7 Key Challenges for Visualization in Cyber Network Defense

Daniel M. Best*
Pacific Northwest National Laboratory
902 Battelle Boulevard
Richland, WA 99352
daniel.best@pnnl.gov

Alex Endert†
Pacific Northwest National Laboratory
902 Battelle Boulevard
Richland, WA 99352
alex.endert@pnnl.gov

Daniel Kidwell‡
Department of Defense
dlkidw2@tycho.ncsc.mil

ABSTRACT

What does it take to be a successful visualization in cyber security? This question has been explored for some time, resulting in many potential solutions being developed and offered to the cyber security community. However, when one reflects upon the successful visualizations in this space they are left wondering where all those offerings have gone. Excel and Grep are still the kings of cyber security defense tools; there is a great opportunity to help in this domain, yet many visualizations fall short and are not utilized.

In this paper we present seven challenges, informed by two user studies, to be considered when developing a visualization for cyber security purposes. Cyber security visualizations must go beyond isolated solutions and "pretty picture" visualizations in order to impact users. We provide an example prototype that addresses the challenges with a description of how they are met. Our aim is to assist in increasing utility and adoption rates for visualization capabilities in cyber security.

Categories and Subject Descriptors

H.5.0 [**Information Systems**]: HCI—*General*

General Terms

Design, Human Factors, Security

Keywords

Visualization, Cyber Security, Defense

1. INTRODUCTION

While the domain of cyber security is rich with opportunities to visualize information, adoption of visual analytic environments for use within cyber security operational settings is extremely low [6]. This lack of adoption is not due to the lack of available visualizations, but for the lack of addressing challenges in cyber security. The mission needs of operational cyber security professionals is such that there must be true impact immediately or there is not much chance for success. All too often visualizations focus on a single data source, are static, or are stand alone applications, which limits the utility of the application and hinders adoption.

To move past "just another packet visualization", a deeper understanding of what cyber subject matter experts (SMEs) need is required. To begin learning about SME needs, we leveraged two user studies. The first, was conducted jointly with Washington State University (WSU) exploring users' need for corroborating information in the decision making process, daily activities, and types of data used in those daily activities. That study identified a concept of "cadence" in cyber security data, which is a rhythm or pattern in the data that defenders used to identify outliers. Building from that study, while designing a situational awareness tool, the opportunity was taken to explore "cadence" further and strengthen understanding of user needs.

The two user studies provide valuable insight into the needs for cyber SME for both collaboration and for situational awareness. In combination, the studies expose seven challenges of the cyber security professional. In turn, these challenges provide guiding concepts when designing an application to be used for cyber security purposes. Some of these challenges are well known in the domain of cyber security. However, we offer several additional challenges identified through our interactions with cyber SMEs. In this paper we discuss the user surveys, their results and identified challenges, and offer visualization impact considerations when designing a solution.

2. RELATED WORK

Adoption and transitioning research into analytic workplaces is an inherently complex and difficult task, consisting of many "chasms" or barriers [12]. In visual analytic technology in general, work has been done to study what creates successful adoption. Previous work suggests that it is critical for novel technology to fit into the culture of the users performing the work, as well as solve one (or more) or their needs [1, 2]. It is likely that these same challenges or barriers

*Cyber security researcher, Visual Analytics

†Scientist, Visual Analytics

‡Sr. Computer Systems Researcher, Systems Behavior

VizSec '14, November 10 2014, Paris, France

http://dx.doi.org/10.1145/2671491.2671497

need to be met for cyber security as well. However, further understanding of the domain is needed to map out "chasms" and barriers to adoption.

Specifically to visualization and cyber security, Fink et al. have found that there is an inherent "distrust" in visualizations [6]; even though studies have shown that visualizations are beneficial to the domain [7]. Through interviews with domain experts, Fink et al. found that visualizations suited for cyber security focus showing the data primitives in an aggregation. For example, showing a collection of network packets in a scatter plot was not considered very helpful for users as much of the data is occluded or aggregated. While this is consistent with information visualization literature [14], it seems that cyber security as a domain requires other visual representations and interactions to properly meet the needs of the users.

To begin understanding users and their needs, several studies have been conducted mapping out roles and responsibilities in cyber security. While user studies in cyber security environments can be challenging due to data sensitivity, research does exist that has helped begin to illuminate the needs of users within the domain. For example, D'Amico et al. performed a cognitive task analysis on a group of analysts to find a common workflow among the group [5]. Such mapping of workflows is helpful to find areas within the workflow where new technology may be able to enhance or improve the current tasks. Further, work has been done to understand how data sources and tools can be mapped to the needs/tasks of users within the cyber domain [8, 3].

