Toward a Visualization-Supported Workflow for Cyber Alert Management Using Threat Models and Human-Centered Design

Lyndsey Franklin* Meg Pirrung† Leslie Blaha‡
Pacific Northwest National Laboratory

Michelle Dowling§
Virginia Tech

Mi Feng¶
Worcester Polytechnic Institute

ABSTRACT

Cyber network analysts follow complex processes in their investigations of potential threats to their network. Much research is dedicated to providing automated decision support in the effort to make their tasks more efficient, accurate, and timely. Support tools come in a variety of implementations from machine learning algorithms that monitor streams of data to visual analytic environments for exploring rich and noisy data sets. Cyber analysts, however, need tools which help them merge the data they already have and help them establish appropriate baselines against which to compare anomalies. Furthermore, existing threat models that cyber analysts regularly use to structure their investigation are not often leveraged in support tools. We report on our work with cyber analysts to understand the analytic process and how one such model, the MITRE ATT&CK Matrix [42], is used to structure their analytic thinking. We present our efforts to map specific data needed by analysts into this threat model to inform our visualization designs. We leverage this expert knowledge elicitation to identify a capability gaps that might be filled with visual analytic tools. We propose a prototype visual analytic-supported alert management workflow to aid cyber analysts working with threat models.

Index Terms: H.1.2 [Information Systems]: User/Machine Systems—Human Factors; H.5.2 [Information Interfaces and presentation]: User Interfaces—User Centered Design

1 INTRODUCTION

Cyber analysts have long been tasked with protecting their networks from the malicious activity of internal and external threats. They must always be searching for indications that a never-before-seen technique is at work against their defenses. The tools at their disposal are often specific to a particular data type, resulting in an expanding number of tools between which an analyst has to switch during the investigative process. Indeed, Silva et al. reported finding 75 unique tools in use during a single cyber exercise [35]. Many cyber analytic tools are designed to be general purpose enough to be reused at different investigation stages but are not always used that way.

In addition to the mental challenges inherent an investigation, cyber networks are never static. They are constantly changing, updating, and generating more data than can possibly be captured or analyzed in a timely manner. They are a prime example of a streaming data system [12, 14] and many tools cannot keep up. Malicious activity is itself a moving target. Threats change over time as adversaries adopt new tactics and make use of newly-discovered resources to evade defenders. Several efforts have been made to understand and model cyber attacks. One such model is the MITRE ATT&CK Matrix [42] which organizes attacker techniques into categories. Unfortunately, few analytical tools directly leverage such frameworks. Work is needed to connect theoretical foundations and frameworks to the available data sources, to the analysts' workflows, and to institutional policies for handling incidents [10, 34].

We argue that visual analytic tools for cyber defense should engage a user-centered design process that combines cyber models, like the ATT&CK framework, with task analyses of subject matter experts in consideration of the end-to-end defense workflow. Herein we present a preliminary design space targeted at meeting open gaps identified by cyber analysts through such an approach. We begin by describing the results of two knowledge elicitation interviews with expert cyber analysts. Our first interview focused on understanding the overall context of their daily tasks, with particular emphasis on what led to and came of investigations into cyber activity. In the second interview, we focused on understanding how specific data types and indicators of network activity were used in practice, mapping the use into the categories of the ATT&CK Matrix. We describe the gaps we discovered and the design spaces that the cyber analysts felt would help with their investigations. We leverage this information as part of a user-centered design process and describe a prototype interface which combines machine learning and visualization to close an alert management capability gap. We conclude by discussing future design and evaluation word toward an integrated human-in-the-loop alert management workflow.

2 RELATED WORK

Cyber security research is a diverse field with significant contributions ranging from theoretical modeling and algorithms to practical tool development. Human-computer interaction, visualization, and cyber workflows have each received significant study [18]. We draw inspiration from recent work on the human factors of cyber security [28] and visual analytic efforts to augment our work with professional cyber analysts.

2.1 Human-Centered Cyber Security

We are not alone in our human-centered approach to the study of cyber security. A number of user-centered design techniques have been employed to develop various tools for cyber security professionals [30]. Others have conducted interviews and field observations of network operation centers [17, 32] and even immersive anthropological approaches [38]. They describe highly-collaborative environments where verbal and co-located interactions are preferred [32]. They also note that because of the preference for verbal communication, a lack of recorded artifacts posed challenges for knowledge transfer between analysts [32]. Other efforts have focused their ethnographic efforts on information technology (IT) professionals practicing security management [9]. Botta et al. made use of surveys and semi-structured interviews with 14 professionals to elicit human, organizational, and technical aspects of security management [9]. Pattern recognition and inferential analysis were identified as two critical skills amongst IT professionals [9]. Botta et al. also describe a number of tasks and tools that constitute IT security management [9]. Our interviews yielded a similar list of tasks in what our cyber analysts referred to as *alerting* tasks (see section 3.1).

*e-mail: lyndsey.franklin@pnnl.gov
†e-mail: meg.pirrung@pnnl.gov
‡e-mail: leslie.blaha@pnnl.gov
§e-mail: dowlingm@vt.edu
¶email: mfeng2@wpi.edu

978-1-5386-2693-1/17/$31.00 ©2017 IEEE

D'Amico and Whitley utilized cognitive task analysis to understand the workflow of computer network defense analysts [16]. They described cognitive transformation as a hierarchy of data filters which produces intrusion data sets from raw network data. They also described a triage and escalation process similar to what we encountered in our interviews with cyber analysts, wherein only a few suspicious activities are escalated for further analysis [16].

In other work, a specialized kill chain model was successfully used in defense of a network [25]. The incident response team was able both to mitigate damage from a compromised mobile asset that moved out of their protected environment and to identify compromise indicators that would have been unavailable without the kill chain model [25]. We leverage a similar threat model identified by our cyber analysts.

