Supporting the Cyber Analytic Process using Visual History on Large Displays

Ankit Singh
Virginia Tech
2202 Kraft Dr.
Blacksburg, VA, 24060
sankit@vt.edu

Alex Endert
Virginia Tech
2202 Kraft Dr.
Blacksburg, VA, 24060
aendert@vt.edu

Christopher Andrews
Virginia Tech
2202 Kraft Dr.
Blacksburg, VA, 24060
cpa@vt.edu

Lauren Bradel
Virginia Tech
2202 Kraft Dr.
Blacksburg, VA, 24060
lbradel1@vt.edu

Robert Kincaid
Agilent Laboratories
Santa Clara, CA
robert_kincaid@agilent.com

Chris North
Virginia Tech
2202 Kraft Dr.
Blacksburg, VA, 24060
north@vt.edu

ABSTRACT
Cyber analytics focuses on increasing the safety and soundness of our digital infrastructure. The volume, size and velocity of these datasets make the analysis challenging on current work environments and tools. A cyber analytics work environment should enable multiple, simultaneous investigations and information foraging, as well as provide a solution space for organizing data. As such, various workflow visualization tools are used to help users track their analysis, reuse effective workflows, and test hypotheses. Also, the use of large display workspaces can provide new opportunities for improving visual analytics in cyber security. In this work, we present a prototype workspace for analysts where the analytic process is maintained in the workspace. Thus, we are able to present analysts with visual states of their data throughout the investigation, in which real-time changes can be made to any previous state, and analysts can backtrack through their investigation.

Categories and Subject Descriptors
H5.2 [Information Systems]: Information Interfaces and Presentation – User Interfaces

General Terms
Management, Performance, Design, Experimentation, Security, Human Factors.

VizSec'11, July 20, 2011, Pittsburg, PA, USA.

Keywords
Interaction Styles, Prototyping, Screen Design, User-Centered Design, Large High-Resolution Displays.

1. INTRODUCTION
Cyber security analysts are tasked with analyzing a large volume of structured network data. Their tasks can range from maintaining the security and privacy of a network, detecting intrusions in real-time, or performing a forensic investigation to find out how an attack occurred, and what damages were done. Previously, we had the opportunity to study the processes and patterns in behavior of professional cyber analysts at a government lab to see how they approached these tasks and managed the vast quantities of data. Through semi-structured interviews, as well as a user study, we found that they primarily turned to conventional tools, such as spreadsheets and simple command line tools [1]. These tools address specific needs in the course of an investigation, but analysts found it challenging to maintain awareness of their process and history. It is from this previous work that the motivation and design of the prototype presented in this paper is grounded.

One of the key challenges stemming from these ad hoc workflows is that maintaining a history of the process is enormously difficult. We identified three forms of information that the analysts found important to keep track of:

- Previous versions or views of the data
- The data transformations previously applied to the data
- A record of the hypotheses explored and corresponding findings

The first two are a reflection of the particular pattern of investigation that we observed from the analysts. Investigation for them is a process of filtering, projecting, selecting, and cross correlating data. Each one of these transformations results in a different collection of data. While no data is actually destroyed, the process of applying these transformations can still be considered destructive, since the intermediary views are lost.

There are a number of reasons to want this information to remain available. One reason is to validate any eventual solution or insight. Maintaining this history can provide transparency, so the process can be checked for errors as well as meaningfulness. Another reason would be to maintain multiple tracks of an investigation that stem from a common view. Finally, availability of previous versions can serve as a sanity check or a reference once the data set has been heavily transformed.

In this paper, we present a prototype for maintaining a visual history of the cyber analytic process using a large, high-resolution display (LHRD) (shown in Figure 5). Designing our prototype for use on a LHRD allows us to display a large quantity of information at a very detailed level. Thus, the core idea of this tool is to use the large, high-resolution display to maintain a visual history of an investigation. The increased space afforded by the large display allows analysts to maintain previous versions or views of the dataset (in this case, Excel spreadsheets). Since these previous views are still fully functional (i.e., running instances of tools), analysts can return to a prior state of the investigation and spawn a new copy to create "branches" in the investigation (Figure 3). Each such branch can represent different paths (or hypotheses) explored during the investigation. Thus, analysts are free to investigate the dataset while the tool maintains the history of their analysis in the workspace.