While the research community continues to build a better understanding of needs for cyber security analysts, there is still a need to ensure visualizations fit within the cyber security analysts' culture and solves their needs. There are many examples of successful visualizations in other domains, raising the question - what is it about cyber security that accounts for the apparent lack of adoption for visualizations? We explore what challenges cyber analysts face in cyber security to increase understanding of the domain, so that we may enable analysts to overcome those challenges. Visualizations, if properly designed, have the ability to meet the needs of users. We utilize user studies and prototype assessment to draw out what needs must be met to accomplish successful adoption of visualization in cyber security.

3. UNDERSTANDING USERS

Two user studies, consisting of focus groups, and informal discussions, and semi-structured interviews were held to explore the challenges facing cyber security. The user studies were leverage to deeply explore a particular concept, while focus groups and informal discussions were used to vet concepts and validate approaches. The first user study included in this work, conducted jointly with WSU, explored users' (both cyber SMEs and general computer users) challenges in cyber security awareness and how they are informed about cyber security threats and maintain awareness. From that user study, the concept of SMEs leveraging "cadence" to identify suspect activity was identified. To further explore the defender challenges, and the concept of "cadence", an additional study was designed and conducted. From the two studies, seven challenges, and three roles were identified (describe below) which contribute to understanding necessary to build useful visualizations.

3.1 User Study Design

Both studies were designed with a similar purpose of tracking experience and familiarity with concepts being discussed while allowing for users to provide feedback that was not necessarily mapped out prior to conducting the research. For our purposes, a short survey was used followed by a semi-structured interview. The survey provided context while assessing the answers of the semi-structured interview; while the semi-structured interview allows participants to answer open ended questions designed to investigate how participants maintain situational awareness, their experience in cyber security, how they conduct their investigations (cyber SMEs), and how abnormal behavior is identified.

Users were solicited through signup sheets that provided background on the study being conducted, contact information, and a privacy statement. Solicitations for volunteers targeted groups in the organization that work with cyber security related data, however no volunteers were turned away. The main Pacific Northwest National Laboratory (PNNL) campus, where the study was conducted, has approximately 4000 staff members; although a significantly smaller population identifies as a cyber SME.

4. RESULTS

From the semi-structured interviews and the focus group discussions, a view of the types of users, their roles and responsibilities, and the challenges they face was constructed. Each role has an important part to play in the overall cyber security of a given organization. While traditionally, the focus of cyber security visualization is on cyber defenders, it is important to take into consideration the other roles and responsibilities that may effect the overall design of an application. Operations and research, while having similar needs regarding data, have very different responsibilities and challenges.

4.1 Users & Roles

Several different high-level groups of users were identified through the user studies, each of which has different roles and responsibilities, risks, and challenges that are important to them. Roles were identified by participants self describing or through descriptions obtained during the semi-structured interviews. Information gathered in the user studies was then combined into groups roles. These groups include:

- Business Office: The business office staff are responsible for the oversight of programs and business systems that are necessary to keep the business running. While cyber security is a concern to this group of participants, it is primarily a concern by how it affects the business systems. Staff in this group deal more with policy, procedures, and insuring systems continue to function correctly.

- Cyber Defense: Staff in this role are responsible for analysis of the network, scanning for vulnerabilities, detecting and mitigating intrusions, and investigating events and alerts to find compromise and determine issues. Risks to these users are associated with the volume of traffic, vulnerabilities, and attacks that they must deal with on a daily basis. It is difficult to remain ahead of the wave of information to keep out of a purely reactive state. A deeper look into the cyber defenders

can be seen in D'Amico's paper on roles in computer network defense [4].

- Cyber Intelligence: Cyber intelligence (CI) staff are less concerned with the standard daily attacks and exploits and become involved when information on the network is at risk for loss or exfiltration. While spread of an attack is of importance, the also need to understand what is being targeted and why at a much deeper level than the security staff.

Each group identified has a different perspective that may influence the result of cyber security decisions. A balance is established for an organization between maintaining absolute security of a network, allowing some information to be gathered, and conducting business. While that balance may shift over time, none of the groups are removed from the equation completely. For example, there are known security measures that will make a network more secure (such as strict group policies and application white listing), however they in turn make it difficult to conduct business. To thwart the spread of a malicious user, a natural instinct may be to wipe and rebuild any system associated; however, understanding the who, what, and why of an attack can be equally, if not more, important to a business. Cyber security defenders do not work in isolation; providing them with a tool that is useful should include features to interact with the other competing views in the cyber security decision making process.

