Kyle Boyer
Erik Sturcke
CMSC636 - Data Visualization
Dr. Jian Chen, UMBC

Term Project - Literature Review

The practice of visualizing network data has existed as long as there has been network data readily available. Even still, Shiravi brings up that relatively speaking, Security Visualization is a young term. [5, 1]. This contributes to the situation we are currently in. There has been such an explosion of data in such a short amount of time, in the internet age, that we are left with more data than we can process effectively. This applies to data about the network itself as well. As Mansmann mentions, there are far too many incidents and anomalies to try to dissect them individually. We need a better method. [3, 1].

The method that has been proposed is to visualize the data, graphically. This has been done in a multitude of ways, with different effects and filters applied for clarity. For example, one of the more interesting approaches applies heuristic functions to filter the data in real time. [4, 3]. This method emphasizes changes in the data, but discards outliers. This way of encoding the data takes advantage of a technique entitled “edge bundling.” [1, 1]. A geographic approach has also been suggested, but this method can only see large patches of data, rather than individual points. [3, 3].

A large challenge with network data visualizations is the massive overload of incoming data. [4, 1]. The main way that current visualization ideologies handle this is by filtering data, and combining trends. Edges are bundled, groups of anomalies are recorded, and individual data entries are discarded. This can be very beneficial, as it allows a user to focus on the data that ‘matters.’ Unfortunately, this comes with the side effect that lots of data is effectively lost.

Our project focuses on detecting anomalies in network data, similar to the work of Shiravi, Shiravi, and Ghorbani. However, the project that we propose emphasizes outliers. We want to be able to zero-in on outliers in the data, and scrutinize them, rather than discard them. Outliers in network data must result from some event. IE: something caused them to be there. Irregular behaviour is often the most interesting and valuable information in a network dataset. [3, 1]. Currently, there are few systems that handle outliers elegantly, and are able to discern some sort of useful information from their presence. We hope to fill that void with our research and development.

Paper list:

1. Holten, Danny. : Hierarchical Edge Bundles: *Visualization of Adjacency Relations in Hierarchical Data*
2. Livnat, Yarden. Argutter, Jim. Moon, Shaun. Foresti Stefano. : *Visual Correlation for Situational Awareness*
3. Mansmann, Florian. Keim, Daniel. North, Stephen. Rexroad, Brian. Sheleheda, Daniel. : *Visual Analysis of Network Traffic for Resource Planning, Interactive Monitoring, and Interpretation of Security Threats*
4. Shiravi, Hadi. Shiravi, Ali. Ghorbani, Ali. *IDS Alert Visualization and Monitoring through Heuristic Host Selection*
5. Shiravi, Hadi. Shiravi, Ali. Ghorbani, Ali. *A Survey of Visualization Systems for Network Security*