2.2 Visualizations for Cyber Security

Cyber security visualizations range from dedicated displays of very specific information to powerful, exploratory visualizations. Several enterprise security tools provide visualizations as a feature (such as [37] and [2]). Research efforts can also range from specialized (e.g., [3] or [6]) to ambitious (e.g., [31] or [24]). Many are efforts to visualize machine-learning components of specific types of network activity. One problematic aspect of cyber security visualization research is that it often does not address the common tasks cyber analysts and defenders routinely perform. Best et al. have noted that despite the availability of visualizations, simple command-line tools are still commonly used [7]. They lamented the ease in developing "just another packet visualization" without addressing deeper needs of cyber analysts. The cyber analysts targeted by such efforts mistrust visualizations. They have significant concerns about misleading data representation and objections to a typical visualization system's lack of interoperability with other tools [17]. Our research goals include producing useful visualizations for cyber security; however, we heed the warnings of [7] and begin this work by conducting interviews with cyber analysts.

3 Interviews with Cyber Defenders

Our expert knowledge elicitation process began with semi-structured, group interviews with four cyber analysts. Three of the analysts had 5 to over 15 years of cyber security defense and analysis experience. The fourth had 7 years of general information analysis experience and had begun working as a cyber analyst within the past two years. Most of the analysts had computer science and IT backgrounds but had completed their institution-specific cyber security training "on-the-job," rather than as a formal process. These cyber analysts were co-located and used similar tools to complete their work protecting a medium-sized network of approximately 20k active IP addresses and 4k users. These common tools included security information and event management (SIEM) software, log viewers such as Nagios [2], threat detection systems like FireEye [1], malware analysis tools such as Yara [4], and software center configuration management (SCCM) software to list a small sampling. The majority of tools used by the analysts were text and command line-based investigation support programs or custom, self-created scripts. The team's regular duties include network policy work, tool development, network architecture design, and cloud infrastructures/non-traditional research network support. We met as a team, because incident response and forensic analysis of cyber data are duties shared across the team, even though analysts typically worked independently.

3.1 Daily Workflow and Task Elicitation

Our first meeting with the cyber analysts focused on identifying their daily activities, workflows, and standard go-to tools. We met for approximately two hours in a collaborative space with whiteboards and Post-It notes, away from the experts' regular workplace (and distractions). We began with a pre-defined set of questions (see Table 1) and asked follow-up questions for clarification. Similar to [9], we focused on understanding routine tasks, particularly their data inputs and outputs. The group interview was advantageous for our purpose as the experts began to talk comfortably in detail with each other about their work and even sharing techniques with each other. This allowed us to focus on note-taking and using Post-It notes to externalize our own notes for the experts to see. As the discussion continued, the experts also joined us in writing down tasks and tools onto Post-It notes. After an hour of discussion and note-generation, we re-focused our experts and asked them to help us group tasks and tools into meaningful categories. Four high-level categories of analytic and defense tasks emerged from this first interview: alerting, thresholding, threat hunting, and reporting.

For Each Identified Task

- What triggers this task?
- What displays and data are used?
  - List specific tools used
  - List specific data sources needed
- Who else is involved with this task?
  - People you provide reports to?
  - People who you delegate to?
  - People who you consult with?
- What steps/actions do you take to complete this task?
  - How long does it take to complete?
  - How do you know when it's done?
  - What are the pain points?

About All Tasks

- Are there tasks that you wish you could do better?
  - Which tasks get rushed?
  - Which tasks are high vs. low priority?
  - Which tasks get ignored until there's a problem?

Table 1: Knowledge elicitation questions for understanding the daily tasks performed by our cyber experts. These formed a template and were repeated for every task discussed during interview 1.

3.1.1 Alerting

Alerting tasks involved the immediate or near-immediate investigation and response to specific network activity that violated some pre-defined rule or set of rules. Alerts arrived in a variety of ways. Most often, analysts received a notification in a rule-based network tool. These are prone to high volumes of false positive alerts amongst the true positive alerts, creating large numbers of alerts competing for analyst attention and resources. Analysts could also be tasked by each other with an investigation of unusual behavior that had been noticed during other tasks (e.g., threat hunting, described below).

3.1.2 Thresholding

Thresholding tasks include maintenance activities such as process optimization, cleaning data, designing new alerts, and, infrequently, behavioral analysis. Tasks in this category are crucial for improving the quality of alerts and data that analysts work with in their other tasks. In particular, thresholding tasks are the process through which analysts exclude "known good" activity from triggering alerts (reducing false alarms). That is, thresholding tasks often result in new or modified network rules that reduce the amount of false alarms that cyber analysts have to respond to.

3.1.3 Threat Hunting

Threat hunting tasks involve the long-term analysis of network activity for trends and targeted exploration of activity to find indication of the "known bad". The goal of threat hunting is to identify known

malicious behavior patterns or activity that cannot be automatically detected through rule-based alerts. Threat hunting can also include activities such as vulnerability assessment, which involves testing a system for known attack patterns.

3.1.4 Reporting

Reporting tasks often conclude an investigation or other tasks and involve authoring reports and the creation of any charts/figures/tables needed to supply evidence to a report. This process frequently involves finding some way to capture slices of data and analytic activity to present results from multiple tools in a single document. Reports are often compiled into summary presentation materials for presentation to a manager or superior.

Alerting, thresholding, and reporting tasks build from each other in a predicable cycle. Alerts serve as a launching point for investigations. Through the analysis of a given incident, cyber analysts learn more about the root cause of the incident and ways in which similar incidents can be detected earlier in the future. This knowledge is reported at the conclusion of an investigation, as lessons learned, and results in new rules or modifications to existing rules. That is, *alerting* tasks trigger investigations which generate knowledge that is *reported* on and leads to better *thresholds*. This workflow forms the general narrative of a cyber analyst's daily tasks. The cyber analysts described threat hunting tasks as ones completed "as time allows" and considered it unfortunate that they did not spend more time pro-actively looking for threats to their networks. Often, threat hunting tasks are rare, semi-annual activities included with vulnerability testing efforts. Alerting, thresholding, and reporting tasks tend to have higher organizational priorities than general threat hunting-improving how those tasks are completed would free analysts to spend more time threat hunting. These results are consistent with the findings in [16] which suggest that the majority of cyber analysts spend their time handling incidents (alerting), producing technical documents (reporting), and security configuration administration (thresholding). In the next section, we examine how investigations are conduced in greater detail.