4.2 Challenges

A major focus of the user studies was to identify challenges in cyber security to explore how visualizations may enable users to meet those challenges. There was not a prescribed set of challenges that users chose from, instead challenges were compiled from study results and then validated with cyber SMEs. Some of these challenges are not surprising, involve the "big-data" problem (velocity, volume, and variety), and have been previously documented in the literature (such as challenges 1-3). However, other identified challenges hold opportunities for further research in visual analytics to enhance the abilities of the personnel defending our networks. By understanding the challenges faced by cyber defenders, a visualization can be designed to help address them, increasing the tool's usefulness for the user. Challenges identified in the study are discussed below.

4.2.1 Lots of Data

The common "big data" challenge also holds true for cyber SMEs. This challenge is exhibited in the large volume, variety, and velocity seen in many domains that have extreme amounts of data to sift through to obtain insight. The challenge this poses to cyber security is that users will only compare a small subset of data (often defined as a small temporal window) when analyzing an event. Even with an aggressive sub-setting work-around, query times and computational complexity of correlation algorithms contributes to other issues of interactive exploration and investigation of potential threats. An alternative culture that has emerged is spending a significant amount of effort formulating queries, so as to avoid placing too much load on the server, which is used by many defenders simultaneously.

4.2.2 Lots of Data Sources

The users we spoke to voiced their concern of fusing many data sources, including FireEye, Solara, HPGary, Mandient, Antivirus, web tools, system events, and network traffic. The diversity and quantity of this information provides challenges. Often, users can only look at short temporal subsets of the information when they directly query for it. The dynamic nature of this data means that the analysis becomes discrete and focused on one particular time-frame, and comparing to previously observed (or not observed) events is challenging. Further, understanding the implications of anomalies in each of these data sources requires expertise that not all defenders possess. This could also include understanding when events or alerts between these data sources are actually correlated, and are caused by the same threat or event.

4.2.3 Data Sources Not Linked

The data sources, (e.g., Netflow, firewall logs, system logs, routing information, ACLs, ARP tables, process lists, web data and logs, packet data, streaming data, encrypted data, etc.) prove challenging not based on their unstructured nature (as most are structured), but in correlating the different and diverse data. Each data source has structure, but synthesizing the meaning of correlations between the datasets can be challenging. Often, users have to understand that events in one set of logs exhibit mirroring events in another set of logs, and so on.

4.2.4 Data Quality

Users voiced their concern about the quality of the data they use for their analysis. These concerns include storage (log size limit truncating data), delivery (data delivered via UDP), invalid or corrupt data, and confidence in historic data. Of these, the last concern is of particular interest, because it hints at the possibility that big data (or just more data) does not necessary indicate that the users will have more confidence in the information (based on all the above concerns). This contradicts how confidence can be calculated statistically, so it reveals an opportunity to leverage how users create this confidence based on their domain expertise.

4.2.5 Cadence of the Network

Users tasked with monitoring the status of a network exhibited the ability to understand the health of a network based on the "usual broken things" that they regularly spot on the network. That is, there is always particular events or threats on a network, and the ability for them to understand the "cadence", or pattern, of these events enables them to understand the network status. This opens the opportunity of analyzing a network similarly to how a living ecosystem is monitored, where a balance occurs between understanding the impact of an event, and how the ecosystem as a whole adapts/responds, and at times, self-heals. The defenders we spoke to have an inherent understanding of what constitutes a typical amount of errors, as well as general data-production of a network. They understand the times that specific data sources update or populate, machine reboot schedules, server latencies, and other qualitative information or knowledge about an inherently quantitative, digital network.

4.2.6 Progression of Threat Escalation

At a high level, the progression of how events get handled is: detect, investigate, resolve. That is, the event gets detected, an analyst is tasked with investigating the event further; if a threshold (of either certainty, or time to gather more information) is reached, an action is taken (such as rebuilding a machine). The role of the analyst in this situation is balancing the potential impact of spending additional time to investigate the threat with the cost of rebuilding the particular user's machine. This is an area where more research can be done to understand how technology can be used to monitor this process, and help users recover their provenance to use for future investigations, or to help them report their process for their given case. For example, understanding whether the similar symptoms observed on two machines are caused by the same threat, and if that's the case, what actions were taken on the prior machine to counter the threat? Did that action work?