2. RELATED WORK

There are a number of tools that maintain history. For example, CzSaw [2, 3] is an intelligence analysis tool that provides lookup to past actions visually in a linear timeline. The user can go back and replay the whole process and possibly the whole history on a similar dataset. Similarly, Tableau [4] is a data analysis tool that also provides lookup in linear time based on the visualization. Users can click on thumbnails of previous states and resume their work. Essentially these tools which support visual history let the user see previous actions and resume them if they want to follow a different path.

These tools, and others like them, implement visual history through the use of thumbnail screenshots that provide a visual reminder about the state of the system at the moment in history. Our prototype provides a similar facility, but instead of thumbnails, we use actual working, "live" copies of the tool being used. In other words, the visualization of the history is simultaneously the working environment, making the process history a more fundamental part of the analytic process. Our target is not to record history, but to create a dynamic environment that allows the analyst to work directly with his or her process.

3. LARGE HIGH RESOLUTION DISPLAYS

There have been many display configurations that have been described as being "large, high-resolution displays". For the purpose of this work, we have used a desktop display consisting of six 30" panels (Figure 1). Each panel has a resolution of 2560x1600, for an overall display resolution of 7680x3200. While this is made of a collection of individual monitors, this is not just a multi-monitor system; this is a fundamentally different environment [5]. The important feature of this display is that the resolution is high enough that it exceeds the user's visual acuity [6]. In other words, there is no position from where a fixed user can make use of every pixel of the display.

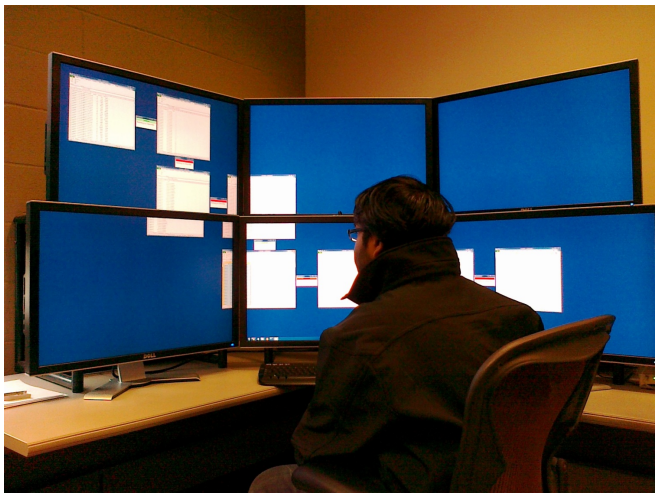

Figure 1: Large Display Workspace

This feature of the display is important because it changes the way we address tool design. Rather than being a large collaborative space, or a medium for single, data intensive visualizations, the strength of this display is in becoming an environment for the user [7]. In other words, we expect the user to work on sections of the display, and employing *physical navigation* (turning, or moving back and forth) to access different regions over time [8].

While it is tempting to suggest that this could easily be replaced by some form of virtual desktop implementation [9], we maintain that large, high-resolution displays provide a significantly different experience. Our work has demonstrated that physical navigation is more efficient [8], and that it changes the user's perspective of the environment [5]. The physical movement of the user provides cognitive advantages that help to maintain coherency and speed access.

We have also demonstrated that the large, high-resolution display biases the user to perceiving it as an implicitly spatial environment [7]. Because the display has such a large number of available pixels, this space can be used to work with conventional tools, while the surrounding space adds extra contextual information not available on a conventional display. It also offers opportunities to further instrument the surrounding space with additional visual cues, which add even more metadata to the space.

In our prior work with intelligence analysts, we leveraged the space as a free form semantic layer that the analysts could use to externalize relationships between reports. In this work, we still intend the space to be used to help externalize the analyst's thought process, but in this case, we are taking a more direct approach and trying to capture the actual flow of the investigation by capturing the key decision points. This change reflects the needs of the more data-centric cyber-security analysts.

