FloVis: Leveraging Visualization to Protect Sensitive Network Infrastructure

DRAFT

J. Glanfield, D. Paterson, C. Smith, T. Taylor, S. Brooks, C. Gates[†], and J. McHugh[‡]
Dalhousie University, Halifax, NS, Canada; CA Labs[†], and University of North Carolina[‡], Chapel Hill, NC

Computer networks have become critical to NATO operations. Much of NATO's computer traffic runs over civilian networks, and NATO computers are accessible to a wide variety of malicious activities. The scale of the network traffic involved makes monitoring and analysis difficult, and the rapid deployment of computer systems to new areas places additional stresses on operators and analysts. We have developed an extensible suite of visualization tools, FloVis, to aid system administrators and system security officers in understanding the traffic that passes over their networks. The suite is useful for both defensive purposes as well as for evaluating and understanding the effects of offensive information operations. This paper describes FloVis and provides examples of its capabilities.

FloVis is a visualization framework that was built with the aim of providing the necessary machinery to allow security analysts to leverage the benefits of data visualization while attempting to detect malicious network behavior (Taylor *et al.*, 2009). This is accomplished not only by providing new and interesting visualizations, but by allowing these visualizations to synergize their unique perspectives to provide further insight into network data. FloVis was developed to promote:

1. *Extensibility*: The integration of additional visualizations is seamless.
2. *Inter-visualization communication*: Visualizations may communicate without prior knowledge of each other's existence.

In its current state, FloVis consists of a supporting framework and the following plug-ins:

OverFlow. This visualization provides a high-level overview of network usage by organizations or enterprises that may be hierarchical in nature (see Figures 1(a) and (b)). It focuses on the administrative relationships rather than being network centric as are the other FloVis components. It is purposefully high-level in order to provide motivation for more-detailed analysis of network entities (*e.g.*, hosts or subnets) with more detailed visualizations (see Glanfield *et al.,* 2009). In using Overflow, the analyst defines the organizational structure and incorporates the address ranges assigned to each component. The visualization displays inter-organizational traffic patterns and volumes. Unexpected or unusual communication patterns can lead to further investigation of the specific network entities involved.

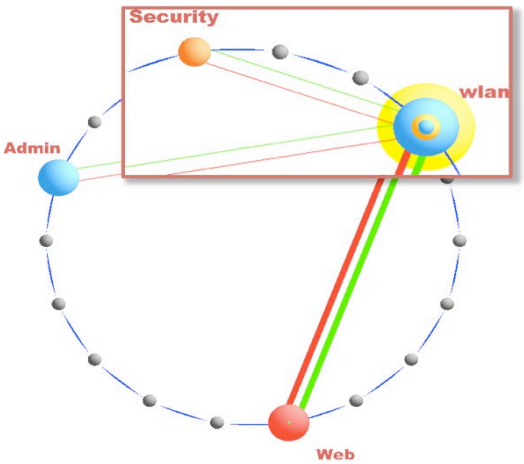

Figure 1(a). An organizational overview of a network.

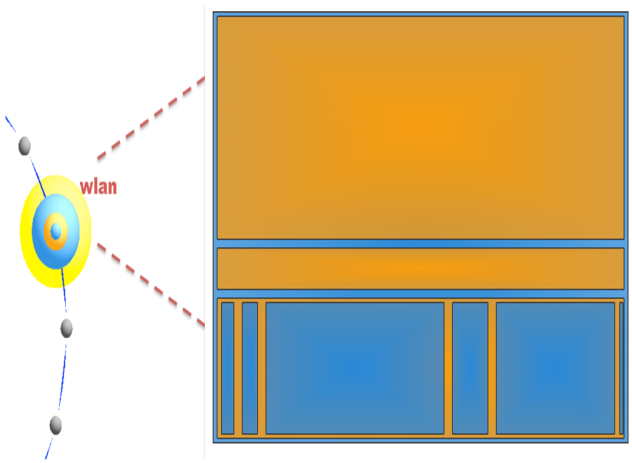

Figure 1(b). A secondary visualization in OverFlow of an organization.