4.2.7 Balancing Risk and Reward

There are two factors that analysts balance during the investigation phase of understanding an event: confidence of threat and cost of gathering additional information. Simply put, the analyst's goal is to increase their confidence of understanding the threat, while minimizing the cost (in terms of time, potential threat, etc.) needed for gathering this additional information. If data can be provided to an analyst to inform them about potential risks or rewards for a current line of inquiry, it allows them to understand where the boundaries lie to justify a particular action. This particular challenge can be aided by visualization approaches, because it involves quickly sharing to others the context surrounding an investigation in order to describe to them the need to utilize extra time for investigation, observation, and other analysis prior to removing a machine from the network. This is a subjective decision, and the process of making the decision can be aided.

5. CHALLENGE DISCUSSION

From the defender challenges, potential impacts to a visualization were identified to inform the development of an application targeting situational awareness for cyber SMEs. These impacts go beyond our single solution and can be applied to other applications to help meet the needs of cyber defenders.

5.1 Lots of Data

Lots of data impacts any visualization tool in many different ways. The clearest of which is the speed of interaction. While not specifically a visual element, the speed at which users can use the tool to answer questions directly influences their view of the tool. Any visualization must have quick interaction to be utilized by users. Preserving interaction speed has been accomplished various ways throughout the years, most often the mantra of "details on demand" (DOD) is leveraged [14].

At the core of the challenge the visualization must provide the answers to the user's most common question at the default level. While "overview first, details on demand" may work in many cases, if a user must retrieve details every time, after only a brief moment on the overview, then time is wasted on the overview. Instead, a starting view exposing the majority of what the analyst needs and overview or further details on demand provide greater utility. A visualization is used to inform; if the user cannot get their information needs immediately addressed by the visualization, then it misses the mark. Perhaps approaches for visual analytics, such as the one by Keim et al., can more appropriately describe and meet the needs of cyber security [10]. For example, they advocate for analyzing first to show what is important, and using the visual analytic system to explore further.

5.2 Lots of Data Sources

No single data source will suffice for answering the questions that users ask in a cyber security visualization. Requiring users to open many different data sources, completely disjoint, is not a valid solution and will greatly decrease the utility of the tool. When possible, the visualization should include multiple data sources and provide some mechanism to tie information together. Lessons can be learned from Security Information and Event Management (SIEM) tools leveraging some of the best practices in that space. Some ways to integrate with multiple data sources are 1) show the data in the tool and link and 2) use the data sources for some analytic process and visualize the result of that process (ensure the users know about the sources being leveraged).

5.3 Data Sources are Not Linked

Data sources not being linked is highly related to the challenge of having many sources, and the impact for that challenge applies here as well. However, the linking of sources has more bearing on this challenge. To address it appropriately, the visualization must link sources in meaningful ways. A simple, but effective, mechanism is to tie through a common parameter between data sources. The primary overlapping parameter is time, the visualization should make it easy to compare, search, and visually line up time elements of multiple data sources. If using coordinated views, whenever possible, the time axis should be at the same scale to ensure visual similarity. Other known overlapping attributes such as IP address, port, or similar should be linked visually through brushing and linking. Linking by visually distinguishing a selection based on an attribute will allow for the investigative process.

5.4 Data Quality

The lack of confidence in the data is not surprising based on the descriptions of data quality issues. Many are simple network up-time issues; others are storage limitations, or hardware failures. Incomplete data is an expected state and needs to be accounted for in the visualization. Confidence in data plays an important role in the day-to-day decision making process of the cyber analyst. Throughout the various sources that are used in the visualization, the lack of data or the confidence in correctness should be conveyed to the user. An example of this impact would be to indicate missing data in a way that users can tell data is missing rather than filling with a default value (e.g., '0' for a numeric field). The default value may have different meaning than an alternative representation, such as marking the background of a chart with black for any time periods with missing data. Confidence in data is a much more complex concept to account for. One element is to highlight the age of the information, since SMEs expressed that they have less confidence in older data sources. Other confidence measures may need to be provided by the analysts themselves, either

by annotating a data source or through a source configuration mechanism. Whichever the process for gathering the confidence measures, that confidence should be conveyed to the users of the application.