4. THE CYBER ANALYTIC PROCESS

The primary motivation for our approach is rooted in our prior study of professional cyber-security analysts at a government lab, which provided several key insights into how the analysts work through the large amount of raw data to reach a possible

Figure 2: Mockup Visual History Prototype for a Large Display Workspace

solution [1]. In the study we provided the analysts with the data from the VAST 2009 challenge [10], which consisted of network traffic logs and physical access data for each employee in a fictitious embassy. Our observations of their process yielded a couple of interesting points.

Many of the analysts relied heavily on Excel for its ability to handle raw data, and produce quick charts and graphs based on their pivot tables. While other cyber security tools were also used, many analysts informed us that nearly all of their investigations involved maintaining their data in Excel. This was further evidenced by each of their expertise when handling dat in Excel. Thus, we chose to base our prototype on Excel.

With regards to history of the analytic process, we saw the analysts exhibit the following behaviors. First, the analysts were careful to **save multiple versions of the data** during important stages of their investigation. When asked further about the purpose of maintaining multiple versions of the files, the analysts commented that it provided a simple way of backtracking if needed. For example, one analyst saved versions of the data as "v1.1", "v1.2", "v2.1", etc. as a way to keep track not only keep track of the previous versions of the files, but also maintain a record of what hypotheses have been explored. The analyst made it clear that this is how he performs his everyday work, and was not simple an artifact of this challenge dataset.

In the study, we provided the analysts with a large, high-resolution display. Thus, one goal of the study was to observe how the additional display space could be used to aid the analysts in their task. The large, high-resolution display used by the analysts during the study proved beneficial in the following ways. As we observed in their work environment, the analysts tended to keep **many windows open simultaneously**. In this study, analysts opened several windows, including Excel and Spotfire visualization for the prox data, Excel and visualization for the IP data, the office map, a Windows Explorer window to access the files, and the scenario description document. The number of windows and tools open increased as their investigation progressed, each tool representing an important piece of their investigation. Some analysts also opened an internet browser to search for specific ports, a calendar to the relevant dates, and attempted to open multiple Excel windows to keep track of previous versions of their files as they made edits.

We observed that this additional use of tools (and display space) was essential to their sensemaking process. As the analysts became more deeply engaged in their investigation, they would leverage the flexibility of the space. First, the information was all visible, relieving them from stepping through the taskbar to switch windows. In addition, this persistent space allowed for **physical cross-referencing**. We observed the analysts physically pointing (with their hand) to data in one window and physically pointing to related data in another window, so they could look back and forth to identify potential connections. This was much more efficient than stacking windows on top of each other.

Analysts commented that they enjoyed being able to rapidly switch between windows by simply moving to another region of the display where the window was visible. This was also important to their investigation because it allowed them to rapidly confirm a finding between datasets. In their regular office setup, they would often take notes (e.g. jotting down an IP or port number), because they knew it would take a long time to re-find the information later after they had switched to other windows. With this display setup, however, they could easily **re-find information** (look-up previously found information) by rapidly glancing back to its window.

The workspaces created by the analysts formed a **synthesis space** that allows the analyst to organize information spatially to reflect their understanding of the scenario. The display flexibly conforms to the mental models of the users. This enables us to go beyond information visualization and begin to understand how analysts use space itself as a problem-solving medium. They also used the space to point out and explain to us the different findings they had encountered in different windows on the screen.

As a result, we designed a mockup (shown in Figure 2) that uses the display space to track the analytic process. It offers visual history that can provide orientation and traceability over the life of an investigation. This visual history provides analysts a means for easily retracing their steps when it comes time to produce a report and share their findings.

Using the additional space of a large display, analysts can "fork off" instances of their tools, and pursue branching hunches in parallel. Windows along each branch of the history tree are running instances of their tools or windows enabling the user to easily backtrack to an earlier state and remain oriented to the entire task. The larger windows are "key frames" that mark a state in the investigation that an analyst deems particularly

important. They might be branch points where the analyst can create a new instance of the tool he is using to pursue a new hunch. They are analogous to the saved versions of the data currently used by analysts to retain the state of an important view into the data. The tree-structure matches the numbered versioning used by one of the analysts. The size of each history window is proportional to the age of its most recent use or to the frequency it is consulted. Seldom used windows slowly become smaller unless the user refers to them by moving the mouse over them, clicking, or resizing them. They never disappear until the user deletes them or the branch they live on, so the user can easily review his thought process and regain orientation quickly when switching among branches.