*FlowBundle***.** Displays communications between network entities and reduces occlusion by using hierarchical edge-bundling (Holten, 2006). The display arranges 512 entities around a circle that is divided into two sections by the border across which communication was observed (see Figure 2). Typically, the entities are subnet or host addresses with 256 points allocated to each side of the border, but other relationships including host port usage can be displayed. The limitation to 512 points allows unambiguous identification of individual connections, however a sliding window allows any consecutive 8 bits of the entity ID (*e.g.*, IP address) to be selected. For example, if the inside network is a single /24, it is possible to view connections between outside /8s and inside hosts. By sliding the outside window, connections from /16s within a given /8 or the /24s within a given /16, etc., can be displayed. Connection line transparency is a rough indicator of traffic volume. Given that the OverFlow plug-in has identified questionable inter-organizational traffic, FlowBundle could be used to identify the subnets or hosts involved.

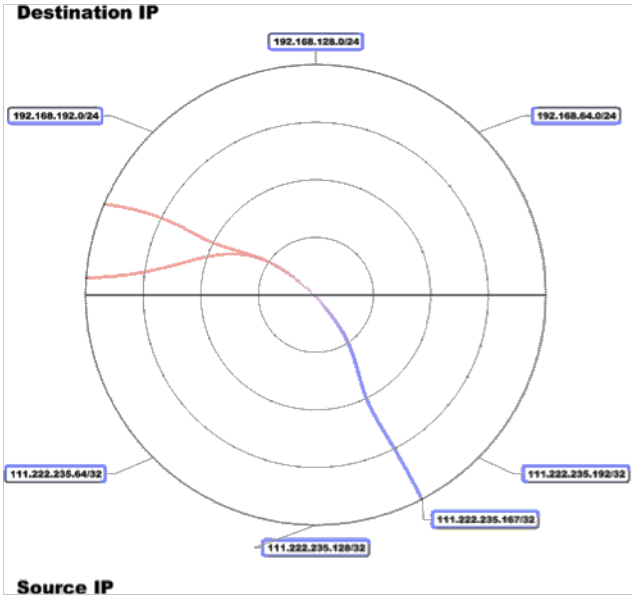

Figure 2. Our host-of-interest communicating with two subnets.

NetBytes Viewer. This plug-in allows detailed analysis of host behaviors over time. It displays an impulse plot in three dimensions that describes port or protocol volumes over time (Taylor *et al.*, 2008). To avoid the occlusion problems that often accompany static 3D plots, the NetBytes Viewer plot can be rotated and rescaled by the user. 2D finder lines allow precise identification of specific impulses in the time / volume and port, protocol / volume planes (see Figures 3 and 6). The viewer is particularly useful in examining the behavior of a compromised machine since the behavior of the machine prior to and after compromise can easily be compared. In addition, unexpected behavior changes associated with a compromise can be detected. These might include bot behaviors or other malicious activity.

Activity Viewer. This visualization shows categorical entity activity as a function of time, using distinct colors to describe a limited number of categories (Taylor *et al.*, 2009). The choice of categories is arbitrary. Any small set of behaviors that can be derived from the available data is suitable. One example uses client / server behavior, another shows hosts' responses to scans. The categories of individual entities are plotted against time in a simple two-dimensional grid, with the entities listed along the vertical axis and time along the horizontal axis (see Figure 4). If a given entity exhibits one or more of the categorical activities during a given hour, the corresponding square is given the color of the activity that causes the most concern. In an operational setting, the categories might correspond to the roles assigned to individual hosts. Hosts behaving in manners consistent with their assigned role would be given colors that identify the role and indicate

normal activity. Hosts that appear to deviate from the role would be given colors that indicate the nature and extent of the deviation. This could be compressed into the common three category "stoplight chart" with green indicating normal, yellow questionable, and red clearly bad. Since some role shifting and deviation from expectation is often observed, the time series of colors allows quick identification of hosts that are deviating from past behavior. The entities need not be hosts. Subnets or organizational units could be used, as well.

