Interactive Visual Classification and Analysis on Network Activity

Siming Chen Xiaoru Yuan

Key Laboratory of Machine Perception (Minister of Education), and school of EECS, Peking University, Bejing, China

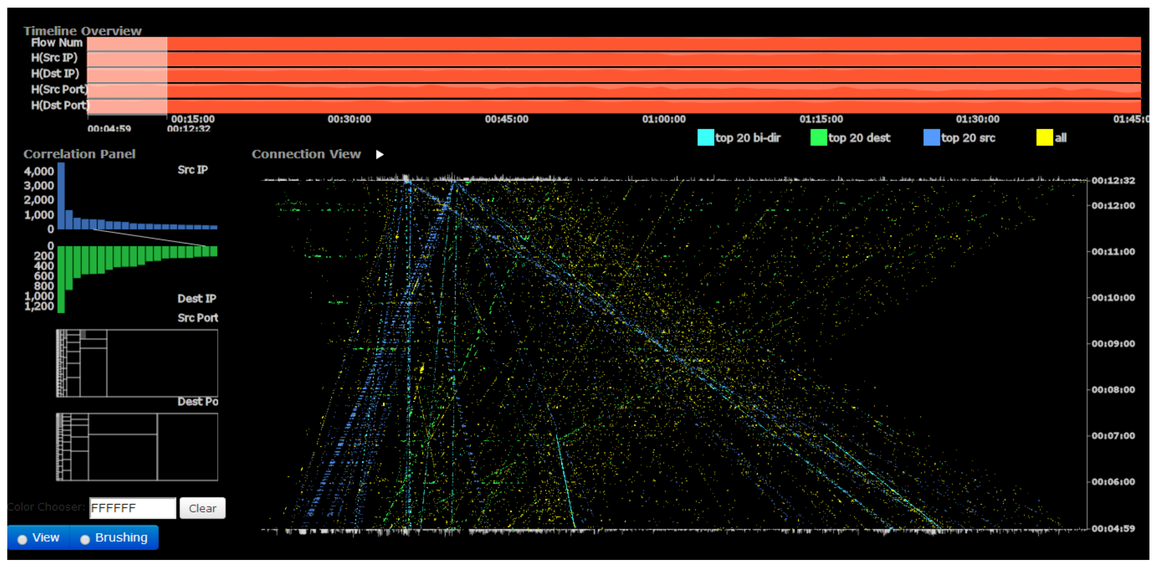
Introduction

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.

Challenges

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.

System Interface



Timeline Overview

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.

Dynamic Correlation View

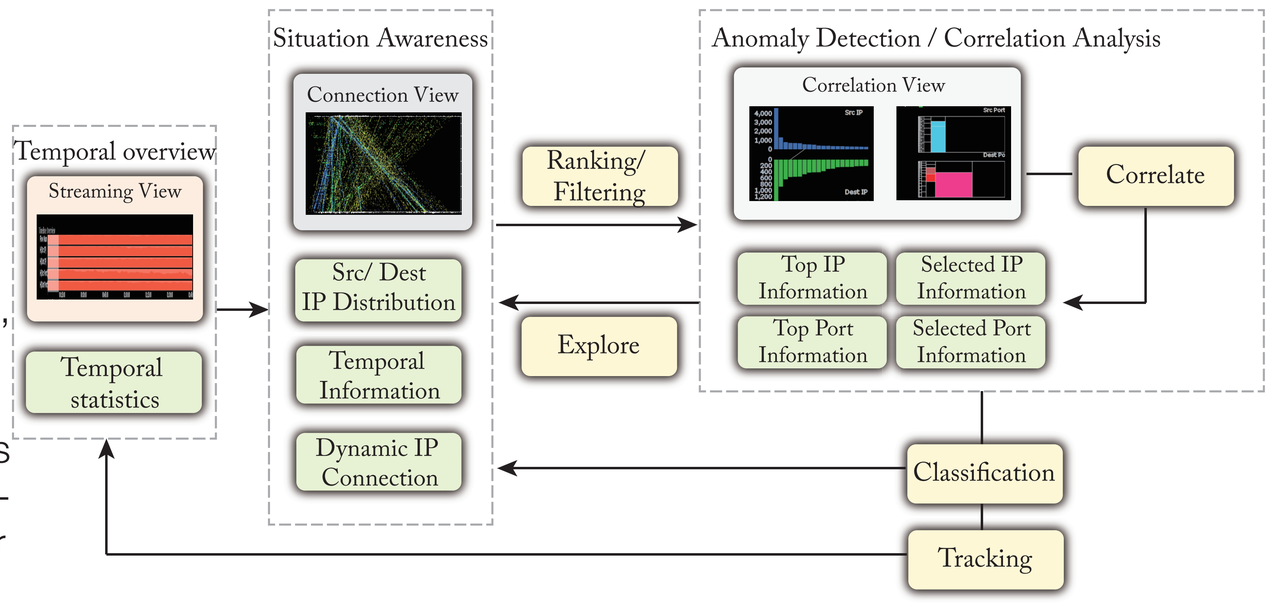
Dynamic Connection View shows how network flow come from the source IP to destination IP. The Y axis is time axis while the top layer is the source IP and the bottom layer is the destination IP.