The goal of this approach is to obtain the benefits of the natural spatial process and interaction history that occur during analysis, and overcome the spatial stagnation and process loss that occurs while managing multiple windows and file versions. The approach transforms temporal intra-window interaction into spatial inter-window interaction.

This approach flexibly enables users to visually represent their thought processes as any number of these paths, write their thoughts using forms provided for each state appended to every window along the paths, and to freely interact with any part of this visual workspace, as all the interactions are still intact. Thus, the workspace becomes one of not only representing the history, but rather *maintaining it* within the context of the analysis.

The final report is the most common way cyber analysts "share with their supervisors and fellow analysts." The number and quality of reports may contribute to analysts' performance evaluations, so analysts are motivated to present them clearly. Maintaining temporal and logical orientation over the lifetime of an investigation would help analysts clearly state their conclusions and how they arrived at them. Our approach gives the analyst a live report of his or her thought process, which is easy to share and would also be easy to understand given the space to write the thoughts.

Another benefit of externalizing the process is that cyber-security analysts frequently encounter variants of the same problems. In our study, we learned that solutions to these problems become part of the analyst's "toolkit" that he or she can apply to future problems. The visual history of the process can simplify the process of recovering and reusing these.

5. WORKING WITH HISTORY

As stated previously, our goal is not to merely provide a graphical history of the analytic process, but to provide analysts with a workspace where maintaining the previous states of information becomes a part of their process. There are a couple of key advantages to this approach.

First, it allows us to fully explore the nature of a state in the process. This is especially important when working with a complex, multifaceted tool such as a spreadsheet. By providing a full working version, the analyst can go back and not just see a image of what the state of the investigation was, but actually work with the tool to see exactly what the configuration or settings may be.

Another key advantage is that the analyst can use previous threads of an investigation as references for a later thread. This is especially important for cyber-security analysts. As stated earlier, the analysts tend to build a library of sub-processes or "queries". Many investigations will have parallel tracks in which the same operations will need to be applied, so it is very useful to be able to exactly mirror a previous sub-process.

Finally, the analyst can arrange the space to add extra information about the process. For instance, the analyst could move dead-ends in the investigation to a remote location in of the display to indicate that this was a trail that was previously followed, but was not fruitful. This will allow the analyst to visually separate branches in the investigation from parallel paths in the "live" investigation.