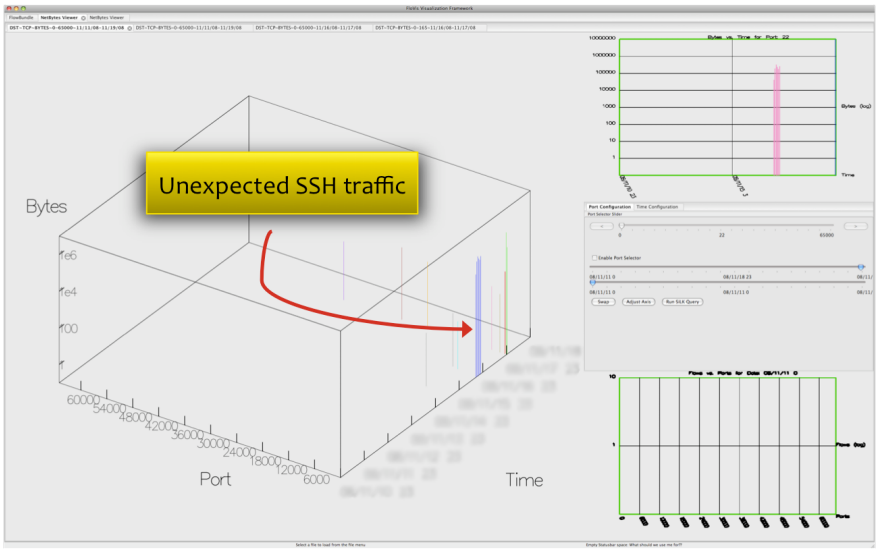

Figure 3. Port 22 traffic occurring over a few hours. The top right of the display shows hundreds of thousands of bytes over a short time period.

The following two plug-ins are experimental and we are in the process of refining them and evaluating their utility for network monitoring and analysis:

FlowBurst. This visualization displays large amounts of hierarchical, numeric network data in a radial space-filling diagram and it allows analysts to quickly obtain an overview of network traffic, visualizing and comparing up to three traffic attributes concurrently. The data can be explored interactively using hyperbolic distortion to shift the center of the hierarchy.

FlowCal. A calendaring visualization (inspired by van Wijk and van Selow, 1999) that displays multiple time series of a quantitative network property, e.g., daily series of port 80 volumes measured hourly. We know that traffic patterns exhibit diurnal patterns, and the basic display allows us to superimpose a large numbers of days of data on a single display, linking each day's line to an accompanying calendar. Clustering algorithms can then be applied to the lines to group them by a variety of similarity measures. When the lines are grouped, each group is assigned a color and the calendar entries for each group are colored accordingly. Once the obvious groupings are accounted for, e.g., weekdays vs. weekend days, smaller groups and singular cases should arise. Among these we hope to find subtle indications of abnormal activities. Since most network activity exhibits fairly strong daily patterns, we suspect that it will be necessary to remove daily patterns before clustering on other time scales is effective.

In our system, the plug-ins are designed to interact through a number of data properties, enabling both drill down to examine smaller groupings of entities in greater detail as well as pivoting that allows data values of interest in one view to be used as the targets of investigation in others. This interaction is illustrated in the following two case studies.

Case #1

This example shows how three of the plug-ins offer unique perspectives of the same data. We drill further into the data via each plug-in and discover undesirable traffic.

Using OverFlow to provide a visual breakdown of important subnets, we found unexpected traffic between two organizations ("Security" and "wlan," see Figure 1(a)) belonging to a conference network.

OverFlow displays a tree-map (Johnson and Shneiderman, 1991) in order to show volume quantities within levels of a hierarchy. Based on this secondary view of an organization (see Figure 1(b)), it was decided to further investigate the host represented by the large orange block in the tree-map since that block corresponds to the largest portion of traffic. Further drill down with the FlowBundle plug-in reveals a single host communicating with two subnets, which belong to the Security organization (see Figure 2), is responsible for all of the suspicious traffic. The network administrators knew beforehand that there was to be no communication between Security and the public network ("wlan") organizations. Hence, the nature of this traffic is of interest.

The specific nature of the traffic can be determined by using the NetBytes plug-in, as it provides specific port and volume information. With NetBytes, we were able to discover that communication occurred over port 22 and that the volume was large enough to suggest that significant data transfer had occurred (see Figure 3).

In spite of the fact that we drill into the data across multiple visualizations, context is retained through FloVis' multi-tabbed and multi-window displays, which allow the multiple views to be open simultaneously.

Case #2

In this example we show how the use of the Activity Viewer allows us to find anomalous behavior by alerting our attention to a change in a host's pattern of behavior. In Figure 4, we see that a host exhibits suspicious server and client activity on the same port. This behavior is of concern because it is unusual for the given host based on the previously observed behavior.