5.5 Cadence of Network

The visual impact for cadence is that 1) the temporal events need to visually correlate and 2) user events and systems events need to be distinguished whenever possible. Visually correlating temporal events can be resolved in a similar way as the linking of multiple data sources. When visually representing events along a timeline, it is important to have the various data sources be on the same timeline for quick review by users. Distinguishing between user and system events is a difficult problem to solve during the data analysis portion. Knowing that a packet belongs to a user-requested communication rather than the host system's many processes will take great insight to determine. Assuming that the data can be separated, the visualization should distinguish between user and system. One technique to expose the difference is to create two visual elements per data source, one for the machine and another for the user, maintaining the same temporal scale. Another mechanism would be to utilize varying color or iconography for the two types of data for a given data source. This would allow the information to be compressed if space is limited due to the multiple data sources as mentioned in other challenges.

5.6 Progression of Threat Escalation

The challenge of threat escalation progression is one of workflow, experience, and provenance. Incorporating solutions for this type of impact is not as visually intensive as is human computer interaction (HCI) intensive. Users need a mechanism to record state, annotate findings, and to have those annotations shared and brought up automatically where possible. For example, if a user annotates that a particular pattern is of interest then that pattern should be highlighted for different data sets. Also, transitioning between the stages of investigation should be supported by the application as long as it is within the scope of the application. If the purpose of the application is to allow for exploration and identification of threats, then allowing the user to track that they are working on a hypothesis, recording supporting information, transitioning to an incident, and transitioning to remediation should be supported.

5.7 Balancing Risk and Reward

The visual impact of balancing risk and reward is a difficult challenge to limit to a few elements in the application. The goal of the visualization is to present available data to assist in the confidence building process. The data must be presented "as is" without misrepresentation, since that skewing of the data may directly impact the confidence building process. On the risk side, a potential means to convey information is to provide more than simple IP addresses. Often applications stop at the IP address, where analysts need the purpose of the machine, who uses it, and where it is located. Providing this information seems like a simple endeavor, however the correlation is difficult due to the complexity of today's network structures. Elements of confidence and data linking can be used to assist with conveying supporting data for risk. For example, host information for an IP address that is often used for the VPN gateway should show lower confidence than an IP that is statically assigned to a server that has been registered.

6. SEQUESTOR APPLICATION

The aforementioned studies were, in part, support the SEQUESTOR project at the PNNL; which has the need to present cyber security related data to network analysts for awareness. SEQUESTOR's primary goals are to model the behavior of human-computer pairs, take soft quarantining actions when behavior deviates from "normal", and present information to analysts for shared situational awareness and sense-making, model steering, and model development. The results from the studies, along with many prototype views, were used to design and build an interface to meet the needs of analysts. Close attention was paid to address the identified challenges and to begin helping analysts accomplish their goals in cyber security network defense and research. SEQUESTOR leverages multiple behavioral models and a visualization to provide awareness into those models and enable investigation.

A research component of SEQUESTOR is to develop behavioral models such that the system maximizes the user's ability to work and accomplish their analytic tasks, while minimizing the ability of an adversary to compromise and use a network account or system. In this system, cyber analysts need to be part of the decision making process and adaptation of models in the environment, due to the complex reasoning requirements. Behavioral models traditionally have a high false positive rate, therefore cyber analysts needed a shared representation of the problem space to investigate and adjust behavior model parameters accordingly. One of the foundations of any shared cooperative human-assisted automated work is a shared representation of the problem situation [9, 11]. In human-machine cooperative work, a common finding is that people continually work to build and maintain a "common ground" of understanding of the monitored process in order to support their problem solving efforts or cognitive work [13]. Following are four representation design principles that balance human-machine coordination with teamwork used in exploring and developing SEQUESTOR visual concepts as prototypes for discovery of what's promising [15].

- **Observable**: Feedback needs to be provided to the observer to gain insight into the behavioral models and quarantining response. Data is integrated based on the model of the process and aligned to reveal patterns and relationships. Context must be provided around details of interest, allowing greater understanding. The view should sequence and evolve over time, showing future activities and contingencies.

- **Directable**: The ability to direct/re-direct resources, activities, priorities as situations change and escalate.

- **Team Work with human (analyst) and machine (computer) agents**: The observer needs the ability to coordinate and synchronize activity across agents by having a common ground and a shared frame of reference. Data becomes informative or meaningful based on its relation to other data and the observer's interests and expectations. To assist in teaming, conceptual spaces can be built by depicting relationships in a frame of reference. Coordination begins with seeding

structure and kick-starting initial activity. Other possibilities should then be suggested to the observer as activity progresses to continue coordination. As activities come to a close, alternatives should be pointed out. Additionally, making other agents' models, intent, activities observable assists in coordination and team work. Finally, delegation can occur by re-directing agent resources as situations change.