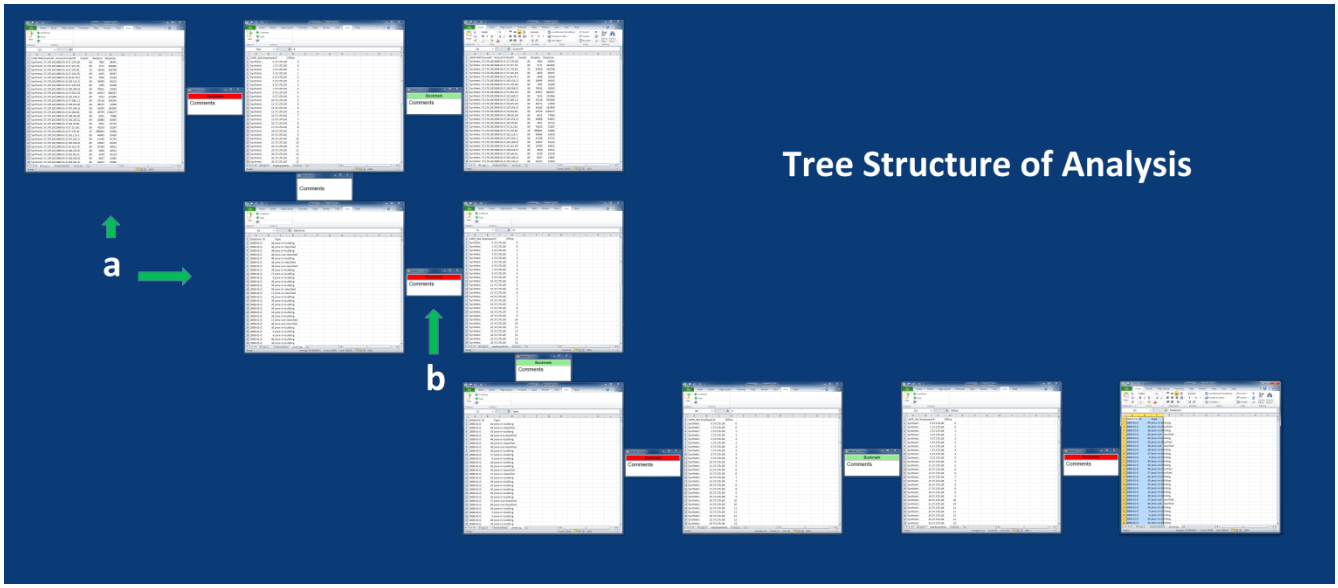

Figure 3: Branching and Visual Linking a. Branches: Individual Excel Instances b. Visual Links with Bookmarks

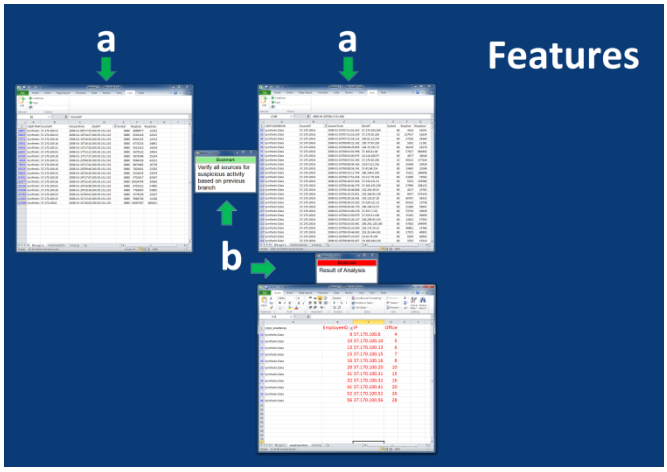

Figure 4: a. Different File Versions b. Comments Form and Bookmarks

6. PROTOTYPE IMPLEMENTATION

Our tool utilizes the large display space to maintain the analytic process. It offers a visual history workspace that can provide awareness and traceability over the life of an investigation. This visual history provides a means for analysts to easily retrace their steps when it comes time to produce a report. Using the additional space of a large display, analysts can "Fork" (Figure 3) instances of their dataset, and pursue different hunches in parallel. Each window is a "key frame" that marks a state in the investigation that an analyst deems particularly important. They are analogous to the saved versions of the data currently used by analysts to retain the state of an important view or state of the data.

We choose to build our prototype using Excel as the analytic tool, as it provides universal functionality to analyze data and do complex data manipulations, as well as simple integrated visualizations. These strengths have made Excel a popular data analysis tool for a number of different types of datasets. Even though Excel provides an environment that empowers the user to search, visualize, and edit information, it naturally does not have any facilities for taking advantage of large, high-resolution displays. For instance, Excel is designed to conserve window usage by opening tabs for each active file, and also each sheet by default in its screen space. While this may be helpful for small displays, this hindered analysts from spreading the various windows out to compare two or more sets of data. We do not recommend this space-conserving behavior because it places limits on how the space on the display can be used. Thus, our prototype is designed to extend Excel to make better use of the available display real estate.

We developed our prototype as an Excel Add-In, which allows it to integrate itself directly into Excel, creating a special entry in the Ribbon toolbar. The entry contains three buttons: "Add", "Continue" and "Fork". The user begins the analysis by using the "Add" button, which opens a selected file and registers it for version history tracking. Whenever the user feels that he or she wants to add a checkpoint (or saved version of the dataset), he or she can click the **"Continue"** button. When this happens, the systems respond in two ways.