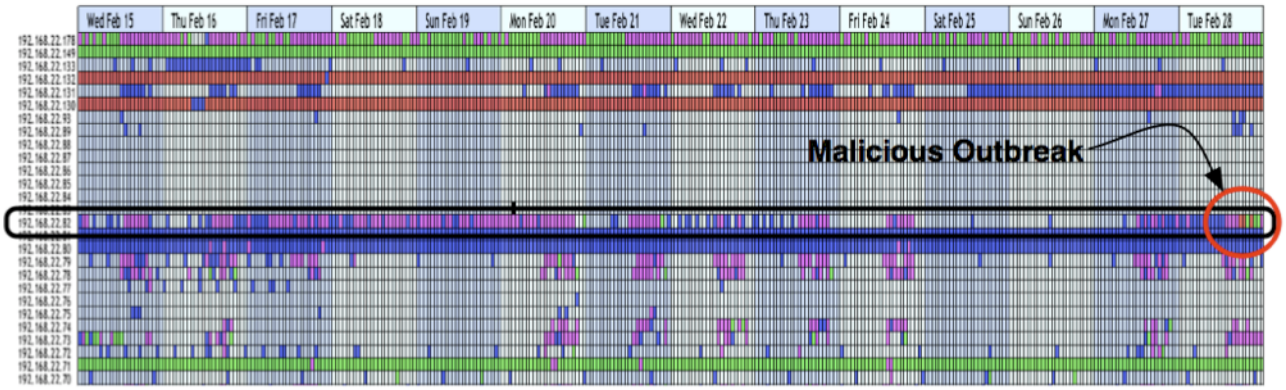

Figure 4. Discovering a change in the pattern of behavior.

Since we are interested in understanding the specific nature of the traffic occurring on our host of interest, we drill down by using the FlowBundle plug-in. Figures 5(a) and 5(b) show us that the host is scanning multiple networks and that it is scanning across a large range of ports, respectively.

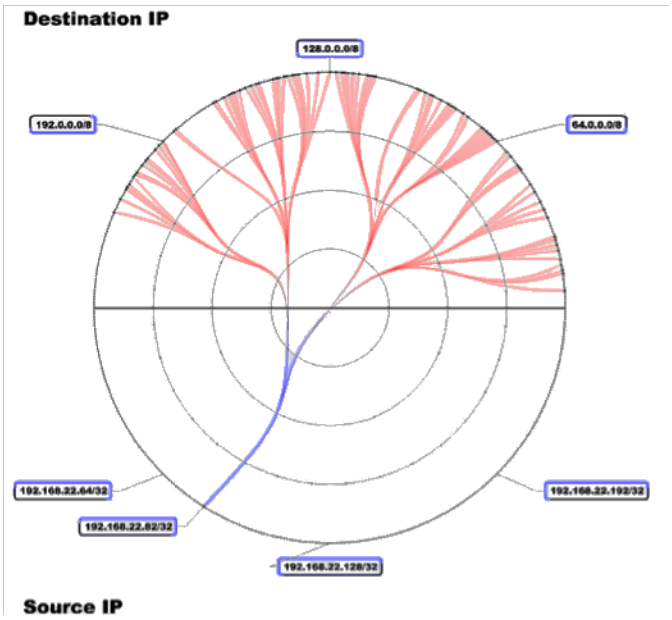

Figure 5(a). Scanning across networks.

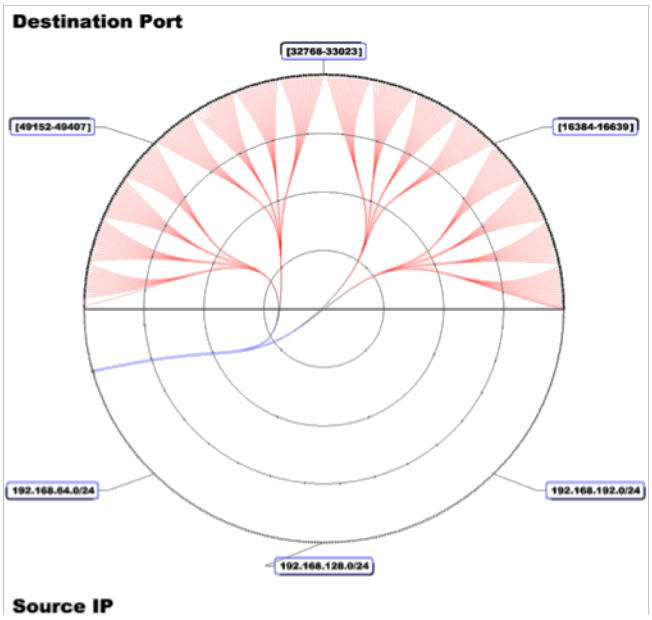

Figure 5(b). Scanning across ports.

