Professional Analysts using a Large, High-Resolution Display

Traffic Mini Challenge Award: Contributions to the VAST Challenge Contest

Alex Endert[1] Christopher Andrews[1] Glenn A. Fink[2] Chris North[1]

[1] Virginia Polytechnic Institute and State University [2] Pacific Northwest National Laboratory

Abstract

Professional cyber analysts were observed as they attempted to solve the VAST 2009 Traffic Mini Challenge using basic visualization tools and a large, high-resolution display. We discuss some of the lessons we learned about how analysts actually work and potential roles for visualization and large, high-resolution displays.

Keywords: VAST contest, large high-resolution displays, visual analytics, cyber analytics, information visualization.

Index Terms: H.5.2 [Information Systems]: Information Interfaces and Presentation—User Interfaces

1 Introduction

Our team's non-traditional approach to this challenge arose from working with the team developing the VAST 2009 dataset. Thus, we knew the solution before the contest was released but retained our independent viewpoint. During the final stages of developing the dataset, we used this study as a means for testing and validating the dataset for accuracy and realism. Four professional cyber analysts from a large government laboratory provided feedback by solving the challenge using the same processes they would use for their everyday work. Since the synthetic data was different in some ways from their normal data, we provided the analysts guidance when appropriate.

2 Experiment Setup

Each of the four analysts was given a two-hour session to explore and analyze the dataset to find the threat. Since our aim was to validate the dataset and to study cyber analysts using large, high-resolution displays rather than to develop special purpose cyber analytic tools, we let them use Microsoft Excel, their normal tool of choice, to display and manipulate the raw data. In addition, we provided the general-purpose visualization tool Spotfire (http://www.spotfire.com). Spotfire is capable of importing Excel data with ease (i.e. click-and-drag) so it was a natural choice to inject visualization into the cyber analysts' processes.

Each analyst was given a two-hour session to solve this challenge. We chronicled their progress via video recording, think-aloud protocol, and an automated tool that captured a screenshot every minute. We followed the study with an interview where we asked a series of questions regarding their experience, as well as their typical workspaces and tools. All analysis was performed using a large, high-resolution display running Windows XP. The display consists of eight 30-inch LCD panels, tiled in a 4x2 configuration (Figure 1). We chose this workspace setup based on previous work where we learned the benefits of a larger display space [1]. The analysts were able to display all of the information relevant to the challenge without minimizing any windows. This meant that they could physically navigate to gain an overview of the dataset, examine details, switch tasks, and rapidly consult multiple views and tools.

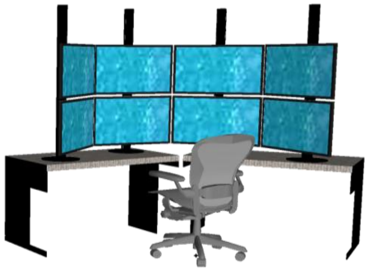

Figure 1. The large, high-resolution display used for this study, totaling nearly 33 megapixels, from [2].

3 Processes of Analysts

All the analysts started the study by performing their personal series of standard searches and questions based on their prior domain knowledge. These included queries on specific IPs, sorting by largest flows, creating pivot tables in Excel to highlight unique IP-to-IP connections, and more. However, this dataset challenged the analysts to go beyond their normal methods to thinking strategically about the problem. This is partly because the challenge required use of some data sources (prox records and office maps) that are not normally available to them. The analysts all commented that finding the proper relationship among all the data sources was very important. Analysts' backgrounds strongly dictated the tools they used. For instance, one analyst was very skilled in Excel, and she performed the majority of her work by creating different views of the data within Excel. To preserve branch points within her investigation, she saved versions of the data, each representing a "working state" of the investigation. The other analysts mainly worked back and forth between the visualization and the data in Excel, with one analyst doing the majority of his work within Spotfire. We believe this occurred due to his previous experience with such a tool, as he felt very comfortable with manipulating the visualization. Keeping the data synchronized between the two tools was difficult, so analysts would often use the visualization for exploration and discovery, then use Excel to "quantify and reconfirm" what they saw. One analyst kept a separate "note file", where she pasted interesting information from time to time.

3.1 Physical Navigation and Correlation

We observed analysts frequently relating information between different tools or windows distributed throughout the space. Figure 2 illustrates a typical workspace layout of an analyst performing this task. Analysts would correlate information by physically pointing at some data in one window and then finding it in another window with a different view. Even after we blanked the screen at the end of the study, they would frequently use phrases such as "this data here" or "what I found over there" while pointing to the different regions of the display where they