First, the existing dataset is copied and loaded in the new window that is placed to the right of the current window. These **file versions** (Figure 4) are numbered on the basis of their perceptual location in the analytical process. Second, a small place-holding window is placed between the original and new windows (Figure 4). This window contains a **notepad**, for the user to record comments about the particular state or window, describing the actions performed in that state. The form also contains a **"Bookmark"** button, which can be used if the analysts wants to come back to review his actions at a particular state. The button toggles the color to red based on whether it is pressed or not. This gives a visual cue to the analyst to remember to return to the particular state later in the analysis.

7. USE CASE DESCRIPTION

We demonstrate the functionality of our prototype in the following use case.

7.1 Dataset Description

In our use case description, we completed the cyber mini-challenge scenario provided for the IEEE VAST 2009 Contest [14]. The VAST 2009 challenge dataset is primarily focused on finding suspicious activity from large chunks of network traffic data combined with physical access data (prox records) for individual employees in a fictitious embassy. The prox records contain data on when an employee entered or exited a building (although this data is not sufficient to infer whether an employee is inside a building because employees can "piggy-back" their way behind other colleagues into the building and avoid using their prox card for entry). It also contains access data from a classified section of the embassy (which does not have any computer or network access) where use of prox cards is strictly enforced. The dataset also includes a map of the embassy offices, which shows that 2 employees share one office. The network traffic data shows IP activity from each of the employee's computer, which contains the destination IP used for connection.

The challenge was to use the network traffic data, prox access records, and physical office locations to determine whether or not there was a malicious insider sending classified information from the embassy. The challenge was designed so that no single source of data was sufficient by itself to solve the mystery. Thus, in order to succeed in solving the challenge, it is essential to tie together multiple "filtered" views of the data. This is representative of the challenging and complex datasets cyber analysts are faced with.

7.2 Approach to the Problem

The problem was approached by first understanding the key points which could lead to a suspicious IP address in the logs. First, we filtered the network traffic data and physical access data by individual employees. Through filtering to a single user and branching it was observed that if a user is in the classified section, no activity should be detected from his or her terminal. We completed this process by maintaining several instances of Excel connected in a path, which revolved around analyzing the network traffic data, in the time intervals when each user was in the classified section. All the instances in these branches and paths could be later reused to analyze the actions of all employees. Comments were added describing each step and important steps were bookmarked for review.

After analyzing these individual employees network traffic information, we identified suspicious activity linked with specific employees' computer terminals (IP traffic while they were in classified areas). That is, the IP address had data sent to them during the timeframes when the employees were in the classified area. These were marked as suspicious IP addresses. Once these IP addresses were detected, a new branch was created from the original dataset. This branch focused on analyzing whether the suspicious IP address was being sent data from any terminals other than those already identified. After coming up with a list of user terminals accessing the suspicious IP addresses, the first branch was reviewed to investigate all terminals that access the suspicious IP addresses, using comments and bookmarks to retrace the logical steps taken in the analysis.

This process was repeated for all identified terminals associated with suspicious activity. The repeating of the process was expedited because we only had to tweak some parameters and follow the process that was had already been used. This was made possible by the fully functional instances of Excel still maintained in the workspace. All results from this step were recorded in a separate window that was branched from the original dataset. Figure 5 shows the screenshot of the display space after the analysis of the data.

After identifying all computer terminals that illegally accessed the suspicious IP address, we turned our focus to narrowing a list of suspicious employees. The embassy access data was analyzed and all the users present in the building when a possible suspicious activity happened were noted, although there was a level of uncertainty here because users could "piggy back" into the building by closely following another employee when they entered the building.

Reporting these findings was simplified using this prototype, because not only was the history still present, but since each step in the process represents a fully functional instance of the Excel with the associated data at that step, we were able to quickly access these stages of the investigation to report the activity observed. For example, some of the stages in the process contained filters or functions that we wanted to report. We were able to simply access these fields within Excel and report on them. Further, while reporting all the successful findings of an investigation is important, we were also able to report on the hypotheses we tested and did not uncover any suspicious activity.

7.3 Advantage of the prototype in analysis

Through our use case we identified several aspects in which the visual history prototype aided the analytic process.

