VACS: Visual Analytics Suite for Cyber Security

Visual Exploration of Cyber Security Datasets (VAST Challenge 2013)

Fabian Fischer*
University of Konstanz
Germany

Daniel A. Keim†
University of Konstanz
Germany

Abstract

Visual exploration of cyber security datasets is an important and highly relevant field of research. To address the cyber security challenge of the VAST Challenge 2013, we utilized our novel *Visual Analytics Suite for Cyber Security (VACS)* to visually explore the given datasets using a combination of different visual representations. *VACS* primarily provides a dashboard view, host-based thumbnail overview and a querying interface to retrieve and drill down to investigate suspicious hosts.

Index Terms: C.2.0 [Computer-Communication Networks]: General—[Security and protection]; I.3.8 [Computer Graphics]: Application—; H.5.2 [Information Interfaces and Presentation]: User Interfaces—

1 Introduction

VACS is a novel visual analytics suite to analyze and visually explore large-scale cyber security datasets. To achieve scalability for large datasets *VACS* makes use of an *ElasticSearch* cluster with five nodes using commodity hardware. *VACS* is a web application using JavaScript, HTML5 and a variety of state-of-the-art toolkits and custom widgets and a mix of interactive client-side visualizations and visual representations generated on the server-side due to performance reasons. In the following, we briefly describe the different elements and describe a basic use case how an analyst can use the system for visual exploration which can lead to a better situational awareness. The system is built on the experiences on past VAST challenges [2] and is reusing techniques developed in the past [1]. While *VACS* is still under further development and ongoing work, we briefly showcase some of the capabilities using the VAST Challenge 2013 as use case.

2 Visual Exploration

The visual analytics system provides a dashboard view and a visual analytics view to identify suspicious host behavior and visually explore the underlying data. Several interactive visualization are integrated to support the user in this analysis process.

Interactive Line Charts are a well-known visual representation for time-series exploration. The analyst can use a dialog to query the different datasets to extract time-series (e.g., network traffic over time, alerts above a threshold, traffic on specific ports, average memory consumption). This simple representation helps to correlate different time-series. The chart is also used to guide the drill-down process to parameterize other visualization with the selected time interval.

Pixel-Based Thumbnails are used as seen on the left part in Figure 1 to visually represent time-series patterns of the different network hosts. The view can be adjusted to show different metrics over time. Several different types are embedded. A list-based example with just a few hosts can be seen in Figure 4. Each colored vertical line represents the value of the time-series. Similar behaving hosts with similar activity patterns can then be spotted and selected for further analysis.

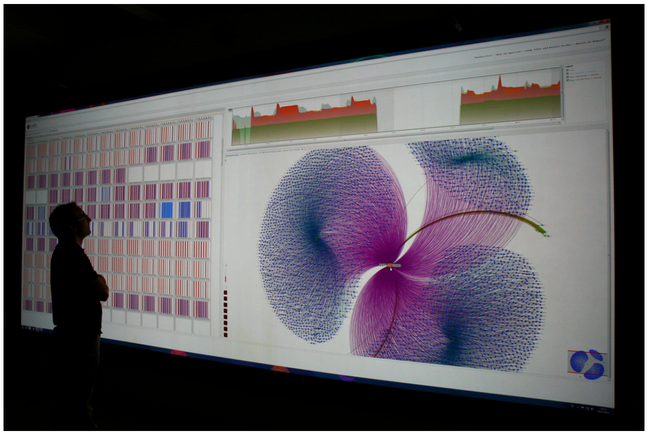

Figure 1: *VACS* used on a large powerwall display. The colored thumbnails on the left represent the different traffic patterns for many network hosts. Several time-series are shown as interactive line charts on the top. The quite cluttered interactive node-link diagram shows the connections between different source and destination ports and other external hosts.

Treemap visualizations are used to show the mostly used ports or the host with the most traffic in a selected time span as seen in Figure 3.

Graph Viewer as seen in Figure 5 helps the analyst to really explore the different connections between different hosts and port communications. This makes sense for shorter time spans or for specific queries. Often the calculated layout is not perfect, so the user is able to interactively modify the node-link representation. An interactive fisheye lens can also help to explore cluttered areas. Color is mapped to the different object types (e.g., IP addresses, source ports, destinations ports).

Hierarchical ClockMap [1] is a visualization technique, which can be used to explore thousands of time-series within their hierarchical structure. We enhanced this approach using SAX [3] to cluster and build the hierarchy which is used to represent all hosts. This helps to spot suspicious hosts and outliers which behave differently from the other hosts in the respective sub network.

Data Exploration Table is another view, which is quite important to the analyst. This view represents the selected or underlying data in a tabular way (when possible) which can then be read in detail, used in reports or exported to other analysis tools.

*e-mail: Fabian.Fischer@uni-konstanz.de
†e-mail: Daniel.Keim@uni-konstanz.de

3 Use Case

An analyst wants to explore the past and the current network situation, because of several reachability and connectivity issues in the company's network. After getting a basic idea using the dashboard about the current situation, he is interested in analyzing the overall incoming and outgoing network traffic. In this analysis he is especially interested to explore the reasons for the connectivity issues which happened several times. In Figure 2 he can clearly identify major traffic peaks on five different points in time. They seem to have slightly different patterns and differ in duration, extend and volume.