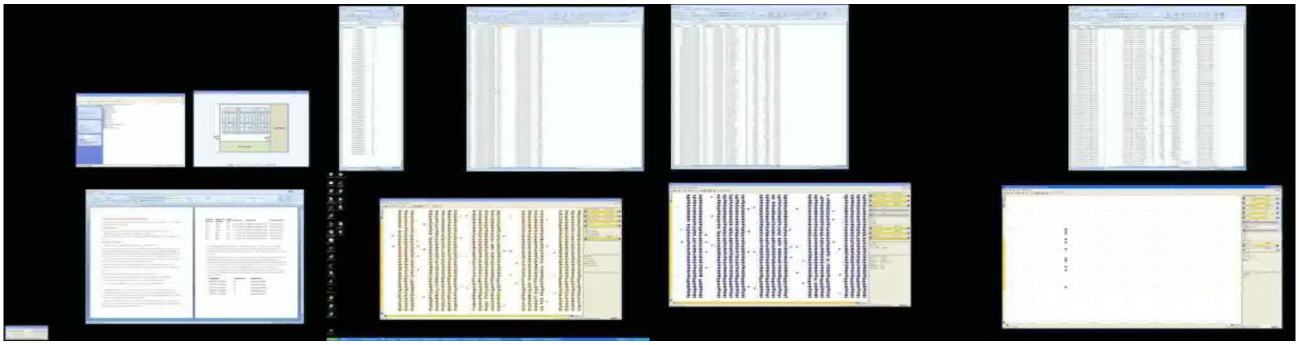

Figure 2. A screenshot of the spatial workspace layout as personalized by the analyst. All the data is spread out in the space, with the analyst being able to correlate from one view to another with ease.

had placed particular views. When the analysts felt they were not discovering interesting findings in one Excel window, they would switch to a different window and manipulate that data. Later, they would relate information from one window to another, immediately remembering the state of the data in a table they had not worked on for some time. This spatial navigation appeared to alleviate the disorientation they reported using small displays that forced minimized or stacked windows.

During one session, we had two analysts collaborating to analyze the dataset. Although the display was set up with only one keyboard and one mouse, the analysts collaborated by taking turns actively working with a tool and statically analyzing a visualization to make sense of the data shown. Then, they would synchronize the information they found from their separate analyses, again physically pointing to data in one view, while the other analyst attempted to find the data in another view.

3.2 From Textual to Visual Investigation Strategies

A challenge analysts faced during this study was the environmental change from a textual approach that concentrated on query-driven manipulation of individual tables, to a visual approach that operated holistically. This did not come easy to most analysts. Often, when we would point out something to them within the visualization, they would glance at it, and then move directly back to Excel and continue their work there. Cyber analysts clearly distrusted visualizations. Our post-study interview revealed that they believe visualizations "hide the data" via over-aggregation. Some claimed they were unable to "save states of what [they were] working on" in visualizations, causing them to be very tentative with their visual exploration for fear of losing their way within the investigation.

A critical point in the analysts' investigations occurred when they made the connection between the prox and IP data. Joining these sources provided a way to easily visualize where an employee was located when their assigned IP was active. The pair of analysts joined the sources manually, with one pointing to the visualization of the IP data, while the other checked physical office locations. Another analyst made the key discovery after two hours of unproductive search in Excel by merely glancing at a visualization of the join, spotting an outlier, and further investigating it. However, as we found out in the interviews, cyber analysts are unaccustomed to joining other data sources with network data. Their job often deals with network and host data, but prox was new to them. Thus, some remarked that joining the two "had never occurred to them".

4 Conclusion

The four professional cyber security analysts performed the task well. Although three out of the four were reluctant to use either visualizations or the additional display space at first, they all became comfortable with the setup as the study progressed. After the study, they remarked how visualization provided them with "interesting findings" much more quickly than working with raw text data. Through this study, we received strong feedback from cyber analysts regarding their view towards visualizations and their general style of work. We learned that cyber analysts:

1. Generally do not work holistically, but are used to dealing with a homogenous set of data (i.e. network traffic data only).
2. Use a personal query set developed through years of experience.
3. Want to save the state of their investigation.
4. Have an inherent distrust of visualizations.
5. Demand deep access to the data, both by detail on demand, but also to edit and manipulate the data in a tabular format.

By observing analysts and their natural analytic process, it became clear that saving a particular state of their investigation is critical. This safety net enables analysts to investigate alternatives without fear of losing their orientation within the investigation. Their current method of saving versions of their data is cumbersome, yet they have trained themselves to do so, further highlighting the importance. In addition to determining a set of requirements for them to perform their work, we were able to gain insight into their view towards large, high-resolution displays as well. They felt that the added display space enabled them to:

1. Physically navigate their workspace to find correlation between the different and diverse data.
2. Physically navigate their workspace to switch between different views of their data with ease.
3. Display a large quantity of data at once, without aggregating or "hiding" as much data.

Overall, the analysts voiced their appreciation of the large, high-resolution display. We feel it was beneficial to their task, as well as applicable to their daily work.

Acknowledgements

The authors are grateful to the Information and Infrastructure Integrity Initiative (I4) and the National Visual Analytics Capability (NVAC) located at Pacific Northwest National Laboratory for their support of this work.

References

[1] Ball, R., C. North, et al. (2007). Move to improve: promoting physical navigation to increase user performance with large displays.

[2] Fink, G. A., North, C. L., Endert, A., Rose, S.: 'Visualizing Cyber Security: Usable Workspaces'. VizSec 2009.