Interactive Visual Classification and Analysis on Network Activity

Siming Chen[1] Xiaoru Yuan [1,2*]

1) Key Laboratory of Machine Perception (Ministry of Education), and School of EECS, Peking University
2) Beijing Engineering Technology Research Center of Virtual Simulation and Visualization, Peking University

ABSTRACT

Monitoring the behavior of hosts and identifying anomaly situation in the streaming network is critical but challenging. There lacks of efficient methods to quickly identify and classified the different behavior for IPs and ports in a dynamic scenario. In this work, we propose a visual analytics approach for quickly identifying anomaly situation and tracking the behavior of interested IP/ports from the streaming network flow data. We build up an interactive visual classification and analysis system, providing filtering and sorting methods, as well as correlation exploration. Features can be classified through interactive brushing and be monitored in other analysis stage. Our case study turns our method can efficiently identify anomaly events from the complex global network flow data.
Keywords: Network security, Classification, Netflow visualization

1 INTRODUCTION

Network security is important for personnel, company or even country. How to monitor the network and gain critical information from the large and complex network flow data is critical. Currently the IDS (Intrusion Detection System), IPS (Intrusion Protection System) can generate alert, but the false alert rate is high. It is challenging to identify and correlate abnormal events pattern from the overall network situation, which should be tightly involved with people. What's more, the real situation always requires the dynamic and correlation analysis for the streaming network data, which is challenging to understand the temporal dimension in the network flow.

In this work, we proposed an interactive system for visual classification and analysis on network activity. It provides dynamic visualization for streaming network flow data, as well as correlation and exploration methods to gain situation awareness in network security. With multiple sorting and filtering methods provided, users are able to identify abnormal IP behavior or aggregation network behavior. Users can classify different type of network events through brushing, and track them dynamically by adding them to the observation list.

Sorts of security visual analytics system are aiming at detecting anomalous activity and discovering trends and patterns [2]. There are works visualizing the network flow based on parallel coordinates [3]. However visual clutter can't be avoided by these methods. Bunch et. al [1] introduced the Netflow observatory. This work provides good overview about the dynamic scene. To further engage users, our methods provide users with a platform to filter, rank and correlate the data. Interactive brushing helps classifying data and enables monitoring the IP/ports of interest in any stage, which improve the situation awareness of the network.

2 SYSTEM WORKFLOW

The data we deal with is the netflow records, including source IP, destination IP, source port, destination port, flag and other attributes. User starts exploration from the timeline (Figure 1-top). By selecting a range of time, dynamic connection view (Figure 1-right) would show the overall connection in the network in this time range. Each dots represent one network connection (netflow record). It shows the connections from source IP (top) to destination IP (bottom), which can also be updated dynamically along the time. Users are able to select/filter/sort IPs to view the details of the connection, including the connected IP and port usage. Also users can trace the connections, starting from the selected IP. By assigning customized color for each selection, users can monitor it in the overview (Figure 2).

*e-mail: {csm, xiaoru.yuan}@pku.edu.cn

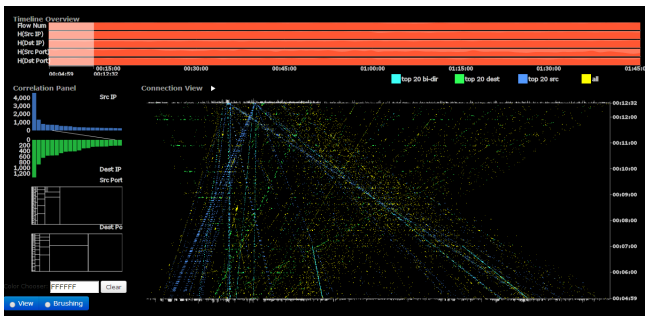

Figure 1: System overview, including the timeline overview (top), dynamic connection view (right-bottom), correlation view (left-bottom).

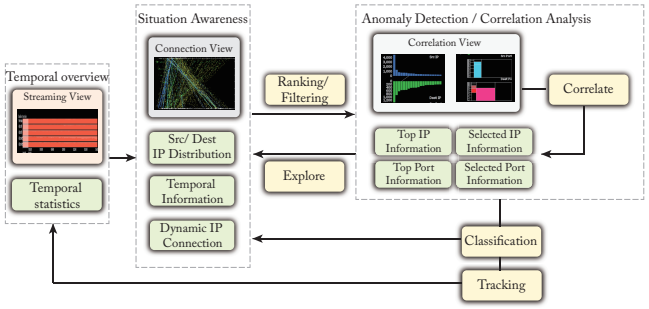

Figure 2: System pipeline, supporting correlated analysis and classification for network situation awareness.

3 SYSTEM DESCRIPTION

Our system is composed of three parts, including the overview timeline, dynamic connection view and the correlation view (Figure 1).

3.1 Timeline Overview

In the timeline overview, we provide four horizon graphs for time selection. We use Shannon entropy for the entries of anomaly detection, which measures diversity of features over time. Entropy measures the distribution's degree of dispersal or concentration of features. In the timeline overview, we have four entropy histograms corresponding to source/destination IP and Port. Entropy measures the distribution's degree of dispersal or concentration of features. It is used as a starting entry for selecting a time range. Timeline view