- **Resilience**: The ability to anticipate and adapt to potential for surprise, error, and failure of sensitive strategies and tactics. The observer needs to be able to explore outside current boundaries and thresholds, and to overcome brittleness of automata. The system should make shared cognitive work observable (yours and others problem situation), and revise focus and avoid fixation.

The SEQUESTOR visualization (SEQViz) is designed to provide shared situational awareness and investigative capabilities necessary for cyber analysts to leverage the SEQUESTOR models and soft adaptive quarantining technologies. Early on in the design process, a number of conceptual sketches were developed, in tandem with the user studies, which aligned with the project's initial requirements. However, once the study concluded, all the conceptual sketches were assessed against the challenges identified and scored based on their ability to meet each challenge. From that evaluation, additional conceptual prototypes were designed in order to meet not only project requirement but to also address as many identified challenges as possible. We then evaluated any new concepts to ensure the challenges were addressed in some manner. The concept that best fit to requirements and addressed all the challenges, at least in part, was then developed further into a functional prototype. While the utility of the prototype visualization is still being assessed; there is a great potential, for any cyber network defense tool, to meet the needs of defenders by mapping their challenges into initial concepts to ensure those needs are met.

6.1 Concept Sketching

Initial design sketches of SEQViz started with familiar visual metaphors in cyber security network visualization; such as graphs, bar charts, line chart, etc. A turning point for concepts occurred when one of the very first concepts (Figure 6.1) that leveraged a node-link diagram with a progressive disclosure capability was reviewed by the client. The client requested to think beyond node-link and other "easy-fit" concepts, and help tell the story of how the user's system got into a particular state, and to provide adequate information to allow the analyst to reason about what is happening on the network.

In total, nine concept sketches were developed both as a composite of multiple views and as an overall application design. The sketches were informally evaluated by cyber security defenders and analysts both at PNNL and at the client organization. In most cases, sketches were purposely left as copies of hand drawn designs to elicit feedback from participants and to allow them to draw on the images provided to them (rather than refined representations). As feedback was gathered more views and designs were explored. Once the research on potential concepts was complete, we evaluated each on their ability to meet the seven challenges identified

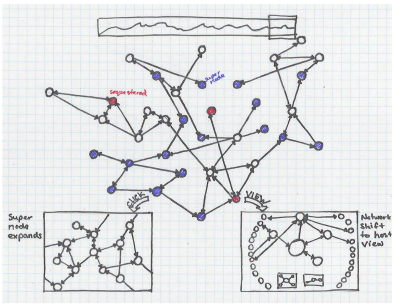

Figure 1: Initial sketch that led us to abandon "easy" solutions

in the user studies. Conceptual designs include:

1. Model of Models: Visualizes multiple behavior models for an awareness of how raw data matches up to established patterns.

2. Concentric Circles Views: Displays communication patterns between systems using concentric circles

3. SEQUESTOR Level / Radial: Highlights an overall view of the network, indicating the threat level for a given user-device pair.

4. Network Overview: A high-level starting point for investigation. The user-device pairs are separated into categories and subcategories based on their behavior.

5. Galaxy: Visually groups systems and provides indication of interaction between the clusters.

6. Host Attribute View: Groups systems together based on function and provides indications of behavior. Overall threat level is shown as a radial plot, and trends are shown using difference and candle plots.

7. Node-Link View: Provides an overall view of network connectivity, data sources used by SEQUESTOR, and drill-down to a single host.

8. Network Cadence: An exploration of using common visual glyphs, including musical elements, to indicate the cadence or "tempo" of the network.

9. Network (with Icons) View: Provides not just connectivity information but also the current activities of the systems on the diagram. Like-activity n-grams are grouped together, providing a rudimentary clustering in this view.

The assessment of concept sketches meeting the challenges indicated that "Lots of Data","Lots of Data Sources", and "Linking Data Sources" were well represented across the prototypes, but the remaining challenges were not. This is likely due to the timing of the sketch design and completion of the user study where the big data concepts were known a priori while cadence, threat escalation, importance of balancing risk and reward, and mixed data quality were not as well understood.

6.2 Culmination

The lack of a single design that met all the challenges, and several challenges not addressed at all; it was clear that additional prototypes were needed to meet the identified challenges and the needs of the client, and to provide a viable solution for situational awareness. The initial prototypes were assessed on how they met the challenges, and what features best addressed the challenges from the various views. The assessments, along with the knowledge gained from the studies, provided us with the resources necessary to design the current functional prototype.