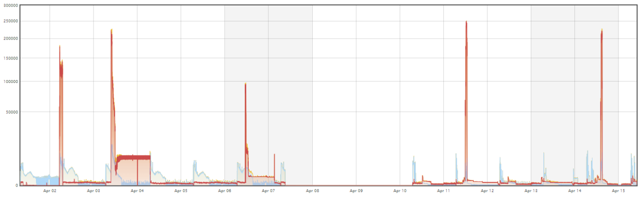

Figure 2: Interactive line charts show the overall incoming and outgoing network traffic. Different normalizations help to focus on peaks or low-traffic periods. Five enormous peaks are standing out.

The analyst can answer more questions by selecting the different high peaks using a rectangular selection. Additionally, he can add more related time-series to the line chart (e.g., different ports, different critical servers, number of alerts). Loading the treemap visualization even show which hosts (or ports) are responsible for the selected peak. With the help of this visual exploratoin possibilities he can distinguish between wide-spread denial-of-service attacks or very specific attacks on specific ports or just an ongoing company campaign with many legitimate connections. Further suspicious hosts can also be identified using the thumbnail glyphs in Figure 4 while the node-link diagram in Figure 5 helps to explore the aggregated connections of different hosts and attacks.

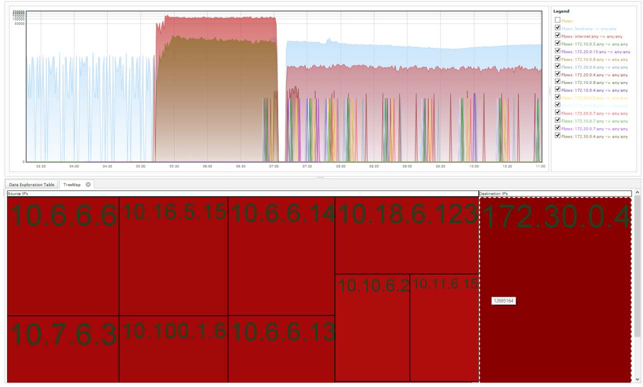

Figure 3: A treemap is loaded with underlying data from the selected time span to identify the top talkers (e.g., IP address with most activity in that time) or to get an overview of involved ports.

4 Conclusions

We made use of *VACS* to successfully explore the VAST Challenge 2013 datasets and show how the system can be used to analyze such security-related datasets. The web-based application uses a distributed database cluster to achieve horizontal scalability and combines different novel and state-of-the-art visual representations to assist the analyst in achieving situational awareness. We identified and provide means to explain unusual happenings in the network.

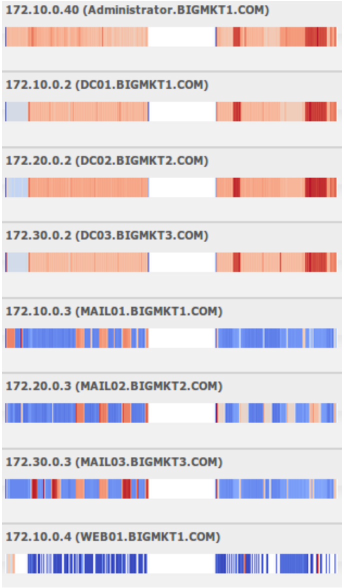

Figure 4: Example of some pixel-based thumbnails to represent the time-series of a metric of interest for the different network hosts.

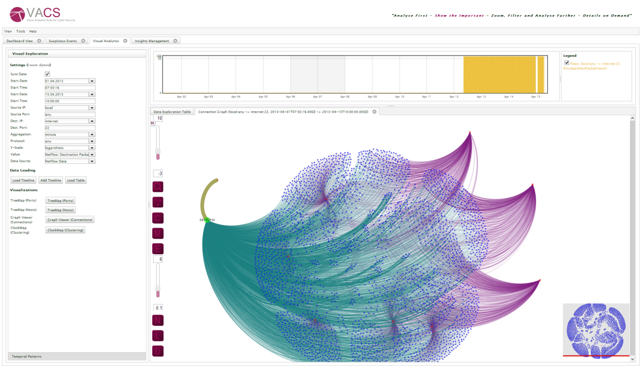

Figure 5: An interactive node-link diagram helps to analyze the aggregated connections between hosts and ports.

The system, especially the overview glyph representations, are best used on a large display as can bee seen in Figure 1. The *Visual Analytics Suite for Cyber Security* is still work in progress and tries to combine different visualization and analytics components to a easy-to-use network security suite.

Acknowledgements

The research leading to these results has received funding from the European Commission's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 257495, "Visual Analytic Representation of Large Datasets for Enhancing Network Security" (VIS-SENSE).

References

[1] F. Fischer, J. Fuchs, and F. Mansmann. ClockMap: Enhancing Circular Treemaps with Temporal Glyphs for Time-Series Data. In *Proceedings of the Eurographics Conference on Visualization (EuroVis 2012)*, 2012.

[2] F. Fischer, J. Fuchs, F. Mansmann, and D. A. Keim. BANKSAFE: A Visual Situational Awareness Tool for Large-Scale Computer Networks. In *IEEE VAST*, pages 257–258. IEEE Computer Society, 2012.

[3] J. Lin, E. Keogh, L. Wei, and S. Lonardi. Experiencing SAX: a novel symbolic representation of time series. *Data Min. Knowl. Discov.*, 15(2):107–144, Oct. 2007.