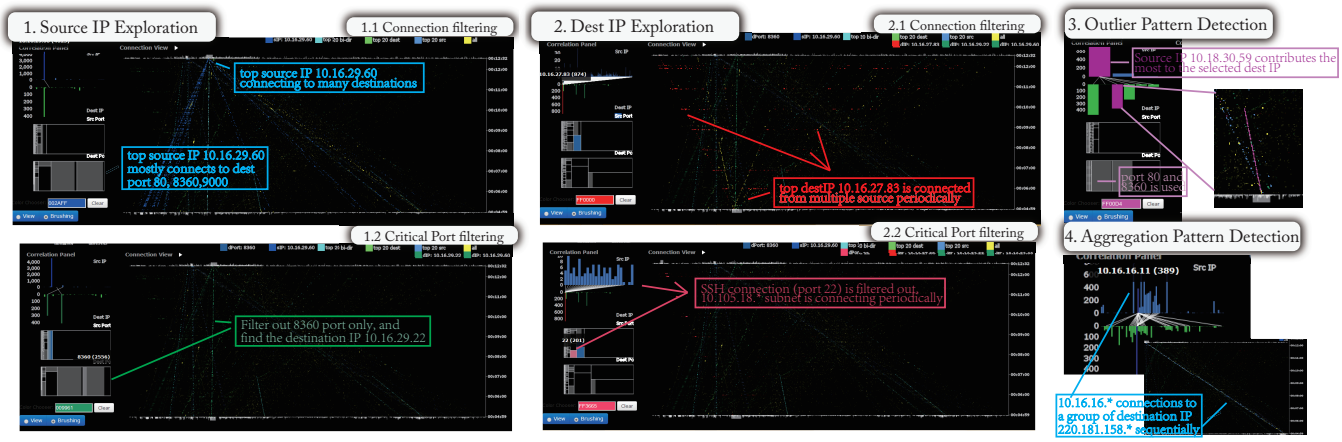

Figure 3: Exploration example of the system. Sorting and filtering are supported to identify IPs and ports of interest (1,2). Correlated analysis further investigate the behavior of related IP and ports (3,4).

also support streaming fetching the netflow data. Time range selection is supported to show all the detailed connection information in the Dynamic Connection View.

3.2 Dynamic Connection View

Dynamic Connection View shows how network flow come from the source IP to destination IP. The Y axis is time axis while the top layer is the source IP and the bottom layer is the destination IP. Each connection is a pixel, position of which is the interpolation of source IP and destination IP. Histograms of source and destination IPs indicate the overall connection number. Color can be customized in the later stage, but initially it highlighted the top-20 connections in the network. Users can clearly view the animation of the connection visualization.

These designs consider analyzing the connection IP and time at the same time for the overview. With the consistency of connection, users can understand the overall trends and avoid the clutter by using only dots to replace the links of connection. To aid understanding of the connection, users are able to brush the source IP and destination IP. Connections including the selected IPs would be highlighted, while others would fade up. Through brushing, users are able to filter the IP or interest, which would be dynamically updated in the correlation view.

3.3 Correlation View

The design goal of the Correlation View is to support correlation and classification through interactive brushing. The Correlation View has two histograms of the source and destination IP, with two treemaps of the source and destination port. These views provide distribution information for the netflow records from different perspective. Users can select IP/port, or compare two features by clicking and hovering. When one feature (IP or port) is selected, other IPs/ports which are connected to/from it would be updated.

Users can assign customized color to classify different connection types and anomaly events based on the interactive exploration. Classification can be applied combining with users' domain knowledge. Selected IP/ port can be tracked when the data is dynamic streaming. Users can better understand the overall trend based on the original classification, and further find outlier or improve the classification.

4 Case Study

This case illustrates the usage of the system (Figure 3). Initially, we started from the overview of a selected 10 minutes time range, with around 10,000 netflow records (Figure 1). Firstly, we can find source and destination IPs with largest connection amounts, indicated by the blue and green color (Figure 3-1). With the treemap view, we found port 80, 8360 and 9000 are correlated with these IPs. After classifying the dominating connections, we filtered the top destination IPs and found a periodical connection to a large group of source IP, which may indicate a DoS attack or other aggregating behaviors (Figure 3-2). We can brush it as red to indicate the finding. Besides IP filtering, we also filtered the top ports from the treemap view and found a large group of SSH behavior (port 22, pink) of 10.105.18.*.

After finding IP and ports of interest with ranking and linking methods, we can further explore the outlier pattern with correlation function. In the correlation panel, by clicking the most connected IP and ports, we found IP 10.18.30.59 using port 80 and 8360, connected to the top-connection IPs we just found (Figure 3-3). Also more aggregation patterns can be classified through iterative exploration (Figure 3-4). Based on these color assignment, we can observe these IPs/ports' behavior in the later time stage.

5 Conclusion and Future Work

We provide an interactive visual classification and analysis system on network activity, enabling identifying the network anomaly events from dynamic network. With global understanding of the network, users can explore patterns and anomaly through suites of interactions such as filtering, sorting and correlation. Customized color can be assigned in any interaction for further comparison, analysis and feature tracking in the dynamic network.

For the future work, firstly, current testing scenario is for around 1 hour range for analysis due to the large connection amounts, we need to consider the scalability issue when time range increase. For the color assignment, currently we set the color assignment in a FIFS (First In First Serve) manner. But multi-attribute assignment problem might be an interesting direction for the future.

Acknowledgements

The authors wish to thank reviewers. This work is supported by NSFC No. 61170204.

References

[1] L. Bunch, J. M. Bradshaw, and M. Vignati. The netflow observatory: An interactive 3-d event visualization. In *Proceedings of the Tenth Workshop on Visualization for Cyber Security*, 2013.

[2] A. Shiravi, H. Shiravi, and A. A. Ghorbani. A survey of visualization systems for network security. *IEEE Transactions on Visualization and Computer Graphics*, 18(8):1313–1329, 2012.

[3] S. T. Teoh, K. liu Ma, S. F. Wu, and T. J. Jankun-kelly. Detecting flaws and intruders with visual data analysis. *IEEE Computer Graphics and Applications*, 24:27–35, 2004.