SEQViz has two coordinating views, the network overview view and the model view. Both views expand and collapse to allow greater detail to be shown for the view of interest while keeping the other view in context. As part of the model detail view, there are several sub-components (such as property and user information, and threat triage) to provide more detail and functionality to assist with analysis.

6.3 Interface Components

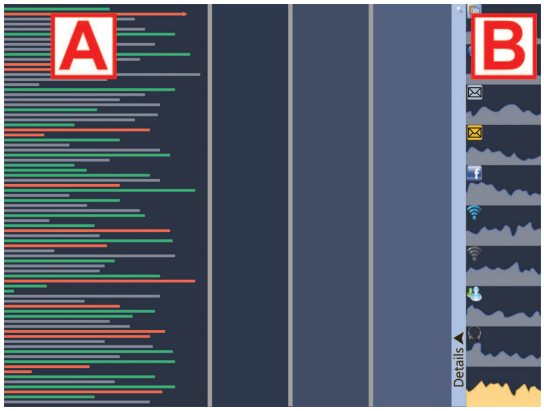

Figure 2: SEQViz showing the network overview presentation

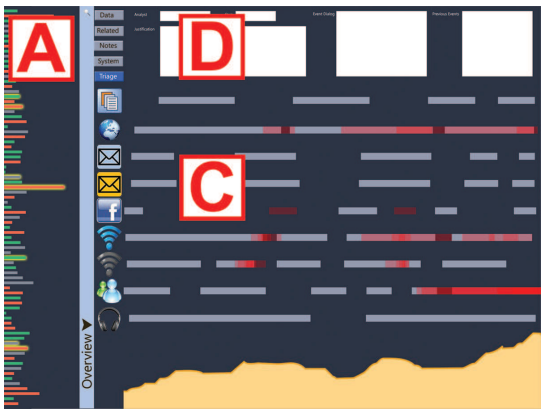

Figure 3: SEQViz showing the detailed model view and supporting information view

6.3.1 *Network Overview*

The Network Overview (A), in both expanded and collapsed views, depicts individual user:device pairs, or groups of similarly categorized user:device pairs (such as by job role). The Network Overview provides high-level awareness of what is occurring on the network and enables correlation of activity between different user-device pairs through the use of brushing and linking to visually tie items together.

6.3.2 *Model Thumbnail View*

The Model Thumbnail view (B) is shown when the Network Overview is expanded. The thumbnail shows temporal trends of each of the models based on the selected user-device pairs. When no user-device pair is selected, the thumbnail models show the historical information of the network as an aggregate.

At the bottom of the Model Thumbnail view is the overall model thumbnail. This chart depicts the historical trend of the overall model (a combination of the other models). Again, this model reflects information based on the current selection.

6.3.3 *Model Detail View*

When the Model Thumbnail view is expanded to show the Model Detail view, the overall Network View is collapsed and the model information displayed is more detailed (C). Instead of thumbnail plots for each model, a plot showing activity and the level of alert for each model is displayed. Again, at the bottom the overall model graph is shown but at greater detail. Selecting a model in the Model Detail view can overlay the contribution of that model to the overall model to provide analysts with greater insight into how the model is built.

6.3.4 *Detail Depth Sub-View*

At the top of the Model Detail view is a sub-view (D) to provide greater detail and functionality to the application. Various views and information are shown, depending on the selection, and the selected depth tab. Detailed raw data, triage support for workflow, notes about investigation, and information about models are all possible sub-views that can be shown in this space.

6.4 Addressing the Challenges

The coordinated views provided by the visualization allow for users to manage the amount of data that is available to them. If a user wishes to maintain overview awareness, they can expand the network overview and either monitor the network as a whole, or a subset of systems. Alternately, a user may maintain awareness at the model detail level while keeping the network overview in context (allowing them to catch unexpected threat progression).

Multiple data sources are brought in, linked, and represented in the detail view to provide context for the behavioral models. Data sources are intended to go beyond "traditional" cyber data sources, such as property tracking data and user information, to help build context around a given computer system. For the data sources, to provide an accurate picture of activity, age and quality of data can be represented in the detail view sub-component through the use of visual indicators such as saturation. Lining up several different data sources on a common timeline allow analysts to recognize the cadence for a given system and compare quality concerns across data sources. Making data available allows risk and reward to be assessed by determining who the user is, what type of system is being triggered, and what

type of activity is currently happening on the system.