3.2 General Investigative Workflow Elicitation

After we discussed the general daily patterns of their regular duties, we asked the cyber analysts to elaborate on the workflow of investigations, specifically. Investigations are a process of establishing the root cause of an alert and collecting relevant evidence of an event. They invoke a sensemaking [27, 33] process over noisy and incomplete data, frequently in the form of event logs. One cyber analyst lamented, "most of our data is functionally useless." At the same time, the analysts felt that they did not have enough information but that "[they] also [did not] know what information [they] needed." This speaks to an underlying reason most cyber security professionals find it critical to have access to raw data [7]. That is, without knowing what will be important as evidence later, analysts conducting an investigation are constantly torn between filtering away noise and preserving a signal.

3.2.1 Alert Triage

The cyber analysts described investigations as a reactive workflow that begin with the arrival of an alert indicating a specific network rule had been violated in some way. Many alert systems are based on manually curated, static rules. This has the unfortunate side effect of generating a high volume of false positive alerts for benign activities. Thus, analysts first go through a process of verifying and validating that a given alert needs additional attention. A typical alert will provide enough information for analysts to discern which rules were violated and which network resources are involved. Knowing this allows analysts to perform their initial triage [13] about the severity of the alert. Analysts will formally indicate that the alert has been seen and indicate that it is not a threat. False positives are often immediately resolved in the alert-system.

Alerts requiring additional attention are then initially scoped by responding analysts to establish 1) what happened and 2) who is involved. This scoping process determines the level of response to a given alert. Organizational priorities, policies, and postures factor in at this stage of investigation [34]. Our experts noted that cyber analysts occasionally encounter incidents which they feel warrant full investigation. But if the incidents do not yet rise to an organizational priority level, they will remain un-investigated.

Alerts that meet organizational priorities for complete investigations are typically handled by a single cyber analyst. This individual begins to collect additional information about network users and resources involved in the original alert. Context is often missing in initial alerts. An analyst will spend considerable time collecting data about network users to establish an understanding the typical behavior of specific network users. This baselining process is labor intensive and requires human judgment: what is "normal" behavior for one network user may be completely out of character for another. It also requires analysts to actively search for patterns of user behavior similar to the activities that caused the initial alert that may have not met a strict rule definition.

3.2.2 Alert Investigations

Investigations will often branch out from the original data contained within an alert. For example, analysts investigating why a new piece of software is violating firewall rules (or attempting to) will often need to learn where the new software came from (license vendor, open source, etc.). This will expand the investigation to include other logs from a given network resource including registry information and process logs to establish when the new software was installed, authentication logs to learn which network user installed the software, and network data flow analysis to learn from where the software was installed. Following network information backwards in time is a manual process requiring analysts to search for subtle indications of activity that may be intentionally hidden. The process of discovering where new software comes from may trigger other information needs, such as the need to evaluate whether the software contains known malware signatures. Often, intermediate investigative results suggest additional lines of inquiry to analysts who must distill out a timeline of interrelated events that encompass much more than the original alert.

Through the investigative process, analysts will often make use of additional tools to externalize and organize their thinking. Whiteboards and Post-It notes are frequently used to keep track of both tasking and important pieces of information. While analysts usually complete investigations alone, they do consult each other from time to time. Most often they consult each other when they feel they have reached a "dead end" with a given stage of the investigation. The externalized portions of their investigation are used then to brief other analysts and explain the investigative analysts' process and thinking. Other tools are also leveraged to organize an investigation. Threat models are useful tools for guiding an analyst's interpretation of the collected data. Our cyber analysts referenced one such framework, the ATT&CK Matrix [42], as particularly useful for characterizing threat behaviors. We will discuss the analysts' use of this frame work in greater detail in the next section.

The concluding stages of an investigation are reached when an analyst has reasonable confidence that s/he has discovered the underlying causes of an alert and have an understanding of how to resolve the issues. Ideally, an analyst would be able to restore network resources to an operating state similar to their pre-alert conditions (albeit hardened to prevent additional rule violations). In some circumstances, analysts reach the end of an investigation without clear future directions. This can be a combination of policy directives or priorities (i.e., the underlying causes are not deemed worthy of

additional action). In other situations, analysts find that they do not have access to the data needed to complete the investigative process. When this happens, analysts may trigger other processes to acquire their needed data (such as obtaining additional permissions). They may also decide that the situation calls for continued monitoring to see if it progresses or escalates.

3.2.3 Alternative Investigative Processes

We asked the cyber analysts how their investigative process might differ under alternative circumstances. In particular, we asked whether or not they would proceed differently if they suspected insider activity (as opposed to an external threat). The cyber analysts indicated that the artifacts generated by the actions of a malicious insider were not measurably different from those of outside attackers. In fact, the evidence can look exactly the same. For example, a common attack pattern involves outsiders stealing the legitimate credentials of an authorized network user through techniques such as phishing. An outsider leveraging an authorized person's credentials would appear, in network traffic, as if the authorized person was operating in the network. The difference between malicious insider and outside attack is negligible at the network level. According to one cyber analyst, "the key to establishing the possibility of a malicious insider is the ability to link people, not just network credentials or identity, to [network] locations." Often, this requires data beyond what is captured in network activity alone [20, 21] and would halt an investigation until proper approvals were in place.

The cyber analysts described a workflow with activities that were both manual and labor intensive. We noted how much depended on an analyst's initial response to alerts and their immediate triage. In the next round of interviews, we focused on these alerting tasks and responses and structured our follow-on interviews to elicit how specific data led analysts to arrive at decisions about an incident.