The NetBytes plug-in provides yet another perspective by displaying port patterns over time. Thus, we can determine precisely when the scanning activity commenced (see Figure 6).

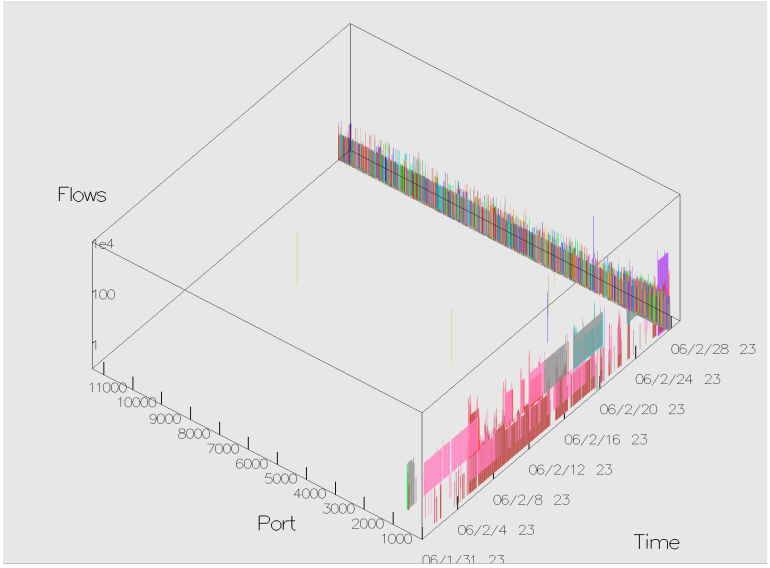

Figure 6. Port-traffic patterns over time.

Comments

We have described how the various components of the FloVis framework work in tandem to allow an analyst to visually drill into network data and explore anomalous network behavior. Should this extended abstract be accepted, our final paper will contain further detailed descriptions of each plug-in and will include additional cases to demonstrate further utility.

This material is based upon work supported by the Department of Homeland Security under Contract No.

N66001-08-C-2032. We also wish to acknowledge the support of Ron McLeod of TARA, CA Labs, and NSERC in this research initiative. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Department of Homeland Security.

References

J. Glanfield, S. Brooks, T. Taylor, D. Paterson, C Smith, C. Gates, J. McHugh. OverFlow: An Overview Visualization for Network Analysis. Accepted to the 6th International Workshop on Visualization for Cyber Security. Atlantic City, NJ. October 11, 2009.

D. Holten. Hierarchical Edge Bundles: Visualization of Adjacency Relations in Hierarchical Data. IEEE Transactions on Visualization and Computer Graphics, 12(5):741–748, 2006.

B. Johnson and B. Shneiderman. Tree-Maps: A Space-Filling Approach to the Visualization of Hierarchical Information Structures. In *VIS '91: Proceedings of the 2nd conference on Visualization*, pp. 284–291, Los Alamitos, CA, USA, 1991. IEEE Computer Society Press.

T. Taylor, S. Brooks and J. McHugh. NetBytes Viewer: An Entity-based NetFlow Visualization Utility for Identifying Intrusive Behavior. In Goodall et al. (eds.), *Mathematics and Visualization (Proceedings of VizSEC)*, Springer-Verlag, August, 2008.

T. Taylor, D. Paterson, J. Glanfield, C. Gates, S. Brooks, J. McHugh. FloVis: Flow Visualization System. In *Proceedings of the Cybersecurity Applications and Technologies Conference for Homeland Security* (CATCH). Washington, DC. March 3-4, 2009.

J. J. van Wijk and E. R. van Selow, E. R. Cluster and Calendar Based Visualization of Time Series Data. In *InfoVis '99: Proceedings of the 1999 IEEE Symposium on Information Visualization,* pp. 4-9, San Francisco, CA, USA, 1999. IEEE Computer Society Press.