The multiple views providing context to a potential threat addresses many of the identified challenges from the user studies. To address threat escalation, we propose integrating reporting and progression into the tool in the detail view. Cyber analysts would start an event, provide notes, and move into identified states of investigation. This allows tracking of provenance and reduces context switching while leveraging the application.

7. CONCLUSION

Several of the challenges identified in the user study are common knowledge for cyber visualization developers, and can be seen in other domains that work with big data (volume, velocity, and variety). However, the additional challenges identified are equally important to consider to ensure an application meets the needs of cyber analysts and defenders. Being able to leverage human cognition to identify anomalies, progress threats through a workflow, and provide risk and reward supporting information allows analysts to accomplish their responsibilities and provide great utility. SEQViz may have been successful without meeting the challenges identified; however it would run the risk of being "just another tool" and end up unused. By considering all of the challenges, we are able to design a visual analytic environment to enable analysts to meet their needs and accomplish their goals.

Figure 4: Early version of working SEQViz application

8. ACKNOWLEDGMENTS

The research described in this document was sponsored by the U.S. Department of Defense (DoD) and Department of Energy (DoE) through Pacific Northwest National Laboratory (PNNL). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Government.

9. REFERENCES

[1] N. Chinchor, K. Cook, and J. Scholtz. Building adoption of visual analytics software. In J. Dill, R. Earnshaw, D. Kasik, J. Vince, and P. C. Wong, editors, *Expanding the Frontiers of Visual Analytics and Visualization*, pages 509–530. Springer London, Jan. 2012.

[2] K. A. Cook, L. Boek-Peddicord, L. Dudney, L. Nowell, and J. Scholtz. Lessons learned and best practices in technology transition. Technical report, Pacific Northwest National Laboratory, 2008.

[3] A. D'Amico, L. Buchanan, J. Goodall, and P. Walczak. Mission impact of cyber events: Scenarios and ontology to express the relationships between cyber assets, missions, and users. Technical report, Dec. 2009.

[4] A. D'Amico and K. Whitley. The real work of computer network defense analysts. In J. Goodall, G. Conti, and K.-L. Ma, editors, *VizSEC 2007*, Mathematics and Visualization, pages 19–37. Springer Berlin Heidelberg, 2008.

[5] A. D'Amico, K. Whitley, D. Tesone, B. O'Brien, and E. Roth. Achieving cyber defense situational awareness: A cognitive task analysis of information assurance analysts. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 49(3):229–233, Sept. 2005.

[6] G. Fink, C. North, A. Endert, and S. Rose. Visualizing cyber security: Usable workspaces. *VizSec*, 2009.

[7] J. Goodall. Visualization is better! a comparative evaluation. In *6th International Workshop on Visualization for Cyber Security, 2009. VizSec 2009*, pages 57–68, 2009.

[8] J. Goodall, A. D'Amico, and J. Kopylec. Camus: Automatically mapping cyber assets to missions and users. In *IEEE Military Communications Conference, 2009. MILCOM 2009*, pages 1–7, 2009.

[9] B. J. Grosz. Focusing and description in natural language dialogues. Technical Report 185, AI Center, SRI International, 333 Ravenswood Ave., Menlo Park, CA 94025, Apr 1979. Published in Elements of Discourse Understanding: Proceedings of a Workshop on Computational Aspects of Linguistic Structure and Discourse Setting.

[10] D. A. Keim, F. Mansmann, J. Schneidewind, J. Thomas, and H. Ziegler. Visual data mining. chapter Visual Analytics: Scope and Challenges, pages 76–90. Springer-Verlag, Berlin, Heidelberg, 2008.

[11] J. C. McCarthy, V. C. Miles, and A. F. Monk. An experimental study of common ground in text-based communication. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '91, pages 209–215, New York, NY, USA, 1991. ACM.

[12] G. A. Moore. *Crossing the Chasm: Marketing and Selling Disruptive Products to Mainstream Customers*. HarperCollins, Aug. 2002.

[13] E. S. Patterson and D. D. Woods. Shift changes, updates, and the on-call architecture in space shuttle mission control. *Computer Supported Cooperative Work*, 10(3-4):317–346, Dec. 2001.

[14] B. Shneiderman. The eyes have it: a task by data type taxonomy for information visualizations. In *Proceedings of the IEEE Symposium on Visual Languages*, pages 336 –343, 1996.

[15] D. D. Woods. Visualizing function: The theory and practice of representation design in the computer medium. Technical Report CSEL 2000-TR-02, Cognitive Systems Engineering Laboratory, Institute for Ergonomics, The Ohio State University, February 2000.