3.3 Mapping Data to the ATT&CK Matrix

During the first interview, the cyber analysts made reference to the MITRE ATT&CK Matrix as one framework they used to provide structure to their investigations and analyses. The matrix is a threat modeling methodology and suite of models for the various phases of an adversary's lifecycle and has variations reflecting several major operating systems such as Windows, Mac OS, and Linux [42]. Similar to other attack pattern frameworks [39, 41, 43], it can provide context for describing attacks and help identify gaps in institutional defenses. It contains behavioral signatures that could emerge in cyber data during an attack. For example, "brute force" is one such technique found in the "Credential Access" category of the matrix. It is less clear from the matrix itself which data inputs and analytic outputs should be leveraged in the investigative process. The available data types may be constrained by the structure of the network and institutional policies; the outputs may be constrained by available tools, computational capabilities, and analyst expertise. This is the motivation for the second group interview: create a realistic mapping between available data sources, the MITRE ATT&CK Matrix, and the analysts' workflows to understand the capability gaps.

During the second interview, we asked two of the cyber analysts from our first interview to help us identify specific data that they would need to describe malicious network activity in the context of the ATT&CK matrix. We again met for two hours in a collaborative space with whiteboards and Post-It notes. We asked the cyber analysts to think about their process for using the ATT&CK Matrix. For a given analysis, we wanted to identify the following things:

1. What data did analysts need to assert an attack is or was occurring?
2. What data would support determination of the attack category?
3. What tool support did they have for investigating that data?

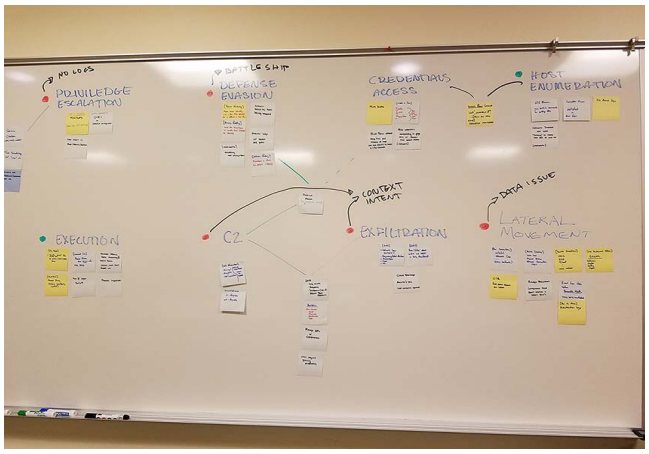

Figure 1: We asked cyber analysts to match available (to them) data types with portions of the MITRE ATT&CK Matrix [42]. Using whiteboards and Post-It notes allowed us to draw links when data or tools overlapped and to highlight gaps.

Our goal was to understand what information analysts needed to support alert triage and other decisions common to investigations. This interview let us identify specific tool gaps within the context of an investigation. We discuss several of these gaps as needs and constraints on visual analytic designs in the next section.

Table 2 lists several categories of the ATT&CK Matrix together with examples of some of the data or indicators that cyber analysts use to verify an attack. Different attack categories will have indicators in different types of data. Table 2 also mentions example tools used to examine this data. Event logs were repeatedly cited as the data type used across all parts of an investigation and all attack categories. Host data and network data were leveraged as evidence for activity on a network, while user data provided context to that activity. SIEM tools such as Splunk [37] were cited as the go-to tools for viewing and analyzing logs of all types. Our cyber experts leveraged Splunk for investigations of all types of attacks.

There were a number of attack categories which shared either data types or tool support. Persistence and Privilege Escalation can both be identified using task scheduling and service creation logs. Cyber analysts noted that the techniques allowing threats to persist themselves in a network could be modified to attempt to escalate privileges over time.

The protocol abuse technique was common to both Defense Evasion and Command & Control, implying similar data in these categories. For example, knowing a given port is open for one type of service allows attackers to make use of that port for other services (protocol abuse). The use of the legitimate port makes it more difficult for defenders to find the abuse and gives attackers the opportunity to issue commands or "talk out." Policy information about intended use of ports and documented exceptions are helpful here.

Credential Access and Host Enumeration (discovery) could both be identified by finding signs of process monitoring and domain reconnaissance in command histories. These techniques enable attackers to listen for the host names of other network assets and the credentials used to legitimately access them.

Exfiltration is a difficult stage of attack to detect but it does share some common data features and techniques with Command & Control. At such late stages of an attack, rogue devices can sometimes be found, along with the execution of user access controls and commands with non-human traits (too fast, too frequent, too large, too random, etc.)

During this interview, we learned about several investigative techniques that existed outside of the matrix and data types. As one

Category	Indication	Tools
Persistence	registry key changes, process execution, autoruns	host logs, SIEM dashboards
Privilege Escalation	new users in admin Groups, presence of masking software	manual analysis (indirectly captured)
Defense Evasion	software/binary padding, network encoding, failure of security tools	software signature matching tools, application log analysis
Credential Access	credentials unsecured in files, brute force attempts	data-loss prevention tools
Host Enumeration	presence of network scanners, file access logs	log analysis, SIEM dashboards
Lateral Movement	authentication logs, rogue processes, open share on host	manual analysis, SCCM software
Exfiltration	canary files, cloud storage use, rogue devices	frequencies/statistical analysis of network logs
Command & Control	data obfuscation, rogue devices, connectedness in-degree/out-degree	PCAP analysis tools
Execution	process injection, script autoruns, remote admin tools	program logs, SCCM software

Table 2: MITRE ATT&CK Categories, Example Data or Indications, and Example Tools

senior analyst said in the first interview, "[we] use human analysts because we don't have [automated] measures for everything yet." Manual investigations are often needed to understand what event logs were capturing. For example, discovering process abuse (Defense Evasion indication) requires an understanding of the intended use of a given port before malicious activity can be identified. Other signs of malicious activity, such as network traffic without accompanying host logs, typically require deeper analysis and forensics.

In terms of tool support, SIEM tools (e.g., Splunk [37]) provide great flexibility when it comes to viewing multiple data types in a single interface. The cyber analysts asserted that they typically knew from the nature of the alerts they received what stage a potential attack would be in. Often, the cyber analysts had their own, customized views within their log aggregation tools for viewing data once they had established where their analysis needed to begin. Creating such custom views and sharing them with the rest of the team is a hallmark of expertise on a given attack category or technique.