Correlation View

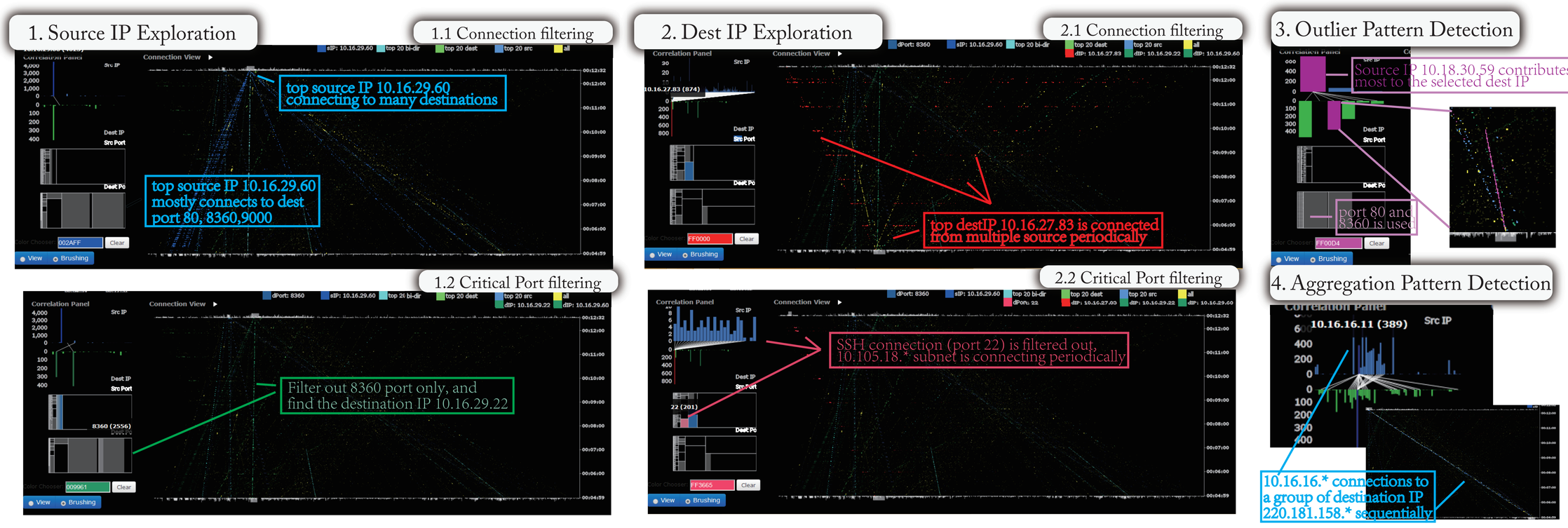
The Correlation View has two histograms of the source and destination IP, with two treemaps of the source and destination port. These provide distribution information for the netflow records

Visual Analytics Pipeline

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. By selecting a range of time, dynamic connection view 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.



Case Exploration with Interactive Operation



Step 1: We can find source and destination IPs with largest connection amounts, indicated by the blue and green color (Case 1-1). With the treemap view, we found port 80, 8360 and 9000 are correlated with these IPs.
Step 2: 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 (Case 2-2).
Step 3: 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 found (Case 3). Iterative investigation can be performed for further aggregation methods (Case 4).

Acknowledgements

This work is supported by National NSFC Project (No. 61170204) . The authors want to thank IEEE VAST Challenge 2014 Committe and reviewers.

Contact
xiaoru.yuan@pku.edu.cn http://vis.pku.edu.cn
VisPKU