Comments and Bookmarks. We used comments and bookmarks to highlight key steps and maintain a log of actions taken, which allowed us to easily review actions and regain awareness of the context surrounding the specific state of the data. This was particularly useful when we needed to repeat a series of actions taken with a new set of parameters. For example in our use case, bookmarking helped review the steps involving the identification of a suspicious IP, which we were then able to apply to different employee's IPs when needed.

Original Dataset Access. Because all of the files in the VAST challenge data set were available on the display space, finding them and starting a new analysis was as simple as looking at them and clicking the "Fork" button. Several times we had to discard a path, following a new path was as trivial as to fork

Figure 5: Use Case on VAST Challenge Dataset

from the original dataset or the dataset from where the path was forked.

Increased Display Space. The availability of the large display space provided more than adequate space to open new windows to record new data and branch the analysis.

Awareness. The use of a workflow based design helped to review and remember the entire process by simply looking at the whole screen space, resulting in a greater level of overview and awareness of the status of the investigation. This helped to identify issues with assumptions made during the investigation, and also helped to validate the process taken when it came time to record the findings.

8. DISCUSSION

Through developing this prototype, we were able to demonstrate how an initial approach to maintaining the analytic process persistently visible in the workspace can aid analysis. With our prototype, users can maintain an awareness of the progress of their investigation through the tree-like layout. In subsequent work, we hope to present an empirical user study to better analyze the potential benefits of visualizing cyber security analytical history on large, high-resolution displays, and explore the following opportunities and challenges.

The current prototype raises several interesting challenges and opportunities to investigate further. In the current implementation, the layout is determined by either the "fork" or "continue" operation, placing the window below or to the right of the window from which the operation was invoked, respectively. From there, analysts have the ability to reposition the windows to locations meaningful in the context of the investigation. While this is an initial approach to creating a visual representation of the versioning behavior observed, it does raise questions regarding window management. For example, how should the system respond when analysts "fork" more than one time from a given window? Additionally, giving the user control over the layout will break the tree-like structure of the process, which opens opportunities for directly manipulating the sequence of operations by changing the order of the windows, copying specific windows from one branch to another, etc.

In addition to exploring the layout challenges with this prototype, the question of when to "continue" or "fork" is equally interesting. That is, in the current implementation the user has to explicitly click on the control in order to cause these actions to happen. However, form the observations of cyber analysts we found that they (and most likely any users of Excel) frequently used the "Ctrl+S" (Save) and "Ctrl+Shift+S" (Save As) shortcuts during their investigation. These commands may provide an opportunity to couple the "Continue" and "Fork" operations to. For example, when an analyst chooses to "Save" a important stage of the investigation, he may be indicating that a significant change has been made over the previous version of the data, thus implying that a "Continue" operation should occur. Further, when choosing to "Save As", he may be indicating that a new branch of hypothesis is being explored, thus invoking the "Fork" operation may be helpful.

While LHRDs provide users with additional space for showing high-detail representations of information (in this case, full instances of tools), the complexity of many investigations will ultimately cause the analyst to run out of space. When this occurs, the system can respond in a number of ways: adding virtual space and allowing the user to pan and zoom, collapsing less used branches into icons, merging multiple steps of a branch into a single window, etc. The cognitive implications of these actions need to be explored before determining the optimal design decision.

Additionally, we would like to implement brushing and linking to the live visual history prototype in order to track specific data points through the different data views [11]. This will allow analysts to track data points throughout their investigative process and allow for upstream and downstream exploration. For example, when analysts select a data point in an earlier window, being able to see the dependency of that data point in subsequent stages of the data allows them to make judgments and gain further insight into the data.

Beyond brushing and linking, there are more broad implications of this work that relate to workflow management tools, such as Vistrails and Taverna [12-15]. These tools focus on presenting users with a visual representation of the data transforms and manipulations. Thus, users are able to save their set of interactions (e.g., filters, data manipulations, etc.) and re-apply them to a later investigation or different data. We would like to extend our prototype to create a hybrid of visual history and workflow tools, where users can benefit from taking portions of the visual history and create a reusable workflow. This will likely be very useful for cyber analysts because they frequently reuse analytic steps with different data parameters.