Our second interview provided insight into specific data needed during an investigation as well as demonstrated one way that information could be organized. We combine this mapping with the results from our first interview about daily and investigative workflows to establish an informed design space for tools supporting the investigative process. We next describe this design space.

4 Capability Gaps in Cyber Defender Support

The cyber analysts identified three stages of an attack as being particularly difficult to detect: Privilege Escalation, Defense Evasion, and Lateral Movement. A lack of log files which directly captured Privilege Escalation techniques was one aspect that made its detection difficult. The presence of certain open source software on network resources and the introduction of new users to user groups with administrative privileges were two cues that the analysts made use of in their decisions. The analysts noted a gap in support for finding correlations between such indirect cues.

Defense Evasion is a naturally difficult problem for cyber analysts. Attackers obscure their malicious activity in the hopes of avoiding detection. Evasion techniques constantly evolve making it difficult to build or maintain rules to alert defenders. One analyst commented that "it's like playing Battleship [45] with pieces that move." Failures of automated security tools such as anti-virus or SCCM are potential indicators of Defense Evasion, where attackers may sabotage known defense capabilities hoping to avoid detection. Root cause analysis is needed to discern exactly why security tools fail, and there is a gap in the support for correlating security failures with other activities.

The cyber analysts also identified Lateral Movement as a difficult attack category to detect. Discovering Lateral Movement typically involves looking through massive volumes of data for small signals. It can be captured in a wide variety of logs including process, connection, and authentication logs. Few support tools are able to process such a volume of data and help analysts identify patterns suggestive of this attack category.

Other themes which implied capability gaps were brought up repeatedly across both interviews. The lack of baselines for judging network user behavior was the most lamented gap across interviews and analysts. The definitions of "normal", "reasonable", or "acceptable" behaviors are fluid in real networks across time and between users, even with respect to a single network user. Critically, tools to support collecting data to form baseline models of network users would be of great value to cyber defenders. Understanding a user's baseline behavior, in comparison to both themselves at previous points in time and to other users at any point in time, can help cyber defenders determine what is "reasonable" for a specific user.

The cyber analysts also indicated that some form of "social network analysis" focusing on the relationships between credentials and access to network resources would be useful. Specifically, they needed to be able to answer the following question easily: if a given network credential were compromised, what network resources would be become accessible to malicious users and therefore at risk? Having this answer would benefit defenders in two ways; (1) it would let them anticipate further compromised assets and (2) it would let them pro-actively deploy additional logging or defensive tools. As one cyber analyst in our first interview remarked, "We don't know enough about our internal environment to know what someone would be after." Social network analysis of network resources and credentials would provide insight into where internal or external threats are headed next.

Alerts themselves can be improved. As one participant remarked, "Good alerts are transposable, composable, and shareable." A desirable system would let defenders specify semantically explicit alerts in such a way that rules are flexible and can evolve as threats change over time. They should also be flexible to reflect the changing analyst capabilities and knowledge.

Other needs were highlighted as general workflow support needs. For example, when discussing office arrangements and the role of collaboration in daily work, the cyber analysts unanimously agreed that they needed more whiteboards. Their specific need was for a shared place to brainstorm, share notes, collaborate, and "hold [their] brains during interruptions." No software tools currently offer a shared think space that easily connects to the varieties of tools and approaches used by the team of analysts. Easy task-tracking that

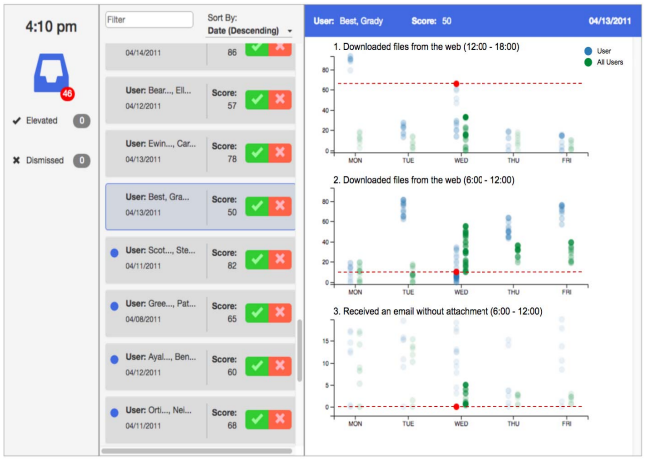

Figure 2: An example interface designed using the inbox metaphor with machine-learning and visualization components.

could be seen by others and used as a common think space must currently be satisfied by whiteboard space.

5 Design of an Baseline-Oriented Alert Management Workflow

We now discuss how we have leveraged the lessons learned from our interviews with cyber experts into the design of a prototype interface supporting alert management. This prototype will be designed to suit two specific research needs. It will 1) provide a platform we can use to continue to study trade-offs in design choices and their effects on the speed and accuracy with which cyber analysts can perform their work, and 2) provide an example tool design that we can continue to augment with additional features at the continued direction of cyber experts. Our prototype is not intended to be adopted as-is by expert cyber analysts in their work. Rather, it is intended as a tangible workspace in which we can explore alert management workflow characteristics in future empirical research.

The need for appropriate baselines was a theme that presented itself several times during the course of our group interviews, particularly with respect to the triaging decisions of alerting tasks. We chose to scope our early design space and prototype to responding to the following combination of needs: an alert management tool that provided baseline information to support the efficient triage of network alerts. To accomplish this, we leverage an inbox metaphor along with machine-learning and visualization components to display alerts and information relating to them. The goal was to keep the number of alerts presented to a human tractable, consistent with attention and memory limitations. The inbox approach leverages a familiar information management structure, requiring minimal learning on the part of a user. Other research has indicated that machine learning can support typical inbox management tasks [15] making the combination an attractive first-choice for prototype development. Our initial prototype is shown in Figure 2, and we describe its various components in the following sections.

