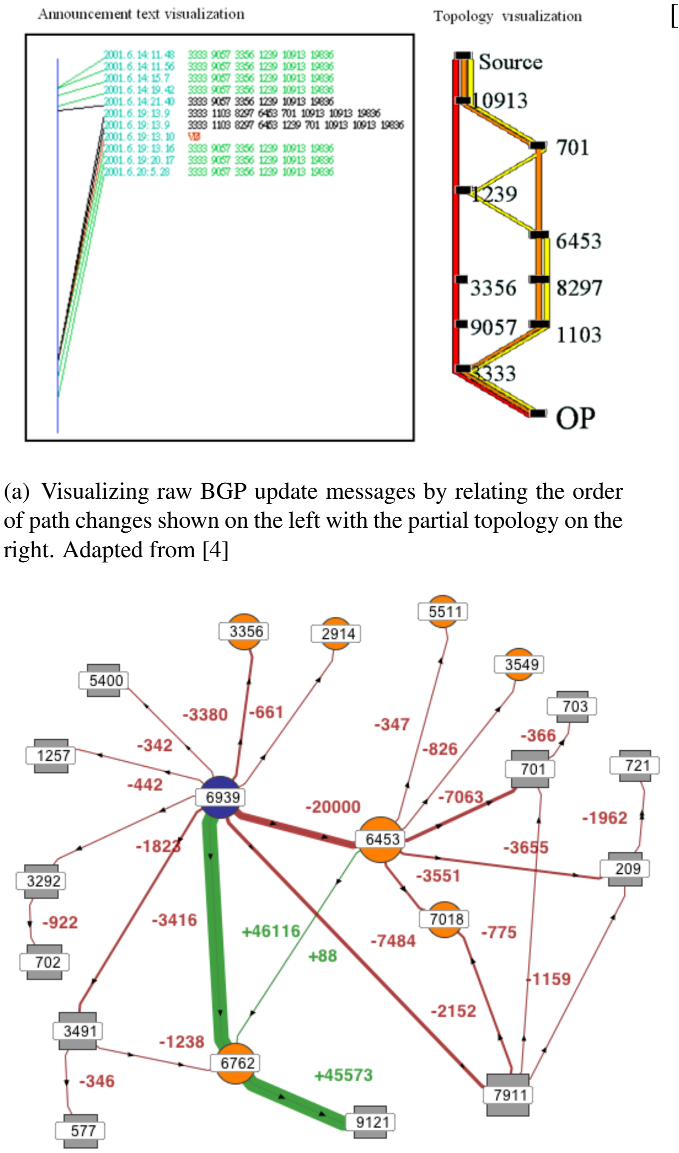
3 Conclusion

Visualization techniques have evolved over the past two decades. With the ever increasing size and complexity of the data to be visualized and the need for human intelligence to evaluate the data, the importance of visualization in security will only increase in the coming decade. We believe looking for new ways to abstract the raw data and answering specific questions from this abstracted data can result in new directions for visualization. We also feel there is a lot of potential for visualizing composite data sets and relating events across different data sets.

References

[1] Alfonso Valdes and Martin Fong. Scalable Visualization of Propagating Internet Phenomena. In *Workshop on visualization and data mining for computer security*, 2004.

[2] Glenn A. Fink, Paul Muessig, and Chris North. Visual Correlation of Host Processes and Traffic. In



```
2001.6.14:11.48   3333 9057 3356 1239 10913 19836
2001.6.14:11.56   3333 9057 3356 1239 10913 19836
2001.6.14:15.7    3333 9057 3356 1239 10913 19836
2001.6.14:19.42   3333 9057 3356 1239 10913 19836
2001.6.14:21.40   3333 9057 3356 1239 10913 19836
2001.6.19:13.9    3333 1103 8297 6453 701 10913 10913 19836
2001.6.19:13.9    3333 1103 8297 6453 1239 701 10913 10913 19836
2001.6.19:13.10   [illegible]
2001.6.19:13.16   3333 9057 3356 1239 10913 19836
2001.6.19:20.17   3333 9057 3356 1239 10913 19836
2001.6.20:5.28    3333 9057 3356 1239 10913 19836
```

(a) Visualizing raw BGP update messages by relating the order of path changes shown on the left with the partial topology on the right. Adapted from [4]

(b) Large scale route hijacking caught by Link-Rank on abstracted data. About 46,000 routes moved from various destinations (red paths) to the hijacker, 9121 (green path) as seen by observation point 6939 (blue node). Adapted from [3]

Figure 2: Examples of raw data and abstracted data visualization

Workshop on visualization and data mining for computer security, 2005.

[3] M. Lad, D. Massey, and L. Zhang. Visualizing internet routing changes. In *IEEE Transactions on visualization and Computer Graphics, special issue on visual analytics*, to appear, 2006.

[4] S. T. Teoh, K.-L. Ma, and S. F. Wu. A Visual Exploration Process for the Analysis of Internet Routing Data. In *Proceedings of IEEE Visualization*, 2003.