9. CONCLUSION

We previously identified through a user study that cyber security analysts require support for accessing previous states of the data set, recalling transformations previously applied to the data, and records of hypotheses explored along with the corresponding findings. We addressed these needs by constructing a prototype that displays multiple views of the same data set at different points in the analytic process, complete with user-created annotations to mark the differences between views and track progress.

This Excel-based prototype takes advantage of the large, high-resolution display screen size by creating a live, interactive history space that can be accessed through physical navigation. By combining the data with the flowchart, we have provided users with an immersive analytic environment. Our prototype has established an interactive method of encouraging exploration of different investigative decisions.

We hope that this notion of maintaining the visibility of all stages of the workflow will foster a greater awareness of the data set, producing more efficient and effective cyber security analysts.

10. REFERENCES

[1] G. A. Fink, C. L. North, A. Endert, and S. Rose, "Visualizing Cyber Security: Usable Workspaces," in *Proc. of Intl Workshop on Visualizing Cyber Security (VizSec 2009)*, 2009, pp. 45-56.

[2] D. Dunsmuir, M. Z. Baraghoush, V. Chen, M. E. Joorabchi, S. Alimadadi, E. Lee, J. Dill, C. Qian, and C. D. Shaw, "CZSaw, IMAS & Tableau: Collaboration among Teams," presented at the Proceedings of IEEE Visual Analytics Science & Technology 2010, Salt Lake City, Utah, 2010.

[3] N. Kadivar, "Capturing and supporting the analysis process," *Visual Analytics Science and Technology (VAST), IEEE Symposium on,* pp. 131-138, 2009.

[4] J. Heer, "Graphical Histories for Visualization: Supporting Analysis, Communication, and Evaluation," *IEEE transactions on visualization and computer graphics,* vol. 14, pp. 1189-1196, 2008.

[5] L. Shupp, C. Andrews, M. Dickey-Kurdziolek, B. Yost, and C. North, "Shaping the Display of the Future: The Effects of Display Size and Curvature on User Performance and Insights," *Human-Computer Interaction,* vol. 24, pp. 230 -- 272, January 2009.

[6] B. Yost, "Beyond visual acuity the perceptual scalability of information visualizations for large displays," ed, 2007, p. 101.

[7] C. Andrews, A. Endert, and C. North, "Space to think: large high-resolution displays for sensemaking," presented at the Proceedings of the 28th international conference on Human factors in computing systems, Atlanta, Georgia, USA, 2010.

[8] R. Ball, C. North, and D. A. Bowman, "Move to improve: promoting physical navigation to increase user performance with large displays," presented at the Proceedings of the SIGCHI conference on Human factors in computing systems, San Jose, California, USA, 2007.

[9] M. Ringel, "When one isn't enough: an analysis of virtual desktop usage strategies and their implications for design," presented at the CHI '03 extended abstracts on Human factors in computing systems, Ft. Lauderdale, Florida, USA, 2003.

[10] *VAST Challenge Dataset*. Available: http://hcil.cs.umd.edu/localphp/hcil/vast/index.php

[11] A. Buja, J. A. McDonald, J. Michalak, and W. Stuetzle, "Interactive data visualization using focusing and linking," presented at the Proceedings of the 2nd conference on Visualization '91, San Diego, California, 1991.

[12] S. P. Callahan, "Managing the Evolution of Dataflows with VisTrails," *Data Engineering Workshops, International Conference on,* pp. 71-71, 2006.

[13] S. P. Callahan, J. Freire, E. Santos, C. E. Scheidegger, Cl\, \#225, u. T. Silva, and H. T. Vo, "VisTrails: visualization meets data management," presented at the Proceedings of the 2006 ACM SIGMOD international conference on Management of data, Chicago, IL, USA, 2006.

[14] D. Hull, "Taverna: a tool for building and running workflows of services," *Nucleic acids research,* vol. 34, pp. W729-W732, 2006.

[15] T. Oinn, "Taverna: lessons in creating a workflow environment for the life sciences," *Concurrency and computation,* vol. 18, pp. 1067-1100, 2006.