5.1 Leveraging an Inbox Metaphor

Alerts can arrive at any time. Our experts lamented in both interviews that they were frequently interrupted by new investigative demands. Institutional polices can also affect the prioritization of alerts and force analysts to stop what they are doing and switch to a new task with potentially different context. This can be disruptive to the investigative process: ill-timed interruptions increase cognitive workload and can degrade cyber-cognitive situation awareness [8,29]. An alert management system is needed to support alert prioritization, minimize disruption, and easily integrate into the analyst's workflow.

To support triage of a stream of incoming alerts as well as management of persistent alerts that have already arrived, we make use of an interface metaphor: an e-mail inbox (which is itself a metaphor from earlier days of paper-based communication [26]). User-interface metaphors provide a means of enabling users to leverage knowledge from other domains to streamline working in a new domain [5]. A standard inbox-like interface provides a central place to easily visually identify newly-arrived alerts as compared to past alerts.

In our prototype (Figure 2), alerts appear in the central column with blue circles along the left side to denote that the given alert has not been "read" yet. This blue circle disappears once an alert has been viewed to distinguish between read and unread alerts. Alerts arrive with information about the network user involved in the alert and a severity score from the machine-learning component of the prototype (based on the work in [44]). Detailed alert information is displayed in the right-hand column once an alert has been selected. Alerts can be triaged through "elevate" and "dismiss" options allowing analysts to both view and take some action on alerts.

Other features from the inbox metaphor allow for alert management to be incorporated into our prototype. These features include: (1) sorting, which can be derived from both complex, automated prioritization rules based on machine analytics and human prioritization inputs based on experience, (2) categorization of alerts into sub-folders, and (3) searching and filtering for ease of navigation through potentially long lists of alerts. The inbox metaphor also enables us to organize alerts by rules that cyber defenders themselves could customize, similar to a spam filter in an e-mail program. Additional messaging could further be incorporated into the same system, providing a one-stop integration point for information flowing through the investigative process. In this way, we can demonstrate how future iterations of this inbox approach should integrate easily into an analyst's workflow.

5.2 Leveraging Automated Machine Analytics

Because of the overwhelming volume of data that cyber analysts must contend with each day, it is desirable to have some automatic component which summarizes, prioritizes and/or highlights anomalies. In our designs, the triage of alerts benefits from an indication of the severity of the alerting behavior. For the task of anomaly detection support, we have selected a deep learning auto-encoder algorithm by Tuor et al. [44]. We leverage the feature space on which this algorithm operates to assist with alert prioritization, to structure the visualization component, and to support the explanation of a given alert. In brief, this algorithm uses an unsupervised deep learning auto-encoder to learn which patterns of network activity are common. It is adept at detecting novel patterns or anomalies. This is particularly attractive in a cyber security context as cyber defenders are able to craft rules for known attack patterns; they need help in spotting new, emerging threats that look nothing like the attacks with which they have experience.

Tuor and colleague's algorithm uses aggregations of network activity over time to find anomalous patterns. Every user-activity pair fed into the algorithm is assigned and "anomalousness" score, indicating the relative familiarity/novelty of the feature combinations for that user. Unfortunately, the raw machine-learning output scores which drive the detection of anomalies are on an arbitrary scale and can be confusing to the humans who would make use of them. Thus, we translate the raw scores into percentiles, with 100 being the most anomalous patterns found; this is used to help rank the alerts presented in the inbox to the user, like an importance ranking. In Figure 2, the scores can be seen between the alert meta-data and the triage decision buttons (checkmark and X) within the gray boxes. We note that the data in this example is the open-source CERT dataset [36].

5.3 Leveraging Interactive Visualizations

There is tension between providing a visualization that is easy to read at a glance and providing the specific and nuanced detail needed to complete an analysis. In fact, the claim has been made by some cyber security experts that visualizations are an unnecessary middle step that interferes with a defender's ability to get from alerts to the data needed [7, 17]. At the same time, providing the right kinds of visualization, particularly proper context for decisions, does indeed improve performance [11, 19].

Rather than designing a new visualization system, we have chosen to include a visualization component in our alert management support tool. With this visualization component, we provide the means to support a cyber defender in need of quick comparisons of relevant data pertaining to a given alert with appropriate baselines. Returning to our earlier email inbox metaphor, the visualization component provides a reading pane to display the data behind the alert and juxtaposes it with baseline information. By selecting an alert in the inbox to view in an adjacent pane, cyber defenders should be able to quickly discern what triggered an alert and how far out of the ordinary it is. Based on the lessons from our interviews with cyber experts, we desire to highlight three aspects of a the relevant user's behavior with respect to a selected alert: (1) how anomalous the given user's activity is compared to other network users' activities on the same day, (2) how anomalous the given user's activities are on that day compared to their own rolling 5-day history, and (3) how the user's pattern of activity compares to all network activity within the rolling 5-day history. Consistent with the needs of our cyber experts, visualization provides critical evidence for whether an alert should be elevated for deeper investigation, monitored for further related alerts or ATT&CK indicators, or discarded as a false alarm.

We use the feature space defined on the CERT data for the autoencoder [44] to structure the visualization component of our prototype and describe why a given alert was generated. A total of 408 features are available and defined by activity type and time of day, which gives us flexibility in providing semantic descriptions of what in a given network user's behavior was found to be anomalous. The visualization component makes use of interactive, small multiples to display the top three features from the machine-learning component which produced the alert. Each component in the visualization displays the data for a single feature across a rolling 5 day history and displays the network events of the user associated with a given alert as well as a sampling of similar network users over the same time period. This allows a cyber analyst to compare the unusual behavior of a particular network user against themselves and their peers as an "activity baseline". A table of raw event-records is paired with each scatter plot and provides cyber defenders with their much-desired instant access to details. Clicking on specific events in the scatter plot will apply highlighting to this table allowing cyber defenders to quickly locate details.

We chose to use a variation of scatter plots to keep our initial prototype visualization as simple as possible while maintaining our ability to provide specific details and options for interactivity. The overall design of our prototype would also allow us to substitute more advanced and richer-featured visualizations. For example, something like the CAR Exploration Tool [40] could be used to directly tie the ATT&CK Matrix into alert response. Others have also suggested complete visualization systems dedicated to exploring alerts in detail [11, 19]. Our immediate needs, however, are for a prototype that we can begin using to empirically evaluate if a visualization-support workflow improves speed and accuracy of alert disposition decisions. This simpler visual component provides a tractable starting point against which we can continue to empirically evaluate more advanced visualizations in future work.

6 Discussion

Recent work defining system requirements to support cyber-cognitive situation awareness has argued that successful support of analysts requires user interfaces and support tools that integrate the analyst's goals and goal-directed tasks [22, 23]. Although a number of tools exist to support cyber analysts, many lack a way for analytic activities to directly connect to their goals [18]. Our work began with expert knowledge elicitation to better understand how the regular tasks of cyber analysts are impacted by data and their need for tool support. We have learned that the goal of triaging system alerts is intimately related with a need for systems that support an understanding of baseline activity for users on network. These baselines offer critical context for understanding what triggered an alert. When integrated with frameworks like the MITRE ATT&CK Matrix, this context helps analysts determine if the indicators fit a malicious pattern or constitute a false alarm.

There is a strong push and many incentives for increasing automation in cyber defense activities. But, as many previous researchers have noted, this can be problematic in the cyber threat space with moving targets and dynamically changing networks. Automated rule-based systems can quickly become obsolete and cannot usually adapt themselves to the changing network. Thus, human analysts will continue to have a strong role to play in cyber investigations. As we have shown herein, machine learning can be used to suggest anomalous activities (alerts) that adapt to the history within that network's data. The human expert then makes the critical decision about alert importance; these decisions might be tracked, in future, by a mixed-initiative system to refine the machine learning so that it tracks changes in network threats according to the analyst's decision process. Additionally, attack models could further refine the mixed-initiative support for threat disposition decisions.

We suggest that an effective, streamlined alert management workflow can be constructed by leveraging an interactive visualization interface and appropriate machine learning and analytics. Use of the inbox metaphor suggests novel ways of tracking the analytic provenance for alerts and provides a platform into which we can develop additional features. Automated support in this task can be easily included in our prototype. Our prototype system also integrates visualizations of the relevant alert and baseline activities called for by the experts we interviewed, and supports interactions to facilitate alert triage and information management. This prototype is intended to be used for empirical research as called for in [18] and to stimulate further discussion about integrated user interfaces. Following initial evaluation by cyber analysts with simulated malicious attacks, we want to consider design options to improve alert evaluation efficiency. Options include more sophisticated sorting functions, user-defined folder creation, and mixed-initiative recommendations of triage actions (e.g., grouping alerts according to similarity with previously triaged alerts). Each option will have an effect on how the cyber analyst utilizes the list of alerts and, therefore, will impact the efficiency in addressing them. Importantly, if we build on a system that provides the baseline visualizations needed by the cyber experts, then we can support the workflow they desire.

Acknowledgments

We thank Joseph Cottam for his technical and editorial support. The research described in this document was sponsored the U.S. Department of Energy (DOE) through the Analysis in Motion Initiative at Pacific Northwest National Laboratory. 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.

References

[1] FireEye, Inc. Fireeye, 2017.

[2] Nagios Enterprises, LLC. Nagios, 2017.

[3] M. Alsaleh, A. Alqahtani, A. Alarifi, and A. Al-Salman. Visualizing phpids log files for better understanding of web server attacks. In *Proceedings of the Tenth Workshop on Visualization for Cyber Security*, pp. 1–8. ACM, 2013.

[4] V. M. Alvarez. Yara, 2017.

[5] P. Barr, R. Biddle, and J. Noble. A taxonomy of user-interface metaphors. In *Proceedings of the SIGCHI-NZ Symposium on Computer-Human Interaction*, pp. 25–30. ACM, 2002.

[6] D. M. Best, S. Bohn, D. Love, A. Wynne, and W. A. Pike. Real-time visualization of network behaviors for situational awareness. In *Proceedings of the seventh international symposium on visualization for cyber security*, pp. 79–90. ACM, 2010.

[7] D. M. Best, A. Endert, and D. Kidwell. 7 key challenges for visualization in cyber network defense. In *Proceedings of the Eleventh Workshop on Visualization for Cyber Security*, VizSec '14, pp. 33–40. ACM, New York, NY, USA, 2014. doi: 10.1145/2671491.2671497

[8] J. P. Borst, N. A. Taatgen, and H. van Rijn. What makes interruptions disruptive?: A process-model account of the effects of the problem state bottleneck on task interruption and resumption. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 2971–2980. ACM, 2015.

[9] D. Botta, R. Werlinger, A. Gagné, K. Beznosov, L. Iverson, S. Fels, and B. Fisher. Towards understanding it security professionals and their tools. In *Proceedings of the 3rd symposium on Usable privacy and security*, pp. 100–111. ACM, 2007.

[10] D. M. Cappelli, A. P. Moore, and R. F. Trzeciak. *The CERT guide to insider threats: how to prevent, detect, and respond to information technology crimes (Theft, Sabotage, Fraud)*. Addison-Wesley, 2012.

[11] B. C. Cappers and J. J. van Wijk. Understanding the context of network traffic alerts. In *Visualization for Cyber Security (VizSec), 2016 IEEE Symposium on*, pp. 1–8. IEEE, 2016.

[12] K. K. Cook, E. R. Burtner, B. P. Kritzstein, B. R. Brisbois, and A. E. Mitson. Streaming visual analytics workshop report. Technical report, Pacific Northwest National Laboratory, Richland, WA, March 2016.

[13] R. J. Crouser, L. Franklin, A. Endert, and K. Cook. Toward theoretical techniques for measuring the use of human effort in visual analytic systems. *IEEE transactions on visualization and computer graphics*, 23(1):121–130, 2017.

[14] A. Dasgupta, D. L. Arendt, L. R. Franklin, P. C. Wong, and K. A. Cook. Human factors in streaming data exploration:challenges and opportunities for information visualization. In *Eurographics Conference on Visualization (EuroVis) 2016*, vol. 35, 2016.

[15] M. H. Dredze. *Intelligent email: Aiding users with AI*. PhD thesis, University of Pennsylvania, 2009.

[16] A. DAmico and K. Whitley. The real work of computer network defense analysts. In *VizSEC 2007*, pp. 19–37. Springer, 2008.

[17] G. A. Fink, C. L. North, A. Endert, and S. Rose. Visualizing cyber security: Usable workspaces. In *6th International Workshop on Visualization for Cyber Security, VizSec 2009*, pp. 45–56. IEEE, 2009.

[18] U. Franke and J. Brynielsson. Cyber situational awareness–a systematic review of the literature. *Computers & Security*, 46:18–31, 2014.

[19] J. R. Goodall, W. G. Lutters, P. Rheingans, and A. Komlodi. Focusing on context in network traffic analysis. *IEEE Computer Graphics and Applications*, 26(2):72–80, 2006.

[20] F. L. Greitzer and D. A. Frincke. Combining traditional cyber security audit data with psychosocial data: towards predictive modeling for insider threat mitigation. *Insider Threats in Cyber Security*, pp. 85–113, 2010.

[21] F. L. Greitzer and R. E. Hohimer. Modeling human behavior to anticipate insider attacks. *Journal of Strategic Security*, 4(2):25, 2011.

[22] R. S. Gutzwiller, S. Fugate, B. D. Sawyer, and P. Hancock. The human factors of cyber network defense. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 59, pp. 322–326. SAGE Publications Sage CA, Los Angeles, CA, 2015.

[23] R. S. Gutzwiller, S. M. Hunt, and D. S. Lange. A task analysis toward characterizing cyber-cognitive situation awareness (CCSA) in cyber defense analysts. In *2016 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, pp. 14–20. IEEE, 2016.

[24] C. Humphries, N. Prigent, C. Bidan, and F. Majorczyk. Elvis: Extensible log visualization. In *Proceedings of the Tenth Workshop on Visualization for Cyber Security*, pp. 9–16. ACM, 2013.

[25] E. M. Hutchins, M. J. Cloppert, and R. M. Amin. Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains. *Leading Issues in Information Warfare & Security Research*, 1(1):80, 2011.

[26] R. J. Jasper and L. M. Blaha. Interface metaphors for interactive machine learning. In *Proceedings of Human-Computer Interaction International: Augmented Cognition 2017*. Vancouver, Canada, 2017.

[27] G. Klein, B. Moon, and R. R. Hoffman. Making sense of sensemaking 2: A macrocognitive model. *IEEE Intelligent Systems*, 21(5):88–92, 2006.

[28] V. F. Mancuso, J. C. Christensen, J. Cowley, V. Finomore, C. Gonzalez, and B. Knott. Human factors in cyber warfare ii: Emerging perspectives. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 58, pp. 415–418. SAGE Publications Sage CA, Los Angeles, CA, 2014.

[29] G. Mark, V. M. Gonzalez, and J. Harris. No task left behind?: examining the nature of fragmented work. In *Proceedings of the SIGCHI Conference on Human Factors in computing systems*, pp. 321–330. ACM, 2005.

[30] S. McKenna, D. Staheli, and M. Meyer. Unlocking user-centered design methods for building cyber security visualizations. In *Visualization for Cyber Security (VizSec), 2015 IEEE Symposium on*, pp. 1–8. IEEE, 2015.

[31] T. Nunnally, K. Abdullah, A. S. Uluagac, J. A. Copeland, and R. Beyah. Navsec: A recommender system for 3d network security visualizations. In *Proceedings of the Tenth Workshop on Visualization for Cyber Security*, pp. 41–48. ACM, 2013.

[32] C. L. Paul. Human-centered study of a network operations center: experience report and lessons learned. In *Proceedings of the 2014 ACM Workshop on Security Information Workers*, pp. 39–42. ACM, 2014.

[33] P. Pirolli and S. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of International Conference on Intelligence Analysis*, vol. 5, pp. 2–4, 2005.

[34] G. Silowash, D. Cappelli, A. Moore, R. Trzeciak, T. J. Shimeall, and L. Flynn. Common sense guide to mitigating insider threats, 4th edition. Technical Report CMU/SEI-2012-TR-012, Carnegie-Mellon University, Pittsburgh, PA, 2012.

[35] A. Silva, J. McClain, T. Reed, B. Anderson, K. Nauer, R. Abbott, and C. Forsythe. Factors impacting performance in competitive cyber exercises. In *Proceedings of the Interservice/Interagency Training, Simulation and Education Conference, Orlando FL*, 2014.

[36] Software Engineering Institute CERT Division. Insider threat tools, 2017.

[37] Splunk, Inc. Splunk, 2017.

[38] S. C. Sundaramurthy, J. McHugh, X. S. Ou, S. R. Rajagopalan, and M. Wesch. An anthropological approach to studying csirts. *IEEE Security & Privacy*, 12(5):52–60, 2014.

[39] The MITRE Corporation. Common attack pattern enumeration and classification, 2017.

[40] The MITRE Corporation. Cyber analytics repository, 2017.

[41] The MITRE Corporation. Malware attribute enumeration and characterization, 2017.

[42] The MITRE Corporation. Technique matrix–ATT&CK, 2017.

[43] The OASIS Cyber Threat Intelligence Technical Committee. Structured threat information expression, 2017.

[44] A. Tuor, S. Kaplan, B. Hutchinson, N. Nichols, and S. Robinson. Deep learning for unsupervised insider threat detection in structured cybersecurity data streams. In *Artificial Intelligence for Cybersecurity Workshop at AAAI*, 2017.

[45] Wikimedia Foundation, Inc. Battleship (game), 2017.